ACADEMY OF SCIENCES OF THE ESTONIAN S.S.R.

INSTITUTE OF LANGUAGE AND LITERATURE

Институт языка и литературы Академии наук Эстонской ССР. **Труды XI МКФН.** Том 4. На английском, русском, французском и немецком языках. Редактор Ю. Викс (Ü. Viks). Художник-оформитель Х. Пузанов (H. Puzanov).
MB—04373

Trükkida antud 01.06.87. Formaat 60×90/8. Ofsetpaber. Ofsettrükk. Tingtrükipoognaid 45,7. Arvestuspoognaid 47,2. Trükiarv 2000. Tellimuse nr. **2423**. Hind rbl. 4.70. Trükitud EKP Keskkomitee Kirjastuse trükikojas. 200090 Tallinn, Pärnu mnt. 67 a. Köidetud Tallinna Raamatutrükikojas. 200006 Tallinn, Laki 26.

CONTENTS

INDEX OF AUTHORS

DURATION: MEASUREMENTS, PHONOLOGICAL FUNCTIONS, THEORETICAL IMPLICATIONS

ILSE LEHISTE

Department of Linguistics
The Ohio State University
Columbus, OH 43210-1229, U.S.A.

ABSTRACT

The paper starts with a survey of the linguistic functions of duration. A specific case is then discussed in greater detail: the durational structure of Estonian disyllabic words with a three-way quantity contrast. Measurements show that the durations of the two syllables exhibit the following typical ratios for the three quantities: Quantity 1 - 2/3, Quantity 2 - 3/2, and Quantity 3 - 2/1. Listening tests, using white noise signals, were given to 28 English-speaking and 28 Estonian-speaking listeners. The results showed that both groups perceive duration ratios of 2/3 as distinct from duration ratios of 3/2 and 2/1, but that they do not use duration ratios to separate the latter two quantities.

DURATION: MEASUREMENTS, PHONOLOGICAL FUNCTIONS, THEORETICAL IMPLICATIONS

The general theme of this session is "The role of phonetics in linguistic theory". This is a vast topic; I doubt whether I could do justice to it in anything shorter than a book-length manuscript. There are two alternatives--to survey the field at a very general (and probably somewhat superficial) level--or to narrow the topic so it could be treated at reasonable depth. I have decided in favor of the second alternative, and after a very brief survey, I will treat one subtopic in somewhat greater detail. A considerable part of the past thirty years, I have studied the temporal organization of spoken language; but there is still very much left to discover, and it so happens that I also still have something new to say.

Let us start with a survey of the various functions of duration. I have treated this topic before in a fair number of publications, and I have summarized the results of my own work, and that of others, in a paper entitled "The many linguistic functions of duration" (Lehiste 1984). While new details have been added to our knowledge since that time, the general picture does not seem to have changed, at least not as far as I am aware, so I will summarize briefly the findings presented in that paper. References to work by other scholars, up to the year 1984, will be found in that publication; I have added some more recent references to the bibliography of the current paper.

As a start, I would classify the linguistic functions of duration as follows.

1. Duration serves to establish the identity of a segment.
2. Duration serves to specify the meaning of a word. According to structuralist terminology, this would be called phonemic duration or quantity.
3. In many languages, duration is a strong cue for stress and emphasis.
4. Duration serves to indicate the position of a linguistic unit within a higher-level linguistic unit:
    - the position of a segment within a syllable
    - the position of a syllable within a word
    - the position of a word within a phrase and a sentence
    - the position of a sentence within a unit of discourse
5. Duration functions in establishing boundaries between linguistic units.

When we talk about the role of duration in establishing the identity of a segment, we are dealing, first of all, with non-contrastive, intrinsic duration. It is well known that some sounds are longer than others, all other factors being kept constant. There are also language-specific durational phenomena at this level: subphonemic differences between sounds can serve to identify the sounds, even though such differences are not used independently for contrastive purposes. For example, in English duration serves as a strong perceptual cue distinguishing certain inherently long and short vowels, and duration of a syllable nucleus serves as a cue to the voicing or voicelessness of a postvocalic consonant.

There are also many languages in which duration can be independently contrastive at the segmental level; at least that is the traditional way of analyzing oppositions between long and short vowels and consonants in languages like Finnish. I believe, however, that in most cases contrastive segmental duration is further modified by durational patterns that apply at the next higher levels--at least at the level of syllables and words. Contrastive segmental duration is integrated into durational patterns that apply at higher levels.

I will come back to that later in the current paper, when I present some new data. Let me mention just now that in Estonian polysyllabic words, the durational pattern is normally distributed over a disyllabic sequence. Thus the durations of the first and second vowel in minimal triples like *sada* - *saada!* - *saada* have an inverse relationship: lengthening of the first vowel is accompanied by shortening of the second vowel. The relationship between the durations of the two syllables appears to play a strong role in the listener's perception of the identity of the word; the durational pattern has the whole disyllabic sequence as its domain.

One result of the temporal relationships just referred to is to keep the duration of disyllabic words approximately constant--at least there is a clear tendency for maintaining something that might be called the "temporal integrity of the word". This means also that the duration of a syllable depends on the total number of syllables within the word: the tendency to keep the duration of the word close to some average level causes the syllables to become shorter when there is a larger number of them in the word.

Studies by Nooteboom (1972) and Lindblom and Rapp (1973) have shown for Dutch and Swedish respectively that duration of a stressed long vowel is longest in monosyllables and decreases systematically with the addition of further unstressed syllables. Tarnóczy showed that for Hungarian already in 1965 (Tarnóczy 1965). In a study which I published in 1975 (Lehiste 1975a) I showed that in English, a stressed syllable nucleus is longest in a monosyllabic word and shorter in polysyllabic words; I showed likewise that a stressed syllable is longer than an unstressed syllable in the same position, and that a syllable in final position is longer than the same syllable in non-final position.

It is thus clear that the position of a syllable within a word influences its relative duration. As part of the same study, I also found that the duration of a test word depends on the length of the frame in which it appears: test words were longest in the shortest frame, and shorter in two long frames used in the study. The way the duration of the test words interacted with the duration of the frames shows that the speakers integrate the test words into the utterance at the level at which the time program for the whole sentence is generated.

I got similar results in my first study of paragraph structure (Lehiste 1975b): sentences were longer when they were produced in isolation, and shorter when they were part of a paragraph--which indicates that the temporal planning extends to units larger than a single sentence. Furthermore, sentences in final position within a paragraph were longer than the same sentences when they occurred in initial or medial position within a paragraph.

This lengthening is part of a more general process which I have called pre-boundary lengthening. Pre-boundary lengthening is also used extensively to indicate the position of syntactic boundaries within a sentence. I have carried out several studies of pre-boundary lengthening, relating it to the rhythm of the spoken utterance. In this context, I would like to review briefly my first study dealing with the disambiguation of syntactic ambiguity (Lehiste 1973). In that study, listeners were able to identify correctly such sentences in which difference in meaning was correlated with a difference in surface syntactic bracketing. Successful disambiguation was achieved when the speakers had increased the interstress interval that contained the relevant boundary. The speakers had used several ways to achieve the same aim; the most straightforward one was the insertion of a pause, but equally successful were other means like the lengthening of one or more segmental sounds preceding the boundary, i.e. pre-boundary lengthening.

My further studies of the relationship between syntactic and rhythmic structure of English sentences lead to the postulation of a connection between rhythm and syntax that operates in the following way. Speech is a rhythmic activity, as are most motor activities performed by human beings. Stressed syllables carry the greatest amount of information; therefore, attention has to be focussed on the stressed syllables. This is facilitated by setting up an expectation as to when the next stressed syllable is likely to occur. Producing sentences in such a way that stressed syllables occur at regular intervals contributes to optimal perception by the listeners whose attention is cyclically directed to the points in time at which the stressed syllables can be expected to be found (Martin 1972, Cutler and Darwin 1981). Furthermore, a disruption of the expected pattern--namely, lengthening of an interstress interval--can be used to convey crucial information about syntactic structure: the placement of a syntactic boundary. At least in English, the syntactic structure of a sentence is thus to a considerable extent manifested in the timing pattern of that sentence when produced orally by a native speaker of the language. Timing appears to me to be primary; whatever other cues may be present, they play a less effective role.

I base this claim on a study in which it was shown that syntactic boundaries can be effectively recognized when the test sentences have been reduced to monotone, thus eliminating any possible contribution from fundamental frequency (Lehiste, Olive and Streeter 1976). In a later study (Lehiste 1983), I confirmed these results from the point of view of perception, and provided additional evidence from the point of view of production.

In this brief survey of the linguistic functions of duration, I have discussed the function of duration in establishing the identity of a segment; I have talked about contrastive duration at the word level; I have also discussed the ways in which duration functions to indicate the position of a linguistic unit within a higher-level linguistic unit, and the ways in which duration functions to establish boundaries between linguistic units. I have not talked much about duration as a cue to stress and emphasis-mainly because by now this function of duration appears to be generally known and accepted. (For a recent treatment of the topic, cf. Beckman 1986.) There may be other linguistic functions of duration that I have unintentionally overlooked. But I hope the general picture is clear: duration plays a part at a number of levels, and no linguistic description of a language is complete without reference to the function of duration within the system.

I would like to return now to a very specific case in which the role of duration appeared to me to need further study. This is the question of the three-way contrast between disyllabic word structures in Estonian--a topic that has considerable theoretical interest.

In my first extensive study of segmental and syllabic quantity in Estonian (Lehiste 1960), I made the observation that the factor that determined whether a disyllabic word was in quantity 1, 2 or 3 was word structure--more specifically, the ratio between the durations of the first and the second syllable. Listeners assigned the word to quantity 1, when speakers had produced the word in such a way that the ratio was approximately 2 to 3; the word was assigned to quantity 2, when the ratio was approximately 3 to 2, and to quantity 3, when the ratio approximated 2:1. The average durations of the syllables analyzed in the study were 106 and 151 msec for words in quantity 1, 295 and 187 for words in quantity 2, and 435 and 195 msec for words in quantity 3.

Duration of the first syllable is, of course, contrastive at the syllabic level; the average durations of the first syllable can be classified into the three categories of short, long, and overlong, as has been traditional in Estonian phonetics and phonology. And the durational differences are accompanied by different fundamental frequency patterns. All three factors are phonetically present; several linguists have been interested in determining the hierarchy of importance among these three factors, and much work has been done in the description of the three-way quantity opposition in Estonian. Let me mention here just the most recent papers by Arvo Eek and several other scholars associated with the Institute of Language and Literature of the Estonian Academy of Sciences and Tartu University (cf. Eek 1983, which contains these references).

In his very thorough study of 1983, Eek related the ratios between syllable durations to speech tempo and to fundamental frequency patterns within the disyllabic sequence. Basically, words were heard as being in quantity 1, when the ratio of the second vowel and the first vowel was equal to or larger than 1.2. The word was assigned to quantity 2, when the ratio V2:V1 was between 0.57 and 0.81, and to quantity 3, when the ratio was equal to or smaller than 0.43. Differences in tempo and in Fo played important roles. According to Eek's study, quantities 1 and 2 differ primarily in duration, since Q1 could be turned into Q2 and vice versa by manipulation of duration alone. Additional phonetic features are required for the perception of Q3.

I had a problem with the ratios described by Eek: they are presented as having a fairly large range of values, and these values appeared too precise and too complex. Already in 1960, I had described the ratios in terms of simple numbers: 2:3, 3:2, 2:1. It seemed intuitively obvious to me that contrastive structures would be based on simple notions; and it appears that there is some experimental support to this idea. I would like to summarize now a paper by Dirk-Jan Povel entitled "Internal representation of simple temporal patterns" (Povel 1981).

Povel started from a study by Fraisse (1946), who had discovered a remarkable phenomenon in the production and perception of durations. Fraisse found that subjects who were asked to produce temporal patterns by tapping basically used only two durations; the longer duration was typically approximately twice as long as the shorter duration, with a ratio of 2:1. Povel investigated the limitations present in the perception of temporal sequences by having subjects imitate sequences of 150-msec beeps whose onset intervals were varied in a systematic fashion. The duration ratios of the intervals between beeps were relationships numerically expressible as 1:4, 1:3, 2:5, 1:2, 3:5, 2:3, 3:4, and 4:5. (Note that Povel always presented the shorter duration first, resulting in ratios smaller than 1.) The results of two experiments yielded the finding that the only duration ratio that was correctly reproduced was 1:2 (i.e. .50). The errors in production were systematic: there was a tendency toward the 1:2 interval ratio, so that smaller ratios were increased and larger ratios were made smaller. For example, a ratio of .40 was reproduced as .45, and a ratio of .66 was reproduced as .49. Under certain special conditions set up for a third experiment, subjects were also able to imitate interval relations of 1:3 and 1:4 accurately in the contexts used in the experiment.

Povel carried through his experiments at Indiana University, and his subjects were presumably native speakers of American

English. It is legitimate to ask whether similar results would be obtained, if the subjects were speakers of a language in which duration plays a contrastive role. The experiments which I am about to report were carried out to test precisely this question: are there any differences in the perception of durational ratios that are correlated with the linguistic use of duration in the native language of the subjects being tested. The experiments were carried out in collaboration with Dr. Robert Fox at the Ohio State University,. and will be described in more detail in a joint publication (Fox and Lehiste, in preparation).

Let us recall here that in Estonian, there exist sets of three minimally contrastive disyllabic words, consisting of the same segmental sounds. One measured characteristic of such words is the durational ratio between the two syllables, which I had already in 1960 observed and formulated as ratios 2:3 for words in Q1, 3:2 for words in Q2, and 2:1 for words in Q3. A considerable literature has grown up in the meantime; the work of Eek is particularly significant in this context (cf. Eek 1983). There is no doubt that measurements do not yield very precise ratios, and that there is a certain amount of variation to be found under different speech conditions. Other phonetic factors are likewise present in spoken utterances, such as the duration of the first syllable nucleus itself (relative to some possible average internal standard) and the fundamental frequency contour applied to the disyllabic sequence. Which of these phonetic factors is contrastive needs to be established by means of listening tests--measurements alone are not enough.

The theoretical interest of the problem is at least two-fold. There have been linguistic schools that claim that all linguistic oppositions are binary; sounds can be short or long, there are no three-way durational contrasts. Fraisse's and Povel's findings seem to support this point of view. If speakers of a language with a three-way quantity opposition likewise can only identify durational ratios of 1:2, the three-way opposition must be manifested by other means. If, however, speakers of such a language can identify additional durational ratios, especially such that occur in their native language, then it is true that the native language of a subject influences his performance in psychoacoustic tests.

Our experimental procedures differed considerably from those used by Povel, since we did not just want to replicate his experiment, but wanted to use stimuli that could be directly related to Estonian disyllabic word patterns. We used pairs of noise bursts with controlled durations. The ratios that we employed were those found in Estonian disyllabic words: 2:1, 3:2, 2:3, and 1:2, numerically equal to 2.0, 1.5, .66, and .5. In Povel's study, the first temporal interval was always shorter than the second; he seems to have assumed that the ratios 2:1 and 1:2 are perceived in the same fashion, but since in Estonian, quantities 1 and 2 contrast, having the ratios 2:3 and 3:2 respectively, we felt that this assumption would not be justified. The ratio 1:2 was included for symmetry's sake, even though it is not regularly found in Estonian disyllabic words.

Each experimental trial consisted of presenting two such paired signals, separated by very short pause. The subjects were asked to state whether the duration ratio of the first noise sequence was the same as or different from the durational ratio of the second noise sequence. There was a 500-msec pause between each experimental trial.

We wanted to ensure that subjects were comparing duration ratios and not, for example, the durations of the first noise burst in each sequence. (This would be comparable to assigning an Estonian disyllabic word to a quantity category on the basis of the duration of the first syllable.) With this concern in mind, a second factor was introduced in constructing the experimental tokens: overall duration of the noise sequences. In particular, in half of the noise sequences, the duration of noise bursts 1 and noise burst 2 summed to 350 msec. In the other half of the sequences, the noise burst summed to 450 msec. Sequences with the same overall duration (LONG-LONG or SHORT-SHORT) alternated, in random order, with sequences of different overall durations (SHORT-LONG or LONG-SHORT). It was hoped that in this way duration ratio differences among the experimental trials would not be confounded with noise burst duration differences. These factors were explicitly discussed in the instructions, and examples of duration ratio differences vs. overall duration differences were included at the start of the stimulus tape. Equal numbers of "sames" and "differents" were included in the experimental tape, and subjects were also informed of this fact so that their responses would not be skewed in one direction or another. Guessing was encouraged. In the actual administration of the test, subjects were asked to encircle the appropriate letter (standing for "same" or "different") on the test sheet.

The test was first administered to 28 subjects at The Ohio State University in Columbus. These subjects were native speakers of English, with minimal exposure to languages in which duration plays a contrastive role. The same test, using identical tapes, but appropriately translated instruction and test sheets, was administered to 28 subjects in Tallinn. (The help of colleagues Arvo Eek, Mati Hint, and Kullo Vende in carrying out the tests is gratefully acknowledged.) These subjects had Estonian as their native language, and they were tested in Estonian. The analysis of the responses was carried out in Columbus. Detailed results will be presented in a separate publication (Fox and Lehiste, in preparation); below are some preliminary results.

The results are presented in the form of five tables. The first four tables have the same structure. The ratio of the first sequence is indicated on the vertical axis, the ratio of the second sequence on the horizontal axis; the numbers in the cells of each matrix represent the percentage of "SAME" responses. Table 1 gives the responses of English listeners to stimuli in which both sequences had equal durations.

Table 1

Percentage of "SAME" responses given by English-speaking listeners to pairs of stimuli consisting of a sequence of two noise bursts. Ratio of the noise burst durations of the first sequence is indicated on the vertical axis, ratio of the stimuli of the second sequence on the horizontal axis. Both sequences had equal total duration.

		Ratio of second sequence			
		1:2	2:3	3:2	2:1
Ratio	1:2	93.9	77.7	18.3	11.6
of	2:3	78.6	92.3	33.0	17.9
first	3:2	18.3	14.7	93.6	91.1
sequence	2:1	14.2	15.2	89.2	93.3

Table 2 presents the same information for English subjects reacting to stimuli in which the two sequences had different durations.

Table 2

Percentage of "SAME" responses given by English-speaking listeners to pairs of stimuli consisting of a sequence of two noise bursts. Ratio of the noise burst durations of the first sequence is indicated on the vertical axis, ratio of the stimuli of the second sequence on the horizontal axis. The sequences differed in duration.

		Ratio of second sequence			
		1:2	2:3	3:2	2:1
Ratio	1:2	63.8	50.9	15.2	17.4
of	2:3	51.3	60.7	22.3	13.8
first	3:2	12.1	16.5	69.0	62.1
sequence	2:1	11.6	10.7	52.2	75.4

Table 3 shows the responses of Estonian listeners to stimuli in which both sequences had equal durations; Table 3 thus corresponds to Table 1.

Table 3

Percentage of "SAME" responses given by Estonian-speaking listeners to pairs of stimuli consisting of a sequence of two noise bursts. Ratio of the noise burst durations of the first sequence is indicated on the vertical axis, ratio of the stimuli of the second sequence on the horizontal axis. Both sequences had equal total duration.

		Ratio of second sequence			
		1:2	2:3	3:2	2:1
Ratio	1:2	96.7	89.7	10.7	9.8
of	2:3	85.7	95.5	18.8	8.9
first	3:2	6.7	12.9	97.6	93.3
sequence	2:1	5.8	6.3	93.3	98.7

Table 4 presents the responses of Estonian listeners in cases in which the sequences differed in duration; Table 4 thus corresponds to Table 2.

Table 4

Percentage of "SAME" responses given by Estonian-speaking listeners to pairs of stimuli consisting of a sequence of two noise bursts. Ratio of the noise burst durations of the first sequence is indicated on the vertical axis, ratio of the stimuli of the second sequence on the horizontal axis. The sequences differed in duration.

		Ratio of second sequence			
		1:2	2:3	3:2	2:1
Ratio	1:2	61.6	52.2	9.8	10.3
of	2:3	45.5	53.7	12.9	8.9
first	3:2	8.9	14.2	68.1	53.5
sequence	2:1	7.6	6.3	52.2	73.8

Let us compare first Tables 1 and 3 with Tables 2 and 4. The cells starting at the top on the left and descending diagonally show the identification as "SAME" of signals in which the ratios were in fact identical (e.g. cases in which the first sequence and the second sequence both had ratios of 1:2). Correct recognition was evidently more difficult in cases when the sequences differed in duration: the percentages in the cells constituting the diagonal are considerably lower in Tables 2 and 4, also indicating, among other things, that the two groups of listeners reacted to the differences in overall sequence duration in the same general fashion.

The results presented in Tables 2 and 4 reflect listeners' reactions to ratios in cases in which the overall duration of the stimuli provided a conflicting cue: they were identifying sequences as "same" in spite of the fact that overall durations were clearly different. On the basis of durations alone, the listeners should have identified all the stimuli serving as basis for Tables 2 and 4 as "different"; and if they had been simply guessing, the scores would have been close to 50%. It is obvious that in many cases, listeners were able to identify ratios correctly even when the signals differed in duration; the statistical significance of these results will be discussed in detail in the forthcoming publication referred to earlier (Fox and Lehiste, in preparation).

Let us look now at the four tables from the point of view of successful discrimination between the four ratios. Here the results are likewise quite clear: the listeners, both English-speaking and Estonian-speaking, recognized only two contrastive patterns--sequences that had a first element that was longer than the second element, and sequences that had a first element that was shorter than the second element. This result emerges from the fact that ratios 1:2 and 2:3 are not distinguished from each other, the same being true for ratios 3:2 and 2:1. The percentage of "correct positive" decisions is somewhat higher than the percentage of "incorrect positive" decisions, but the difference appears not to be statistically significant.

What about the difference between the linguistic backgrounds of the two groups of listeners? Table 5 provides some information that is relevant in the present context.

Table 5

Average percentages of "SAME" responses given by English-speaking and Estonian-speaking listeners to pairs of stimuli consisting of a sequence of two noise bursts. "Correct positive" refers to cases in which duration ratios that were actually identical were identified as "SAME". "Incorrect positive" refers to cases in which duration ratios of 1:2 and 2:3 on the one hand, and 3:2 and 2:1 on the other hand, were identified as "SAME". "Wrong" refers to cases in which ratios of 2:1 and 2:3, or 1:2 and 3:2, were identified as "SAME".

	Average "correct positive"	Average "incorrect positive"	Average "wrong"
English-speaking	80.3	69.1	16.4
Estonian-speaking	80.6	70.7	9.9

This table presents average percentages, calculated on the basis of the data presented in Tables 1-4. Average "correct positive" decision refers to cases in which, for example, the ratios of 2:1 and 2:1 were identified as "SAME". "Incorrect positive" refers to cases in which, e.g., the ratios 2:1 and 3:2 were identified as "SAME". Average "wrong" decision gives the percentage of "SAME" decisions involving pairs of opposite durational ratios (e.g. 2:1 and 2:3). And it is here that a difference between English-speaking and Estonian-speaking listeners emerges: the Estonian-speaking listeners appear less likely to call such ratios "same". The difference of 6.5 percentage points is in fact significant--and it is the only significant difference between the two groups of listeners.

Let us return now to the theoretical questions that were raised at the beginning of the paper. The listeners seem in fact to have been capable of distinguishing between shorter and longer signals, and to have been able to decide whether the first or the second member of a sequence was longer. Under the conditions of this experiment, the listeners did not distinguish between the ratios 1:2 and 2:3 on the one hand, and the ratios 3:2 and 2:1 on the other hand. The linguistic background of the listeners did not have any effect on this aspect of the outcome; but Estonian listeners were much less likely to confuse the ordering of longer or shorter elements within a sequence than were English-speaking listeners.

From the point of view of Estonian prosody, the following conclusions may be drawn. The results clearly show that words in Q 1, with a duration ratio of 2:3, are perceived as distinct from words in quantities 2 and 3, with duration ratios 3:2 and 2:1. These two long quantities, however, are not distinguished on the basis of duration ratio. Since under normal conditions listeners do indeed recognize the difference between words in quantities 2 and 3, other phonetic factors must provide the decisive information. Fundamental frequency contours are the most likely candidate, but further research may bring new information and new ideas. The present experiment suggests that Estonian should rightfully be considered an accent language, in which other phonetic factors besides durational ones play a significant role. The experiment also demonstrates that phonetics does indeed provide crucial information that must be taken into account when questions of linguistic theory are to receive satisfactory solution.

BIBLIOGRAPHY

Mary E. Beckman (1986), Stress and Non-Stress Accent. Foris Publications: Dordrecht, Holland/Riverton-U.S.A.

R. Carlson and B. Granström (1986), "A search for durational rules in a real-speech data base". Phonetica 43:140-154.

Elizabeth Couper-Kuhlen (1986), An Introduction to English Prosody. Max Niemeyer Verlag: Tübingen.

Anne Cutler and Christopher J. Darwin (1981), "Phoneme-monitoring reaction time and preceding prosody: effects of stop closure duration and fundamental frequency". Perception and Psychophysics 29:217-24.

Arvo Eek (1983), "Kvantiteet ja rohk eesti keeles (1)". Keel ja Kirjandus 26, 9:481-489, 10:549-559.

Robert A. Fox and Ilse Lehiste (in preparation),"Perception of duration ratios".

P. Fraisse (1946), "Contribution a l'étude du rythme en tant que forme temporelle". Journal de psychologie normale et pathologique 39:283-304.

Ilse Lehiste (1960), "Segmental and syllabic quantity in Estonian". American Studies in Uralic Linguistics, Bloomington, pp. 21-82.

Ilse Lehiste (1973), "Phonetic disambiguation of syntactic ambiguity". Glossa 7, 2:107-122.

Ilse Lehiste (1975a), "Some factors affecting the duration of syllable nuclei in English". Salzburger Beiträge zur Linguistik 1:81-104.

Ilse Lehiste (1975b), "The phonetic structure of paragraphs". A. Cohen and S. G. Nooteboom (eds.), Structure and Process in Speech Perception. Springer-Verlag: Berlin- Heidelberg-New York. Pp. 195-206.

Ilse Lehiste (1983), "Signalling of syntactic structure in whispered speech". Folia Linguistica 17, 1-2: 239-245.

Ilse Lehiste (1984), "The many linguistic functions of duration". James E. Copeland (ed.), New Directions in Linguistics and Semiotics. Rice University Studies, Houston, Texas. Pp. 96-122.

Ilse Lehiste, Joseph P. Olive, ànd Lynn A. Streeter (1976), "Role of duration in disambiguating syntactically ambiguous sentences". Journal of the Acoustical Society of America 60:1199-1202.

Björn Lindblom and Karin Rapp (1973), "Some temporal regularities of spoken Swedish". Papers from the Institute of Linguistics 21, University of Stockholm.

J. Martin (1972), "Rhythmic (hierarchical) vs. serial structure in speech and other behavior". Psychological Review 79:487-509.

S. G. Nooteboom (1972), Production and Perception of Vowel Duration. Doctoral Dissertation, University of Utrecht.

Dirk-Jan Povel (1981), "Internal representation of simple temporal patterns". Journal of Experimental Psychology 7, 1:3-18.

T. Tarnóczy (1965), "Can the problem of automatic speech recognition be solved by analysis alone?" Rapports du 5e Congrès International d'Acoustique, Vol. II, Conférences générales. Liége: D. E. Commins. Pp. 371-387.

BITE-BLOCK SPEECH IN THE ABSENCE OF ORAL SENSIBILITY

PHILIP HOOLE

Neuropsychological Department
Max-Planck-Institute for Psychiatry
D-8000 Munich 40

ABSTRACT

The ability of a patient suffering from loss of oral sensibility to produce acoustically accurate vowels in the presence of a bite-block, both with and without additional auditory masking, was examined. The results indicated that in the absence of oral afferent information articulatory compensation was forced to rely on auditory feedback.

INTRODUCTION

Bite-block experiments have been a popular means of investigating the articulatory system's compensatory abilities, especially regarding the speed with which compensation is achieved and the necessity for various forms of feedback. Lindblom, Lubker and Gay /4/ reported for isolated vowels almost perfect articulatory compensation for the presence of a 22 mm bite-block, even when formant measurements were made at the first glottal pulse. The question of whether production of bite-block vowels suffers when sensory information from the oral region is suppressed was addressed by Lindblom, Lubker and McAllister /3/ and Gay and Turvey /1/. The former reported distorted formant values when the bite-block condition was combined with anesthesia of the oral mucosa; the latter also reported distortion, but only when sensory deprivation also included temporo-mandibular nerve-block. The results of these two experiments were interpreted by Perkell /6/ as demonstrating the motor system's dependence on afferent information to mark out an orosensory frame of reference.

In /1/ one subject was able to approach normal formant values over the course of several syllables, presumably by using auditory information. This led to Kelso and Tuller's /2/ logical extension of the paradigm, with auditory information now being eliminated as well through masking with white noise. For their 5 subjects, including, remarkably, Gay and Turvey's subject just mentioned no significant vowel distortion was found, even under these more difficult conditions.

These results thus cast doubt on Perkell's concept of an orosensory frame of reference underlying compensatory behaviour.

Using a different paradigm (unexpected electrical stimulation of orbicularis oris) Linke /5/ has reported undisturbed spontaneous speech but reduced compensatory abilities in a patient suffering from absence of trigeminal afferent information bilaterally following surgical treatment for trigeminal neuralgia.

These conflicting results impelled us to perform a bite-block experiment with a patient from our clinic who showed substantial deficits in oral sensibility.

SUBJECT

Three years prior to the experiment reported here the patient (male, aged 29, native German speaker with some Bavarian dialectal influence) suffered closed-head trauma and whiplash injury to the cervical cord in a sporting accident. For about the first month afterwards he was only capable of monosyllabic utterances, but subsequently his articulatory abilities recovered rapidly, being essentially normal six months after the accident. Substantial sensory deficits for the oral region were observed immediately after the accident, with no signs of subsequent improvement. Immediately prior to the experiment we examined the patient's oral sensibility in detail. In all speech structures where detailed testing was possible, namely lower and upper lip, tongue tip and blade, and mucosa of the oral cavity, thresholds for light touch, two-point discrimination, temperature and vibrotactile perception were raised so substantially as to be unmeasurable with our custom-developed equipment for assessment of oral sensibility. No forms at all could be recognised in a 12-form test of oral stereognosis. Less formal testing techniques also revealed substantial deficits in the pharyngeal region. As far as the speech system is concerned, the sensory deficits of our patient were thus probably more severe than those of Linke's patient and possibly also than those of the subjects in /1/, /2/ and /3/. It is perhaps also relevant to point out that in contrast to these subjects the sensory deprivation no longer constituted a novel experience for our patient.

Regarding the patient's articulatory abilities we have mentioned above that they recovered quickly, and at the time of the experiment he had for a considerable period no longer been considered dysarthric. Intentional mobility of the tongue for non-speech tasks had remained impaired, however (e.g. moving the tongue along the outer surface of the upper lip on command); yet it is important to note that the patient described by Linke showed very similar problems while also apparently having undisturbed speech articulation.

PROCEDURE

We endeavoured to replicate the procedure followed in /4/ as closely as possible, regarding vowels produced, mode of elicitation and size of bite-block (although we restricted our investigation to the larger-size bite-block, i.e. 22 mm). The patient was asked to produce nine repetitions (in three sequences of three) of the German vowels /i:/, /u:/ and /ɑ:/ under the following conditions and in the following order:

(1) initial unperturbed (IU)
(2) perturbed by white-noise at 80 dB delivered over headphones (WN)
(3) perturbed by a 22 mm bite-block between the lateral incisors (BB)
(4) perturbed by both white-noise and bite-block (WN/BB)
(5) final unperturbed (FU)

(Abbreviations used in Table 1 in brackets)

The subject was asked to produce the vowels as accurately as possible and without delay following presentation of a card with the target vowel triad.

The order of the triads in conditions 1 and 5 was randomized, while in the perturbed conditions all 9 tokens of a particular vowel were spoken as one sequence with the headphones or bite-block being removed briefly between each sequence. The order of the vowels was arbitrarily chosen as /ɑ,u,i/ in condition 2, /i,ɑ,u/ in condition 3 and /u,ɑ,i/ in condition 4.

The order of the perturbed conditions was so chosen that any learning effects would lead to a conservative result in the combined perturbation condition, i.e. would tend to underestimate the actual degree of disturbance (if any).

RESULTS

Vowel articulation was assessed by measuring the first two formants using an LPC-based procedure. In contrast to earlier investigations the main results again adopt a conservative approach to measurement since average values for the steady-state portions of the vowels were determined (an exception is the first vowel in the simple bite-block condition, see below). The results for each token are displayed in Figs. 1-3 for /i/, /u/ and /ɑ/ respectively, with the means for each condition being given in Table 1. The range for the initial unperturbed condition is also indicated in the Figures. The results will first be presented and discussed for each condition individually, followed by assessment of the results of the experiment as a whole.

White-noise condition

In this condition /i/ and especially /u/ show evidence of centralisation: for /i/ mean F1 is raised by 19 Hz and F2 lowered by 97 Hz; for /u/ F1 and F2 are raised by 89 and 119 Hz respectively. On the other hand /ɑ/ is relatively unperturbed. Under this condition the patient is, of course, effectively speaking without afferent information of any kind. The fact that /ɑ/ is less perturbed may reflect the fact that it is nearer than /i/ or /u/ to a neutral "setting", particularly for speakers of Bavarian. There is no evidence of systematic changes in the articulatory configuration in the course of the sequences under this condition.

Table 1

Steady-state F1 and F2 values in Hz averaged over each vowel in each condition

	/i/		/u/		/ɑ/	
	F1	F2	F1	F2	F1	F2
IU	273	2137	311	793	626	1083
WN	292	2040	400	912	628	1119
BB	291	2099	338	865	687	1187
BB/WN	332	2021	418	1176	717	1219
FU	266	2175	298	799	672	1068

Simple bite-Block condition

In the bite-block condition the main question is less whether compensation is achieved but rather how fast it occurs. In previous investigations compensation was virtually instantaneous, i.e by the first glottal period. To put the following figures into perspective we cite the estimates given in /4/ for the formant shifts to be expected for /i/ and /u/ in the complete absence of compensatory behaviour with a bite-block of this size:

For /i/ F1 +250 Hz, F2 -300 Hz
For /u/ F1 +300 Hz, F2 +500Hz.

Looking at /i/ and /u/ in these terms the subject shows clear compensatory behaviour since means over all vowels in the sequences are quite close to the initial unperturbed condition with F1 for /i/ raised by 17 Hz and F2 lowered by 27 Hz while for /u/ F1 and F2 are raised by 27 and 72 Hz respectively, i.e these formant values are all less perturbed than in the white-noise condition.

However, if we adopt as criterion for success that both F1 and F2 should be within the normal range then in the case of /u/ this criterion is only reached in the last vowel of the sequence and only 5 of a total of 18 F1 and F2 values are within the range of the initial unperturbed condition. Particularly striking is the fact that the last four vowels show a progressive and increasingly successful approach to the normal region.

For /i/ there is a fair amount of variability, but three of the nine vowels fulfil the criterion, with 10 of 18 F-values within the normal range. Note, however, that all these remarks apply to the measurements made in the steady-state portion of the vowel. For /i/ and /u/ we also measured a frame of 25 ms at the onset of the first vowel in each sequence. These values are indicated by squares in Figs. 1 and 2. The onset of the first /i/, in particular, was rather hesitant, being characterized by laryngealized, low intensity phonation. The precise values were:

/i/ F1 390 Hz, F2 2005 Hz
/u/ F1 421 Hz, F2 1152 Hz

Clearly these are a long way off target.

This subject is thus capable of compensation, but it is certainly not instantaneous, requiring tenths of seconds, or even seconds for complete success. This suggests a reliance on auditory information.

The results for /ɑ/ are somewhat puzzling. We had expected that the bite-block would cause virtually no articulatory disturbance. However the disturbance is, in fact, greater than for /i/ and /u/. F1 and F2 deviate upwards by 61 Hz and 104 Hz respectively, with no sign of an approach to the normal range over the course of the sequence. Auditorily the /ɑ/ productions sounded considerably fronted. This may provide the clue as to why no compensation is apparent. Unlike /i/ and /u/ the distortion caused by the bite-block would not, in the German vowel system, push the vowel into a different phonological category. It is probable that normally this low, back vowel can be realized acceptably with very little jaw opening, hence the observed distortion with the bite-block in place.

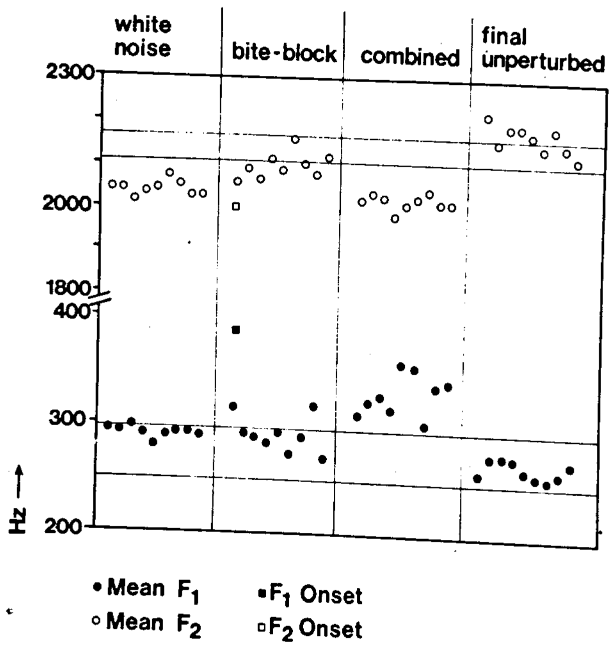


Fig. 1: Formant frequencies for all tokens of /i/ in each condition except initial unperturbed. Range for this condition indicated by horizontal lines.

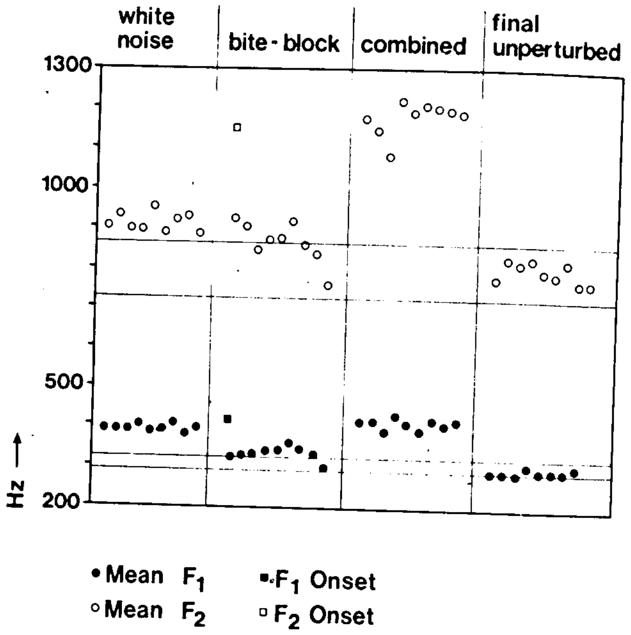


Fig. 2: Formant frequencies for all tokens of /u/ in each condition except initial unperturbed. Range for this condition indicated by horizontal lines.

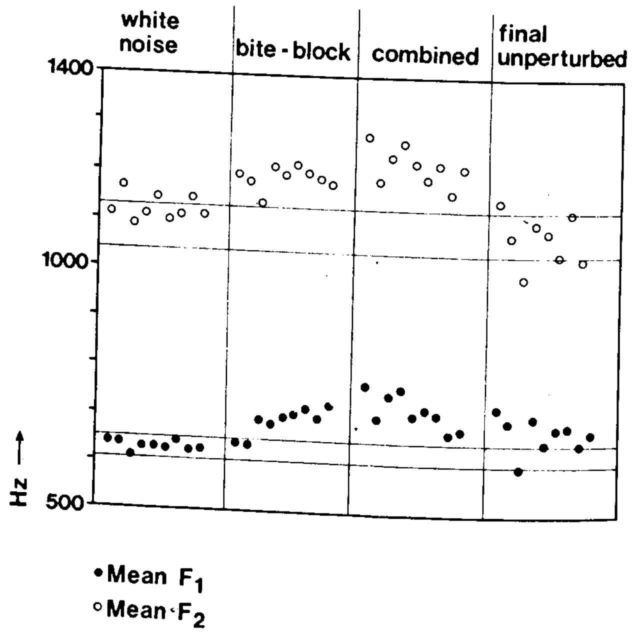


Fig. 3: Formant frequencies for all tokens of /ɑ/ in each condition except initial unperturbed. Range for this condition indicated by horizontal lines.

Combined white-noise/bite-block condition

Bearing in mind the interpretation offered above for the /i/ and /u/ results, it is to be expected in this combined condition that these vowels should be even more distorted. Figs. 1 and 2 show that this is indeed the case. The means in Table 1 show F1 for /i/ raised by 59 Hz and F2 lowered by 116 Hz, while F1 for /u/ is raised by 107 Hz and F2 by as much as 383 Hz. This continues a tendency for /u/ to show greater disruption than /i/.

The distortion is substantial, and there is no evidence of compensation improving over the sequence. It is also interesting to note that these mean values for /i/ and /u/ are quite close to the values measured at the onset of the first bite-block vowel, thus reinforcing the interpretation that the subject's compensatory behaviour was guided by auditory feedback.

For /ɑ/ the distortion is about the same as in the simple bite-block condition but with much increased variability.

Final unperturbed condition

Turning, finally, to this last, control condition it is again noticeable that /i/ and /u/ exhibit similar behaviour since the values tend to cluster around the extreme of the initial normal range opposite to the "perturbed" region. This suggests that the subject has indeed been trying to compensate, and may even be rather slow in turning off the compensatory behaviour. The results for /ɑ/ are again somewhat different, with a weaker tendency to depart from the perturbed region of the F1/F2 space. This again suggests that in the case of /ɑ/ simply less effort was made to compensate, and that apparently the distorted productions were still considered phonologically acceptable. One could also note that the greater distortion for /u/ than /i/ suggests that the subject followed a strategy of tongue-fronting when trying to cope with the perturbed conditions. This may, in addition to the greater jaw opening, have contributed to the unexpectedly large distortions for /ɑ/.

GENERAL CONCLUSIONS

The results for /i/ and /u/ are clearly very different from those obtained by Kelso and Tuller /2/. Our results strongly suggest that success in this type of perturbation experiment crucially depends on intact oral sensibility. Afferent information seems, as suggested in /6/, to be used to establish a frame of reference for motor commands. When sensory information is unavailable and when the natural geometry of the vocal tract is disturbed by a bite-block the necessary recalibration of the frame of reference fails to take place.

It might have been expected that information from the temporo-mandibular joint would be more important for the establishment of this frame of reference than information from the oral mucosa. The results in /1/ and /3/ suggest that this is not the case. This fact may, however, provide a line of attack for explaining the major discrepancy between our results and those in /2/, as well as the minor discrepancy between those of /1/ and /3/ regarding the amount of sensory deprivation necessary to cause vowel distortion.

The reduction in afferent information was clearly substantial in all reported experiments; it would thus be singularly unhelpful to simply put the different results down to surprisingly large effects of rather subtle differences in amount of sensory deprivation. We would like to conclude with a more concrete proposal:

In the reported experiments it is generally unclear to what extent anesthesia included the pharyngeal region. In our patient substantial sensory losses extended as far down as the laryngeal level. Recalling the unexpected amount of disturbance for the back vowel /ɑ/ we suggest that information from the pharyngeal region may have a prominent rôle to play in maintaining the integrity of the orosensory frame of reference as a whole.

REFERENCES

/1/ Gay, T. J. & Turvey, M. T. (1979): "Effects of afferent and efferent interference on speech production: Implications for a generative theory of speech motor control". Proceedings of the Ninth International Congress of Phonetic Sciences, 2: 344-350.

/2/ Kelso, J. A. S. & Tuller, B. (1983): " 'Compensatory articulation' under conditions of reduced afferent information: a dynamic formulation". J. of Speech and Hearing Research, 26: 217-224.

/3/ Lindblom, B., Lubker, J. & McAllister, R. (1977): "Compensatory articulation and the modeling of normal speech production behavior". In: Articulatory modeling and phonetics. Carré, R., Descout, R. & Wajskop, M. (eds.).

/4/ Lindblom, B., Lubker, J. & Gay, T. (1979): "Formant frequencies of some fixed-mandible vowels and a model of speech motor programmming by predictive simulation". J. of Phonetics 7: 147-161.

/5/ Linke, D. (1980): "Vorprogrammierung und Rückkopplung bei der Sprache". In: M. Spreng (ed.) Interaktion zwischen Artikulation und akustischer Perzeption. Stuttgart, Thieme.

/6/ Perkell, J. (1979): "On the use of orosensory feedback: An interpretation of compensatory articulation experiments". Proceedings of the Ninth International Congress of Phonetic Sciences, 2: 358-364.

ELECTROMYOGRAPHICAL CORRELATES OF SHOUTED AND WHISPERED VOICE

JEAN-FRANCOIS P. BONNOT

Laboratoire de Phonétique et Centre de Recherches en Informatique
et Combinatoire, UA CNRS 1099
Université du Maine, 72017 Le Mans Cedex, France

ABSTRACT

L'examen de l'activité électromyographique des muscles orbiculaire des lèvres et élévateur du voile lors de la production en voix normale de logatomes CVCVCV, nous a permis de mettre en évidence un modèle d'encodage hiérarchisé. L'émission en voix criée et en voix chuchotée de séquences du même type entraîne une restructuration de l'organisation temporelle (2 sujets; muscles étudiés: orbiculaire inférieur, élévateur, palatopharyngien). Ces conditions exceptionnelles présentent un certain nombre de caractères communs: il sèmble en particulier que la programmation opère à plus court terme, par unités de la taille de la syllabe ou du segment; les instructions motrices seraient donc fragmentées, au lieu d'être émises sous forme de "liste". D'autre part, l'élévateur du voile est plus sensible que le palatopharyngien aux modifications dans la situation de production. Une explication physiologique et linguistique est proposée.

INTRODUCTION

In two recent publications (Bonnot |1| and Bonnot et al.|2|), we brought forward a certain amount of experimental evidence supporting the concept of a temporal hierarchical organization of speech production. The utterance (CVCVCV nonsense words) was partly preprogrammed and C1 constituted an encoding reference for the whole item. A local reappraisal of timing arose during phonation, determining a re-structuration of the electromyographical activity (orbicularis oris sup.: OOS, levator veli palatini: LP) on an intrasegmental level. The basic motor controls were thus governed by two components operating in two different temporal fields: the sequencing was in charge of the seriation of the units and depended on the macro-structure. The phasing was related to the micro-structure. Its role was to produce the necessary adjustments and to protect the fluency (Kent |3|, Glencross |4|). This theory, which implies that time is a controlled variable, is compatible with a structural linguistic description because it accounts properly for the translation between an abstract dimension and a superficial one. The preprogrammed component carries out the choice and the transfer of units from the phonemic level to the phonetical level. This process is followed by allophonic specifications.

The model does not exclude biomechanical effects, but subordinates them to the programming requirements of the voluntary movement.

These experiments can be integrated to the framework of a normal use of the possibilities of the vocal tract. As is pointed out by Lubker |5|, apropos of the velopharyngeal mechanism, it is tempting to take up a teleological standpoint. The muscular activity and the articulatory gestures are organized and directed toward the goal of communication. Of course, speech production depends on temporal and physiological "boundary limits". The performer has to take into account the constraints peculiar to the implemented structures. The velocity and the accuracy of the various articulators or of parts of an articulator vary very much, as was shown by Eek |6| and Bothorel |7| among others. Furthermore, the motor task has to be carried out in a well-defined period. For Lubker |5|, "within these boundary limits, speakers have a great deal of variability open to them in their use of the velopharyngeal system." This variability, within or between subject(s), can also be linked to specific configurations of the tractus (post-operative patients, dental prosthesis ...) or just to unusual circumstances, such as local anaesthesia or shouted and whispered voice.

Both latter cases belong, like the pathological ones, to the "extrinsic variability", which is partly independent of the structure of the phonological system and of the "physiological weight" of the articulatory units. However, the point here is that we are within a natural use of the possibilities of the phonatory apparatus and of its motor controls. It can be proposed that the model which is described above undergoes a drastic restructuring when it comes to encounter these requirements.

RESULTS AND DISCUSSION

In order to test this hypothesis, we recorded two male native speakers of French (DA,JFB). The subjects were instructed to read nonsense CVCVCV words in normal, whispered and shouted voice. The consonant was |p t k g R| and the vowel |i u|.

We selected the following parameters: acoustical duration of CVCVCV; duration of the activity of the orbicularis oris inf.(OOI), LP, and palatopharyngeus (PPH); latency time of the same muscles (in the present case, interval between the onset of electromyographical activity and the first periodic oscillation for the initial vowel: |p t k| were of course voiceless. In view of comparison, an identical procedure was used for the voiced |g| and |R|). Student's t test and, in some cases, Cochran's test were applied. The coefficient of variation (C.V.= 100.SD/x) were calculated. With JFB, we could not obtain a good signal for LP during this session.

We first noticed that PPH and LP were acting in a very different way for subject DA. Whereas LP was very sensitive to the three conditions of production, the pattern of PPH remained steadier: in table 1, it can be seen that the normal, whispered and shouted mean values of LP are perfectly separated. On the contrary, there is considerable overlapping for PPH. The statistical comparisons reached a significant threshold in 12 cases out of 15 for LP, but in only 3 cases out of 15 for PPH.

Total duration of LP (in msec). Subject DA

	Normal		Shouted		Whisp.	
	$\bar{X}$	C.V.	$\bar{X}$	C.V.	$\bar{X}$	C.V.
|p	1545	11.52	1417	3.82	1216	11.73
t	1521	9.69	1403	3.09	1226	11.71
k	1529	10.19	1424	5.54	1260	13.48
g	1698	12.56	1498	5.97	1163	12.53
R|	1549	12.93	1450	5.29	1139	10.15

|p	Shouted vs. Normal	ddl: 18 NS
	Shouted vs. Whisp.	ddl: 18 $p < 0.05$
	Whisp. vs. Normal	ddl: 18 $p < 0.001$
t	Shouted vs. Normal	ddl: 18 $p < 0.05$
	Shouted vs. Whisp.	ddl: 18 $p < 0.05$
	Whisp. vs. Normal	ddl: 18 $p < 0.001$
k	Shouted vs. Normal	ddl: 17 NS
	Shouted vs. Whisp.	ddl: 17 $p < 0.05$
	Whisp. vs. Normal	ddl: 16 $p < 0.01$
g	Shouted vs. Normal	ddl: 18 $p < 0.05$
	Shouted vs. Whisp.	ddl: 18 $p < 0.001$
	Whisp. vs. Normal	ddl: 18 $p < 0.001$
R|	Shouted vs. Normal	ddl: 18 NS
	Shouted vs. Whisp.	ddl: 18 $p < 0.001$
	Whisp. vs. Normal	ddl: 18 $p < 0.001$

TABLE 1 A

Total duration of PPH (in msec). Subject DA

	Normal		Shouted		Whisp.	
	$\bar{X}$	C.V.	$\bar{X}$	C.V.	$\bar{X}$	C.V.
|p	1168	6.91	1236	6.58	1226	7.70
t	1156	16.64	1302	11.94	1094	7.38
k	1385	7.68	1464	1.96	1317	8.27
g	1502	9.04	1512	6.95	1354	1.61
R|	1193	15.37	1395	7.61	1167	5.26

|p	Shouted vs. Normal	ddl: 8 NS
	Shouted vs. Whisp.	ddl: 8 NS
	Whisp. vs. Normal	ddl: 8 NS
t	Shouted vs. Normal	ddl: 8 NS
	Shouted vs. Whisp.	ddl: 8 $p < 0.05$
	Whisp. vs. Normal	ddl: 8 NS
k	Shouted vs. Normal	ddl: 8 NS
	Shouted vs. Whisp.	ddl: 7 NS
	Whisp. vs. Normal	ddl: 8 NS
g	Shouted vs. Normal	ddl: 8 NS
	Shouted vs. Whisp.	ddl: 8 NS
	Whisp. vs. Normal	ddl: 8 NS
R|	Shouted vs. Normal	ddl: 8 $0.10 > p > 0.05$
	Shouted vs. Whisp.	ddl: 8 $p < 0.01$
	Whisp. vs. Normal	ddl: 8 NS

TABLE 1 B

NB: the smaller number of items for the "totrl duration of PPH" is due to the following fact: in most cases, for subject DA, te activity of PPH was absent or very weak for the nonsense words CiCiCi. Consequently, we took only into account utterances with |u|. For a detailed discussion,cf.Bonnot |1|.

It can thus be suggested that some muscles are more directly sensitive to those kinds of extrinsic constraints. It can be recognized that PPH plays a role in the narrowing of the velopharyngeal Isthmus (Fritzell |8|; Legent, Perlemuter and Vandenbrouck |9|), but there is no denying that LP is the only one which is responsible for the upward gesture of the velum, and to a great extent for the holding of the closure of the port (see for example Bell-Berti |1|). Even if we consider that Halle's model |11| describing the velar functioning is far from being adequate, we agree with his suggestion that "the distinctive features correspond to controls in the central nervous system which are connected in specific ways to the human motor and auditory systems."

For subject DA, an increase in the acoustical duration was not accompanied by a concomitant lengthening of the electromyographical activity. Whereas the durations were mostly shorter for the normal nonsense words on the acoustical level, on the contrary, they were systematically higher when considering the activity of LP and OOI.

It seems thus that a greater duration is not always straightforwardly correlated with a higher "force of articulation". The datas obtained from speaker JFB brought some support: here it is true that both the acoustical duration and the electromyographical activity of PPH increased from normal voice to shouted voice and finally to whispered voice. However, in both cases, significant differences were found between mean values for whispered vs. normal voice and for shouted vs. normal voice, but never for shouted vs. whispered voice. For example, for the nonsense words with |p t k|, the activity of PPH varied as follows (durations in msec.): normal voice: 1287-1309; shouted voice: 1594-1609; whispered voice: 1644-1729. The superior and inferior limits were separated by 285 msec. for normal voice vs. shouted voice, but by only 35 msec for shouted vs. whispered voice. With the |RVRVRV| items, the differences were 213 and 15 msec.

It must be added that the activity of PPH was remarkably similar for the normal and whispered utterances: the signal was poor and of a very limited amplitude; the shouted items were characterized by a much richer pattern.

This phenomenon underlines again the separation of the levels and suggests that duration is highly conditioned by the constraints inherent to the temporal programming of the sequence. Furthermore, the values of the C.V. were smaller for shouted voice and, to a lesser degree, for whispered voice: it could be that the speaker was "obliged" to reconsider partly his program, and to reduce to a minimum the area of variability.

Total duration of PPH (in msec). Subject JFB

	Normal		Shouted		Whisp.	
	X̄	C.V.	X̄	C.V.	X̄	C.V.
|p	1309	16.83	1609	4.62	1670	11.78
t	1294	13.75	1594	5.10	1644	8.27
k	1287	16.47	1604	6.04	1729	13.57
R|	1458	12.06	1671	6.51	1686	7.49

|p	Shouted vs. Normal	ddl: 16	$p < 0.05$
	Shouted vs. Whisp.	ddl: 16	NS
	Whisp. vs. Normal	ddl: 18	$p < 0.01$
t	Shouted vs Normal	ddl: 18	$p < 0.05$
	Shouted vs. Whisp.	ddl: 16	NS
	Whisp. vs. Normal	ddl: 20	$p < 0.001$
k	Shouted vs. Normal	ddl: 16	$p < 0.05$
	Shouted vs. Whisp.	ddl: 26	NS
	Whisp. vs. Normal	ddl: 28	$p < 0.001$
R|	Shouted vs. Normal	ddl: 18	$p < 0.05$
	Shouted vs. Whisp.	ddl: 16	NS
	Whisp. vs. Normal	ddl: 20	$p < 0.01$

TABLE 2

For DA, the latency times of implementing of LP were shorter in shouted and whispered voice, in comparison with normal voice (12 comparisons out of 15 were significant). For DA and JFB, OOI varied precisely in the same manner, even if all the comparisons did not reach the significant threshold of $p < 0.05$. As could be predicted on the basis of the behaviour of the total durations, the modifications in the latency times of PPH were scarcely noticeable although they followed the same pattern.

It can be concluded that:

(a) A stronger articulatory energy does not necessarily manifest itself through an earlier implementing of muscular activity.

(b) The shouted and whispered utterances can probably be joined together under the same head.

Latency time of LP (in msec). Subject DA

	Normal		Shouted		Whisp.	
	X̄	C.V.	X̄	C.V.	X̄	C.V.
|p	629	24.	405	12.75	370	28.51
t	630	21.91	393	9.98	373	28.90
k	592	23.29	399	13.90	334	22.72
g	717	24.40	491	15.71	336	13.83
R|	476	33.89	296	21.56	357	35.98

|p	Shouted vs. Normal	ddl: 18	$p < 0.05$
	Shouted vs. Whisp.	ddl: 18	NS
	Whisp. vs. Normal	ddl: 18	$p < 0.001$
t	Shouted vs. Normal	ddl: 18	$p < 0.05$
	Shouted vs. Whisp.	ddl: 18	NS
	Whisp. vs. Normal	ddl: 18	$p < 0.001$
k	Shouted vs. Normal	ddl: 18	$p < 0.05$
	Shouted vs. Whisp.	ddl: 17	$p < 0.05$
	Whisp. vs. Normal	ddl: 17	$p < 0.001$
g	Shouted vs. Normal	ddl: 18	$p < 0.05$
	Shouted vs. Whisp.	ddl: 18	$p < 0.001$
	Whisp. vs. Normal	ddl: 17	$p < 0.05$
R|	Shouted vs. Normal	ddl: 18	$p < 0.05$
	Shouted vs. Whisp.	ddl: 18	NS
	Whisp. vs. Normal	ddl: 18	$0.10 > p > 0.05$

TABLE 3

Instead of assuming that in uncommon circumstances, the latency time essentially reflects the global programming of the nonsense word, as it seems to be the case under normal conditions, we suggest that the encoding system works within a shorter temporal window (see also for an acoustical study of French CVC syllables: Rostolland et al. |12|). In our interpretation, the initial latency partly expresses the read-out time of generative encoding rules which are specific to the language in question. However, these rules are dependent on a rhythmic and accentual frame which is modified by the constraints inherent to the shouted and whispered phonation.

CONCLUSION

"Extrinsic variability" constitutes an essential part of a model of speech production, since the encoding modalities are conditioned by the constraints exerted on the muscular (sub-)system(s) and on the articulatory organs. Our conclusions allow to extend the notion of "syllabic segregation" applied by Kent and Rosenbek |13| to "apraxia of speech", and by Kent |14|, to the acquisition of language by children. Close connections can also be established between our results and some works on "expressivity". Fónagy |15| studied the articulatory manifestations of hatred and anger. These sentiments were rendered in the same way in French and Hungarian, by a series of jerky movements (abrupt transitions). In a cineradiographic study, Flament |16| looked into the question of "stylistic emphasis" in French. He came to the conclusion that there was a marked individualization of the articulatory units, in comparison with a neutral context: the coarticulatory link between successive segments was strongly weakened.

It appears that our analysis can be integrated into a more comprehensive body of facts, regrouping a great number of pathological and exceptional conditions. Of course, the patterns will be different according to the severity of the disease or to the weight of the constraint.

All these productions have probably in common to provide the subject with feedback information which is particularly difficult to handle. It could be that the (normal) subjects "tend to achieve some kinethetic-tactile feedback by finding articulatory landmarks" as was proposed by Rothman |17| for deaf adult speakers. Therefore, it can be suggested that the instructions are being split up, instead of being issued in the form a "list".

However, it must be stressed that there is a great variety of possible "amendment procedures" (Glencross |18|) and, consequently, a high flexibility of the matching of various kinds of feedback with the contextual requirements.

ACKNOWLEDGEMENTS

This work was supported in part by the Université du Maine (Le Mans), and in part by the Unité 3 INSERM "Physiologie et pathologie cérébrales". We are very indebted to Mme C.Chevrie-Muller, to M.B.Maton, and to Professor G.F.Greiner for the scientific and logistic support they provided us in the carrying out of the experiments. We also wish to thank our colleague, Mme C.Guillerme, for her able assistance in the english translation of this paper.

REFERENCES

|1| Bonnot J-F.P. Contribution à l'étude phonétique et phonologique de l'organisation temporelle de l'activité électromyographique labiale et vélaire. Coarticulation et processus d'encodage moteur. Thèse de doctorat d'Etat, Strasbourg II, 1986, 2 volumes, 704 pages.

|2| Bonnot J-F.P., Chevrie-Muller C., Arabia-Guidet C., Maton B. and Greiner G.F. "Coarticulation and Motor Encoding in CVCVCV Nonsense Words" Speech Communication, 5, 1986, 83-95.

|3| Kent R.D. "The Segmental Organization of Speech" in P.F.Mac Neilage (ed), The Production of Speech, Springer Verlag, New York - Heidelberg, 1983, 57-89.

|4| Glencross D.J. "Temporal Organization in a Repetitive Speed Skill" Ergonomics, 16, 1973, 765-776.

|5| Lubker J.F. "Some Teleological Considerations of Velopharyngeal Function", in Daniloff R.G.(ed), Articulation Assessment and Treatment Issues, College-Hill Press, 1983, 179-193.

|6| Eek A. "Articulation of the Estonian Sonorant Consonants I. |n| and |l|", Eesti NSV Teaduste Akadeemia Toimetised 19, Koĭde, Uhiskonnateadused, 1, 1970, Institute of Language and Literature, Estonia, USSR.

|7| Bothorel A. Etude phonétique et phonologique du breton parlé à Argol (Finistère-Sud). Thèse de doctorat d'Etat, Strasbourg II, 1978. Atelier national de reproduction des thèses, Lille III, diffusion Breizh, Spezed, 1982, 514 pages.

|8| Fritzell B. "The Velopharyngeal Muscles in Speech", Acta Otolaryngologica, Supplementum 250, Göteborg, 1969, 79 pages.

|9| Legent F., Perlemuter F. et Vandenbrouck C. Fosses nasales, pharynx, Cahiers d'anatomie, Masson, Paris, 1974.

|10| Bell-Berti F. "An Electromyographical Study of Velopharyngeal Function in Speech", Journal of Speech and Hearing Research, 19, 1976, 225-240.

|11| Halle M. "On Distinctive Features and Their Articulatory Implementation" Natural Language and Linguistic Theory, 1, 1983, 91-105.

|12| Rostolland D., Parant C., Takahashi A. et Pandales E. "Durée vocalique intrinsèque et co-intrinsèque en français: contraintes physiologiques et variations temporelles dans des syllabes CVC", Actes des 14e Journées d'Etudes sur la Parole,Paris GALF et E.N.S.T., 1985, 179-182.

|13| Kent R.D. and Rosenbek J.C. "Prosodic Disturbances and Neurologic Lesion" Brain and Language, 15, 1982, 259-291.

|14| Kent R.D. "Sensorimotor Aspects of Speech Developments" in Aslin R.N., Alberts J.R. and Peterson M.R. (eds), The Development of perception: Psychobiological Perspectives, Academic Press, New York - London, 1982, 161-189.

|15| Fónagy I. La vive voix. Essais de psychophonétique, Payot, Paris, 1983, 346 pages.

|16| Flament B. Recherches sur la mise en relief en français. Thèse de doctorat d'Etat, Strasbourg II, 2 volumes, 1984, 1175 pages.

|17| Rothman H.B. "A Spectrographic Investigation of Consonant-Vowel Transitions in the Speech of Deaf Adults" Journal of Phonetics, 4, 1976, 129-136.

|18| Glencross D.J. "Levels and Strategies of Response Organization" in Stelmach G.E. and Requin J. (eds), Tutorials in Motor Behavior, North-Holland Publishing Company, Amsterdam, 1980, 551-566.

THE WORD LEVEL INTERPLAY OF STRESS, COARTICULATION, VOWEL HEIGHT AND VOWEL POSITION IN ITALIAN

MARIO VAYRA

Scuola Normale Superiore
56100 Pisa, ITALY

CAROL A. FOWLER

Dartmouth College, Hanover, NH 03755, USA
Haskins Laboratories, New Haven, CT 06511

ABSTRACT

The experiment investigates the effects of stress and transsyllabic vowel-to-vowel coarticulation in Standard Italian. The study replicates evidence from our previous work on Italian and English of strong influences on unstressed vowels of their flanking stressed vocalic context. In Italian as in English, coarticulation has stronger effects on the front-back dimension than on vowel height. In contrast to English, however, coarticulatory influences in Italian are symmetrical in direction. As for stress, in the present study, we find effects of stress only on vowel opening, not along the front-back dimension. Interestingly, effects of stress on F1 interact with effects of a vowel's position in a word or utterance. We find that a stressed vowel is produced with a decreasingly extreme jaw position throughout the word or utterance. This may point to a suprasyllabic organization of jaw trajectories in Italian speech.

INTRODUCTION

Our study was designed to investigate three aspects of the articulatory organization of Italian speech: vowel-to-vowel coarticulation, word-level compensatory shortening and spectral differences between stressed and unstressed vowels. It was suggested in part by the outcome of our previous work (Vayra, Fowler and Avesani /15/), in which we compared measures of vowel-to-vowel coarticulation and shortening in Standard Italian and English. That cross-language comparison wes of interest in light of evidence linking coarticulatory and durational shortening patterns of English to the presumed rhythmic character of the language.

English is traditionally identified as a "stress timed" language according to a timing typology that classifies all languages into those, such as English, that are said to have a tendency for strong or stressed syllables to be evenly timed, and others, including Italian (called "syllable timed"), in which syllables are said to recur at regular intervals.

The description of English as stress-timed is consistent with several aspects of its prosodic structure. English words are sometimes described as being composed of "feet" consisting of a stressed syllable and zero, one or two following unstressed syllables (e.g. Selkirk /12/; Bolinger /2/).(Following Selkirk and others we will call a foot like that with the stressed syllable first: "left dominant"). Compatibly, measures of coarticulatory influences of stressed on unstressed syllables and of shortening of stressed vowels due to neighboring unstressed syllables are both asymmetrical, they correlate, and both mirror the left-dominant foot structure of words. (Fowler /5/). That is, stressed vowels coarticulate (at least on the front-back dimension) more with, and are shortened more by, following than preceding unstressed syllables. Fowler /5/ has interpreted these findings as evidence that coarticulatory influences by vowels reflect "coproduction"—that is,overlap of the stressed vowels' production by unstressed syllables in the same foot. Because a following unstressed vowel "covers over" the trailing edge of a stressed vowel, the stressed vowel is measured to shorten and it exerts a coarticulatory influence on the unstressed syllable to the extent that the syllable shortens it.

If the coarticulatory and shortening patterns just described for English do, in fact, reflect its presumed foot structure, then they should not be found in languages identified as syllable timed. Instead, shortening should be confined to the syllable and should serve to maintain equal syllable durations. In syllable-timed languages, if vowel-to-vowel coarticulation occurs at all, it should not reflect a foot structure, either left or right dominant.

The findings from previous studies of spoken Italian, including our own, do not support this picture of a syllable timed language. Nor do they give any clear picture of the timing structure of Italian. Researchers have found shortening of a vowel as consonants are added to the syllable (Farnetani and Kori /4/; Vayra, Avesani and Fowler /13/). However, the shortening is not consistently found (see also Bertinetto /1/). Moreover, it is asymmetrical with stroger shortening effects of following than of preceding consonants in the syllable (Farnetani and Kori /4/; Vayra, Avesani and Fowler /13/). Maddieson /9/ reports that this asymmetrical shortening pattern is widespread in the world's languages and is not especially associated with syllable-timed languages.

Just as Italian shows only weak and inconsistent shortening at the syllable level, it also shows weak and inconsistent evidence of a foot structure. Nespor and Vogel /10/ invoke a left-dominant foot structure to explain patterns of syllable stress in Italian words. Compatibly, den Os /11/ (see also Farnetani and Kori /3/; Vayra /14/) report evidence of reduction of unstressed vowels in Italian— ostensibly a characteristic of stressed-timed languages— and Koopmans-van Beinum /8/ finds evidence of vowel reduction in spontaneous spoken Italian, Dutch and Japanese as compared to vowel quality in more formal styles of speech.

Despite these findings, patterns of vowel-to-vowel coarticulation and shortening in Italian do not consistently reflect a stress-timing tendency (Vayra et al. /15/. Among three talkers we examined in our earlier study, one showed an asymmetrical coarticulation and shortening pattern similar to those found in English, one showed the reverse asymmetry in both coarticulation and shortening, and the other showed an asymmetry in coarticulation opposite to that in his shortening patterns. For none of the three talkers were measures of coarticulation and shortening correlated.

Talkers in that study were Piedmontese speakers of Standard Italian. One hypothesis we considered as to why patterns of coarticulation and shortening were idiosyncratic to each talker was that the prosodic differences in their pronunciation reflected the presence in the spoken Standard of ongoing conflictual processes of adaptation — outside Tuscany —to the morphophonology of the Florentine-based Standard system (represented in the orthography). Accordingly, in the present study, we examine patterns of vowel-to-vowel coarticulation among Florentine speakers of Standard Italian. In addition, we looked at effects of stress on vowel quality among these speakers. The experiment was designed to ask whether we would see consistent evidence of a foot structure in in the coarticulatory and shortening patterns of these talkers, and, if so, whether the same talkers would show evidence of vowel reduction in absense of stress.

EXPERIMENT

Our subjects were two female (S and F) and one male (N) native speakers of the Florentine variety of Standard Italian. Each of them produced several tokens each of 18 different bisyllabic nonsense words and 27 trisyllabic nonsense words. The bisyllabic words were versions of "VbV", in which the Vs were /i/, /a/ and /u/ and in which stress was either on the first or the second syllable of the nonsense word. The trisyllables were versions of " VbVbV ", again using all combinations of /i/, /a/ and /u/ for the two Vs ,and using all possible patternings of one stressed and two unstressed syllables. The talkers produced the nonsense words in isolation; we analyzed three tokens of each word type spoken by each talker.

We used the ILS system at Haskins Laboratories to measure center frequencies of F1 and F2 of the vowels. The measures we report were taken from vowel midpoints. In addition, vowel durations were made from waveform displays. These latter measures have not yet been analyzed, however, and so we report our findings on the formant measurements only.

In this report,too, we will present just a subset of our findings using F1 and F2 as measures. We have found, in general, that vowel-to-vowel coarticulatory effects are largely confined to the front-back dimension (that is, to measures of F2) rather than to the height dimension (F1), and so we report just F2 measures of coarticulation. For its part, stress has its major effect on F1 (see also /3/),and so we confine our exposition of stress effects to its effects on F1.

<u>Coarticulation and F2.</u>
We focused on three kinds of findings relating to wowel-to-vowel coarticulatory effects. First, we looked generally for effects of a context vowel (/i/ /a/ or /u/ on F2 of a target neighboring /a/. Next we asked whether any such effect were asymmetrical so that carryover effects of a preceding vowel were larger or smaller than anticipatory effects of a following vowel. If carryover effects are larger than anticipatory effects, then, as in English, coarticulatory effects in Italian and in these types of words in spoken Italian would mirror their presumed metrical left-dominant foot structure (Nespor and Vogel /10/). Finally we asked whether coarticulatory influences are affected by the stress of either the context (coarticulating) vowel or of the target /a/ vowel.

Tables 1 and 2 present the findings that address these issues. All three speakers showed significant and large effects of context vowels on F2 of a neighboring /a/. Across the talkers, this main effect of context vowel accounted for 11-35% of the total variation in F2 of the target vowel in bisyllables and 27-37% of the variation in F2 of the target vowel in trisyllables. Neighboring /i/ raised F2 of /a/ as compared to its value in the context of /a/ and /u/; /u/ generally lowered F2 as compared to its value in the context of /a/.

As for asymmetries in coarticulatory effects, no talker showed a significant asymmetry in either the bisyllables or the trisyllables. However one talker showed a marginal tendency in trysyllables for anticipatory effects of a context vowel to exceed carryover effects (p= .06), and, in general, talkers showed numerical differences in coarticulation favoring an anticipatory over carryover coarticulation (see Table 1). Thus, no talker showed a significant tendency for coarticulatory asymmetries to reflect a left-dominant foot structure. This finding is similar to our earlier findings on speakers of the Piedmontese variety of Standard Italian; however,speakers in the present study were more consistent one with the other than in our previous experiment.

TABLE 1

Bisyllables

	Carryover			Anticipatory		
	/i/	/a/	/u/	/i/	/a/	/u/
S	1671	1625	1653	1632	1563	1537
F	1433	1434	1388	1441	1413	1360
N	1313	1262	1210	1393	1276	1217

Trisyllables

	Carryover			Anticipatory		
	/i/	/a/	/u/	/i/	/a/	/u/
S	1651	1621	1572	1711	1621	1560
F	1406	1337	1330	1443	1337	1282
N	1305	1218	1160	1298	1218	1191

Table 1. F2 of /a/ in the context of preceding (carryover) and following (anticipatory) /i/, /a/ or /u/

Table 2 shows the interaction of stress and coarticulation on trisyllables. In that table, we have subtracted our F2 measure of the target vowel /a/ when it is in the context of /a/ from its value in the context of /i/. A positive difference, then, reflect the expected fronting effect of /i/ on /a/. We have presented the difference scores for three stress conditions separately. In the first column, the target /a/ vowel is stressed; in the second column, the context (coarticulating) vowel is stressed; in the third column neither is stressed. (So, nonwords i'baba and a'baba contributed to the first column of difference scores; 'ibaba and 'ababa contributed to the second column; iba'ba and aba'ba contributed to the third column). The table reveals two interesting findings. One is that there are essentially no coarticulatory influences of unstressed /i/ on stressed /a/. A second is that influences on unstressed /a/ are as large from unstressed neighboring vowels as from stressed neighboring vowels. This interaction between stress and contest vowel was significant for two talkers and marginal (p= .11) for a third. These findings are interesting in showing that,in these words,only unstressed vowels are subject to coarticulatory effects, but they receive coarticulatory influences from neighbors regardeless of their neighbor's stress level. The first finding is similar to effects found inEnglish. Unfortunately, we do not have data on English words comparable to those on which the second finding were obtained.

TABLE 2

	Target /a/	Stressed vowel Context /i/,/a/	Neither
S	17	130	106
F	11	121	164
N	26	84	141

Table 2. F2 of /a/ in the context of /i/ minus F2 of /a/ in the context of /a/. Data average over direction of coarticulatory influences and represent trisyllables only.

Stress and F1.
To examine reduction of unstressed vowels, we looked only at the utterance "aba" and "ababa" with all stress patterns. Table 3 presents our findings.

We find highly significant effects of stress on F1 of /a/, such that stressed /a/ is a more open vowel (with a higher F1) than is unstressed /a/. These are significant for two talkers in the bisyllables and marginal (p = .06) for the third. They are significant for all talkers in the trisyllables. Moreover, the effects of stress tend to be quite substantial, accounting for 5-35% of the total variability in F1 in our analysis of the "aba" words across the three talkers and 60-79% of the variance in F1 in "ababa" words. Thus, as others have found (e.g. den Os /11/), we find that in Italian, as in stress timed languages, unstressed vowels (at least the vowel /a/) are subject to reduction.

An unexpected finding in this analysis was a significant effect of vowel position in the word on F1. In bisyllables, all three talkers had lower F1s for final than for initial vowels; the difference was significant for two talkers and marginal for the third (p = .08). In trisyllables, the effect was significant for all talkers, but it interacted with vowel stress. Table 4 shows this interaction. For all three talkers, F1 of stressed vowels decreases monotonically across the word, while F1 of unstressed /a/ is highest in word-initial position and lowest in medial position. The interaction is significant for two of the three talkers, but the pattern is present in all three sets of means.

TABLE 3

	Bisyllables		Trisyllables	
	Stressed	Unstressed	Stressed	Unstressed
S	1125	925	1093	932
F	1002	916	993	875
N	772	761	801	734

Table 3. Effcts of stress on F1 of /a/.

DISCUSSION

If Italian, like English, has left-dominant foot structure in words, the feet are not reflected, here, in coarticulatory asymmetries. Instead,in the words we examined, coarticulation is largely symmetrical,with a weak, but fairly consistent,numerical tendency to favor anticipatory coarticulation. We have not yet analyzed our measures of durational shortening to determine whether they reflect the left-dominant foot structure or else reflect the coarticulatory near symmetry (or else do neither). Discovering how shortening is patterned in these words may help to clarify the relation of shortening to coarticulation and to metrical structure in words. In particular, it may help to determine whether the convergence of all three patterns in English is or not accidental.

Although coarticulatory patterns in F2 do not suggest a foot structure ostensibly characteristic of stress-timed languages, nevertheless, effects of stress itself on articulation of vowels are similar to its effects in stress-timed languages. Stressed vowels are not subject to coarticulatory influences from neighboring vowels along the front-back dimension and unstressed vowels are less open than stressed vowels. A new finding was that stressed vowels exert no stronger coarticulatory effects on their neighbors than do unstressed vowels.

One way to capture these findings is to suggest that,as compared to unstressed vowels, stressed vowels in Italian speech are relatively impervious to two kinds of influence: coarticulatory influences along the front-back dimension due to neighboring vowels (and possibly to consonants as well), and influences on the height dimension due either to the closed jaw position of neighboring consonants or else to a disposition for the jaw to return to a rest position.

TABLE 4

	Stressed			Unstressed		
	I	M	F	I	M	F
S	1178	1112	989	1144	803	850
F	1052	986	942	972	794	859
N	877	800	725	794	701	705

Table 4. The interaction of stress and position on F1 of syllables in initial (I), medial (M), and final (F) position in trisyllables.

A final interesting finding was of a "position effect" on opening for /a/ across a word. Stressed /a/s were progressively less open in later syllables of words. One hypothesis we have entertained to account for the effect (see Table 4) is that it is an utterance-level (as opposed to word-level) phenomenon that is analogous in some ways to declination in fundamental frequency. That is, just as (other things equal), fundamental frequency declines over the course of an utterance, largely following the decline in subglottal pressure (e.g. Gelfer, Harris, Collier and Baer /6/),so do excursions of the jaw from its rest position decline. Both declination and our position effect, then, might reflect an articulatory system that in some sense "winds up" at the beginning of an utterance and then "runs down" gradually as the utterance proceeds. Perhaps compatible with this view is a weak tendency for our talkers' productions of stressed /i/ and /u/, two closed vowels, to open increasingly across the syllables of a bisyllable or trisyllable.

REFERENCES

/1/ P.M. Bertinetto,"Ancora sull'Italiano come lingua ad isocronia sillabica". Studi linguistici in onore di G.B. Pellegrini, Pisa, 1981.
/2/ D. Bolinger, Intonation and its parts, London,1986
/3/ E. Farnetani, S. Kori"Lexical stress in spoken sentences....". Quaderni del Centro Studio Ricerche Fonetica, I, Padova, 1982, 106-133.
/4/ E. Farnetani, S.Kori "Effects of syllable and word structure on segmental durations in spoken Italian". Quaderni..., III, Padova,1984, 143-187
/5/ C.A.Fowler, "A relation between coarticulation and compensatory shortening". Phonetica, 38, 1981, 35-50.
/6/ C. Gelfer, K. Harris, R.Collier, T.Baer, "Is declination actively controlled?". I.Titze, ed., Vocal Fold Physiology:...,Iowa City, IA, 1985.
/7/ T. de Graaf, F. Koopmans-van Beinum, "Vowel contrast reduction in terms of acoustic system contrast in various languages". Proc. Institute of Phonetics. Amsterdam University, 1984, 41-52.
/8/ F. Koopmans-van Beinum, "Systematics in vowel systems". M.van den Broecke, V. van Heuven, W. Zonneweld, eds., Sound structures...Dordrecht,1984.
/9/ I. Maddieson, "Phonetic cues to syllabification". V. Fromkin, ed., Phonetic Linguistics...., Orlando, 1985.
/10/M. Nespor, I.Vogel, "Clash avoidance in Italian", L.I., 11, 1979, 467-482.
/11/E.den Os,"Vowel reduction in Italian and Dutch", Phonetica, 42, 1985.
/12/E. Selkirk,"The role of prosodic categories in English word stress". L.I , 1980, 563-605.
/13/M. Vayra, C.Avesani, C. Fowler, "Patterns..." M. van den Broecke, A.Cohen, eds., Proc. Xth Int. Congr. Phon. Scien., Dordrecht, 1984, 541-546.
/14/M. Vayra, "Effects transyllabiques de coarticulation... ", XIVè J.E.P., G.A.L.F., Paris, 1985.
/15/M. Vayra, C.A. Fowler, C. Avesani, "Word level coarticulation and shortening in Italian and English speech". Studi di Grammatica Italiana, in press. Also, Haskins Laboratories Status Report, in press.

THE EFFECT OF SYLLABLE STRUCTURE ON VOWEL DURATION

A.C.M. Rietveld

Institute for Phonetics
University Nijmegen
The Netherlands

U.H. Frauenfelder

Max Planck Institut für
Psycholinguistik, Nijmegen
The Netherlands

ABSTRACT

This production study investigated the influence upon vowel duration of syllable structure and the postvocalic consonant. The results obtained showed a differential effect of syllable structure on the measured vowel durations as a function of the postvocalic consonant. The hypothesis that the amount of coarticulation between this consonant and the preceding vowel conditioned this effect was partially confirmed by the results

INTRODUCTION

Vowels are elastic segments; they can be compressed and expanded by the influence of a large number of factors, including among others: the number of syllables in the word, their postion in the word and the position of the word in the utterance, the number and type of the surrounding consonants and the location of the syllable boundary [7] [4] [5].

In this contribution we have investigated the sensitivity of vowel durations to two of these factors and their interaction: a) the syllable structure and b) the postvocalic consonant. The effects of these factors have been analyzed separately in previous research, but their interaction has not received much attention.

The effect of syllable structure on the duration of vowels is well known. Vowel shortening in closed syllables (recently been called Closed Syllable Vowel Shortening (CSVS) by Maddieson [5] and the influence of the number of syllables can both be seen as a tendency towards isochrony. An increase in the number of segments in a syllable and in the number of syllables in a word tends to shorten the segments and syllables involved.
The local environment of the vowel also has an influence on its duration. Thus the postvocalic consonant conditions the duration of the preceding vowels. For example, the feature of voicing of this consonant exerts strong effects upon vowel duration; voiced consonants tend to lengthen the preceding vowel, whereas voiceless consonants have a shortening effect. Another phenomenon may also influence the duration of vowels: the amount of coarticulation. Fowler [1] pointed out that coarticulation may reflect itself in the shortening of the segment that undergoes the effect of coarticulation.

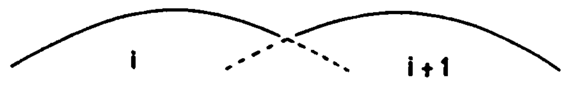


Figure 1 Schematic representation of coarticulation

In figure 1 segment i+1 coarticulates with segment i. Consequently, its acoustical manifestation will emerge during the articulation of segment i; thereby shortening the segment i. Seen in this perspective, when a segment coarticulates with the preceding segment, it tends, all other things being equal, to shorten the latter segment.

The purpose of this paper is to examine the relationship between the above mentioned closed syllable vowel shortening effect and co-articulation. In particular, we want to explore the possiblity that vowel duration varies as a function of the strength of coarticulation which in turn is a function of syllabic structure. According to our hypothesis, the measured vowel durations will depend upon the syllable structure, the strength of coarticulation with the postvocalic consonant and the interaction between these two factors.
We will first briefly examine the relationship between coarticulation and syllable structure. Many experiments have been carried out with the aim of investigating the factors that affect coarticulation. Supporting as well as disconfirming evidence has been found for syllable based models of coarticulation (cf. Sharf & Ohde [9]). If the syllable plays a role in articulation programming - as many researchers think it does (cf. Fujimura & Lovins [2]) - we can expect that a tautosyllabic consonant (i.e. vowel followed by a consonant in the same syllable) will have a stronger coarticulatory effect than a heterosyllabic one (i.e. vowel and following consonant in different syllables). In this case, we may predict that a tautosyllabic consonant like the /r/ in a Dutch word like "peer$den" will coarticulate more strongly than the heterosyllabic /r/ in "pe$ren". As a consequence the vowel /e/ preceding the /r/ in "peerden" will be shorter than that in "peren".

While the Closed Syllable Vowel Shortening hypothesis makes the same prediction, it does not differentiate between the shortening effects as a function of the type of postvocalic consonant. Since /s/ is known to coarticulate much less than /r/ [8] [3], we would expect the position of the syllable boundary around vowel-s-sequences to affect the vowel duration to a smaller degree. In other words, following this reasoning, the amount of closed syllable vowel shortening could, at least in part, depend on the strength of coarticulation between the consonant and preceding vowel.

In the experiment to be reported here we will investigate this hypothesis using duration measurements of vowels preceding four consonants: /s/, /l/ , /r/ and /m/. We may assume that these consonants do not coarticulate to the same extent with the preceding vowel (cf. Sharf & Ohde [9], Klaassen-Don [3]). Klaassen-Don carried out some experiments in which she investigated the identification of consonants on the basis of vowel transitions in VC· and CV-sequences. On the basis of the identification scores she obtained we have scaled (from 0 to 1) the consonants under scrutiny: /s/ and /m/ have a value of 0.05 (small coarticulatory effect), /l/ a value of 0.55, and /r/ 0.80.
To summarize, our line of reasoning is the following:

Assumption I: coarticulation and shortening are positively related (cf. Fowler [1]).

Assumption II: the consonants /s/, /m/, /l/ and /r/ show increasing coarticulatory effects.

Hypothesis: the location of the syllable boundary conditions the amount of coarticulation between the consonant and the preceding vowel: tautosyllabic consonants coarticulate to a greater extent than heterosyllabic ones.

Prediction: the syllable boundaries around /l/ and /r/ have stronger effects on the duration of the preceding vowel than the boundaries around /s/ and /m/.

Subsidiary prediction: As the overlap between segments i and i+1 as shown in Figure 1 increases, the total duration of and i+1 should decrease. If this is true, and again if the syllable structure determines the amount of coarticulation, one would expect the same pattern of differences in VC·duration as that obtained for the vowel duration for the four consonants used. In other words, we could expect increasing differences between the VC duration for hetero and tautosyllabic consonants: /s/ differences < /m/ differences < /l/ differences < /r/ differences.

EXPERIMENT

speech material, procedures

The consonants to be used were /s/, /l/, /m/ and /r/; the vowels were /a/ and /o/.
Eight pairs of bisyllabic nonsense words were constructed, the first member of each pair having a tautosyllabic consonant, the second having a heterosyllabic consonant.

Tautosyllabic		Heterosyllabic Consonant
peer$de	-	pe$ren
poor$de	-	po$ren
peel$de	-	pe$len
pool$de	-	po$len
peem$de	-	pe$men
poom$de	-	po$men
pees$de	-	pe$sen
poos$de	-	po$sen

These word pairs were embedded in a Dutch carrier sentence of the form: "jij moet -------- zeggen" (you should say --------). The carrier sentences and target nonwords were read aloud by 10 Dutch speakers (7 male and three female); each speaker repeated each sentence two times. A number of filler sentences were included in the list. In order to prevent speakers from voicing the /s/ by assimilating it with the following /d/, they were instructed to pronounce /s/ and not /z/; the realization of this instruction was confirmed by auditive control and inspection of the waveform. No instructions were given for the reading tempo. The speech material was recorded in a professional studio, with a tape speed of 19 cm/sec.

Measurements

The duration measurements of the vowels were carried out by means of a speech editing system, which allows visual and auditory segmentation. To that end the target words were digitized (sample frequency 10 kHz) and their waveform displayed on a high resolution screen. Generally the segmentation did not present great difficulties. Changes in the amplitude envelope or variations in the waveform, together with auditory cues, were the main criteria for segmentation and measurements. As each subject produced each target word two times, the total number of vowel durations to be measured was : 10 (speakers) x 16 (words) x 2 (repetitions) = 320.

RESULTS

In figure 2 we present the durations of the vowels, pooled over repetitions, vowel type and subjects.
Both main effects were significant at the 0.05 level. Syllable structure: $F(1,9) = 31.27$, Consonants: $F(3,27) = 11.50$.
Figure 2 shows a clear interaction between syllable boundary location and postvocalic consonant; an analysis of variance, carried out on the mean durations of repeated realizations, resulted in a significant interaction: $F(3,27) = 14.02$ ($p < 0.05$). Shifting the syllable boundary to the right of /m/ and /r/ shortens the pre-consonant vowel dramatically, whereas this effect is much smaller for /l/ and not existent (not significant at the 0.05 level) for /s/. The respective F-values of the post-hoc comparisons between the two syllabic conditions for /m/, /r/, /l/ and /s/ were: 23.99, 17.44, 5.67 and 1.10, the latter being not significant ($p> 0.05$, df1=1, df2=27).
The rank order of magnitudes of the effects for /s/, /l/ and /r/ is fully in line with the strength

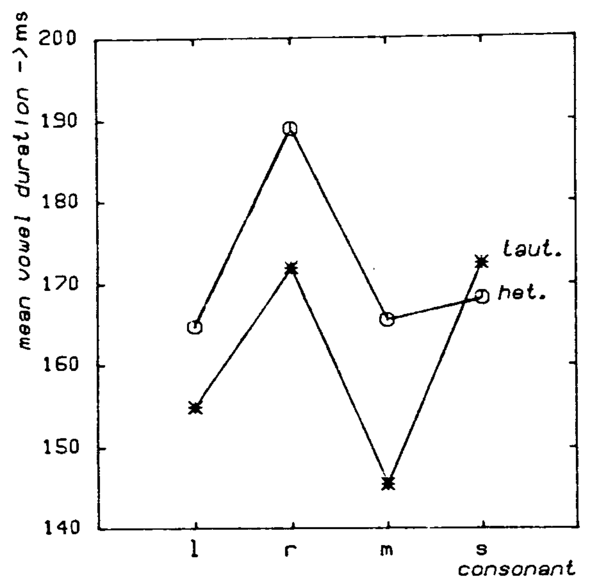


Fig. 2 Mean durations of vowels as a function of the following consonant and the location of the syllable boundary

of coarticulatory effects, as assessed by Klaassen-Don [3]: the effect of /m/ is an exception we will discuss later.
Thus, all three effects under focus in this experiment, viz. syllable structure, consonant type and their interaction were found to have significant effects on the duration of preconsonantal vowels.

We also measured the total duration of the vowel-consonant sequences.

Figure 3 shows the durations of the VC-sequences, pooled over vowels, repetitions and subjects. Two main effects are significant at the 0.05 level: consonant (F(3,27) = 15,27) and vowel (F(1,9) = 5.52). Here too, there is a significant interaction between the factors boundary location and type of consonant: F(3,27) = 20.97.

DISCUSSION

We have investigated the effects of two factors: syllable boundaries and postvocalic consonant on vowel duration. The results of our experiment show main effects of both factors. Preconsonantal vowels are shorter in closed than in open syllables. These results are consistent with the closed vowel shortening hypothesis mentioned in the introduction. However, an interaction between the syllable structure and the postvocalic consonant was also observed. The size of the difference in vowel duration between the two types of syllable stucture (hetero and tautosyllabic consonants) was not constant for the four types of consonants examined. Indeed, at last in three of the four consonants observed the size of this difference corresponded to the amount of coarticulation expected between the /l,r,s/, based on

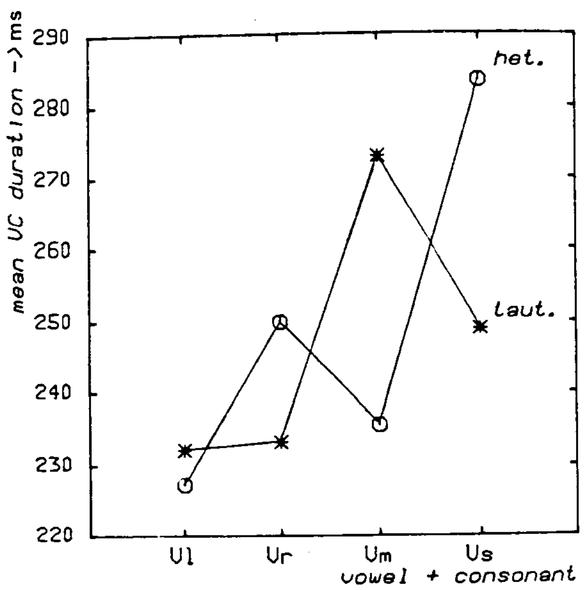


Fig. 3 Mean durations of VC-sequences as a function of consonant type and syllable boundary

Klaassen-Don's [3] results. Only the durations of the vowels before /m/ did not follow the expected pattern, since despite its low coarticulatory measure found by Klaasen-Don [3], large syllable structure effects were obtained. Our hypothesis on the relationship between coarticulatory strength and the effect of syllable structure on vowel duration clearly does not tell the whole story. Further research is needed to identify other factors also playing a role in the determination of vowel duration.
The same can said for the VC durations we measured. Other factors may play a role, and obscure tendencies as they are not equal in their effects for the different consonants involved. These factors are (among others): the lengthening of /s/ when it is syllable initial, the influence of the following consonant in the tautosyllabic condition (cf. Umeda [10]), the lengthening effect of stress, etc. We may, therefore, not be surprised to see that the above mentioned expectation on the basis of vowel consonant overlap is not confirmed by the data given in figure 3.

CONCLUSION

The results obtained in our experiment suggest the two following tentative conclusions: coarticulation and timing phenomena are related, and coarticulation is sensitive to linguistic structure like syllable boundaries. These conclusions were derived from our observation that syllable structure has a differential effect upon vowel duration depending upon the properties of the postvocalic consonant.

Our results show that to arrive at a proper characterization of the acoustic properties of speech, we cannot view speech simply as a linear concatenation of phonetic segments, but we must take into account its linguistic (i.e. syllable) structure. Research in the perceptual domain has also revealed the importance of linguistic structure in determining subject's perception performance. For example, in a phoneme monitoring study [6], French subjects showed a preference for syllabic segmentation. When presented CV or CVC targets (like /ba/ or /bal/) to detect in words whose initial syllable was this CV or CVC (like 'ba$lance' or 'bal$con') they reacted quicker when the syllable structure of the target matched that of the target-bearing word (like /ba/ in 'ba$lance').

This production study represents a first step in identifying the acoustic cues supporting decisions about the identity of segments and syllables. We found that syllable structure influenced vowel duration to varying degrees depending upon the postvocal consonant. The variability in the syllabic inluences could have interesting consequences for studies in speech perception. In particular, since the syllable structure effect in French has only been tested with one class of consonants (liquids), it is important to establish whether this effect generalizes to other types of postvocalic consonants and syllables or whether it depends specifically upon the strong allophonic character of the vowels and liquids used. We are currently conducting phoneme monitoring experiments in French with the aim of determining the role of syllable structure in language perception.

BIBLIOGRAPHY

[1] Fowler, C.A. A relationship between coarticulation and compensatory shortening. Phonetica, 1981, 38, 35-50.

[2] Fujimura, O. & Lovins, J.B. Syllables as concatenative phonetic units, in: Syllables and segments, A. Bell & J.B. Cooper (eds.), North-Holland Pub.Co.pp. 107-130, 1979.

[3] Klaassen-Don, L.E.O. The influence of vowels on the perception of consonants, Diss. Leiden, 1983.

[4] Klatt, D.H. Linguistic uses of segmental duration in English, Acoustic and perceptual evidence, JASA, 1976, 59, 1208-1221.

[5] Maddieson, I. Phonetic cues to syllabification, UCLA Working Papers in Phonetics, 1984, 59, 85-101.

[6] Mehler, J., Dommergues, J.Y., Frauenfelder, U. & Segui, J. The syllables role in speech segmentation. Journal of Verbal Learning and Verbal Behaviour, 1981, 20, 298-305.

[7] Nooteboom, S.G. Production and perception of vowel duration, A study of durational properties of vowels in Dutch, Diss. Utrecht, 1972.

[8] Ohman, S.E.G. Perception of segments of VCCV utterances. Journal of the Acoustical Society of America, 1966, 40, 979-988.

[9] Sharf, D.J. & Ohde, R.N. Coarticulation and articulatory disorders, in: Speech and Language, Vol. 5, N.J. Lass (ed.) New York: Academic Press, 1981, 513-247.

[10] Umeda, N. Consonant duration in American English, JASA, 1977, 61, 846-858.

ARTICULATORY DYNAMIC ORGANIZATION OF WORD PRODUCTION ACCORDING TO CINEMA X-RAY PHOTOGRAPHY DATA (METHODS OF INVESTIGATION AND RESULTS OF APPLICATION)

SKALOZUB L.G.

Dept. of Filology, Laboratory of Experimental Phonetics Kiev State University
Kiev, Ukraine, USSR, 252017

The definition of the dynamic articulatory organization of words according to the cinema X-ray photography data is based on the comparative analysis of regularly resumed micromovements (impulses), superglottal (tongue) and glottal (the pre-laryngal part of the pharynx) ones, in speech organs. The analysis resulted in the definition of modal articulatory indications of the syllable and the word as well as in the consonantal and the vocalic types of articulatory activity determined by them.

It has been proved that the nature of the word entirety can be defined on the segmental level: syllables as segments of CVCV word types unite due to interconnection of modal indications. This makes up the structural nature of the word.

In home linguistics word has been studied both as an entirety and as an intersystematic unit for the last decade.

The entirety of word is understood as the entirety of a syntagmatic unit; the analysis of phonemes and establishment of regularity in their distribution and combination are considered to depend on the entirety of the word and its inherent morphological and syntactical characteristics. Investigations in a number of languages carried out by L.G.Zubkova resulted in a series of universal and typological regularities /1/. Phonemic structure of the word is acknowledged as one of its universal properties. It has been proved that distribution of phonemes, first of all of consonants, and the concrete specificity of their positions, regarded in the unity with their phonetic and morphological properties, is of typological nature.

Finding out regularities of the organization of word entirety in the speech process is one of scantily explored problems of linguistics and psycholinguistics, as well as of general and experimental phonetics. Detection of dynamic characteristic properties of all phenomena in a linguistic system is an indespensable condition in defining the entity of language. In the process of speech activity, and first of all of its production and perception, language exists as a social as well as a dynamic phenomenon, as an individual speech experience of any person speaking a living language, the experience which is to be investigated. /10/

A method of cinema X-ray photography worked out in the Laboratory of Experimental Phonetics in Kiev Shevchenko State University serves to investigate the perceptive and articulatory sound manifestation of the word in its dynamics, and not in statics. /4, 5, 7/

The experimental material presented by cinema X-ray photography enables us to investigate the last link of speech production - the direct process of syllable formation and uniting syllables into words as entireties. The object of research made by means of the cinema X-ray pictures is in fact transient: it reflects results of the production of the syllable as a most complicated component of speech activity, and at the same time, characteristics of sound segments (traditionally called articulatory characteristics) manifest themselves dynamically in it.

A linear succession of cinema X-ray frames obtained at the rate of 48-50 frames per second, was transformed into comparative schemes, which made it possible to observe at the interval of every 20 msec the changes occurring in the superglottal part of the speech apparatus.

The detection of dynamic articulatory tendencies working in the organized entirety of Russian CV syllables in disyllabic words was the main objective of the cinema X-ray photographs presented here.

Methods of processing of the experimental material as well as segmentation of the articulatory continuum were carried out in accordance with the modes described by the author. /6, 7, 8/

The definitions of the notions "articulatory dynamics", "consonantal and vocalic activity types", "modal and qualitative indications"used in this paper are based on the analysis of the micromovements occurring regularly in the superglottal and glottal speech organs, but being differently performed in the articulations of consonants and vowels.Conventional abbreviations TI and PLI mean superglottal and glottal activity accordingly. The superglottal activity manifests itself by impulsive muscular contractions of the tongue (in the form of clasping and unclasping movements). The glottal activity is relized by impulsive downward and upward movements in the prelarynx zone of the pharynx (PLI). The superglottal activity is different with consonants and vowels: with consonants it is intermittant and frequently resumed, which results in the development and the establishment of a certain form of the tongue (the consonant activity type); with vowels it is respectively continuous and leads to the muscular contraction of the whole tongue, and consequently to the increase of the superglottal resonator (the vocalic activity type). The glottal activity (PLI) is realized uninterruptedly from the beginning of the segment (the syllable, the word) and up to its completion. /6, 9/

The rate of downward and upward movements in the pre-laryngal part of the pharynx (PLI) is lower with consonants and higher with vowels. The rate of movements (contractions) of the tongue (TI) at the beginning of a CV-syllable is higher during the consonant and lower during the vowel articulation.

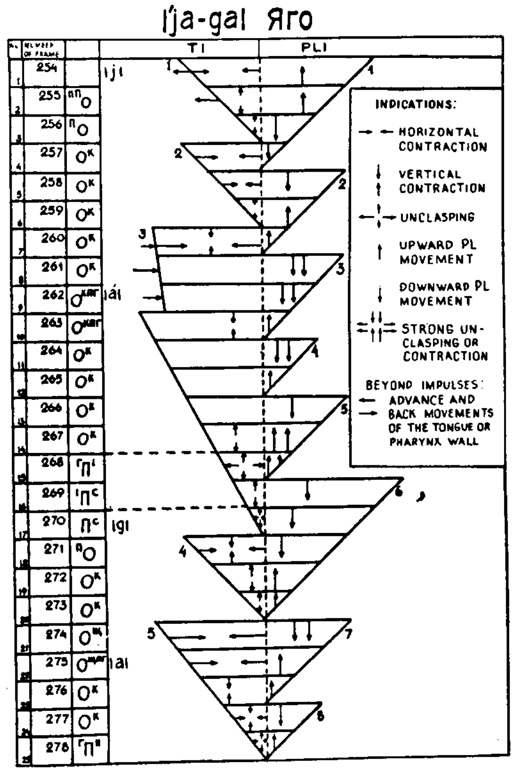
The correlation of such parametres of the tongue (TI) and the pre-laryngal impulses (PLI) as the rate of development, duration, amplitude, synchronization - nonsynchronization of matching have been basic ones for the definition of a new notion - that of modal articulatory indications of syllables and words. It's according to these indications that the consonantal and the vocalic types of articulatory activity are discerned. /7/

The superglottal activity (TI) is responsible mainly for qualitative, and the glottal activity (PLI) for modal vocalic indications of the syllable and the word as entireties. The notion of articulatory dynamics implies first of all modes of development of articulatory efforts, modal indications. /6, 9/

The articulatory dynamics of words of the ʼCVCV and CVʼCV types was analysed and described by comparing the modes the consonantal and the vocalic activities are realized. The results of the analysis are illustrated with the words /ʼjaga/ and /jiʼga/.

According to the obtained data, the CVCV model is characterized 1) by articulatory tension within the first syllable, which manifests itself both in the contrastive dynamics of its components and in the greater duration of the syllable peak (the vowel); 2) by relative separateness of the dynamics of the Ist and the 2d syllables, which is evident from the contrast of interruptedness - uninterruptedness of the TI cycles and from the rate difference of PLI at the syllable boundary.

Though vocalic activity comes to its maximum at the end of the first syllable, the articulation sonority grows towards the end of the word: from interruptedness to greater uninterruptedness of TI, from the greater to the lower uninterruptedness of PLI. The continuity of articulatory dynamics in the word of the CVCV type manifests itself in two ways and can be represented 1) by the TI line in which the growing rate of renewing TI is followed by deceleration and becomes minimal at the end of the first syllable, while in the second syllable the transition from the higher rate to the lower rate of renewing TI is shortened; 2) by the PLI line where the higher rate (interruptedness) in the first syllable is followed by the lower rate and then, in the second syllable, the renewed impulses grow more frequent. Both lines correlate in rate indications. A mutual compensation of the two growing amplification types seems to take place. This makes the articulatory dynamics of the word uninterrupted and entire (Fig.1)



Consequently, the CVCV word type is one of the existing organizations of the uninterrupted integrity of the word. The peak of sonority is next to the second syllable. This syllable completes the growth of amplification developed within the word, making up its high-rated shortened final phase. The end of the first syllable, the moment of its peak is accompanied by the growth of amplitude of PLI and their greater uninterruptedness, which matches the uninterruptedness of TI. There is every reason to regard these facts as manifestation of the growing tension, as indications incarnating the unity of syllables as well as stress in the word of the CVCV type. Maximum of activity, both superglottal (tongue activity) and glottal, which is prosodic by nature, are indispensable to the realization of stress.,

In the CVCV word type the unity of syllables is attained due to a special correlation of the development of TI and PLI.

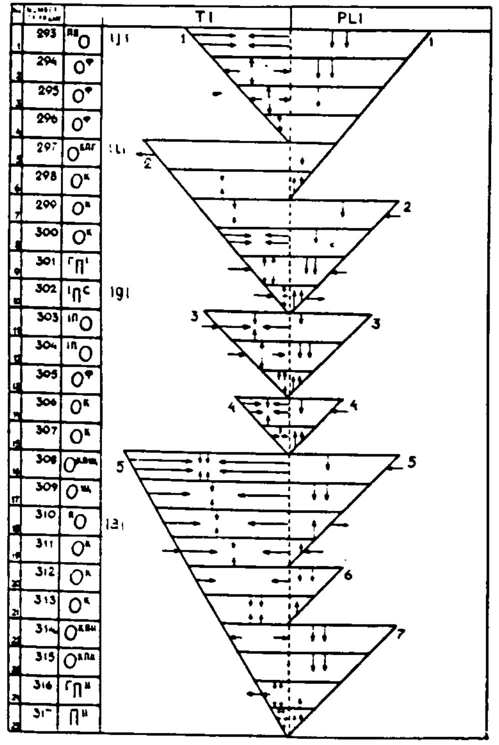
The integrity, the unity of the components in the first syllable manifests itself in that a lower rate and duration of TI matches with a greater duration and lower rate of the development of PLI in the initial part of the syllable, changing further into some other correlation in which the greater uninterruptedness (duration) of TI matches the greater interruptedness (brevity) of PLI. Such development favours the initial part of the syllable to become a vocalic threshold of the word peak, owing to a synchronic development of TI and PLI, their rate increasing.

The correlation of the first and the second syllables in the CVCV word type is of a peculiar type.

The organization model of articulatory dynamics in the CVCV word type resembles the impulse developing according to the scheme of rising sonority: a harmonious vocalic beginning, when the first syllable conditions the development of a strong consonantal-vocalic peak completed by the final vocalic segment of the word - .the vowel.

The amplification of the peak, i.e. the prominence of the stressed syllable in the CVCV word type is attained 1) due to a contrastive, mutually compensating relation of syllables according to their modal indications: the connection of the growing tension of the tongue (TI) and the pre-larynx (PLI) in the first syllable leads to a more rapid development of the intersyllabic transition. It also promotes the development of modal indications in the second syllable, which seem to compensate the absence of contrasts and the relative equivalence of TI and PLI in the first syllable (Fig.2); 2) owing to a synchronic development of TI and PLI at the head of the second syllable and to the growing rate of development of these impulses connected with synchronization; 3) due to further nonsynchronous yet contrastive development of TI and PLI, as compared with the final syllabic segment; a maximum of vocalic intensification of TI is observed, which means a continuous and the most durative in the word process. Thus the vocalic component of the syllable increases as well as the vocalic completion of the word; 4) owing to the growing number of amplifications of PLI (three PLI's correlate with a single TI which unites the end of the consonant and the vowel).

|ji-'gal ЯГА



The growth of amplification and duration of the components in the second syllable is caused by different reasons: the consonant grows longer owing to synchronous actions of the superglottal and glottal amplifications and their speeding-up before they unite with the vowel; duration and sonority of the vowel increase in consequence of lengthening the impulse of the tongue activity (the phase of the tongue contraction becomes relatively uninterrupted) and a simultaneously growing frequency of the pre-laryngal impulsation.

In the word of the ĆVCV́ type the unstressed syllable differs by a relative uninterruptedness of TI and PLI. Something like their mutual "seisure" takes place: the beginning of PLI matches with the end of TI. The beginning of the syllable is notable for a greater rate of TI and a lower rate of PLI; their end is distinguished by the reverse correlation. Here the consonant shares with the vowel the function of syllable formation to a greater extent than the consonant of the second syllable. Consequently, the beginning of the word is realized as growth of vocalic activity. Further, continuity changes into interruptedness, that is into a synchronous inclusion of TI and PLI going at a higher rate. The second segmentof the word, which involves the beginning of the consonant and the moment of its joining the vowel, makes up the beginning of the peak of the word organization. The final word segment is represented by a vowel, which dynamics differs by the combination of maximum continuity of TI (the maximum lengthening of the contraction phase of the last TI in the word) and a simultaneous maximum interruptedness (frequency) of the PLI in the word. This results in the amplification of the vowel, that is of the end of the second syllable, and accordingly in the amplification of vocalic characteristic and in the sonority of the word end. Within the word of the CVCV type, as well as in the syllable, acts a dynamic model, i.e. an impulse made up by the following scheme of matching superglottal (tongue) and glottal (prelaryngal)amplification(Fig.2).

Thus pre-laryngal amplifications play the leading part in making up the integrity of the word: its common ascending line (the first stage of development) going from higher to lower interruptedness, that is from lower to higher frequency of occurance, unites the initial syllable and the consonant of the second syllable. Their following resumption adjoins the vowel to the consonant of the second syllable, making it more durative and sonorous. The sonority of the consonant in the second syllable depends on the first ascending line of the development of PLI. Thus voicing of consonants in the Russian speech is an indication stipulated by the organization of word as entirety.

The second stage of the resumed amplifications of PLI is marked by a lower rate and a greater amplitude of impulses. This stage serves to complete the prominence of the end of the word.

All said above enables us to maintain that the articulatory organization of the word as an entirety (the CVCV́ and CV́CV models) can be defined (at the segmental level) as a phenomenon realized in the correlation of modal indications of syllables; within syllables it is realized in the correlation of consonantal and vocalic activity. The unity of syllable components as well as syllables depends on modal indications.

The analysis which has been carried out gives reason to maintain that the described models (schemes) of word articulation are originated and produced as entireties. These entireties have their own maxima, and the shapes of their development before and after the maxima are not identical. The segments of each model are mutually conditioned. Thus isomorphic character of both models manifests itself. There is every reason to speak, on the one hand, about isomorphism of the articulatory formation of syllables within words, and, on the other hand, about the organization of words.

The segment that embodies the articulatory maximum is most likely the defining segment in each model. The models 'CVCV and CV'CV differ in the way of their modal organization and position of that segment.

Changes in the rate of glottal and superglottal impulsation and ways of their matching are the crucial mechanism of realization of the segmental organization of the word, the dynamic and articulatory structure of the word, as a unit of speech production.

REFERENCE

I. Бондарко Л.В. Структура слога и характеристика фонем. - Вопросы языкознания, 1967, № I, с.45.

2. Зубкова Л.Г. Сегментная организация слова. - М., 1977, с.7-28.

3. Зубкова Л.Г. Фразовые признаки сегментной организации слова в свете универсальных закономерностей речеобразования. - В кн.: Фонетические единицы речи. Сб.науч.трудов. - М., 1982, с.I00-109.

4. Лийв Г., Эак А. О проблемах экспериментального изучения динамики речеобразования: комплексная методика синхронизированного кинофлуографирования и спектрографирования речи. - Изв. АН Эст. ССР. - 1968. Т.I7. Биология, с.78-I02.

5. Прокопова Л.И., Родзаевский А.П., Тоцкая Н.И. Применение рентгенокинематографий при изучении речевых артикуляций. - Журнал ушных, носовых и горловых болезней. - Киев, 1964, № 3, с.80-89.

6. Скалозуб Л.Г., Хоменко Л.М. Артикуляторная динамика слова (об артикуляторном выражении ударения). - Рус.языкознание, 1986. Вып.I2, с.I25-I33.

7. Скалозуб Л.Г. Динамика звукообразования (по данным кинорентгенографирования). - Киев, 1979. - I3I с.

8. Скалозуб Л.Г. Артикуляторная динамика речеобразования. - Автореф.дис. ... докт.филол.наук. Киев, 1980. - 44 с.

9. Скалозуб Л.Г. Артикуляторная динамика слогообразования. - В кн.: Экспериментально-фонетический анализ речи.Проблемы и методы. Вып.I. Л., 1984, с.28-29.

I0. Щерба Л.В. Языковая система и речевая деятельность. - Л., 1974. - 427 с.

ACOUSTIC VARIATION AND TYPES OF PALATALIZATION

ARVI SEPP

Dept. of Dialectology
Institute of Language and Literature
Tallinn, Estonia, USSR 200106

ABSTRACT

Dialectal variation in Estonian consonant palatalization (as a secondary articulation) can be accounted for in terms of location of the maximum effect of palatalization in the time dimension ("prepalatalized" vs "postpalatalized"). The same acoustic property can be used to describe cross-linguistic variation in palatalization. The effect of palatalization is manifested mainly in rise of the frequency of the second formant.

1. Specification of the Feature "Palatalized"

Palatalization in the sense of a secondary articulation (International Phonetic Alphabet (IPA): s̡ t̡ n̡ l̡ and other consonants /1: 13/) is approached in this paper.

Palatalization has been treated as a single feature both in the IPA transcription system and in the distinctive feature system of Jakobson, Fant, and Halle (the feature Sharp) /2: 31/. In later distinctive feature systems palatalization has been defined as a particular feature combination: [+high, -back] in Chomsky and Halle /3: 306/ or High, Front in Ladefoged /4: 80/.

Jakobson, Fant, and Halle propose acoustic correlates to their distinctive features. The feature Sharp, by contrast with the feature Plain, "manifests itself in a slight rise of the second formant and, to some degree, also of the higher formants" /2: 31/. Hence, the two features are relational, based on comparison of the palatalized (or sharp) consonants with the nonpalatalized (or plain) consonants, everything else (context, speaker) being equal. The feature Sharp is defined as inherent (vs prosodic), without any reference to sequence. "No comparison of two points in a time series is involved" /2: 13/.

Fall of the frequency of F1 (formant one) can compensate for F2 rise in perception /5: 6/. Ladefoged /4: 75/ also uses frequency of F1 to specify his acoustic features Height and Backness which together define palatalization in his distinctive feature system. See also /6/ and /7/ where concrete lists of simple physical parameters and their usual lack of one-to-one relation with linguistic categories (features) have been presented. Measurements /8: 220/ /9: 150/ /10: 61/ /11: 3/ /5: 3/ /12: 10/ have shown that the frequency of F2 is indeed the main (the sufficient) acoustic parameter, whose values differentiate palatalized consonants from nonpalatalized ones. However, contextual and inter-speaker variation has not been satisfactorily accounted for yet.

Contextual variation has been considered neglectable. In comparison with values for nonpalatalized consonants, F2 frequency values for palatalized consonants, measured at the terminal point (beginning or end) of a vowel formant transition, are less variable over

(1) different vowel environments
(2) different consonants of the same "point of articulation"
(3) consonants of different "point of articulation" series (labials, dentals, velars) /8: 223/.

Inter-speaker variation. Two separate threshold values have been proposed for male vs female speakers. In Russian, the crossover points of F2 distributions for palatalized and nonpalatalized consonants occurred at 1700 Hz for males, and at 1900-2000 Hz for females. (Measurements were taken within the consonant or at the beginning of a transition from the consonant to the vowel.) /11: 4/.

Relational values differ less than absolute values for high and low voices. E.g. the percentage by which F2 frequency of palatalized consonants exceeds that of nonpalatalized consonants of the same speaker was found to be approximately 30% in Estonian /9: 146/. Although such values (involving comparison with nonpalatalized consonants) show less inter-speaker variation, they are more context variable than the absolute values (because of the contextual variation of nonpalatalized consonants). Normalization for F3 frequencies has been attempted in some cases /11: 4/.

Palatalized and palatal. It is not clear wether the acoustic difference between palatalized and palatal consonants (IPA: c ʎ ɲ ɟ ç j) is that of degree or whether any new acoustic parameter is involved. The two sets of consonants rarely contrast within one language. University of California, Los Angeles, Phonological Segment Inventory Database shows palalized dental/alveolars and palatals of the same manner class contrasting in Irish: n̡ and ɲ /13/. It has been proposed to distinguish in IPA three degrees of palatalization /14/, e.g. ɲ n̡ and (Estonian) ń.

2. Accounting for Dialectal and Cross-Linguistic Variation

Dialectal (or cross-language) comparison may complicate acoustic descriptions by showing consistent differences, not accountable for in terms of acoustic properties (parameters) used to specify established linguistic features (distinctions). This may point to the lack of cross-linguistic acoustic invariance of the distinctive features (categories) (see discussion of "alveolar" in /15/ /16/). Or it may indicate that distinctive features (categories) cannot account for all audible (noncontrastive) differences /7: 500/.

2.1. Estonian dialects. Both in terms of linguistic distribution of palatalized consonants /17/ and acoustically, Estonian exhibits a variety of palatalization types. Preliminary analysis of spectral characteristics of palatalization in VC(V) (vowel consonant (vowel)) sequences (frequencies of formants were calculated using linear prediction analysis /18/), reveals that palatalization in Estonian dialects varies with respect to the following (acoustic) features:

(1) Location of the maximum effect ("focus") of palatalization in the time dimension.
(2) Characteristics of coarticulation in the sequence VCV in different vowel combinations.

The "focus" is defined either (1) as the point (interval) in which the frequency of F2 of palatalized consonants maximally exceeds that of nonpalatalized consonants (in phonologically (near-)minimal pairs) or (2) as the point of maximum F2 in symmetrical vowel environments. Location of the "focus" on the time axis can be measured towards the left or right of the V-C boundary. An earlier "focus" on the time axis is accompanied by shorter transition to the following vowel. See Fig. 1 for two (extreme) types of palatalization, contrasting in the location of the maximum effect (A) at the transition from the consonant to the following vowel ("postpalatalized") (B) at the transition from the preceding vowel to the consonant ("prepalatalized").

2.2. Russian vs Estonian. In both languages palatalization can distinguish word meaning. Russian v̡es "weight" vs v̡es̡ "entire". Estonian palk+ vs pal̡k "beam". Russian has been the model language for the acoustic study of palatalization. Dentals, labials, labiodentals, and in more restricted environments (not word-finally) also velars palatalize in standard Russian. In standard Estonian, palatalization is limited to the position immediately after the vowel of the primary-stressed syllable, and only alveodentals t s n l can be palatalized. F2 frequency values corresponding to palatalization, assuming a single value per (phonemic) segment, overlap in the two languages /19/. Time location of maximal F2 frequency change is a more stable characteristic of the difference in palatalization between Russian and standard Estonian. The difference can be expressed in terms of the percentage by which the frequency of F2 of palatalized consonants exceeds that of nonpalatalized consonants (a) at the beginning and (b) at the end of the consonants:

	Russian s̡	Estonian alveodentals
(a)	12	24
(b)	42	3

(based on data from /5/ and /9/)

In Russian palatalization is manifested most prominently at the release of the consonant and at the transition to the following vowel, if any vowel follows (although F2 frequency in the preceding V and within C are influenced to a lesser extent). In standard Estonian, the maximum effect appears at the transition from the preceding vowel to the consonant (presence of the preceding V is obligatory, word-initial C does not palatalize). The difference in palatalization between Estonian dialects as well as between standard Estonian and Russian is similar to that, expressed with special features in /13/, nasalized : nasal release (postnasalized); aspirated : preaspirated. A general need for acoustic specifications to be time-varying /7/ /20/ and context-sensitive /21/ or in terms of longer units /22/ has been admitted.

2.3. In summary, one and the same acoustic parameter, the frequency of F2, distinguishes palatalized consonants from nonpalatalized consonants in all known cases (although contextual effects and normalization for high and low voices have not been sufficiently elaborated).

+"wages"

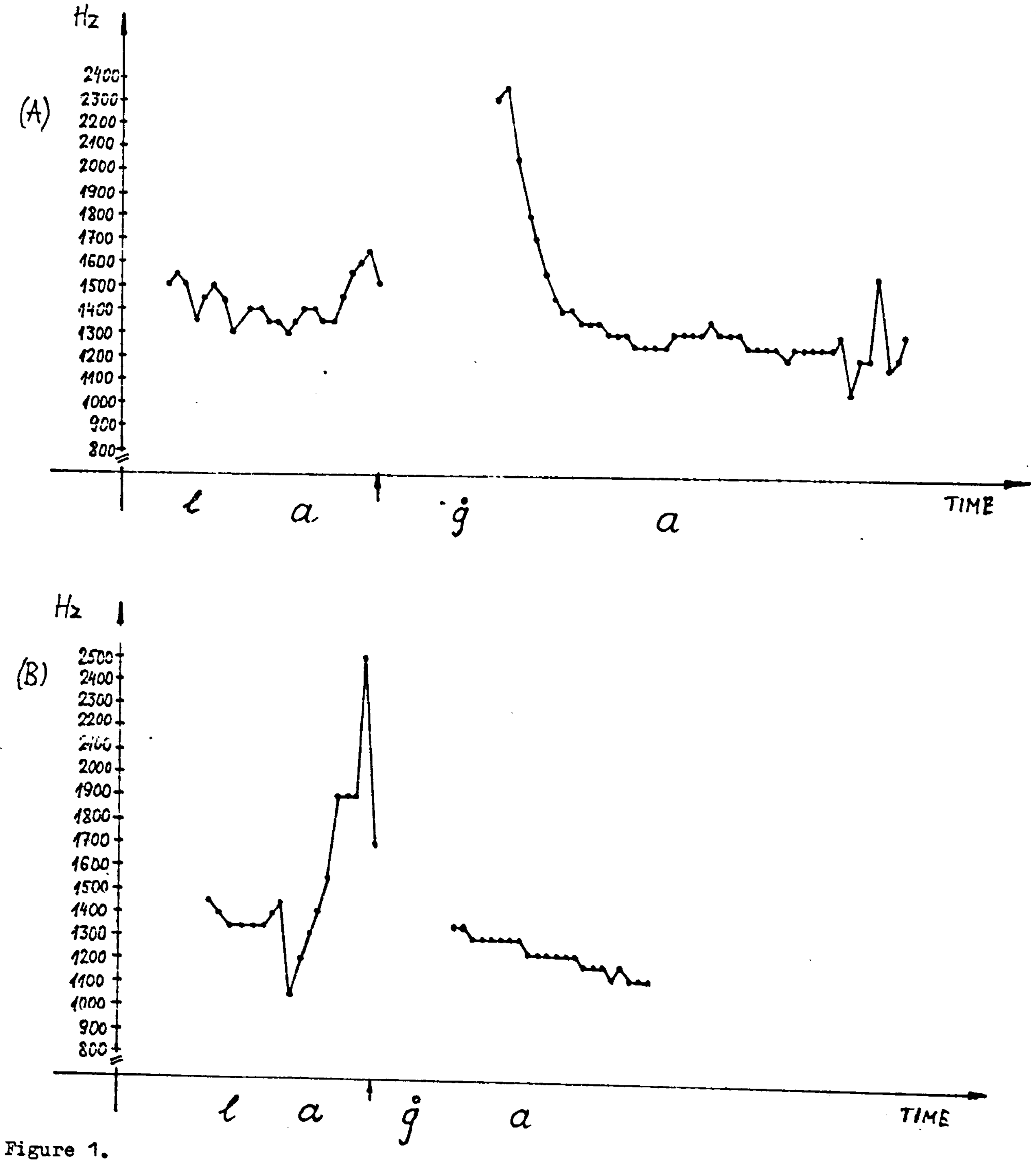


Figure 1.

F2 trajectories of the word lag̊a "broad" in Estonian dialects. Words pronounced in isolation, slow tempo, female speaker in both cases. ↑ points to V-C boundary (the last frame where FO appeared before voiceless eonsonant).
after ten milliseconds.

(A) "postpalatalized" g̊ (Võru dialect)
(B) "prepalatalized" g̊ (Tartu dialect)

To account for dialectal and cross-linguistic (non-contrastive) differences, time-varying values of the same parameter must be considered.

References

/1/ "The Principles of the International Phonetic Association, being a description of the International Phonetic Alphabet and the manner of using it," Department of Phonetics, University College, London, 1949.

/2/ R. Jakobson, G. Fant, M. Halle, "Preliminaries to speech analysis", Cambridge (Massachusetts): Massachusetts Institute of Technology, 1952.

/3/ N. Chomsky, M. Halle, "The sound pattern of English", New York: Prentice-Hall, 1968.

/4/ P. Ladefoged, "Preliminaries to linguistic phonetics", Chicago: University of Chicago Press, 1972.

/5/ M. Derkach, G. Fant, A. de Serpa-Leitão, "Phoneme coarticulation in Russian hard and soft VCV-utterances with voiceless fricatives", Royal Institute of Technology (Stockholm), Speech Transmission Laboratory Quarterly Progress and Status Report 1970 (2-3): 1-7, 1970.

/6/ G. Fant, "Descriptive analysis of the acoustic aspects of speech", Logos 5 (1): 3-7, 1962.

/7/ P. Ladefoged, "What are linguistic sounds made of", Language 56 (3): 485-502, 1980.

/8/ G. Fant, "Acoustic theory of speech production", The Hague: Mouton, 1970.

/9/ I. Lehiste, "Palatalization in Estonian: some acoustic observations", In: Estonian poetry and language. Studies in honor of Ants Oras (editors V. Kõresaar, A. Rannit): 136-162, Stockholm: Vaba Eesti, 1965.

/10/ G. Liiv, "Preliminary remarks on the acoustic cues for palatalization in Estonian", Phonetica 13: 59-64, 1965.

/11/ V. Shupljakov, G. Fant, A. de Serpa-Leitão,"Acoustical features of hard and soft Russian consonants in connected speech: a spectrographic study", Royal Institute of Technology, Speech Transmission Laboratory Quarterly Progress and Status Report 1968 (4): 1-6, 1969.

/12/ A. Eek, "Acoustical description of the Estonian sonorant types", Estonian Papers in Phonetics 1972: 9-37, 1972.

/13/ I. Maddieson, "UPSID: UCLA Phonological Segment Inventory Database: Data and Index", University of California, Los Angeles, Working Papers in Phonetics 53, 1981.

/14/ W. K. Matthews, "Palatalization in Estonian", Le Maître Phonétique, Troisième Série (100): 29-32, 1953.

/15/ P. Ladefoged, Z. Wu, "Places of articulation: an investigation of Pekingese fricatives and affricates", Journal of Phonetics 12: 267-278, 1984

/16/ A. Jongman, S. E. Blumstein, A. Lahiri, "Acoustic properties for dental and alveolar stop consonants: a cross-language study", Journal of Phonetics 13: 235-251, 1985.

/17/ T.-R. Viitso, "Läänemeresoome fonoloogia küsimusi", Tallinn: Eesti NSV Teaduste Akadeemia, 1981.

/18/ M. Mihkla, H. Kaldma, M. Piirmets, "Speech analysis on the basis of a minicomputer", Estonian Papers in Phonetics 1980-1981: 60-65, 1982.

/19/ M. M. Vihman, "Palatalization in Russian and Estonian", Phonology Laboratory, Department of Linguistics, University of California, Berkeley. Project on Linguistic Analysis. Reports Second Series 1: V1-V32, 1967.

/20/ D. Kewley-Port, "Time-varying features as correlates of place of articulation in stop consonants", Journal of the Acoustical Society of America 73 (1): 322-335, 1983.

/21/ K. Suomi, "The vowel-dependence of gross spectral cues to place of articulation of stop consonants in CV syllables", Journal of Phonetics 13: 267-285, 1985.

/22/ O. Fujimura, "Remarks on speech synthesis", In: Abstracts of the Tenth International Congress of Phonetic Sciences: 113-119, Dordrecht: Foris Publications, 1983.

LES INDICES ACOUSTIQUES DU TRAIT DE VOISEMENT DANS LES OCCLUSIVES DU FRANÇAIS PARLÉ À MONTRÉAL

BENOIT JACQUES

Université du Québec à Montréal
Montréal, Québec, Canada
H3C 3P8

Résumé

La distinction des occlusives entre sourdes et sonores lorsqu'elles sont en position initiale de syllabe peut reposer sur plusieurs indices. En français, on est amené à considérer le rôle prépondérant du VOT à cause de la tenue voisée des occlusives sonores. D'autres indices peuvent aussi contribuer à cette distinction; ce sont entre autres la durée de l'intervalle silencieux, la variation F_0 de la voyelle suivante, la pente et la durée des transitions, en particulier celles de T_1.
Nous avons vérifié l'importance de chacun de ces indices sur un vaste échantillon de 1302 consonnes occlusives initiales de syllabe accentuée et inaccentuée prononcées par quatre sujets francophones nés et vivant à Montréal. À partir de mesures de fréquences et de durées faites sur des sonagrammes, des compilations statistiques ont été effectuées. Les résultats montrent que, parmi les indices qui contribuent à la distinction sourde-sonore, le VOT conserve un rôle prépondérant.

La distinction des occlusives entre sourdes et sonores lorsqu'elles sont en position initiale de syllabe peut reposer sur plusieurs indices acoustiques. En français, l'on est amené à considérer comme indice principal le VOT selon la définition qu'en donnent Lisker et Abramson [1], parce que le voisement débute avant la rupture de l'occlusion et est compté en valeurs négatives pour les occlusives sonores, et qu'il débute après la détente ou coïncide avec celle-ci pour les occlusives sourdes; il est alors compté en valeurs positives ou il est égal à zéro.
Divers travaux portant sur le français et d'autres langues, notamment ceux de Fischer-Jørgensen [2], Serniclaes [3], Slis et Cohen [4], Santerre et Suen [5], Jeel [6] ont souligné la contribution d'autres paramètres dans la caractérisation du trait de voisement. Ces paramètres sont la durée de l'intervalle silencieux, la variation de F_0 au début de la voyelle suivante, la pente et la durée des transitions vers la voyelle suivante, notamment celles de T_1.

Toutefois, les données disponibles ont souvent été obtenues à partir d'études effectuées sur des corpus limités quant à leur dimension ou encore à partir d'expériences de synthèse. Notre recherche poursuivait donc un double but: 1° vérifier le rendement du VOT et des autres indices acoustiques dans la distinction de voisement des occlusives dans le français de Montréal, 2° vérifier ce rendement sur un grand corpus compte tenu qu'une consonne donnée n'est jamais prononcée plusieurs fois de la même manière.

MÉTHODOLOGIE

Quatre (4) informateurs de sexe masculin âgés de 20 ans, nés et élevés à Montréal et n'ayant pas complété d'études au-delà du niveau du secondaire ont prononcé 558 phrases de quatre syllabes. Ceci a permis, entre autres, pour les besoins de la présente étude, la production de 1302 consonnes occlusives sourdes et sonores dans les trois positions suivantes:

1- Accentuée et intervocalique (AI), illustrée par /b/ dans l'énoncé «Ta paire de bottes» [tapɛrdabɔt];
2- Inaccentué et intervocalique (II), illustrée par /p/ dans le même énoncé;
3- Initiale absolue de phrase (IA), illustrée par /t/ suivi de /a/ dans le même énoncé.

De plus, comme la dénivellation et la durée de T_1 sont dépendantes de la voyelle adjacente, il fallait qu'il y ait une certaine régularité dans les contextes vocaliques. Aussi, les consonnes étaient-elles suivies de façon systématique des voyelles /i/, /e/, /ɛ/, /A/, /ɔ/, /o/ et /u/.
L'enregistrement des phrases a été effectué dans les conditions les meilleures et on a tiré de chacun un sonagramme à bandes larges et un autre à bandes étroites. Le premier à servi à l'étude du VOT, de la durée de l'intervalle silencieux, ainsi que pour l'étude des T_1 de la voyelle suivante. Le second a servi à l'étude des variations de F_0 au début de la voyelle suivante.
Pour toutes les mesures effectuées sur les tracés à bandes larges et étroites, des analyses multivariées ont été effectuées par ordinateur. Les tableaux montrent les moyennes obtenues (moyennes générales et moyennes par position ou contexte, selon le cas) exprimées en centisecondes ou en Hz, le nombre de consonnes et les écarts-types par rapport aux moyennes.

RÉSULTATS

1. Le VOT

Le tableau 1 montre que le VOT a des valeurs négatives pour les consonnes occlusives sonores et des valeurs positives pour les occlusives sourdes. Les valeurs positives sont faibles dans les cas de /p/; on peut considérer que, pour cette consonne, le début du voisement coïncide avec l'explosion. Par contre, /b/ montre des VOTs négatifs importants, et ce dans les trois positions étudiées.

Tableau 1
VOT des consonnes occlusives

	N	M	S		N	M	S
/p/	**254**	**+1.1**	**2.9**	**/b/**	**201**	**-7.2**	**2.9**
AI	106	+0.47	3.9	AI	50	-8.1	1.62
II	104	+1.77	2.04	II	115	-7.2	3.2
IA	44	+1.02	0.68	IA	36	-6.04	3.2
/t/	**276**	**+3.6**	**2.8**	**/d/**	**236**	**-5.1**	**3.5**
AI	96	+4.1	2.6	AI	76	-5.5	3.9
II	128	+3.2	3.0	II	99	-5.5	3.05
IA	52	+3.5	2.6	IA	61	-4.0	3.7
/k/	**216**	**+3.65**	**1.95**	**/g/**	**141**	**-5.8**	**3.6**
AI	65	+4.3	1.6	AI	75	-6.1	3.5
II	103	+3.4	2.2	II	49	-5.7	3.2
IA	48	+3.3	1.8	IA	17	-5.1	4.9

N = Nombre de consonnes
M = Moyenne
S = Écarts-type

Dans le cas des dentales et vélaires sourdes /t/ et /k/, le VOT montre des valeurs positives, de 3.2 cs. ou plus. Le début du voisement marque donc un retard par rapport à la détente, ce retard étant légèrement plus important lorsque la consonne est dans une position accentuée que lorsqu'elle est en position inaccentuée. Les occlusives sonores /d/ et /g/, pour leur part, ont des VOTs négatifs, mais les valeurs atteintes sont moins importantes que celles montrées par le VOT de /b/ dans des positions comparables. Pour ces deux consonnes, c'est en position initiale absolue que le début du voisement anticipe le moins sur l'explosion.

Par ailleurs, l'examen des écarts-types permet de constater qu'il y a peu de recouvrement dans les ensembles, ce qui signifie que, dans la grande majorité des cas, le VOT pourra être un indice suffisant servant à distinguer les occlusives sourdes des sonores. Toutefois, on doit noter que le VOT a un comportement différent selon le lieu d'articulation des consonnes: aussi, il y a lieu de considérer séparément la paire /p ~ b/ des paires /t ~ d/ et /k ~ g/. Les écarts-types montrent en effet que /p/ peut avoir un VOT négatif, un début de voisement qui anticipe légèrement sur l'explosion. Néanmoins, la distinction /p ~ b/ est sauvegardée la plupart du temps à cause des valeurs négatives importantes du VOT de /b/. Dans le cas des paires /t ~ d/ et /k ~ g/, il y a davantage d'équilibre entre les valeurs positives du VOT des sourdes et les valeurs négatives du VOT des sonores. Des possibilités de neutralisation subsistent toutefois surtout en position initiale absolue.

2. La durée de l'intervalle silencieux (SI)

La durée de l'intervalle silencieux ne peut contribuer à la distinction de voisement que pour les consonnes en position intervocalique. En position initiale, en effet, l'absence de segment précédent empêche de déterminer le début d'un tel intervalle. Le tableau 2 montre que, en moyenne, l'intervalle silencieux des occlusives sourdes excède légèrement en durée celui de leurs homorganiques sonores. Toutefois, exception faite de la paire /p ~ b/ en position accentuée intervocalique, la différence de durée est inférieure à 1 cs. et tous les écarts-types sont supérieurs à cette valeur. Par conséquent, un grand nombre des consonnes sonores de notre échantillon ont un SI plus long que leurs homorganiques sourdes. La durée supérieure du SI des occlusives sourdes est donc une tendance qui se dégage à partir d'un ensemble d'occurrences relativement vaste.

Tableau 2
Durée de la tenue des occlusives (cs.)

	N	M	S		N	M	S
/p/	**210**	**8.9**	**2.14**	**/b/**	**165**	**8.06**	**1.56**
AI	106	9.5	2.5	AI	50	8.2	1.6
II	104	8.4	1.6	II	115	8.02	1.5
/t/	**224**	**7.31**	**2.1**	**/d/**	**175**	**6.49**	**1.53**
AI	96	7.4	2.4	AI	76	6.6	1.5
II	128	7.2	1.8	II	99	6.4	1.56
/k/	**168**	**7.3**	**1.82**	**/g/**	**124**	**6.8**	**1.65**
AI	65	7.44	1.84	AI	75	6.9	1.6
II	103	7.2	1.8	II	49	6.6	1.76

3. La variation de F_0 au début de la voyelle suivante

Le tableau 3 montre la variation en Hz de F_0 au début de la voyelle suivante. Une quantité affectée du signe "plus" (+) indique une pente positive, à savoir que F_0 est plus aigu au point de contact avec la consonne et qu'il baisse ensuite. Dans le cas contraire, la quantité est affectée d'un signe "moins" (-). On peut observer dans ce tableau que le ton fondamental d'une voyelle suivant une occlusive sourde est plus élevé au point de contact avec cette consonne. Ce comportement se manifeste dans les trois positions étudiées. Par ailleurs, si la voyelle suit une occlusive sonore, la variation de F_0 apparaît comme étant très faible. On serait donc amené à conclure que la variation de F_0 constitue un indice important dans la distinction des occlusives entre sourdes et sonores. Toutefois, les écarts-types par rapport aux moyennes indiquent que les ensembles se recouvrent largement, de sorte que plusieurs consonnes, parmi les sourdes et parmi les sonores font varier F_0 de la voyelle suivante dans le sens contraire de celui indiqué par les moyennes.

Si la variation de F_0 peut ainsi contribuer à la distinction de voisement des occlusives, son rendement en tant qu'indice demeure limité, beaucoup de réalisations individuelles allant dans le sens contraire de celui attendu.

Tableau 3
Variation de F_0 au début de la voyelle suivante (±Hz)

	N	M	S		N	M	S
/p/	**293**	**+3.7**	**7.5**	**/b/**	**190**	**+0.15**	**5.8**
AI	98	+2.2	8.2	AI	47	+0.06	6.3
II	74	+5.4	6.8	II	107	+1.1	5.4
IA	41	+4.1	6.8	IA	36	-2.6	6.3
/t/	**246**	**+5.0**	**6.2**	**/d/**	**186**	**+0.66**	**8.5**
AI	86	+4.8	7.5	AI	66	+1.6	12.0
II	99	+6.2	5.2	II	62	-4.5	5.4
IA	51	+3.8	6.3	IA	58	-0.7	5.6
/k/	**185**	**+5.5**	**7.6**	**/g/**	**114**	**+0.5**	**6.2**
AI	57	+4.6	7.8	AI	60	+1.1	6.2
II	80	+7.4	8.4	II	17	+0.43	5.9
IA	48	+4.3	5.7	IA	37	-1.5	6.9

4. La transition T_1 de la voyelle suivante

Le tableau 4 montre pour les consonnes occlusives la dénivellation en Hz et la durée en centisecondes de T_1 selon la voyelle suivante.

Tableau 4
Dénivellation et durée de la T_1 de la voyelle suivante

	Dénivellation (Hz)			Durée (cs.)	
	N	M	S	M	S
/p/	**224**	**12**	**35**	**0.35**	**1.2**
i	33	0	0	0	0
e	14	11.4	29	0.14	0.36
ɛ	36	13	38	0.43	1.13
A	27	33	61	1.05	1.97
ɔ	62	16	39	0.42	1.97
o	21	3.8	17	0.05	0.21
u	31	3	16	0.14	0.8
/b/	**197**	**20**	**47**	**0.48**	**1.08**
i	23	3.4	16.6	0.04	0.2
e	14	2.8	10.6	0.07	0.26
ɛ	34	39	62	0.97	1.5
A	24	45	70	0.87	1.33
ɔ	41	36	50	0.96	1.36
o	26	0	0	0	0
u	35	0	0	0	0
/t/	**248**	**14.6**	**43.3**	**0.39**	**1.18**
i	42	6.6	23	0.24	0.84
e	52	2	17	0.06	0.41
ɛ	29	11	28	0.46	1.18
A	43	47	77	1.2	1.75
ɔ	18	28	52	0.78	1.4
o	24	0	0	0	0
u	40	8	32	0.15	**0.53**
/d/	**193**	**33.4**	**60**	**1.36**	**5.2**
i	44	6	24	1.35	7.6
e	43	17	39	0.40	0.88
ɛ	13	50	75	1.7	1.98
A	38	78	80	3.03	8.05
ɔ	22	65	72	1.65	1.77
o	8	16	45	0.5	1.4
u	25	9.8	29.7	0.34	0.96
/k/	**199**	**27.6**	**58**	**0.8**	**1.5**
i	30	1.3	7.3	0.3	0.18
e	16	5.3	21	0 12	0.5
ɛ	25	82	91	2.4	2.6
A	30	51	76	1.1	1.5
ɔ	51	27	53	0.96	1.5
o	7	5.7	15	0.21	0.56
u	40	8	24	0.27	0.8
/g/	**123**	**54**	**63**	**1.7**	**2.0**
i	16	5	20	0.12	0.5
e	15	17	48	0.46	1.24
ɛ	16	87	65	2.7	2.04
A	22	65	100	2.4	2.09
ɔ	31	106	69	3.2	2.01
o	12	13	36	0.5	1.24
u	11	0	0	0	0

En général, le VOT positif d'une consonne non voisée entraîne une diminution de la dénivellation et de la durée de T_1. Toutefois, on a vu que le VOT de /p/ est bref et l'on peut constater que la T_1 de la voyelle suivante n'est pas vraiment diminuée en durée ou en dénivellation du fait que la consonne soit sourde.

La dénivellation et la durée de T_1 de la voyelle suivante contribuent donc à la distinction entre occlusives sourdes et sonores, surtout dans le cas des consonnes non labiales.

Pour ces consonnes, /t ~ d/ et /k ~ g/, ces paramètres ont un certain rendement à condition que la voyelle qui suit ne soit pas une voyelle dont le F_1 est bas, ce qui est le cas de /i/ et de /u/. Lorsque le contexte s'y prête, la durée de T_1 contribue à la distinction sourde-sonore parce que les transitions les plus longues sont toutes le fait de la consonne sonore. La dénivellation de la transition contribue aussi à cette distinction, toutefois, l'on note des recouvrements d'ensembles qui sont de nature à limiter le rendement de cet indice.

CONCLUSION

Les relevés effectués sur 1302 consonnes occlusives indiquent une tendance générale pour les sourdes à comporter un VOT positif et un intervalle silencieux plus long. De plus, F_0 est plus aigu au début de la voyelle suivante et les T_1 de cette voyelle ont des dénivellations moins fortes et des durées plus brèves. Les occlusives sonores, par contre, ont un VOT négatif et un intervalle silencieux plus bref. De plus, F_0 au début de la voyelle suivante ne varie guère et les T_1 de cette voyelle ont des dénivellations plus fortes et des durées plus longues.

Toutefois, il s'agit de tendances générales et non de comportements systématiques; dans bien des cas, en effet, un indice donné varie dans le sens contraire de celui attendu. On doit également noter l'influence du lieu d'articulation, de l'accent, de la position ainsi que du contexte vocalique.

En tenant compte de ces faits, l'on peut conclure que le VOT, la durée de l'intervalle silencieux, la variation de F_0 au début de la voyelle suivante, la dénivellation et la durée de T_1 de cette même voyelle sont tous des indices acoustiques pouvant contribuer à la distinction des occlusives entre sourdes et sonores lorsqu'elles sont en position initiale de syllabe. Parmi ces indices, même s'il ne peut à lui seul toujours rendre compte de la distinction de voisement, le VOT reste néanmoins prédominant, parce que c'est l'indice qui présente le moins de possibilités de neutralisations.

Références

[1] Leigh Lisker et Arthur Abramson, «A Cross Language Study of Voicing in Initial Stops: Acoustical Measurements», Word, 20, 1964, pp. 384-422.

[2] Eli Fischer-Jørgensen, «Les occlusives françaises et danoises d'un sujet bilingue», Word, 24, 1968, pp. 112-153.

Eli Fischer-Jørgensen, «"PTK" et "BDG" en position intervocalique accentuée», Albert Valdman (dir.), Papers in Linguistics and Phonetics to the Memory of Pierre Delattre, La Haye, Mouton, 1972, pp. 143-200.

[3] W. Serniclaes, «La simultanéité des indices dans la perception du voisement des occlusives», Rapports d'activités de l'Institut de phonétique, Université libre de Bruxelles, 7/2, 1973, pp. 59-67.

W. Serniclaes et P. Bejter, «Différences interlinguistiques dans le traitement perceptif des indices de voisement», Rapports d'activités de l'Institut de phonétique, Université libre de Bruxelles, 12/1, 1977-1978, pp. 83-94.

[4] I. H. Slis et A. Cohen «On the Complex Regulating the Voiced-Voiceless Distinction I, II», Language and Speech, 12, 1969, pp. 80-102 et 137-155.

[5] Laurent Santerre et Ching Yee Suen, «Why Look for a Single Feature to Distinguish Stop Cognates», Journal of Phonetics, 9, 2, 1981, pp. 163-174.

[6] Vivi Jeel, «An Investigation of the Fundamental Frequency of Vowels after Various Danish Consonants, in Particular Stop Consonants», Annual Report of the Institute of Phonetics, University of Copenhagen, 9, 1975, pp. 191-211.

INVARIANT ACOUSTIC CORRELATES FOR PLACE OF ARTICULATION IN CATALAN VOICELESS STOPS

JOAQUIM LLISTERRI

Laboratori de Fonètica
Universitat Autònoma de Barcelona
Bellaterra, Barcelona, Spain

MARTIN WEST

Dept.ofAppliedAcoustics
University of Salford
Salford M5 4WT, UK.

ABSTRACT

The results of an acoustic analysis of Catalan voiceless stops are presented and discussed in terms of the invariant correlates for place of articulation within this class of sounds. Both gross spectral shape- a combination between frequency peaks in the burst and starting frequency of the formant transitions - and temporal parameters - specially VOT - can be considered as invariant acoustic correlates that distinguish between labial, dental and velar place of articulation in Catalan.

1. INTRODUCTION

The theory of acoustic invariance proposed by Stevens and Blumstein claims that a particular phonetic dimension must show invariant properties in the speech signal across all languages [1]. Invariant acoustic correlates for place of articulation have been found so far for English, French and Malayalam stop consonants [2] but there is still a need to carry out research on other languages. Catalan is a Romance language with three non aspirated unvoiced stops contrasting bilabial, dental and velar place of articulation ; its vowel system contains seven vowels in stressed position : [i], [e], [ɛ], [a], [ɔ], [o], [u] and a schwa [ə] appearing only in unstressed contexts. The purpose of this paper is to give the results of an acoustic analysis of Catalan stops, reviewing both frequential and temporal parameters that could provide invariant correlates for place of articulation in this language.

2. ACOUSTIC ANALYSIS

2.1. SUBJECTS AND UTTERANCES

Three male and two femalenative speakers of Catalan from Barcelona were asked to read a list of 24 carrier sentences containing bisyllabic words with initial voiceless stop-vowel groups in stressed position, except for the schwa that was unstressed. The eight Catalan vowels were combined with [p], [t] and [k], giving a total of 120 utterances studied.

The recordings were made in anechoic conditions using a Revox A77 tape recorder and a Sennheiser MD 441N cardioid microphone, placed at constant distance from the mouth.

2.2. METHOD

The recordings were low-pass filtered at 5kHz, sampled at 10 kHz and stored on a PDP-11 computer. 14 coefficient LPC spectra were calculated using a 12.8 ms Hamming window positioned at the consonantal release and automatically moved along the signal in 2.5 ms steps until the steady-state of the vowel was found. Formant frequencies were extracted using an automatic peak-picking program. Temporal information was obtained from manual measurements of digitised oscillograms, and checked for accuracy in the waveform display of a Brüel & Kjær 2033 narrow band analyser. Narrow band spectra were also used to check the accuracy of the steady-state formant frequencies for vowels.

2.3. RESULTS

2.3.1. Release Burst

The release of Catalan voiceless stops is accompanied by a short burst of acoustic energy. Its duration values averaged across all speakers and vowel contexts are given in Table 1:

	Min	Max	Mean	SD
[p]	1.1	11.02	3.71	2.17
[t]	1.88	19.22	6.71	3.67
[k]	4.3	46.4	14.12	8.85

Table 1: Duration values for burst in ms.

It can be observed that burst duration is greater from labial to velar place of articulation; this same trend is observed in the data reported also for Catalan by Martí [3] ([p]: 8.6; [t]: 13.6; [k]: 20.5ms); differences in the absolute values might be explained by the fact that Martí made his measurements on spectrograms. However, this has not been observed for Castilian Spanish : [p]: 15.38; [t]: 15.62; [k]:21.86ms (Poch [4]), a language which shows the same contrasting places of articulation for stops as Catalan.

Frequency values were obtained for the first two prominent peaks (K1 and K2) in the LP spectrum with the window center positioned at the burst onset; they are shown in Table 2:

	K1	K2
[p]	1341.91 (321.32)	2060.64 (413.71)
[t]	1787.59 (250.68)	3027.74 (508.62)
[k]	1868.59 (689.86)	2949.49 (674.43)

Table 2: Burst frequency values and standard deviation (in parentheses) in Hz averaged for the two male speakers across all vowel contexts.

2.3.2. Voice Onset Time

Voice Onset Time values have been measured for all the speakers with the following results:

	Min	Max	Mean	SD
[p]	3.68	19.69	11.75	4.09
[t]	7.7.4	44.54	17.35	7.43
[k]	12.27	59.54	28.37	9.76

Table 3: VOT values in ms averaged across all vowel contexts.

It can be seen from the table above that Catalan voiceless stops show a short time interval between the consonantal release and the starting of voicing. The well known correlation between VOT and place of articulation is maintained. These results agree with those previously given by Julià [5] ([p]: 3; [t]: 16; [k]: 35ms) and Martí [3] ([p]: 10.2; [t]: 16.1; [k]: 26.1ms). They follow the same pattern as those reported for Spanish [6] ([p]: 17.18 ms; [t] : 19.75 ms; [k]: 30.01 ms), Italian [7] ([p] :12 ms; [t]: 17ms; [k] : 30 ms), French [8] ([p]:28.5 - 27.6; [t] 31.4-35.4; [k]: 53-51.7) or Modern Greek [9] ([p]: 9; [t] : 16; [k]: 30) although in some cases absolute values might differ.

2.3.3. Formant Transitions

Formant transitions will be described by means of two parameters:starting frequency and duration from the onset to the steady-state of the vowel. Both have been calculated by examining the successive LP spectra starting at the onset of voicing. The average starting frequencies for two male speakers are given across all vowel contexts in Table 4:

	F1	F2	F3
[p]	338.91 (79.73)	1019.02 (390.51)	2066.08 (400.15)
[t]	351.11 (90.04)	1613.5 (313.96)	2640.2 (236.47)
[k]	358.34 (102.47)	1728.99 (596.88)	2326.74 (288.91)

Table 4: Averaged starting formant frequencies in Hz (standard deviation in parentheses) for two male speakers across vowel contexts.

Both these results and Martí [3] show very similar values for F1 in [p], [t] and [k], lower values for F2 and F3 in labial contexts, and higher values for F3 in contact with the dental stop.

Transition durations are presented in Table 5. The extent of the transition has been taken from the onset of the formant to the steady-state of the vowel. The dental consonant tends to show longer F2 and F3 transitions than the labial, while the velar appears to have longer transitions than the other two stops. A similar trend is observed in Martí [3].

	F1	F2	F3
[p]	14.4 (8.2)	20.8 (4.9)	23.2 (13.6)
[t]	18.8 (8.8)	25.1 (12.1)	19.6 (9.2)
[k]	25.6 (9)	24.8 (6.5)	23.6 (8.9)

Table 5: Mean duration of consonant-vowel transitions in ms (standard deviation in parentheses) averaged across all vowel contexts for two male speakers.

3. DISCUSSION

Acoustic data presented so far will be discussed in terms of its potential role as invariant correlate for place of articulation in Catalan stops.

3.1. FREQUENCY CORRELATES

Stevens and Blumstein [10] suggested that an approximation to the shape of the onset spectrum could be obtained by plotting the frequency of the burst against the difference between F3 and F2 frequency values at the onset of transitions. This is shown in Fig 2:

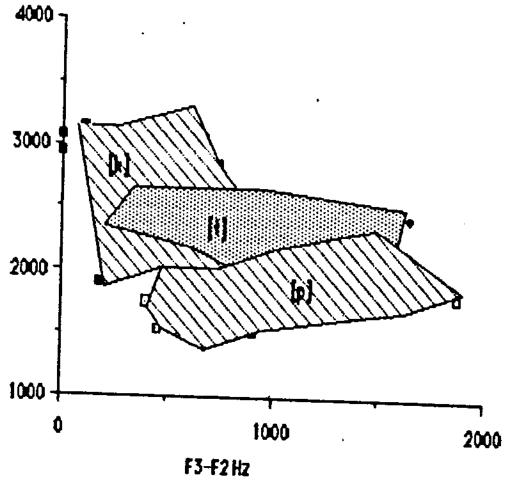


Fig 2: Burst frequency (mean value of the two spectral peaks measured in LP spectra) vs F3-F2 onset formant frequencies.

The degree of overlapping is similar to the one obtained when plotting only the first prominent peak (K1) or only the second one (K2).

It can be seen that this is not a completely satisfactory classification in terms of place of articulation, particularly because it fails to make a clear difference between dental and velar consonants. A more careful examination of the parameters involved shows that the difference between K1 for [p] and [t] is statistically significant (t= 3) (1) but it is not so for [t] and [k]. Similar results are obtained for K2: [p] and [t] show significant differences (t= 4.1), but [t] and [k] are not significantly different.

As for the onset of formant transitions, F2 can distinguish between [p] and [t] (t= 4.7), but again the onset of the F2 transition alone fails to separate [t] from [k]. The results for the onset of F3 transitions show a somewhat different patters: [p] is distinguished from [t] (t= 4.9) and [t] differs from [k] (t= 3.3). Significant differences are also found between F3-F2 onset frequencies for [t] and [k] (t= 2.9) but they do not appear for [p] and [t].

It has not been possible to measure our spectra in terms of the metrics developped by Blumstein and Stevens [11]. An informal observation of data presented in Table 2 and Fig. 2 suggests the predominance of low frequency spectral peaks for labial consonants and a high peak at the starting frequency of the second formant for dental stops. The high standard deviation found in K1 for velars (a range of 2046 Hz for [k] vs. 628 Hz for [t]) seems to suggest the confluence of K1 and K2 in a high-medium range of frequencies. This is in agreement with Blumstein and Stevens templates, the more problematic being the dental consonant.

Examining changes in energy distribution between the burst and the onset of transitions it appears that difference between the frequency of spectral peaks in the burst spectrum and at the onset of F2 and F3 transitions are more significant for [t] (t= 10.9) than for [p] (t= 6.5) or [k] (t= 9.2), suggesting a stronger change in spectral shape for dental place of articulation in front of labial or velar stops [2].

3.2. TEMPORAL CORRELATES

Burst duration has been found to provide a mean of differentiating between the three classes of stops studied: differences are significant for [p]-[t] (t= 4.4) (2) and for [t]-[k] (t= 4.8), the duration of the burst increasing from labial to velar place of articulation.

The same results hold for VOT, with significant differences between [p]-[t] (t= 4.1) and even more for [t]-[k] (t= 5.6). According to these results, both can provide temporal acoustic cues for place of articulation within the class of voiceless stops.

Burst duration and VOT show a strong correlation (Spearman's rho = 0.6 for 40 paired observations) although no correlation has been found between VOT and mean vowel duration in the Catalan stop-vowel groups studied.

The duration of the transitions does not seem to provide a cue for place of articulation, since no significant differences were found between the three classes of stops.

Finally, a strong correlation (Spearman's rho = 0.9 for 40 paired observations) was found between total stop-vowel and vowel duration, although this is independent of place of articualtion.

4. CONCLUSIONS

In summary, Catalan voiceless stops may be characterized by a very brief release burst and a short VOT, both increasing in length from labial to velar and by abrupt transitions into the following vowel. Frequency parameters for the burst or for the starting point of F2 and F3 transitions alone are not able to discriminate among the three places of articulation, but a gross characterization of the spectral shape at the consonantal release and at the onset of transitions combining the parameters measured seem to provide invariant cues for this phonetic dimension. In our data the differenciation is clearer between [p] and [t] than between [t] and [k].

Both burst duration and VOT can discriminate between place of articulation, these two parameters being strongly correlated. Differences in VOT between [t] and [k] are stronger than between [p] and [t]. This seem to suggest an interaction between temporal and frequency acoustic cues that would be used in conjunction for the discrimination of place of articulation independently of other factors studied as vocalic context or speaker variation. This fact agrees with the concept of dynamic invariance proposed by Blumstein [2].

(1) All t values are given for $p \leq 0.05$ and 30 degrees of freedom. (2) All t values are given for $p \leq 0.05$ and 77 degrees of freedom.

REFERENCES

[1] Blumstein, S.E. (1986) "On Acoustic Invariance in Speech" in J.S. Perkell & D.H. Klatt (Eds) *Invariance and Variability in Speech Processes*. Hillsdale: LEA. pp.178-193.

[2] Lahiri, A. et al. (1984) "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study", *JASA*, 76,2: 391-404.

[3] Martí, J. (1986) *Estudi acústic del català i síntesi automàtica per ordinador*. Unpublished PhD. Universitat de València.

[4] Poch, D. (1984) *Las oclusivas sordas del español*. Unpublished PhD. Universitat Autònoma de Barcelona.

[5] Julià, J. (1981) "Estudi contrastiu dels oclusius de l'anglès i del català. Un experiment acústic", *Estudi General*, 1,2: 75-85.

[6] Poch, D. (1985) "Caractérisation acoustique des occlusives de l'espagnol: le problème du VOT", *Revue de Phonétique Appliquée*, 77: 477-490.

[7] Vagges, K. et al. (1978) "Some acoustic characteristics of Italian consonants", *Journal of Italian Linguistics*, 3,1: 69-86.

[8] Durand, P. (1985) *Variabilité acoustique et invariance en français. Consonnes occlusives et voyelles*. Paris: Editions du CNRS.

[9] Fourakis, M. (1986) "A timing model for word-initial CV syllables in Modern Greek", *JASA*, 79,6: 1982-1986.

[12] Stevens, K.N.- Blumstein, S.E. (1978) "Invariant cues for place of articulation in stop consonants", *JASA*, 64,5: 1358-1369.

13] Blumstein ,S.E.- Stevens, K.N. (1979) "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants", *JASA*, 66,4: 1001-1017.

ACOUSTIC CORRELATES FOR PLACES OF ARTICULATION IN SPANISH STOP CONSONANTS

Dolors Poch-Olivé
Laboratori de Fonètica
Universitat Autònoma de Barcelona
Bellaterra - Sapin

ABSTRACT

This work aims at an acoustic description of Spanish plosives. A corpus of utterances taken from fluent speech has been spectrographically analyzed. The results confirns that there are invariant acoustic properties for place of articulation in Spanish plosives. These properties have been found in the first ms of the consonant. They are context independent and can be used so distinguish between places of articulation.

1. INTRODUCTION

The study of invariant acoustic properties present in the phonic chain is one of the main subjects in the research that is taking place at present in the field of phonetics. Stevens and Blumstein have formulated a theory of acoustic invariance that claims that invariance is the result of series of acoustic properties that encompass several components of the sound wave: "/.../ acoustic invariance corresponding to a particular phonetic category or distinctive feature resides in the acoustic signal /.../ but rather is provided by integrated acoustic properties that may encompass several of these components" (Stevens, Blumstein, 1981).
The idea that integrated acoustic properties which manifest themselves in an invariant manner for each place of articulation exist, is accepted in the theoretical development of Stevens and Blumstein. Several researchers had already suggested (Cole, Scott (1974); Stevens (1975); Fant (1973)) that integrated properties associated with a phonetic feature are invariant from an acoustic point of view, that is to say, they are independent of the context in which the feature is produced. According to Stevens and Blumstein (1978) the invariant cues of the place of articulation can be characterized regarding the spectrum at a given moment rather than variations through time, as for example, transitions. Thus, a consequence of formant positions and burst characteristics, the spectrum, taken between 10-20 msec. after the explosion of the plosive way show different characteristics for each place of articulation.
The present study follows this line, and concentrates on Spanish plosives. Its aim is to determine, from an acoustical analysis the invariant elements of the sound wave that might characterize the three different places of articulation.

2. CORPUS

Our corpus of analysis is consists of 200 sentences 8 syllable in which [p t k] always appear in the penultimate syllable which is also the one in which stress falls. This structure has been chosen because it corresponds, in number of syllables and stress position, to the normal structure of Spanish (Navarro Tomás, 1941).
It must be emphasized that the corpus analyzed comprises fluent speech sentences. This is an important fact since the characteristics of speech sounds vary according to the type of statement in which they are inserted (Shoup, Pfeiffer, 1976).
The realizations of [p t k] always appear in VCV sequences, where C represents the allophone which is being studied and V each of the vowels of Spanish.
The corpus has been realized by 6 informants, 4 male and 2 female speakers, without strong dialectal features (standard). For the recording, we have used the method of inserting the utterance to be analyzed into a "sentence-frame" the structure of which is: "He said in a quiet voice: "————", and left".

3. EXPERIMENTAL PROCESS

Utterances have been recorded in an anechoic chamber, with an UHER 4000 Report-L and Uher M-514 microphone. Afterwards, the corresponding

spectrograms have been made with a Voice Print series 7000 sonograph, in broad band and with a linear range of 5000 Hz at the most. Data corresponding to burst, VOT and transitions have been taken out from all these recordings.
The values thus obtained, which correspond to 1200 Spanish utterances, have been subjected to a computerized treatment by means of the SPSS (Statistical Package for the Social Siences).

4. RESULTS

4.1. Burst

The characterization of burst in spectrograms has been realized regarding its duration and the range of frequencies in which its energy spreads. The results obtained show that the values that burst duration reach are similar in male and female voices and also in the different vowels that follow [p t k]. It seems that stress does not affect this acoustic cue either. On the other hand, there is a difference between the values for average duration which have been obtained in each plosive:

- average duration [p]: 15,384 msec.
- average duration [t]: 15,267 msec.
- average duration [k]: 21,868 msec.

The burst of [k] is 29,65% longer than that of [p], and 28,539% longer than that of [t], the difference between [p] and [t] being very small.
From the point of view of the frequences over which energy's spread, for all plosives, the energy of burst is found in a band of frequencies of about 200 Hz. that ranges from 800 Hz. to 1000 Hz. in 65-70% of the utterances. Bursts next to front vowels have a tendency to initiate the energy in bands of frequency which are more acute than bursts beside back vowels, although, in both cases, the significant band is still the one between 800 and 1000 Hz. In respect of the end of burst energy, the limit for [p] lies between 1200 Hz and 4000 Hz. in 86,146% of the utterances. For [t] burst energy ends between 2200 Hz and 5000 Hz. in 93,733% of the cases. And, finally, for [k], between 1200 and 3500 Hz. in 85,443% of the utterances. In this latter case the energy is concentrated between these limits, unlike the cases of [p] or [t] the limits of which can reach 4000 Hz., or even in the case of [t], to 5000 Hz., as mentioned above.
There is a clear difference in duration and concentration of energy for [k] as opposed to the other two plosives.

4.2. VOT

The values obtained for the VOT of [p t k] regarding the different elements of variation have been the following:

- [p]: 17,182 msec.
- [t]: 19,757 msec.
- [k]: 30,014 msec.

although there is an increasing progression between the values for [p] and [t]. It is true that the difference is very small. This is not the case the value of the VOT of [k]. The VOT for [k] has a duration 40,78% greater than that of [p] and 31,905% than that of [t].

4.3. Transitions

Spectrographic representations only show a "continuum" of energy in movement it being difficult or even impossible to establish a border line between "transition" and "formant". And even more if we deal with fluent speech. Therefore, we have decided to determine the tangent between the point where vowel energy begins and the frequency value that has been obtained 20 msec. after the beginning of the vowel. The tangent provides information about the degree of slope of formants at every point under consideration making it possible to measure formant directions and the degree of inclination of slopes.
The results show that the values of the tangents of F1 are always smaller than those of the tangents of the other formants. The slopes are always more marked for F2. The values of the tangents of F3 are always intermediate between those of F1 and F2.
The values of the tangents are always greater when the vowel follows the plosive rather than preceeding it.
The values at which the tangents of the formants for front vowels preceeded or followed by back vowels are higher than those arrived at when the preceeding of following vowel is a front vowel. This implies that the slope of the formant is more marked in phenomenon can occur: the value of the tangent is higher for back vowels when the preceeding or following vowel is front, in which case, slopes are also more marked. This fact leads us to cunclude not that the direction and the value of the slope do not depend only on the contoid under consideration, that is to say

on the place of articulation, but also on the vowel that preceeds or follows the plosive.

The slopes for the formants adjacent to [p] are the only ones that always take the same direction: descendent in all cases. Those of the formants for the vowels adjacent to [t] follow no regular pattern in the case of F1. As for as F2 are concerned , slopes are, in general, descendent when the vowel is front and ascendent when the vowel is back. The slopes of F2 are usually descendent with many exceptions. The F1 slopes for [k] have no regularity and, although in some cases F2 and F3 tend to converge, this is not the general trend so, in relation to transitions, to establish regularities for a velar place of articulation is practically impossible.

5. DISCUSSION

As the results that have been obtained are centered upon the analysis of the behaviour of approximately the first 35-40 msec. of the beginning of plosives, we can make a comparison between these and the results that Stevens and Blumstein have obtained. Their fixed window of analysis of 26 msec. includes practically the same consonantal segment than the one analyzed in this study. For these authors, the spectrum of the velar place of articulation presents more amplitude in high frequency peaks, and the energy is distributed over the entire spectrum. In the results obtained for Spanish consonants, the dental-alveolar one presents the larger distribution in the energy spectrum of the burst and second formant transitions, above, tend to go towards the high zones of the spectrogram thus these data seem to coincide with those of Stevens and Blumstein. According to these authors would give rise to a spectrum in which the frequencies with greater amplitude are the lower ones (a labial place of articulation). In our corpus, this place of articulation shows a smaller amount of diffusion of burst energy than for [t], the minimum values of VOT and a regular tendency of transitions to go to the lower zones of the spectrum such as Stevens and Blumstein propound. Finally, for a velar place of articulation, for which in Spanish the VOT has a longer duration, the data obtained also coincide with those of Stevens and Blumstein who propound a compact spectrum for [k]. In our case, although transitions do not show any regularities, burst energy does, as it is always concentrated in an intermediate zone of the range of frequencies (compact spectrum) and also, as already mentioned, the VOT.

6. CONCLUSION

Viewed from the theory of acoustic invariance, the data which have been obtained in this study for Spanish plosives, allow us to state that there is indeed a series of integrated acoustic properties (not only one) that manifest themselves in an invariant manner for each place of articulation. These properties seem to manifest themselves in the first msec. from the beginning of the burst to the onset of voicing corresponding to the vowel.

REFERENCES

ABRAMSON, A.S.; LISKER, L. (1965), "Voice onset time in stop consonants: acoustic analysis and syntheses", Actes 5 Congrès International d'Acoustique, 1-10.

ABRAMSON, A.S.; LISKER, L. (1972), "Voice-timing perception in Spanish word-initial stops", SRSR, 28/29:15-26.

ABRAMSON, A.S.; LISKER, L. (1973), "Voice-timing perception in Spanish word-initial", Journal of Phonetics, 1:1-8.

DENT, L. (1976), "Voice onset time os spontaneusly spoken Spanish voiceless stops", JASA, 59, sup. 1, S 41.

FANT, G. (1973), Speech sounds and features, Cambridge, MIT Press.

KEWLEY-PORT, D. (1982), "Measurements of formant transitions in naturally produced stop consonant-vocal syllabes", JASA, 72/2:379-389.

KEWLEY PORT, D. (1983), "Time varying features as correlates of place of articulation in stop consonants", JASA, 73/1:322-335.

LLISTERI, J. WEST, M. (1984), "Analysis of stop-vowel transitions in calatan", 11th International Congress on Acoustics, Revue d'acoustique, 279-285.

NAVARRO TOMAS, T. (1941), "El grupo fónico como unidad melódica", RFH, 1:77-107.

NIE, N.H.; HADLA HULL, C.; JENKINS, J.G.; STEINBRENNER, K.; BENDT, D. (1975), Statistical Package for the Social Sciences, SPSS, New York.

POCH OLIVE, D. (1984), Las oclusivas sordas del español, PhD, Universitat Autònoma de Barcelona.

POCH OLIVE, D. (1985), "Caractérisation acoustique des occlusives de l'espagnol: le problème du VOT", Revue de phonétique appliquée, 77:477-489.

SHOUP, J.E.; PFEIFFER, L.L. (1976), "Acoustic characteristics of speech sounds,in LASS, J. (ed.), Contemporary issues in experimental phonetics, New York, Academic Press, 171-224.

STEVENS, K.N.; BLUMSTEIN, S. (1981), "The search for invariant acoustic correlates of phonetic features", in EIMAS, P.; MILLER, J.L. (eds.), Perspectives on the study of speech, Lawrence Erlbaum, New Jersey, 65-99.

QUANTITATIVE COMPARISON OF SPEECH FUNDAMENTAL PERIOD ESTIMATION DEVICES

DAVID M HOWARD IAN S HOWARD

DEPARTMENT OF PHONETICS AND LINGUISTICS

UNIVERSITY COLLEGE LONDON, U.K.

ABSTRACT

Speech fundamental frequency estimation devices are usually designed to suit the application for which they are intended. A technique is described which enables the operation of such devices, which operate in the time domain, to be quantitatively compared. It is shown that the use of this technique enables device operating parameters to be fine-tuned in a rigorous manner.

INTRODUCTION

There are many methods available for the estimation of fundamental period in speech, and these can be separated into the following categories as devices which operate: in the time domain on the speech pressure waveform (Sp), in the frequency domain on Sp, in hybrids of the time and the frequency domains on Sp, and directly from an input gained at the level of the larynx (see [1] for a review). To date no one device exists which reliably estimates fundamental period from speech for all speakers in all conceivable operating conditions. Thus the choice of a device for a particular application must be made with due attention being paid to errors which are not acceptable against those which can be tolerated.

Generally this procedure will involve the implementation of the devices under consideration, in hardware or software, and it is not always clear whether the result is operating as intended with a speech input. Further, many designs require an elaborate optimisation procedure for the particular speakers and set of operating conditions for which the final device is destined. These areas are most time consuming and they often leave the designer the formidable task of weighing up the beneficial effect of altering a parameter to, for example, reduce output frequency doubling errors when it is found that this adjustment also causes an increase in voicing onset deinition errors.

Such problems require a quantitative method which enables device performance to be compared using a speech input, of the type one expects when the device is in use, against a standard. Then the setting up of a device could be achieved with reference to a quantity defined, ideally by the designer, for a particular recorded speech input, and optimisation could be carried out with quantified feedback as to the effects that altering parameters has on device performance. Indeed, if appropriate controls are made available and the requirements of the users can be rigorously defined, then this optimisation procedure could become an automated process. This paper describes such a quantitative technique for the assessment of time domain fundamental frequency estimation device performance, and an illustration is given as to how it can be used to optimise device parameters automatically.

DEVICES STUDIED IN THE TESTS

The technique described below [2] can only be used with devices which are designed to produce a pulsatile output where each pulse corresponds to an epoch of acoustic excitation due to vocal fold closure. Such devices usually operate in the time domain, and here an already established time domain device is made the subject of study. This is a peak-picking device [3] which has been developed as the input speech processing stage of the EPI group [4] hearing prostheses for the totally deaf and profoundly deaf. It is a small battery-powered device which operates in the time domain producing a pulsatile output suitable for these tests. The version used in these tests is a software implementation [5] which is written in C which runs under "UNIX" on the department's Masscomp 5500 computers.

This work also requires a 'standard' against which the operation of the device is based on the laryngograph [6], and the algorithm used to detect period epochs is described in [7]. The laryngograph gains its input directly from the vocal folds by measuring the current passing between two electrodes placed on either side of the throat at the level of the larynx. When the vocal folds vibrate the current flow between the electrodes changes and this is clearly shown in the output waveform from the laryngograph (Lx), and an example is shown in figure 1b. The main advantages of using the laryngograph as a standard, a practice also used in [8], is that it is unaffected by competing acoustic noise, and that the Lx waveform conveys the periodicity associated with voiced sounds in a clearly defined manner which can be simply processed to give a suitable pulsatile output.

DESCRIPTION OF ANALYSES

The methodology used is composed of two parts designed to investigate the one-to-one deviations of the pulse markers generated by the test and reference devices -- thus it can be thought of as a 'micro' level comparison. It is complimentary to a 'macro' level (whole passage input) methodology which is being investigated, and the initiation of these is described in [7]. The two stages, described in detail in [8], are as follows:

1) the jitter distribution which is a histogram of the differences in the times of occurrence of output pulses from the reference and the corresponding time-aligned pulses from the device under test, and

2) the receiver operating characteristic (ROC) which is a plot of the probability of successful detection of a vocal fold closure on comparison with the reference (a HIT) against the number of pulses generated with no corresponding pulses in the reference output (FALSE ALARMS).

The ROC enables a quantitative measure to be gained as device operating parameters are altered. The peak-picking device, under test in this case, has a user-adjustable gain control which essentially determines the threshold level for the generation or non-generation of an output pulse. When this is altered there may be a change in the number of HITS and FALSE ALARMS, and this is shown by the ROC for the device. Each point on the ROC is plotted as the percentage of HITS generated against the number of FALSE ALARMS. As the gain is altered the points on the ROC trace out a curve (see figure 3). As the gain is lowered the number of HITS will increase, but so will the number of FALSE ALARMS. In general just one point for a particular device will specify the position of the ROC curve which indicates how detectable the signal is to the algorithm/device. Device operation can be ranked since those producing outputs highly similar to the reference will have some point on the ROC more closely approaching the perfect performance point (FALSE ALARMS = 0,HITS = 100%).

QUANTITATIVE COMPARISON METHODOLOGY

The data for this work was taken from a passage recorded by a male speaker (JM) in the anechoic room at UCL. A two channel digital (pcm) recording (Sp and Lx) was obtained, and the sentence "We can learn a little something from the birds, he said"
was transferred onto a Masscomp 5500 computer at a sampling rate of 12800Hz using a 12 bit ADC via a suitable anti-aliasing filter. The Sp and Lx waveforms are shown in figure 1a and 1b.

RESULTS

The reference, based on Lx, produces the period markers, and the reciprocal of these are plotted to give a fundamental frequency with time (Fx) trace in figure 1c. The peak-picker also produces period markers, which are not shown here due to lack of clarity on this scale, its outputs for a series of gain settings being shown as Fx contours (see figures 1d to 1h which corresponding to gains of 0.03, 0.1, 0.25, 0.5 and 1.0 respectively). In this manner a visual comparison can be made between the operation of the peak-picker with different gain settings, and the reference, and it can be seen that the gain appears optimum around a value of 0.25 .

This value of gain has been used for the peak-picker in both the jitter histograms shown in figure 2. They are plotted for the peak-picker (test device) against the laryngograph-based method (reference device) for (a) anechoic speech (figure 2a), and (b) anechoic speech degraded with white noise, SNR = 6dB (figure 2b). It can be seen that there is greater deviation from the zero jitter point with noise contaminated speech.

The ROC curves for these two speech input conditions are shown in figure 3. As the peak-picker gain is increased, a curve is traced away from the origin. Ideally optimum gain would result in a point at (hits = 100%,false alarms = 0). In practice, however, the optimum will only

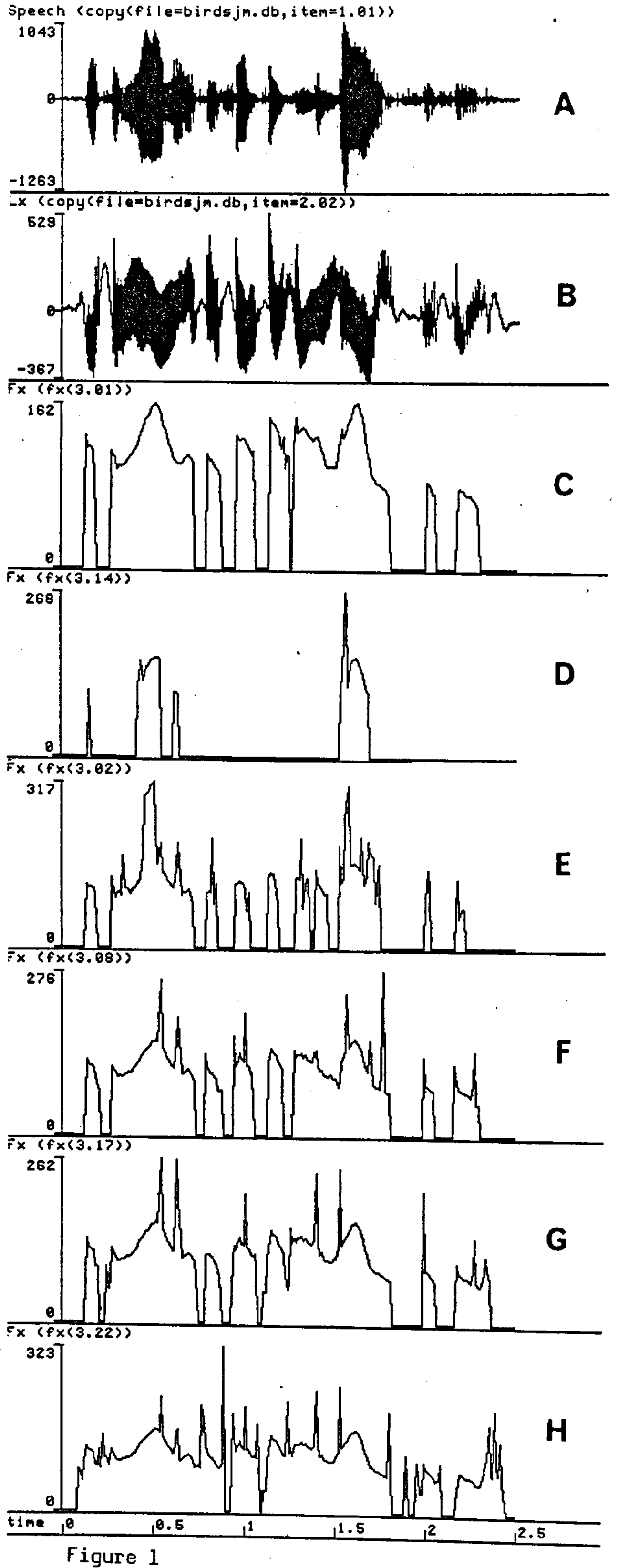


Figure 1
A) speech, B) Lx and C) Fx from Lx, D) to H) Fx from peak-picker.

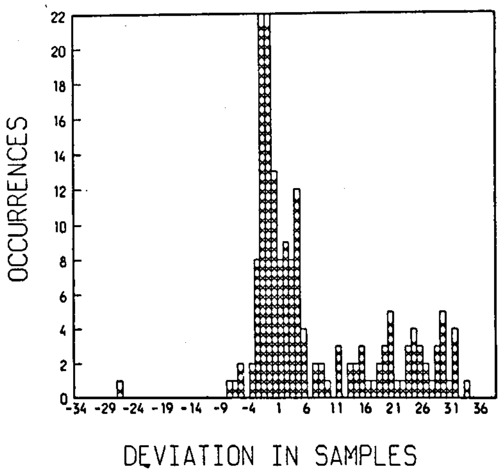


Figure 2a.
Jitter histogram for recording room quality speech .

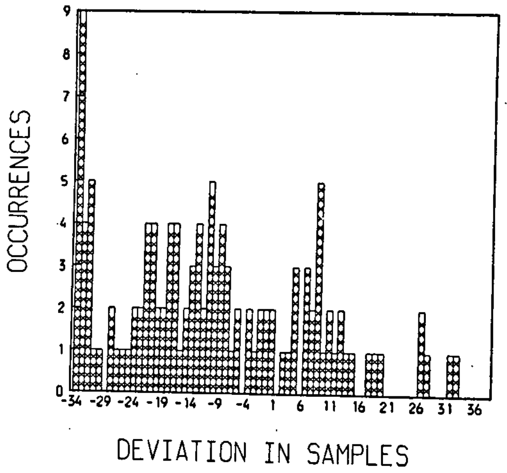


Figure 2b.
Jitter histogram for speech contaminated with uniform density noise (SNR = 6 dB).

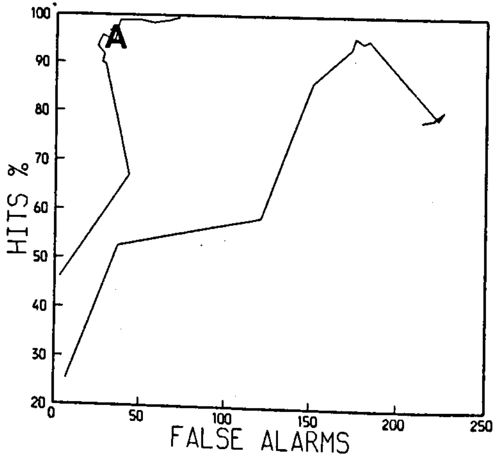


Figure 3
Top curve ROC for peak-picker with recording-room quality speech, lower curve for noise corrupted case.

approach this point and will depend on a trade-off between number of hits required against the error rate. It can be seen, in this case, that point A on the top curve is a good choice for optimum gain (point A corresponds to a gain of 0.25) because a higher gain only results in a marginal increase in the number of hits for a considerable increase in the false alarms. This value also corresponds to the value determined above for optimum gain from the Fx contours (see figure 1c to 1h). With the addition of noise device performance is degraded, and this is shown by its ROC which is below the other curve for all gain settings.

From these results, it can be concluded that the ROC gives a basis for an automated optimisation and assessment technique. In the particular case discussed above optimum gain has been selected by observation of the ROC and of the Fx contours. In practice any parameter could be optimised automatically using the ROC method for the particular application for which the fundamental period device is intended.

ACKNOWLEDGEMENTS

This work was supported by Alvey grant MMI/056 and MRC studentship RS-85-2.

REFERENCES

[1] Hess, W., "Pitch determination of speech signals", Springer-Verlag, Berlin, (1983).

[2] Howard, D.M., Maidment, J.A., Smith, D.A.J., and Howard, I.S. (1986). "Towards a comprehensive quantitative assessment of the operation of real-time fundamental frequency extractors", IEE Conf. Publ., 258, 172-177.

[3] Howard, D.M. and Fourcin, A.J. (1983), "Instantaneous voice period measurement for cochlear stimulation", Electronics Letters, 19, 19, 776-778.

[4] Fourcin, A.J., Douek, E.,Moore, B.C.J., Rosen, S.R., Walliker, J.R., Howard, D.M., Abberton, E.R.M., Frampton, S., "Speech perception with promontary stimulation", An. New York Acad. Sci., 405, 280-294, (1983).

[5] Howard, D.M., "Digital peak-picking fundamental frequency estimation". Speech hearing and language; Work in progress, 2, London: UCL, (1986).

[6] Fourcin, A.J., and Abberton, E.R.M., "First applications of a new laryngograph", Med. and Biol. Illust. 21, 172-182, (1971).

[7] Howard, I.S., and Howard, D.M. (1986). "Quantitative comparisons between time domain speech fundamental frequency estimation algorithms", Proc. Inst. Acoust., 8, 7, 323-330.

[8] Hess, W. and Indefrey, H., (1984). "Accurate pitch determination of speech signals by means of a laryngograph", Proc. ICASSP-84, 1-4.

A LOGARITHMIC SPECTRAL COMB METHOD
FOR FUNDAMENTAL FREQUENCY DETECTION

PHILIPPE MARTIN

Experimental Phonetics Laboratory
300 Huron Str., Toronto, Ontario,
CANADA M5S 2X6

ABSTRACT

The spectral comb method is a fundamental frequency detection algorithm based on the cross-correlation of a modified power spectrum of the speech signal and a spectral comb function with teeth of decreasing amplitudes and variable intervals. In order to reduce the overall computational complexity and obtain a constant frequency resolution, a modified approach is proposed to compute the cross-correlation function, using a logarithmic scale for both the amplitude and the frequency. The cross-correlation is obtained by iteratively summing each spectral peak shifted on the frequency scale by factors of 1, 1/2,...,1/n, and on the amplitude scale by factors of 1, 2,...,n dB according to the comb tooth order n.

INTRODUCTION

Fundamental frequency detection plays an important role in phonetic research, as well as in many aspects of speech analysis, such as speech recognition and synthesis. Although many experimental devices and algorithms have been proposed to date [7], none provide error free results, specially in the case of noise or telephonic recordings. The choice of a specific pitch detector will depend on the application, as its structure will define an implicit articulatory model for Fo detection. Any discrepancy between this model and the real conditions of analysis will lead to errors in the fundamental frequency detected.

Among the numerous methods available for pitch detection, those based on short term spectral analysis of the input signal offer usually a good resistance to noise and provide adequate results even if the fundamental component is absent from the speech input. Despite some drawbacks, essentially due to lower time and frequency resolution which precludes their use for medical applications, spectral pitch detection appears to be quite attractive for phonetic and linguistic research.

Most methods of pitch detection based on the short-time spectrum aim to detect some periodicity of the fundamental frequency harmonics. The popular cepstrum approach [1], for instance, computes the Fourier transform of the logarithm of the power spectrum. Other methods are based on a more direct and computationally efficient direct search for periodicity in the spectrum. Schroeder [2] uses an histogram of subharmonics derived from the spectral peaks, and Fo is taken as the smallest common multiple of the periods of its harmonic components. Harris and Weiss [3] use a high resolution Fourier spectrum and retain the most numerous equal spacing of adjacent peaks as fundamental frequency. Sreenivas and Rao [4] use only high quality peaks (well above the noise level), and compute their approximate highest common factor to obtain the pitch value. Sluyter, Kotmans and Leuwaarden [5], in order to reduce the influence of phase distortion in the peak frequency measurement, utilize a minimum distance criterion to recognize harmonic pattern and the resulting fundamental frequency.

THE SPECTRAL COMB METHOD

By contrast with other pitch detection schemes such as the sieve algorithm, the spectral comb method [6] is based on a direct search of an harmonic structure in the spectrum integrating both the harmonic frequency and amplitude informations. This ensures that a correct value of Fo will be obtained even if no periodicity in the spectrum is found. (Interestingly enough, most frequency domain pitch detection methods such as the cepstrum will fail if the signal has no harmonics, as for a pure tone).

To evaluate Fo, the short-time spectrum |F(w)| is first "groomed" by replacing spectral peaks meeting an appropriate selection criterion by narrow parabola, and by zeroing the remaining of the spectrum. This ensures that non-harmonic related values, usually with low energy, will not interfere in the overall computation.

The groomed spectrum is then crosscorrelated with a spectral comb function C(wp,w) with teeth of decreasing amplitude and variable intervals wp.

$$C(wp,p) = \sum_{n}^{1} An \quad (nwp-w)$$

The maximum of the crosscorrelation function I(wp) is reached when a large number of the comb's teeth coincide with the harmonic peaks of the spectrum. When this value exceeds a voicing treshold, the corresponding tooth interval is taken as 1/Fo.

$$I(wp) = \sum n \exp -1/8 \, |F(wp)|$$

If |F(w)| is represented by m samples, n*(m*m) sums and products are necessary to evaluate n values of the crosscorrelation function I(w). In a 1000 Hz frequency range, with a 4 Hz resolution, this corresponds to
250 * (250*250)= 15,625,000 sums and products.

A FASTER METHOD

Due to the nature of the groomed spectrum and of the spectral comb, many of the operations involved in this computation involve a zero factor. A much more efficient algorithm can be obtained if only non-zero values were to be taken into account. This can be done if I(wp) is evaluated from the peaks of the spectrum only, whose amplitude and position on the frequency axis are the only information retained. The cross-correlation function is obtained by iteratively adding, for each spectral peak of |F(w)|, parabola

- shifted in frequency according to the order of the comb tooth n;

- shifted in amplitude by an appropriate factor proportional to the comb tooth order n.

A LOGARITHMIC SPECTRAL COMB

The use of a linear frequency scale ensures the possibility of using the FFT to evaluate the short time spectrum |F(w)|. On the other hand, since all computations are performed on sampled values, a linear frequency scale creates an uneven frequency resolution in the contribution of low and high harmonic components. Using a logarithmic frequency scale, all harmonic components will have a similar impact on the final cross-correlation result. Furthermore, the operations will only involve additions and substractions.

Starting from a logaritmic short-time spectrum, the cross-correlation with a logarithmic spectral comb function is then obtained by recursively adding n times

- shifted in frequency by log n

- shifted in amplitude by n dB

Again, a more efficient algorithm will proceed from the spectral peaks added recursively after having been shifted by log n on the frequency scale and by n on the amplitude scale (n=1, 2,..., n).
With this approach, assuming that each peak is represented by p values and that n comb's teeth are considered, the total number of additions is reduced to p*n*h , with h= number of harmonics taken into account. With typical values of p=16, n=8 and h=8, we have thus 1024 additions to perform to obtain the cross-correlation function (Each addition involving an extra address calculation).

The price to pay to implement this logarithmic approach is the spectral analysis of the speech input, which has to be obtained either by a relatively high resolution FFT followed by a logarithmic mapping of the frequency scale, or a direct logarithmic DFT. The latter solution would be more easily implemented in hardware form.

Furthermore, the sides of each peak parabola must be constant on the logarithmic frequency scale. This suggests a possible improvement in the computation of the logarithmic DFT : since the width of each peak is proportional to the duration of the time window used, shorter blocks of sampled speech input can be used for higher frequencies. Starting for example at +10 Hz at -20 db below a spectral peak at 100 Hz, the equivalent logarithmic width at 1000Hz will correspond to a frequency width of +100 Hz. Using for instance a Gaussian window, this would imply a duration of the time window equal approximatively to 20 times the period involved, i.e. 20* 1/100 Hz=200 ms and 100 Hz and 20 ms at 1000 Hz. This variable window length roughly corresponds to the time resolution of the ear for pure tones.

CONCLUSION

Using a logarithmic scale for both the amplitude and the frequency scale of the short-time power spectrum, the computational effort to evaluate the cross-correlation function in the spectral comb method is dramatically reduced. Typically, only 1024 sums are necessary, compared to more than 15,000,000 sums and products in the direct approach. This method, which requires the Fourier transforn of the speech input to be logarithmic, seems suitable for hardware implementation leading to reliable real-time operation.

REFERENCES

[1] A.M. Noll, Short-time Pitch Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection, JASA, 36: 296-302, 1969.

[2] M. R. Schroeder, Period Histogram and Product Spectrum: New Methods for Fundamental Frequency Measurement, JASA, 43: 829-834, 1968.

[3] C. M. Harris and M. R. Weiss, Pitch Extraction by Computer Processing of High Resolution Fourier Analysis Data, JASA, 35: 339-343, 1963.

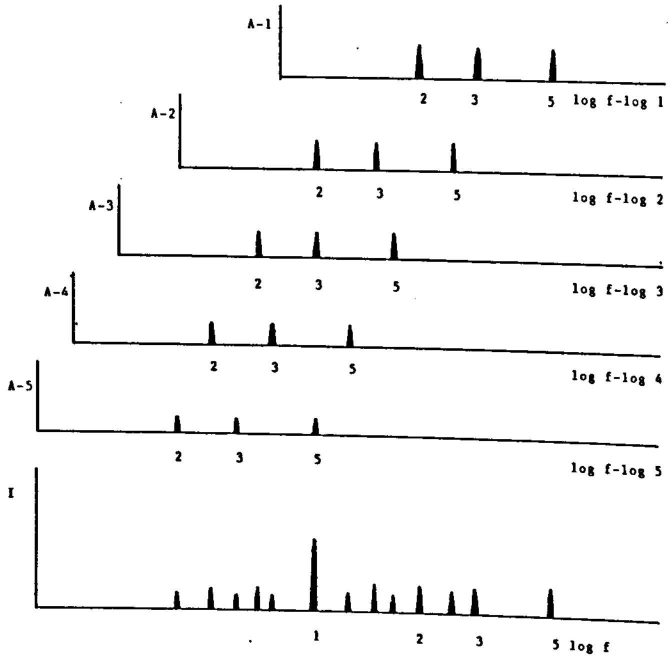
[4] T. V. Sreenivas and P. V. S. Rao, Pitch Extraction from corrupted Harmonics of the Power Spectrum, JASA, 65: 223-228, 1979.

[5] R. J. Sluyter, H. J. Kotmans and A. V. Leuwaarden, A Novel Method for Pitch Extraction from Speech and a Hardware Model Applicable to Vocoder Systems, Proc. of the ICASSP-80, I: 45-48, 1980.

[6] Ph. Martin, Comparison of Pitch Detection by Cepstrum and Spectral Comb Analysis, Proc. of the ICASSP-81, I: 180-183, 1981.

[7] W. Hess, Pitch Determination of Speech Signals, Springer, Berlin, 1983.

[8] Ph. Martin, Spectral Comb Gives Real-Time Pitch Analysis, Speech Technology, Sep-Oct. 1983.



Cross-correlation-function for a 3 components spectrum (200 Hz, 300 Hz and 500 Hz). The maximum is obtained for Fo=100 Hz.

A CLUSTER-SEEKING TECHNIQUE FOR PROSODIC ANALYSIS
(with special reference to Russian sentence intonation)

Leonid A. Kanter, Alexander P. Chizhov, Ksenia G. Guskova

Department of Phonetics, Faculty of Foreign Languages
(English Language Division), Lenin Pedagogical Institute
Moscow, USSR 107140

ABSTRACT

A cluster analysis algorithm proposed by Sammon is used to identify intonational zones which can be correlated with intonemes of Standard Russian.

INTRODUCTION

The need for cluster analysis arises in a natural way in many areas of phonetic research. The goal of clustering methods is to provide a means to discover structure within a complex body of data /4/. With regard to intonology the first use of a cluster-seeking technique was reported in /2/.

This paper attempts to analyse the manner in which intonation contours of five Standard Russian intonation types are located in the space of acoustic parameters.

MATERIAL

The material analysed consists of the test phrase ОН ЗНАЛ [on znal] = "He knew", pronounced in dialogical contexts by sixteen male native speakers of Standard Russian. The speakers were instructed to read the phrase with context appropriate vocal modifications so that they could be identified as belonging to the following five intonation types, or communicative modes: (1) a final statement, (2) a reply statement, (3) a general question, (4) an exclamation,(5) a non-final statement. The test phrase was read twice in each mode, whereupon 160 utterances were produced. Used as test stimuli, the utterances were then listened to and categorized by a group of subjects in terms of the set of intonation types under consideration.

The subsequent instrumental (intonographic) analysis was performed to measure fundamental frequency (F_o), intensity and duration in 80 utterances selected as a result of the foregoing listening tests. Fourteen initial parameters of each intonation contour were analysed:

(1) maximum F_o value within the first syllable;

(2) minimum F_o value within the first syllable;

(3) maximum F_o value within the second syllable;

(4) minimum F_o value within the second syllable;

(5) F_o at the starting point of the first syllable (F_o at the starting point of an utterance);

(6) F_o at the end point of the first syllable;

(7) F_o at the starting point of the second syllable;

(8) F_o at the end point of the second syllable (F_o at the end point of an utterance);

(9) F_o at the last turning point of an utterance;

(10) maximum F_o value between the starting point and the last turning point inclusive;

(11) maximum value of intensity within the first syllable;

(12) maximum value of intensity within the second syllable;

(13) duration of the first syllable;

(14) duration of the second syllable.

METHOD

To reduce variance between speakers the available acoustic parameters were subject to the following normalization procedures. The fundamental frequency parameters were normalized by the formula:

$$\tilde{y}_i^{(j)} = \frac{100 \cdot y_i^{(j)}}{y_{max}^{(j)}},$$

where $\tilde{y}_i^{(j)}$ and $y_i^{(j)}$ are normalised (relative) and non-normalised (absolute) values of the i-th parameter in the j-th utterance respectively; $y_{max}^{(j)}$ is the maximum y value in the j-th utterance.

The intensity and duration parameters were normalised using the formula:

$$\tilde{x}_i^{(j)} = \frac{K\cdot x_i^{(j)}}{x_1^{(j)} + x_2^{(j)}},$$

where K=100 for intensity parameters and K=200 for duration parameters; $\tilde{x}_i^{(j)}$ and $x_i^{(j)}$ are normalised and non-normalised values of the i-th parameter of intensity/duration in the j-th utterance; i=1;2.

In this study use is made of the algorithm of non-linear, non-parametric mapping of vectors from the multidimensional space of parameters on a plane according to Sammon's criterion /3/. This criterion makes it possible to locate points on a plane in a manner whereby distances between them approximate distances between the corresponding vectors (intonation contours) in the multidimensional space of acoustic parameters. The criterion is formulated as follows:

min E, where

$$E = \frac{1}{\sum\limits_{i<j} d_{ij}^*} \sum_{i<j} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*},$$

where d_{ij}^* is the distance between the i-th and j-th intonation contours in the multidimensional space; d_{ij} is the distance between the i-th and the j-th points on a plane; E is the error of approximation, minimised through location of points on a plane.

It follows from the criterion that the location of points on a plane represents (approximates) the location of intonation contours in the space of acoustic parameters, provided that the error of approximation (E) is negligible. Therefore, the clustering of intonation contours grouping with respect to each intonation type can be assessed in terms of the position of the corresponding points on a plane. A cluster of points pertaining to a specific intonation type can be regarded as intonational zone in the space of acoustic parameters.

The experimental data processing was computerised via IKLIPS MV/8000.

RESULTS

The results of this study are displayed in figure 1 below. It reveals the arrangement of intonation contours being analysed on a plane*). The error of approximation (E) was 0.048.

The lines in the figure delimit the clusters of points corresponding to the intonational zones of a final statement, a reply statement, a general question, an exclamation and a non-final statement. The zones in question can be correlated with intonemes of Standard Russian.

CONCLUSION

The cluster-seeking technique used in this paper has been found to be highly effective in analysing intonation. The technique can be regarded as a development of the algorithmic method reported in /1/.

Possible areas of further linguistic research involving the above technique include description of the intonational system of a language in terms of phonological oppositions, the study of a foreign accent in intonation, intonological typology, phonostylistics etc.

REFERENCES

1. L.A.Canter, N.A.Sokolova and A.P.Chizhov, "On an Algorithmic Study of English Intonation". - In: Proceedings of the Ninth International Congress of Phonetic Sciences, vol. 1, Copenhagen: Institute of Phonetics, University of Copenhagen, 1979, p. 369.

2. A.P.Chizhov, L.A.Canter and N.A. Sokolova, "English Intonation in the Space of Acoustic Parameters". - In: Abstracts of the Tenth International Congress of Phonetic Sciences, Dordrecht-Holland/Cinnaminson-U.S.A.: Foris Publications, 1983, p. 360.

3. J.W.Sammon, Jr., "A Non-linear Mapping for Data Structure Analysis". - In: IEEE Transactions on Computers, vol. C-18, № 5, 1969, pp. 401-409.

4. R.C.Tryon, D.E.Bailey, "Cluster Analysis", New York: McGraw-Hill Book Company, 1970, 347 pp.

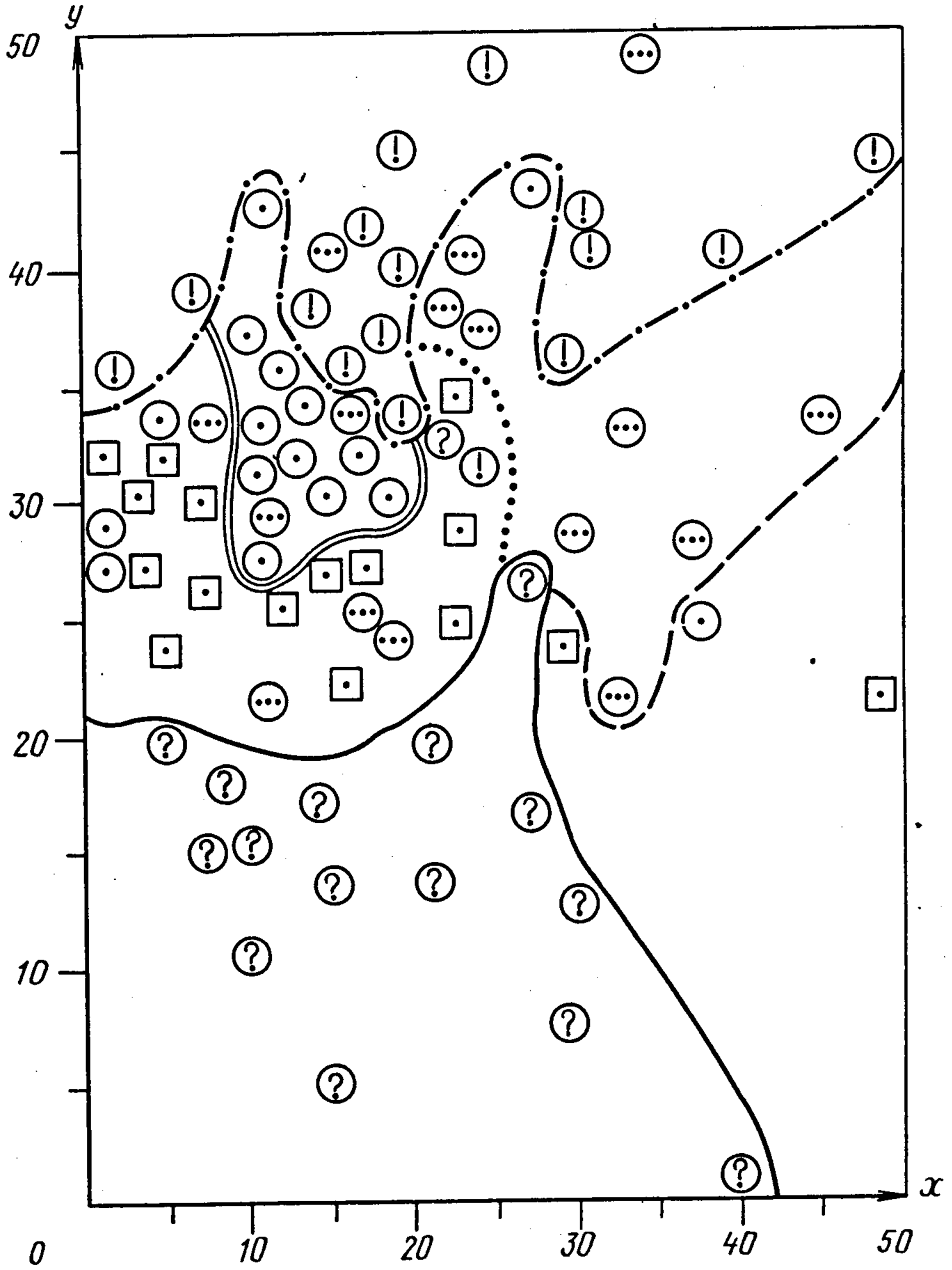


Figure 1. Clustering results reflecting location of intonation contours on a plane.

*) The intonation contours are identified in the following manner:

⊙ - final statement
⊡ - reply statement
(?) - general question
(!) - exclamation
(…) - non-final statement

ВЫДЕЛЕНИЕ ОСНОВНОГО ТОНА, ОСНОВАННОЕ НА МАРКОВСКОЙ МОДЕЛИ

ЛЮДОВИК ЕВГЕНИЙ КУЗЬМИЧ

Институт кибернетики им. В.М.Глушкова АН УССР
Киев, СССР 252207

АННОТАЦИЯ

Учет основных свойств траекторий основного тона на базе марковской модели позволяет существенно повысить надежность определения периода основного тона и признака тон/шум.

ВВЕДЕНИЕ

Чтобы повысить надежность выделения основного тона /ОТ/ предлагается применить к этой задаче статистическую теорию распознавания образов. Рассматриваемая в связи с этим статистическая модель включает две компоненты:

1) модель, описывающую зависимость применяемого первичного описания речевого сигнала от значения признака тон/шум и периода ОТ на отдельном интервале анализа, и
2) модель, отражающую основные свойства траекторий признака тон/шум и периода ОТ на фрагментах речевого сигнала из многих интервалов анализа.

Для построения модели разобьем весь диапазон значений ОТ от 62,5 Гц до 500 Гц на 12 сегментов в равномерной логарифмической шкале. При этом на каждую октаву приходится по четыре сегмента. Во временной области получим 12 сегментов $[T_{s-1}, T_s]$, $s=1,\dots,12$, длина которых достаточно хорошо соответствует максимально возможной разности периодов ОТ на соседних интервалах анализа, сдвиг между которыми равен 10 мс.

Будем исходить из того, что речевой сигнал подвергнут процедуре выравнивания спектра (например, путем обратной фильтрации системой линейного предсказания) и пропущен через фильтр низких частот, после чего вычислена нормированная автокорреляционная функция $R(i)$, $0 \le i \le L$.

Для каждого из 12 сегментов найдем опорное значение ОТ T_s :

$$T_s = \arg\max_{i:T_{s-1}<i<T_s} R(i)$$

и в дальнейшем вместо автокорреляционной функции $R(i)$ будем рассматривать ее "проекцию" $\bar{R}(s) = R(T_s), 1 \le s \le 12$. Кроме того, здесь, говоря о периоде ОТ, как правило, будем иметь в виду номер сегмента s, в который попадает этот период. Номер сегмента s является огрубленной характеристикой ОТ; уточненное значение, соответствующее огрубленному значению s, равно опорному значению T_s.

МОДЕЛЬ ОСНОВНОГО ТОНА НА ОТДЕЛЬНОМ ИНТЕРВАЛЕ АНАЛИЗА

Построение такой модели означает определение зависимости плотности совместного распределения вероятностей отсчетов проекции автокорреляционной функции (ПАФ) $\bar{R} = \{\bar{R}(s)\}_{s=1}^{12}$ от значений признака тон/шум v и периода ОТ s : $p(\bar{R}/s,v)$, где $v=1$ соответствует признаку "тон", а $v=0$ – признаку "шум".

Знание такого многомерного распределения, зависящего к тому же от параметров, является в общем случае довольно сложной задачей. Здесь, однако, допустимо прибегнуть, например, к такому упрощению. Найдем максимум ПАФ $\bar{R}$:

$$s^* = \arg\max_{s:1\le s\le 12} \bar{R}(s) .$$

Объявление s^* огрубленным значением периода ОТ соответствует корреляционному методу выделения ОТ; мы, однако, сделаем некоторый шаг вперед по сравнению с этим методом, если учтем вероятности получения различных значений s^* при фактическом значении s : $P_1(s^*/s)$. Причем здесь можно ограничиться заданием вероятности правильной оценки $P_1(s/s)$ и вероятности ошибки, соответствующей удвоенному периоду ОТ, $P_1(s+4/s)$, полагая остальные ошибки равновероятными с вероятностью

$$P_1(s'/s)=(1-P_1(s/s)-P_1(s+4/s))/10,$$
$$s'\ne s,\ s'\ne s+4.$$

Величина $R^*=\bar{R}(s^*)$ в корреляционном методе служит для определения признака тон/шум; при этом решение принимается в результате сравнения этой величины с некоторым порогом. Мы, однако, определим вероятности значений R^* в зависимости от значения признака тон/шум: $p_2(R^*/v)$, сохранив больше информации, чем используется в корреляционном методе.

Теперь статистические модели вокализованного и невокализованного интервалов могут быть определены, например, следующим образом:

$$p(\bar{R}/s,v=1) = p(R^*,s^*/s,v=1) =$$
$$= P_1(s^*/s)\cdot p_2(R^*/v=1) ,$$
$$p(\bar{R}/s,v=0) = p(R^*,s^*/s,v=0) =$$
$$= C\cdot p_2(R^*/v=0),$$

где $C = 1/12$ – вероятность любого конкретного значения s^* на невокализованном интервале.

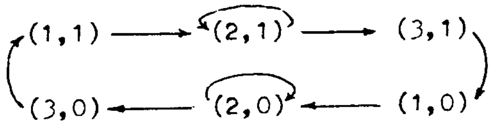
Более точная статистическая модель получится, если учесть зависимость между $\bar{R}(s)$ и $\bar{R}(s+4)$ при фактическом значении ОТ s . Возможны и другие варианты.

МАРКОВСКАЯ МОДЕЛЬ ТРАЕКТОРИЙ ОТ

Модель траекторий ОТ опирается на следующие их свойства. Во-первых, продолжительность вокализованных и невокализованных отрезков речевых сигналов ограничена снизу. Во-вторых, имеет место сильная зависимость между значениями периода ОТ на соседних вокализованных интервалах анализа. Аналогичная зависимость, хотя и в меньшей степени, наблюдается между значениями периода ОТ на интервалах, разделенных невокализованным участком сигнала, причем эта зависимость убывает с увеличением длительности разделяющего невокализованного участка.

Первое из этих свойств хорошо отражается следующей марковской моделью, генерирующей траектории признака тон/шум. Допустим, что минимальная длина вокализованных и невокализованных участков равна 3. Введем по три состояния $(1,v)$, $(2,v)$ и $(3,v)$ для каждого значения признака тон/шум $v=0,1$.

Граф переходов модели имеет вид:



Состояния $(1,1)$ и $(1,0)$ являются начальными при порождении вокализованного и невокализованного участков соответственно, а $(3,1)$ и $(3,0)$ – конечными. Повторение состояний $(2,1)$ и $(2,0)$ позволяет генерировать участки вокализованной и невокализованной речи любой длины, большей или равной трем интервалам анализа. Для полного описания этой модели необходимо задать $P_v(2/2)$ – вероятность перехода по петле в каждое из этих повторяемых состояний, при этом $P_v(3/2) =$ $= 1 - P_v(2/2)$.

Второе из сформулированных выше свойств также может быть описано с помощью некоторой марковской модели; эта модель будет генерировать траектории значений периода ОТ. Модель содержит I2 состояний, соответствующих огрубленным значениям ОТ s. С достаточно хорошим приближением можно считать, что огрубленные значения периода ОТ s и s' на соседних интервалах анализа могут отличаться не более, чем на единицу: $|s - s'| \leqslant 1$. Задав вероятности переходов $P_3(s'/s)$, определяющие среднюю скорость изменения периода в огрубленной шкале, получаем марковскую модель траекторий периода ОТ.

Объединение модели траекторий признака тон/шум и модели траекторий периода ОТ приводит к совокупной марковской модели траекторий ОТ. Возможные переходы и их вероятности представлены ниже:

$$(s,1,1) \to (s',2,1),\ P_3(s'/s),$$
$$(s,2,1) \to (s',2,1),\ P_3(s'/s)\cdot P_v(2/2),$$
$$(s,2,1) \to (s',3,1),\ P_3(s'/s)\cdot P_v(3/2),$$
$$(s,3,1) \to (s',1,0),\ P_3(s'/s),$$
$$(s,1,0) \to (s',2,0),\ P_3(s'/s),$$
$$(s,2,0) \to (s',2,0),\ P_3(s'/s)\cdot P_v(2/2),$$
$$(s,2,0) \to (s',3,0),\ P_3(s'/s)\ P_v(3/2),$$
$$(s,3,0) \to (s',1,1),\ P_3(s'/s),$$

причем во всех случаях $|s - s'| \leqslant 1$.

Состояния с различными значениями s для невокализованных интервалов вводятся, чтобы отразить зависимость значений периода ОТ на вокализованных интервалах, примыкающих с двух сторон к невокализованному участку.

Введем следующие обозначения:

$S_n=(s_n,i_n,v_n)$ – состояние модели на n-м интервале,

R_n – ПАФ сигнала на n-м интервале;

$p(R_n/S_n)$ – условная плотность вероятности параметров ПАФ в зависимости от значения признака тон/шум и периода ОТ;

$P(S_{n+1}/S_n)$ – вероятность перехода из состояния S_n в состояние S_{n+1};

$P(S_1/S_0)$ – распределение вероятностей начального состояния.

В качестве $P(S_1/S_0) = P(S_1)$ естественно взять распределение с равными вероятностями для всех невокализованных состояний и нулевыми вероятностями для вокализованных.

Теперь задача выделения ОТ сводится к задаче определения состояний построенной марковской модели, наилучшим образом соответствующих наблюденному речевому сигналу. Возможны различные постановки этой задачи.

Можно, например, поставить задачу отыскания наиболее правдоподобной траектории ОТ по всей реализации речевого сигнала. Для этого следует максимизировать апостериорную вероятность траектории:

$$P(S_1,\ldots,S_N/R_1,\ldots,R_N) = \prod_{n=1}^{N} P(S_n/S_{n-1})\cdot p(R_n/S_n)/p(R_1,\ldots,R_N). \qquad (I)$$

Эта задача решается следующим алгоритмом динамического программирования:

$$\left.\begin{aligned} I(S',n) &= \operatorname*{argmax}_{S\in Q(S')}(F(S,n-1)+g(R_n,S,S')) \\ F(S',n) &= \max_{S\in Q(S')}(F(S,n-1)+g(R_n,S,S')) \end{aligned}\right\}\ 2\leqslant n\leqslant N,$$

$$S_N^* = \operatorname*{argmax}_{S} F(S,N), \qquad (2)$$

$$S_{n-1}^* = I(S_n^*,n),\ n = N,\ldots,2.$$

В этих формулах:

$g(R_n,S,S') = \ln P(S'/S) + \ln p(R_n/S')$,

$F(S,1) = \ln P(S) + \ln p(R_1/S)$,

$Q(S')$ – множество состояний, из которых возможен переход в состояние S'.

Другим важным вариантом постановки задачи является минимизация вероятности ошибки выделения ОТ на m-м интервале анализа по фрагменту реализации из первых $m+k$ интервалов. Для решения этой задачи, как известно, следует определить, какое из интересующих нас событий имеет максимальную апостериорную вероятность. Мы будем рассматривать I3 событий: одно, означающее невокализованность m-го интервала, и еще I2, соответствующих двенадцати возможным значениям s_m периода на m-м интервале, $1 \leqslant s_m \leqslant 12$.

Для вычисления вероятностей этих событий следует просуммировать апостериорные вероятности состояний модели на m-м интервале, составляющих эти события. Апостериорная вероятность состояния $S_m = (s_m,i_m,v_m)$ определяется следующей формулой:

$$P(s_m,i_m,v_m/R_1,\ldots,R_{m+k}) = P(S_m/R_1,\ldots,R_{m+k}) = \sum_{\substack{S_1,\ldots,S_{m-1},\\ S_{m+1},\ldots,S_{m+k}}} (1/A)\cdot\prod_{n=1}^{m+k} P(S_n/S_{n-1})\cdot p(R_n/S_n),$$

где $A=p(R_1,\ldots,R_{m+k})$. Суммирование производится по всем состояниям на первых $m+k$ интервалах, кроме m-го.

Теперь можно вычислить апостериорные вероятности огрубленных значений периода:

$$P(s_m,v_m=1/R_1,\ldots,R_{m+k}) = \sum_{i_m} P(s_m,i_m,v_m=1/R_1,\ldots,R_{m+k}),\quad 1\leqslant s_m\leqslant 12,$$

и апостериорную вероятность того, что m-й интервал невокализованный:

$$P(v_m=0/R_1,\ldots,R_{m+k}) = \sum_{i_m,s_m} P(s_m,i_m,v_m=0\ /R_1,\ldots,R_{m+k})$$

Выбор наибольшего из этих I3 чисел определяет оптимальное значение признака тон/шум и периода ОТ.

Итак, предлагаемый метод применим и в случае, когда траекторию ОТ можно определять после завершения ввода всей реализации, и в случае оперативного оценивания.

В первом случае ошибки в значениях ОТ практически исключаются. Во втором случае вероятность ошибки после 4-5 вокализованных интервалов также довольно мала.

УПРОЩЕННАЯ СХЕМА: МОДИФИКАЦИЯ КОРРЕЛЯЦИОННОГО МЕТОДА

Ниже предлагается упрощенный вариант, являющийся фактически усовершенствованной модификацией корреляционного метода, поскольку в нем принятие решений осуществляется на основании сумм коэффициентов автокорреляции на траектории ОТ.

Сначала для каждого состояния модели $S_n=(s_n,i_n,v_n)$ определяется вес:

$$d(R_n,S_n) = \begin{cases} R_n(s_n), & \text{если } v_n=1, \\ A - \max\limits_{s} R_n(s), & \text{если } v_n=0. \end{cases}$$

где A – эмпирическая константа. Затем для каждой пары состояний, между которыми возможен переход определяется величина:

$$g(R_n,S_{n-1},S_n) = \begin{cases} d(R_n,S_n), & \text{если } s_n = s_{n-1}, \\ d(R_n,S_n)-D, & \text{если } s_n=s_{n-1}\pm 1, \end{cases}$$

где D – также эмпирическая константа.

Теперь ставится задача поиска такой последовательности состояний S_n, $1\leqslant n\leqslant N$, связанных допустимыми переходами, которая имела бы наибольшую сумму:

$$F = \sum_{n=1}^{N} g(R_n,S_{n-1},S_n).$$

Эта задача решается алгоритмом (2).

Описанный упрощенный вариант проверялся на тестовом материале из 9 фраз, произнесенных двумя дикторами-женщинами и одним диктором-мужчиной; частота дискретизации I6 кГц, порядок обратного фильтра I4, частота среза НЧ-фильтра I,25 кГц, длина интервала анализа 30 мс, шаг – 20 мс. Общий объем речевого сигнала около 800 интервалов анализа. Эксперименты показали высокую надежность алгоритма: не было ни одного ошибочного значения периода ОТ. Были только некоторые сомнения относительно признака тон/шум на стыках вокализованных и невокализованных участков.

ЗАКЛЮЧЕНИЕ

Предварительные эксперименты говорят о перспективности применения марковской модели для надежного выделения ОТ. При этом роль модели будет тем больше, чем менее информативным является используемое первичное описание, что имеет место, например, в случае зашумленной речи.

HOW TO GET PARAMETERS OF THE SPEECH PERCEPTION MODEL FROM THE RESULTS OF PSYCHOACOUSTIC EXPERIMENTS?

ANATOLY VENTSOV

Pavlov Institute of Physiology
Academy of Sciences of the USSR
Leningrad, USSR, 199034

Analysis of the results of many psychoacoustical experiments has made us to conclude that we are dealing with the system adapting to the parameters of a particular set of stimuli and to the instructions, either given by the investigator or generated by the subject. In this case the concept of the continuous psychological scales seems to be invalid and some new approach to the description of the subjective mechanisms of speech perception is needed. To explain usual experimental data a special mechanism including the restricted linear scale and means for projecting the signal parameters onto the scale is proposed. Some preliminary results show the reliability of the suggested mechanism. Possible ways of further detailed investigation of the mechanism are discussed.

Any functional model of speech perception in human must be strictly formulated and efficient, i.e. it must consist of a finite set of well defined quantitative algorithms for speech signal processing and making decisions. On the whole the results of natural speech processing obtained with the model must be similar to those of human speech perception. Apparently these algorithms are to describe the processes of natural speech signal transformation into its internal subjective representation followed by subjective estimation of the results of transformation and making decision. To construct the above algorithms one may study the whole speech perception system both by neurophysiological and psychoacoustical methods. But when interpreting the results of psychoacoustical investigations one must keep in mind that the part of the system resposible for making decisions deals with parameters of subjective but not physical representation of perceived signals.

Analysing the results of many psychoacoustical investigations one has to conclude that we are dealing with the system adapting both to the parameters of a particular set of stimuli and to the instructions, either given by the investigator or generated by the subject. There are reasons to assume that in paired comparison experiments (on sound duration perception, in particular) the subjective estimates of 'longer-equal-shorter' type are more suitable than others. But when instructed to decide 'which of the two is long' (two alternative forced choice procedure), the subjects still succeed in stimulus discrimination. Thus instruction itself may cause a change in the correlation between the internal system of the subjective estimates for each stimulus and subject's responses to the stimulus. It also appears, that only by changing the instruction one can alter the 'observed' accuracy of detecting the deviations in stimulus parameters manifested in the value of the difference threshold when three categorical responses of the 'longer-equal-shorter' type are permitted /2/.

It is common observation that the classical 'point of subjective equalty' obtained in the 'constant stimuli' experiments usually falls into the center of the investigated signal parameter range, regardless of the particular mode of the studied parameter and the range value applied. Hence the obtained 'phonemic boundary' (if speech-like stimuli are involved) and the magnitude of differential limen may depend on the characteristics of a particular set of signals. Sometimes the influence of signal parameters and rigid instructions is so strongly manifested that the subject's responses do not correspond to the real physical characteristics of the signals used. Thus being presented with a steady-state vowel in the given stimulus set the subject gives under certain instruction the response - 'syllable'/1/. It follows from the above that when responding to presented signals with predetermined phonetic categories the subjects do not necessarily perceive those categories. It seems likely that the categories may be used as labels to mark the observed differences in signal parameters which are not directly connected with the given phonetic categories. Therefore in search of the numerical parameters of the speech perception system one should use signals and instructions compatible with the natural conditions of speech perception, because natural speech may be regarded as a very specific set of stimuli processed under almost unknown instructions.

To construct a reliable model of the human speech perception system it is necessary to understand among others the main principles of the system readjustment to changing external conditions whether it is the signal set parameters or the instructions that change. Some preliminary results demonstrating the possibility of these principles investigation were obtained when we studied the perception of speech-like sound duration with the paired comparison method.

The functional model of the subject's behaviour in the experiment with the constant stimuli method must include a unit for the comparison of a tested signal with a 'criterion' whether the identification or paired comparison takes place. The subject uses his internal 'criterion' in the identification task or the 'criterion' given by the investigator in the form of the standard signal in the comparison task. In fact the measurement and comparison of subjective durations take place in neural network, so one must take into account possible signal transformation errors ('noise') and the threshold character of neuron reactions. Therefore at least three independent quantitative parameters of the model should be taken into account: mean square root error (on the supposition of normal distribution) and difference thresholds for positive and negative increments of signal durations. It has been shown in /4/ that the only possibility to get the above parameters is to apply in psychoacoustic experiments 'longer-equal-shorter' or 'same -different' responses. The classical two alternative forced choice procedure cannot give the necessary model parameters.

While studying sound duration discrimination with the stimulus sets containing different standards and the 'same-different' response procedure the familiar results have been obtained: the difference threshold has appeared to grow up with increasing standard duration, the relative difference threshold being almost constant /4/. This effect could be realized in the algorithm of speech signal analysis providing that the continuous nonlinear subjective scale for durations was introduced into the algorithm. But when the subjects were presented with a large set of signals, consisted of several subsets with its own standard each, they developed the idea of an undivided set, for which the single difference threshold (positive and negative) was established /5/. Evidently the rule of relative difference threshold constancy does not work in the case, which is totally inconsistent with the concept of the nonlinear continuous scale for subjective durations.

Suppose however that there exist a restricted linear subjective scale and special mechanism of projecting parameters of the particular set of signals onto it, which provides effective use of the scale, with difference threshold being constant in the units of the scale. Then to obtain an efficient algorithm we must know what is projected onto the scale and how it is done.

Speaking of what is projected let us consider two possibilities. 1) Subjective parameters of a particular set of signals are transformed so that their minimum and maximum values fall onto the initial and final points of the scale respectively. 2) The maximum values of negative and positive differences in the parameter between two compared signals fall onto the above mentioned points of the scale. Accordingly the difference thresholds measured in experiments are not unlikely to turn out a constant if related to the whole experimental range either of signal durations or of differences of compared signal durations.

The sign of duration difference in the signal pair may be changed in two different ways: 1) when signals in a pair are presented in the permanent order (standard-test) throughout the experiment the test is made longer or shorter than the standard; 2) when the test signal always exceeds the standard in duration their order in the pair varies (standard-test or test-standard). When the stimulus set is organized according to the first way the range of signal durations is equal to that of differences in signal durations. When the set is organized according to the second way the duration range is a half of that of duration differences. Now comparing the relative thresholds obtained in both cases we can find out what is projected onto the suggested subjective scale.

The independent values of the difference thresholds were obtained in the experiments with tonal pulses /4/ and steady-state vowels /5/. Then the difference between the positive and negative thresholds was calculated and used as a measure of the subjective differential sensitivity. Let us call it the insensitivity zone.

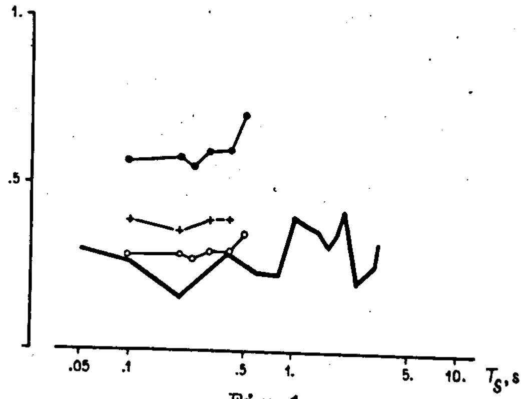


Fig.1

The normalized widths of the suggested zone are plotted on fig.1 against the standard stimulus durations. The normalization was achieved by dividing corresponding values of the zone size either by the stimulus duration range or by the range of the stimulus duration differences. The crosses mark the results obtained when the stimuli have been organized according to the first way mentioned above. The circles mark the results obtained with the stimuli organized in the second way, the filled ones representing the normalization by the range of the stimulus durations and the unfilled ones - by the range of the stimulus duration differences. For the comparison the thick line represents the normalized doubled differential limens for the duration of the gap between two acoustical clicks obtained in /3/, where the experimental stimulus set was organized according to the first way.

As far as one can judge from the picture the normalization by the range of the duration differences brings into a good agreement the results obtained in a quite different experimental conditions. So it may be concluded that it is the difference of the compared signal parameters that is projected onto the suggested subjective scale.

Though the collected data are insufficient to make any final conclusion, they nevertheless show a possible way of studying the processes of creating the internal psychological representation of natural speech signals. It may also be supposed that in search of the mechanisms of projecting the signal parameter onto the proposed subjective scale, one must investigate the dynamics of the subject's responses to one and the same stimulus throughout the experiment.

REFERENCES

/1/ Dechovitz D., Mandler R., "Effects of transition length on identification and discrimination along a place continuum", Hask. Lab. Stat. Rep. on Speech Research, SR-51/52, pp. 119-129. 1977.

√2/ Fernberger S., "Instructions and the psychological limen", Amer. Journ. Psychol., v.43. 1931.

/3/ Getty D.J., "Discrimination of short temporal intervals: A comparison of two models", Percep. & Psychophysics, v.18, pp.1-8, 1975.

/4/ Ventsov A.V., Malinnikova T.G., "Modelling the subjective mechanism of duration comparison", in Issledovanije modelej recheobrazovanija i rechevosprijatija, Leningrad, Nauka, 1981, pp.19-36.

/5/ Ventsov A.V., "What the differential subjective sensitivity on duration depends on?", in Doklady X Vsesojuznoj Akusticheskoj konferencii, Moscow, 1983, sec.X, pp.18-21.

N-DIMENSIONAL METRICAL FORMALISM OF THE PERCEPTUAL SPACE OF POLISH PHONEMES

Wojciech Myślecki, Wojciech Majewski

Institute of Telecomunication and Acoustics
Technical University of Wrocław, Poland

ABSTRACT

This paper presents a method of metrical scaling of nonmetric perceptual space of Polish phonemes perceived by listeners in a presence of additive disturbances and frequency distortions. The experimental material consisted of 10 confusion matrices of Polish phonemes obtained by means of subjective tests for 10 various listening conditions. It was assumed that the confusion matrices estimate the subjective proximity between the phonems. The Shepard's algorithm of N-dimensional analysis of proximity [1,2] was used to establish a space arrangement of investigated phonemes.

INTRODUCTION

The over-all effects of additive disturbances and frequency distortions upon the average intelligibility of human speech are by now well-understood. Most of the existing studies present the results in terms of the articulation score, i.e. the percentage of the spoken words, logatoms, syllables or phonems that the listeners hear correctly. As a consequence, therefore, all of the listener's error are treated as equivalent and no information of the perceptual confusion is available. Perhaps the major reason that confusion data are not a popular subject of investigations is the time and cost of collecting them. An example of the investigations, where phoneme confusion data were used, is the study of Myślecki and Majewski [3] related to the mean conditional entropy of transmission channel (CETC). The base to calculate CETC was a phoneme confusion matrix which, in this case, constituted an experimental estimator of channel matrix for given transmission condition.

The fact that the authors were in possession of the phoneme confusion matrices for various transmission conditions are urged them to a closer look at the problem of mutual configuration of Polish phonemes in the listener's perceptual space.

It was decided to applicate the Shepard's multidimensional scaling technique [1,2] as to obtain a metrical formalism of the nonmetric perceptual space of phonemes. In this technique it is necessary to determine:

- experimental similiarities of investigated objects,
- a strategy of conducting the iterative process,
- optimal values of the constant multipliers for vectors for the approach to monotonicity and to minimum dimensionality,
- number of the iterations before a rotation to principal axes and for terminating the iterative process.

After this it is possible to estabilish a final resolution consisted of:

1° the minimum number of dimensions of the Euclidean space required such that the distances in this space are monotonically related to the initially given proximity measures

2° an actual set of orthogonal coordinates for the points in this minimum space.

The aim of this study is to determine the above mentioned values and solutions for 36 Polish phonemes.

EXPERIMENTAL PROCEDURE

Subjective measurements were carried out for 10 different conditions of speech transmission obtained by means of specially designed model of transmission channel. Masking noise and another additive disturbances of different levels (white noise, overhearing, hum) and frequency distortions (band limiting) were introduced to change a quality of speech transmission.

The test material consisted of phonetically and structurally balanced logatom lists (one or two syllable nonsense word lists) that were read by professional male speaker. For each condition of transmission, i.e. for each measuring point, four lists of 100 logatoms (1520 phonemes) were read. The listening tests were carried out in a listening studio by means of SN-60 *Tonsil* headphones.

Table 1. The transmission conditions and phoneme intelligibility scores for 10 measuring points.

Nº	I_{ph} %	S/N_n* dB$_A$	S/N_d* dB$_A$	Type of disturb.	Band limitation
1	73	-3	X	X	200÷4000Hz
2	75	-3	X	X	600÷2000Hz
3	86	+3	X	X	600÷2000Hz
4	87	+6	+6	hum	400÷2500Hz
5	89	+6	+6	overhear.	400÷2500Hz
6	90	+12	+6	hum	400÷2500Hz
7	91	+6	X	X	200÷4000Hz
8	92	+12	+6	overhear.	400÷2500Hz
9	95	+12	+18	overhear.	400÷2500Hz
10	99	+30	X	X	400÷2500Hz

I_{ph} phoneme intelligibility
S/N_n signal to white noise ratio
S/N_d signal to additive disturbance ratio
* S/N_n and S/N_d were measured independently before band limitation.
X without a disturbance

The listening team consisted of 12 subjects (age: 20÷24) of normal hearing. The measuring procedure was based on ISO recommendation (DP 4870, 1976). As the results of this experimental procedure applied to each of 10 measuring points a phoneme intelligibility and confusion matrix were obtained (data for all listeners and 4 logatom lists have been pooled). Table 1 summarizes the articulation data obtained for all of 10 measuring points. The S/N ratios and band limiting conditions are there also given.

METHOD OF COMPUTATION

Generally, the problem of multidimensional scaling is to find N points whose interpoint distances match in some sense (here, in the rank order sense) the experimental proximity measure (here, the confusions between 36 Polish phonemes). In this study all the computations were carried out in accordance with a program presented in [1,2], where a computing time conserving strategy was adapted. In the choosen strategy the iterative procedures start with larger values of a constant multiplier α for vectors for the approach to monotonicity

$$\alpha_{ijk}=\frac{\alpha[s_{ij}-s(d_{ij})](x_{jk}-x_{ik})}{d_{ij}} \qquad (1)$$

where

$$d_{ij} = [\sum_{k=1}^{N-1}(x_{ik} - x_{jk})^2]^{½} \qquad (2)$$

x_{ik} = coordinate for vertex *i* on axis *k*
s_{ij} = rank of the proximity measure
$s(d_{ij})$= rank of the distance (in N-dimensional space) corresponding to s_{ij}

and a constant multiplier β for vectors for the approach to minimum dimensionality

$$\beta_{ijk}=\frac{\beta[s_{ij}-\bar{s}](x_{jk}-x_{ik})}{d_{ij}} \qquad (3)$$

where
$\bar{s}$=the mean of N(N-1)/2 proximity measures.

Larger values of α and β promote faster though less accurate convergence between the configuration in the perceptual space and its N-dimensional Euclidean formalism. Next, when the criterion for terminating (departure from monotonicity) the iterative process

$$\delta = \frac{2\sum_{j=2}^{N}\sum_{i=1}^{i-1}[s_{ij}-s(d_{ij})]^2}{N(N-1)} \qquad (4)$$

attains its minimum, a rotation to principal axes is performed. The results of the rotation could not be taken as the final solution, but they can however be used to estimate the number of dimensions that can be eliminated. The final solution can then be reached (in a "new" reduced space) by iterating with a small value of α and β set to 0.

CALCULATIONS AND RESULTS

At a start of calculations we have to determine the values of α and β constant multipliers. Shepard [1] has undertaken a systematic exploration of the effects of these two parameters upon a convergence. He used however a set of artificial proximity measures, i.e. a known distance function and a known configuration, so it was not obvious, that his results could be directly adapted to phoneme perceptual space. For α=0.4 and β=0.2, recommended by Shepard, we calculated the δ(n) criterion (4) for matrix № 1 (worst transmission conditions) and for matrix № 10 (best conditions). The obtained curves δ(n), n-number of iterations, are quite similiar for matrices 1° and 10° and pass through a minimum for n=3 and then increase again (see Fig. 1).The comparison of the curves from Fig. 1. with the Shepard's appropriate curve [1] shows their similiarity, hence we can conclude that the investigated configuration has no strong influence on the choice of α and β multipliers.

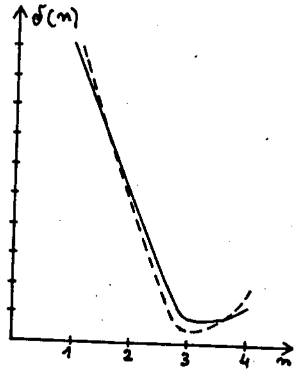


Fig. 1. Departure from monotonicity δ(n) for the worst (dashed line) and best (solid line) transmission conditions.

To the further calculations for theall of 10 confusion matrices it was decided to rotate to principal axes after the third iterations. The results obtained for 10 matrices after the rotation have shown, that the fractions of variance accounted for by the two first (in order of their importance) rotated axes were from 0.991 to 0.998.

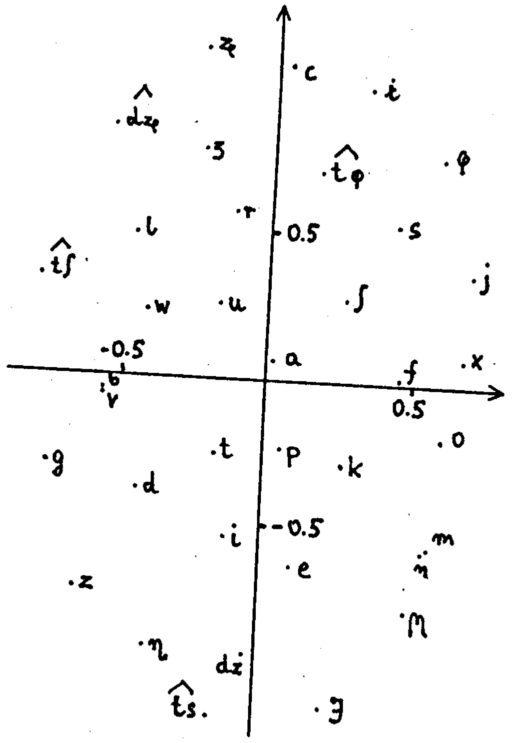


Fig. 2. Final configuration for matrix № 10 (best transmission conditions)

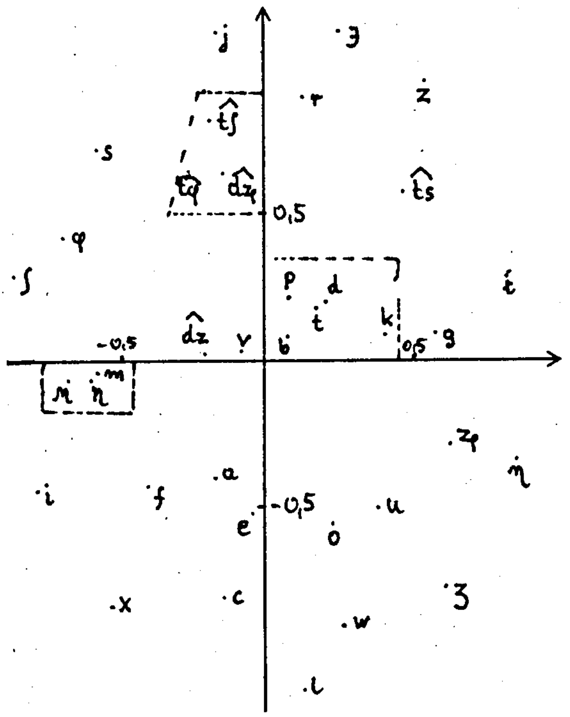


Fig. 3. Final configuration for matrix № 5 (median transmission conditions)

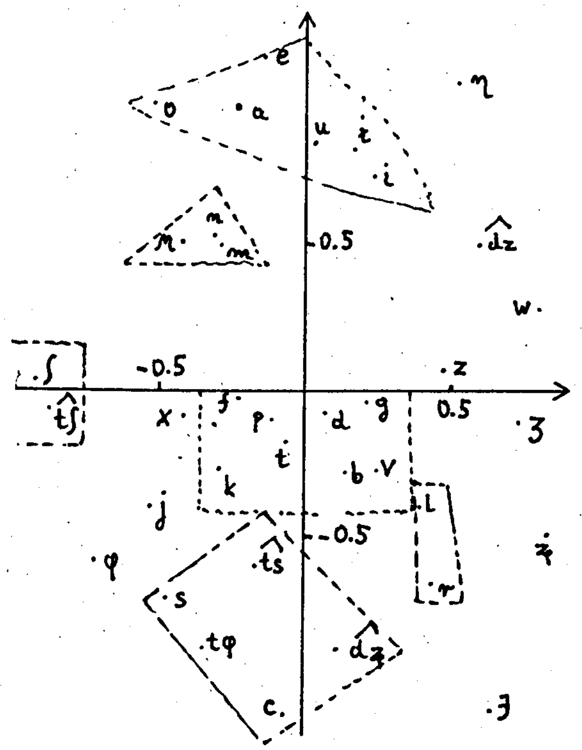


Fig. 4. Final configuration for matrix № 1 (worst transmission conditions)

From this the sufficiently close metrical formalism of perceptual space of 36 Polish phonemes might be inferred to have two dimensions. The final metrical configurations for the all of 10 investigated confusion matrices have been reached in the two-dimensional Euclidean space with the two multipliers setted so that α=0.2 and β=0. The departure from monotonicity δ(n) attained its minimum during the iterations from 49 to 58, and its value was between 1.8*10⁻³ and 2.9*10⁻³. The examples of the final two-dimensional configurations of 36 Polish phonemes for the confusion matrices № 10,5 and 1 are shown in Fig. 2÷4.

SUMMARY

In this study an attention was focused on the application of Shepard's multidimensional scaling method to achieve metric formalism of Polish phoneme perceptual space. It was proved that the strategy and the parameter values recommended by Shepard [1,2] enable N-dimensional scaling of the phoneme perceptual space. The two-dimensional Euclidean space was sufficient for metric representation of arrangement of 36 Polish phonemes with an error, i.e. a departure from monotonicity, less than 0.3%.

REFERENCES

[1,2] Shepard R.N., The analysis of proximity: multidimensional scaling with unknown distance function, Psychometrika, 1962, 27, part I, pp.125-140, part II, pp.219-246

[3] Myślecki V., Majewski V., Relations between subjective and objective measures of speech transmission quality evaluation, Proc. 6-th FASE Congr., 1986, Sopron, Hungary

CAN WE PREDICT F'2 BY MEANS OF A SIMILARITY MEASURE ?

Denis TUFFELLI & Haiyan YE

Laboratoire de la Communication Parlée (ICP Unité Associée au CNRS)
INPG-ENSERG 46, Avenue Félix Viallet
38031 GRENOBLE CEDEX, France

ABSTRACT

The prediction of F'2 is an important aspect for vowel perception. Several prediction models have been proposed in the recent years. In these studies, relationships with the Center of Gravity (in particular with broad band integration) are important. In this paper we propose a new approach for the prediction of F'2 by means of measures of similarity and/or dissimilarity. Several algorithms have been tried including an integration by a critical distance dynamic programming (CDP) and a critical distance transformation (CDT). The evaluation tests are carried out with two kinds of data: vowels formants frequencies and synthetic vowels. The results show that the CDT with a simple euclidean distance give good results. This transformation could retain the phonetic qualities of a sound and give us a good spectral representation for a speech recognition system.

INTRODUCTION

Previous works have underlined two interesting phenomenons: center of gravity of spectral peaks and the F'2 of vowels /1-4/.

The center of gravity (CG) found at Leningrad, is the rough estimate by listeners of two formants F1 and F2 with one formant Fv. In short, a listener hears a first sound made up by F1 and F2 then he is asked to vary the Fv frequency of a second sound in order to find the "best" Fv. If the gap between F1 and F2 is less than 3.5 Barks (so called critical distance) the listener adjusts Fv between F1 and F2 (near the center of gravity), else F1 or F2 is found as the best value for Fv.

With F'2 the principle is similar but the first sound is made up by four formants F1,F2,F3,F4 and the second sound by two formants F1,Fv. The best Fv is called F'2 (effective second formant).

We found that it is hard to simulate both these experiments with a machine "operator" instead of a listener. It is the aim of this paper to describe the machine operators we used.

With CG as a first attempt, we could try to compare two sounds by an euclidean distance D on the spectra Rf(F1,F2) and Tf(Fv) (with pure peaks at F1, F2 and Fv). The best Fv* frequency could be defined as:

$$D(Fv^*) = \min_{Fv} D(Fv)$$

with any gap, between F1 and F2, the result is F1 or F2. So by extrapolation on formants with small bandwiths, we consider that this result is not correct.

Different hypothesis can be made. For instance, with F'2 previous works /5/ have lead to three hypothesis for F'2 perception:

1) after one broad band integration a feature extractor detects F'2 as a parameter for vowel identification.

2) F'2 is a by product of a classification process.

3) F'2 is a by product of a similarity or dissimilarity evaluation of the auditory system.

In this paper we have chosen the 3rd hypothesis. We will use a distance measure between two sounds S1 and S2, so called afterwards D(S1,S2). The basis parameters will be two spectra Rf and Tf with N components on a mel scale.

COMPLEMENTARY TESTS

Of course we measured the machine operators quality by the obtained values on F'2 and CG, but we thought that it was not sufficient. We used two others tests:

The first complementary test is the following one. Listeners were asked to determine the boundaries between vowels pair. That is, for instance for one vowel pair V(F1,F2,F3,F4) and V'(F'1,F'2,F'3,F'4) we generated intermediate sounds Ui by formants interpolations. Therefore these formants had the values:

$$b_i F1 + (1-b_i)F'1$$
$$b_i F2 + (1-b_i)F'2$$
$$\ldots \text{ etc.}$$

Where b_i is a value between 0 and 1.

Then we determined with the listeners the i* value which gives the maximal ambiguity between the two vowels V and V'. We compared the obtained values i* with those given by the operators (for one pair the machine boundary i is roughly determined by the equality D(V,Ui)=D(Ui,V').

The second complementary test is the following checking: with two inputs spectra Rf(F1,F2) and Tf(F1,Fv), the best Fv frequency must be equal to F2 (if the amplitudes are the same). This condition seems obvious but it is not necessarily verified with dynamic programming based algorithms.

ALGORITHMS PRINCIPLES

A Critical Distance Dynamic Programming Algorithm (CDP)

The classic distance measures compare two spectra ,component by component, at the same frequency. If we draw a graph with Rf on the x axis and Tf on the y axis, in this case the followed path is the diagonal. Some people proposed that any kind of paths should be possible /6-8/. They used dynamic programming to get the best path. Each graph node (with coordinates x,y) had a weight which was computed by an elementary distance d(Rx,Ty). The obtained results were not very satisfactory. Following this idea, we propose here that an horizontal or vertical segment is the result of an "integration" (Fig.1). The maximal lenght of such a segment is 3.5 Barks, that is the maximal warping allowed.

To get the best path we try to get a maximum of an inter-spectrum correlation Cxy which is weighted by a distortion term. This term measures the distance to the diagonal. Cxy is a value tied to each point (x,y) and is defined as:

$$Cxy = Rx*Ty*(1 - ((x-y)/alpha)^2)$$

We can see that the distortion term:

$$(1-((x-y)/alpha)^2)$$

is maximum on the diagonal and becomes small when $|x-y|$ tends towards alpha.

If we take pure peaks at frequencies F1, F2 and Fv, during a "machine experiment" of the center of gravity, the best path comes through the horizontal segment (F1,Fv) (F2,Fv) with $Fv=(F1R_{F1}+F2R_{F2})/(R_{F1}+R_{F2})$. Therefore Fv is the mathematical center of gravity. From the best path we can get Fv. Here we don't compute a real distance.

A Critical Distance Transformation (CDT)

The previous technique is an awfully time consuming one. It is the result of a particular interpretation of the human experiments from an algorithmic point of view. Another interpretation can be that the human results are the consequences of a particular preprocessing. We applied it to a spectral preprocessing we are going to describe. Then the distance to use becomes simple (for instance euclidean type on the CDT preprocessed spectra).

Starting from a spectrum Sf, we get the transformed spectrum Sf* from the formula:

$$S^*f = \max_x \sum_{x}^{x+Cd} Sx (1 - ((f-x)/alpha)^2) \quad (1)$$

with $|f-x|$ smaller than alpha.

The distance to use between two spectra Rf, Tf is:

$$D(R,T) = \sum_{i=1}^{N} || Rf^* - Tf^* ||$$

If we take a spectrum Sf which consists of two pure peaks at frequencies F1 and F2 with amplitudes a1 and a2, we have (providing that F1 and F2 are not too far and Fv belongs to some frequency range):

$$S^*f = a1(1-((F1-f)/alpha)^2) + a2(1-((F2-f)/alpha)^2)$$

S^*f is a parabola with a maximum at the frequency Fv which is the mathematical center of gravity:

$$Fv = (a1F1+a2F2)/(a1+a2)$$

Of course one can find always the center of gravity with more than two spectral peaks between F1 and F2. At last one can demonstrate that the resulting distance is almost linear, in some particular cases, with the gap between spectral peaks.

The maximum in the formula (1) does not seem necessary, but without it we have a too broad integration in our first experiments. Others experiments are necessary.

RESULTS

The alpha and Cd parameters, of the previous section, were tuned to get the best results. Generally the tuning is very difficult because there are sharp discontinuities when two formants are integrated or not.

Moreover we made some modifications on the previous formulas to improve the results, for instance with the CDT algorithm:

- when we worked on real signals, the input LPC spectra were too "soft" for this technique, we had to add a formants enhancement procedure.
- The parabolic terms, like $(1-((f-x)/alpha)^2)$, had to be slightly modified. We improved the continuity of the curves and we introduced a slight dissymmetry in the computation.
- We introduced also a slope term in the expression of the distance D.

The CDP algorithm was very difficult to tune. We had to introduce sizable modifications (Moreover the second complementary test is not verified).

Results with spectral peaks

We use the results of a previous experiment which has been carried out essentially with nine swedish vowels (in Hz) /2/:

	F1	F2	F3	F4	F'2(human)
u	310	730	2250	3300	730
o	400	710	2460	3150	720
O	360	1690	2200	3390	1720
a	580	940	2480	3290	960
y	255	1930	2420	3300	2010
U	280	1630	2140	3310	1730
e	375	2060	2560	3400	2370
ae	605	1550	2450	3400	1960
i	255	2065	2960	3400	3210

The estimated F'2 by CDP and CDT (with peaks as inputs) are as following (in Hz):

	F'2(CDP)	E_{abs}	F'2(CDT)	E_{abs}
u	742	0.08	740	0.06
o	725	0.04	720	-0.01
O	1830	0.4	1880	0.58
a	949	-0.07	950	-0.05
y	2084	0.24	2200	0.58
U	1774	0.16	1800	0.24
e	2216	-0.45	2340	-0.10
ae	1938	-0.07	1770	-0.66
i	3097	-0.24	2980	-0.50
E_{tr}		0.19		0.31
E_m		-0.45		-0.66

Where E_{abs} is the absolute error in Bark.
E_{tr} is the total mean absolute error in Bark.
E_m is maximal error.

We obtained also similar results with the center of gravity. We found that with our methods it is possible to get good results for F'2 or center of gravity. But it is very difficult to obtain good results for both of these experiments (F'2 and CG). More some F'2 values from previous experiments (Bladon & Carlson) seem incompatible and may be these values are also language depending. At last the employed energies are not sufficiently well defined and are difficult to reproduce. For further work it is necessary to get more accurate values from human experiments.

Results with synthetic sounds

We have tested the CDT algorithm with synthetic vowels by using an another criterion (first previous complementary criterion), because this method seems to us promising. This criterion is a correlation coefficient with respect to human phonetic judgements. The comparison procedure is described in /10/.

In Fig.2 one can find a LPC spectrum and a CDT filtered spectrum of the same signal. We can see that higher-spectral components are well integrated. With input FFT spectra, the results are similar.

The correlation coefficient of CDT euclidean distance with respect to human phonetic judgement are between 0.87(test X) and 0.895(test ABX) for the 11 french vowels. This means that CDT has retained a great deal of phonetic information. By comparison the Itakura distance obtained, with this method, the values 0.88(test X), 0.91(test ABX). As an example one can find on figure 3, the distance behaviour between two vowels.

CONCLUSIONS

This study is just a try to predict some perceptual parameters (Center of Gravity and F'2) by means of a measure of similarity. These methods can give us a precise estimation of these parameters. Through this study, we can see that modelization of perceptual phenomena can be conducted by different ways.

The advantage of our methods is that a priori knowledges about formants are not necessary. So they can be applied to any spectra, even consonants. The application of these methods to speech recognition is more delicate and is to be tested. The CDP algorithm does not seem well adapted for that.

The phenomena of F'2 is very closely linked with human phonetic judgement. A preprocessing (similar to CDT) which can not only retain but also enhance F'2 parameter will be certainly a better and robust preprocessing for speech recognition.

ACKNOWLEDGEMENT

The authors wish to thank J.L. Schwartz and M. Mantakas for the fruitfull discussions about speech perception.

REFERENCES:

/1/. L.A. CHISTOVICH, "Central Auditory Processing of Peripheral Vowel Spectra" J. Acoust. Soc. Am. 77(3), 789-805, 1985

/2/. R. CARLSON, B. GRANSTROM, G. FANT, "Some Studies Concerning Perception of Isolated Vowel" Speech Transmission Lab. QPSR 2-3, 1970

/3/. A. BLADON, "Two-formants Model of Vowel Perception: Shortcoming and Enhancement" Speech Communication 2, 305-313, 1983

/4/. K.K. PALIWAL, D. LINDSAY, W.A. AINSWORTH, "A Study of Two-formant Models for Vowel Identification", Speech Commun. 2, 295-304, 1983

/57. J.L. SCHWARTZ & P. ESCUDIER, "Le Système Auditif Humain Comprend-il un Mécanisme d'Integration à Large Bande ?" 14 JEP, Aix-en-Provence 1986

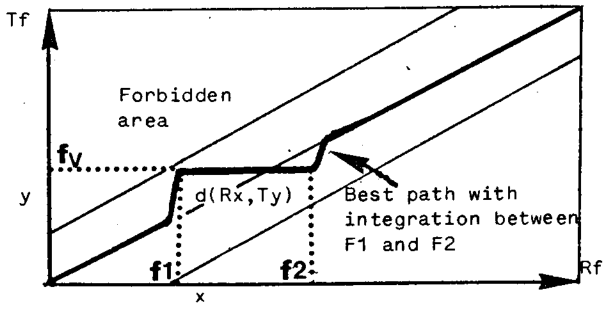
/6/. H. MATSUMOTO & H. WAKITA, "Frequency Warping for Nonuniform Talker Normalization" Int. Conf. Acous. Speech Sig. Proc., pp566-569, 1979

/7/. K.K. PALIWAL, W.A. AINSWORTH, "Dynamic Frequency Warping for Speaker Adaptation in Automatic Speech Recognition" Journals of Phonetics 13, 123-134, 1985

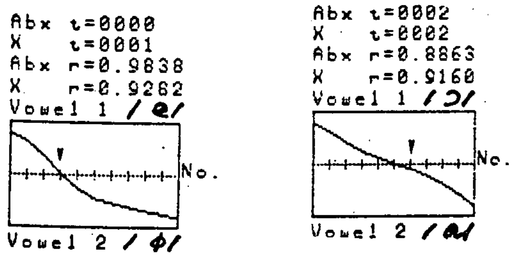
/8/. M. BLOMBERG & K. ELENIUS, "Nonlinear Frequency Warping for Speech Recognition" Int. Conf. Acous. Speech Sig. Proc., 49.2, 1985

/9/. L.A. CHISTOVICH, V.V. LUBLINSKAYA, "The Center of Gravity Effect in Vowel Spectra and Critical Distance Between the Formants: Psychoacoustical Study of the Perception of Vowel-like Stimuli" Hearing Reseach 1, pp 185-195, 1979

/10/. D. TUFFELLI & H. YE "Distortion Measures Evaluation Using Synthetic Sounds and Human's Perception" Montréal Symposium on speech recognition, (1986)



A path with "integration"
Fig.1.



Two examples of distance behaviour (with a CDT preprocessing) between a vowel pairs (V and V'). A comparison can be made with human phonetic judgements (cf /10/ for details). Here we have two vowel pairs /e/-/Φ/ and /ɔ/-/a/. t is an error number with respect to the perceptual boundary. r is a correlation coefficient between distances and perceptual data. Abx and X are two kinds of experiments. The zero crossing points are the discrimination points of the distance D. The arrows are human perception boundary. On the x-axis the numbers of the intermediate sounds Ui (from V on the left to V' on the right). On the y-axis the value D(Ui,V')-D(V,Ui).
Fig.3.

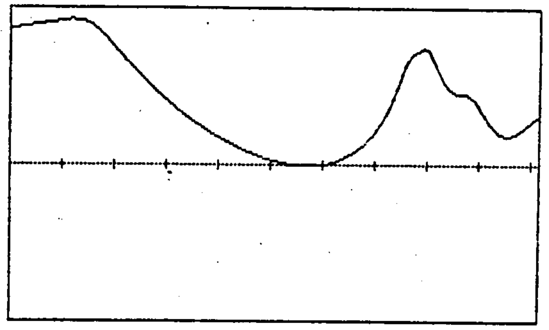


Fig.2a. LPC spectrum of a /i/
y axis: dB, x axis: Mel scale

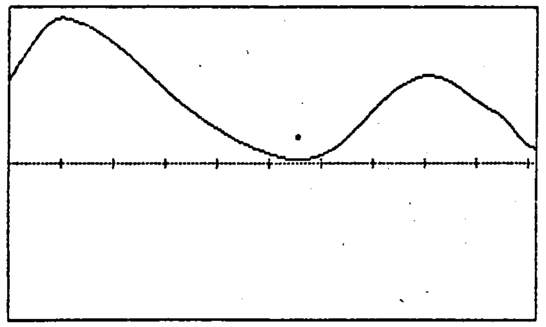


Fig.2b. CDT spectrum of the same /i/
y axis: dB, x axis: Mel scale

RESYNTHESIS AND MATCHING EXPERIMENTS ON AN AUDITORY THEORY OF MALE/FEMALE NORMALISATION

R.I. DAMPER*, R.A.W. BLADON**, R.W. HUKIN* and G.N.A. IRVINE*

* Department of Electronics and Computer Science
University of Southampton
United Kingdom

** Phonetics Laboratory
University of Oxford
United Kingdom

ABSTRACT

Variability between speakers, particularly those of different sexes, poses problems for speaker-independent speech recognition. Recently, it has been suggested that much of this variability could be minimised using a suitable computational model based on known or assumed details of human auditory processing. We are attempting to test this notion experimentally by resynthesising speech which has been processed by the model and studying its perceptual nature.

INTRODUCTION

Current approaches to speech recognition are characterised by the use of signal and pattern processing techniques which are "general" in the sense that little account is taken of the fact that the input (speech) has some very particular properties. As a consequence, spectral representations are typically used in which the coordinates are decibels (relative to some reference level) and logarithmic hertz-frequency, in spite of perceptual evidence that the human auditory system uses a loudness-density versus tonality representation. It is now widely held that the exploitation of knowledge about human speech processes (production and perception) is a pre-requisite for further, significant advances in speech technology, embracing recognition, synthesis and coding. Indeed, there have been several recent attempts to embody at least some of the current understanding of auditory perception into computational models ("auditory models"). The hope is that such models may prove to be more effective as pre-processors for recognition and coding than are traditional speech analysers.

One area where conventional signal processing and statistical pattern matching techniques have proved inadequate is in the handling of speaker variability such as arises from speaker sex and age differences. This sort of variability poses clear problems for speaker-independent recognition. Recently, Bladon and his coworkers [1] have suggested that many of these differences could be minimised (in vowel spectra at least) using a suitable "auditory normalisation" model. In Bladon's model, a perceptually-motivated "auditory spectrum" (obtained by transformations of the spectral coordinates and convolution with a filter intended to represent peripheral frequency analysis) undergoes linear shifts in the tonality (bark scaled) dimension. We believe that claims for the normalising potential of the model are, to some extent, testable by resynthesising speech direct from the bark-shifted auditorv spectral representation.

For instance, resynthesising a one-bark-incremented version of a male vowel spectrum, but with voicing appropriate to a female speaker, should induce listeners to report no change in perceived vowel quality. On the other hand, playback of the "incremented" vowel with the male voicing retained should yield shifts in perceived quality. Indeed, it may even prove possible to effect an automatic transformation of male to female speech, or vice versa. We are attempting to substantiate these ideas experimentally and this paper reports on the early stages of the work.

The paper is structured as follows. First, previous work on auditory models and speaker normalisation is reviewed. The implementation of one particular model (essentially that due to Bladon **et al**) is then described. Subsequently, the resynthesis operation is described and a number of problems identified; the most important being that certain of the "forward" (acoustic-to-auditory) transformations effect a data-reduction and so are inherently non-invertible. Finally, some early results of listening experiments using the resynthesised speech are presented.

AUDITORY MODELS AND NORMALISATION

There is considerable variability in the acoustic realisations of the same speech sounds by different speakers [2]. Thus, the human auditory system has the ability to perceive as phonetically equivalent vowels of markedly different formant (and voicing) structure. This normalisation process implies an ability to make allowances for different vocal tract sizes and shapes. In attempting to mimic this ability in model systems, we might take either of two somewhat different approaches. One possibility is to adopt a speech production viewpoint whereby some dimensional scaling is effected according to supposed vocal tract characteristics. The alternative speech reception point of view leads us to search for an explanation of normalisation ability on the basis of known or assumed details of auditory processing i.e an "auditory model". For instance, the hypothesis of Potter & Steinburg [3] that a particular pattern of stimulation on the basilar membrane might be identified as a given sound, within limits independant of displacement along the membrane, is one possible mechanism for normalisation.

Auditory models are generally based, at least in part, on the concept of the auditory filter originally proposed by Fletcher [4]. He suggested that the peripheral auditory system behaves as if it contained a bank of filters, with a continuum of centre frequencies. The output of such a filter bank is usually termed an 'excitation pattern' since it is meant to represent the degree of activity (or excitation) evoked by a particular sound at some unspecified level of the auditory system. Schroeder suggests that the excitation pattern, E(z), could just as well be thought of as mean-squared amplitude of the basilar membrane motion at place z [5]. His model uses a rather broad auditory filter shape estimated from the somewhat dated masking experiments of Zwicker [6]. More recent evidence from experiments taking into account factors such as off-frequency listening suggests that filter shape should be much narrower [7, 8].

The auditory modelling approach to speaker normalisation is exemplified by the work of Bladon **et al** [1]. In this model, the spectral frequency axis is transformed from hertz to bark prior to filtering using the filter shape described by Schroeder (see [5]). Because of the broadness of these filters, there is a "smearing" of the spectrum with a substantial loss of resolution rendering different realisations of the same vowel more alike and removing much of the fine detail due to voicing. Following a conversion from intensity to loudness density to yield an "auditory spectrum", a linear shift in the bark dimension is effected. From the data presented, it is apparent that such shifts can have a normalising effect, by bringing vowel spectra for male and female speakers into reasonable coincidence. Following this work, Holmes [9] attempted to investigate the perceptual effect of bark-scaled shifts in formant frequencies using a speech synthesis-by-rule system. Preliminary results suggest that, for some vowels at least, an approximately constant bark difference between F_1 and F_2 is necessary to maintain phonetic quality.

The principal objection to the Bladon model is the use (following Schroeder) of a wideband auditory filter. Klatt [10] has observed that male and female speech can be made to look similar merely by increasing the bandwidth of the analysis filter in the spectrogram. Thus, caution must obviously be exercised to ensure that vowel identity is preserved when the variance is reduced in this way. There is little virtue in making the same vowel from different speakers appear more alike if different vowels from the same speaker also look more alike. It is only to be expected that representations preserving gross features only of the spectrum shape would be more likely to improve similarity between male and female vowel spectra, since a lot of information (whether relevant or not) has been discarded. It is important to know, therefore, what information is left in the smoothed spectrum representation. One way to discover this might be to conduct listening experiments with speech resynthesised directly from the auditory spectrum. Such resynthesis also offers a means of studying the perceptual effect of bark-scaled shifting, much as Holmes has done, but with real (rather than synthetic) speech.

One difficulty with this approach is apparent. If the auditory system really does perform a frequency smearing operation, then the resynthesised speech will naturally be subjected to this operation. I.e. the speech will be smeared "twice", hence possibly invalidating the idea of testing by resynthesis. Evidence that the smeared, auditory representation is adequate to retain vowel identity is given below. Of course, it may be that a second application of the smearing has relatively little effect, most of the data reduction being done on the first application. One early priority, therefore, must be to compare smoothed and unsmoothed speech for perceptual differences.

IMPLEMENTATION DETAILS OF THE MODEL

An auditory model based closely on that described by Bladon **et al** [1] has been implemented on a DEC MicroVAX computer. As well as "forward" acoustic-to-auditory transformations, some provisional "inverse" auditory-to-acoustic transformations have also been included to allow resynthesis.

Forward Transformations

The excitation patterns for the auditory model are computed as follows. The power spectrum S(f) for the input speech is computed over (Hamming weighted) time windows of approximately 32 ms using an FFT algorithm. The windows are advanced in steps of 8 ms for each new segment. The power spectrum (with units of V^2/Hz) is then transformed to a critical band density (with units V^2/bark) using the formula:

$$S(z) = S[f(z)] \cdot \frac{df}{dz}$$

The mapping between frequency, f, and critical band number, z, is approximated by the expression due to Traunmuller [11]:

$$f = \frac{1960(z + 0.53)}{(26.28 - z)}$$

Thus, the critical band density is computed from the spectrum by the f -> z mapping followed by multiplication with the density conversion factor.

Next, an excitation pattern is computed from the critical band density by convolution with the auditory filter frequency response. The specific filter used at this stage is Schroeder's (as described in [5]) but we intend to investigate the use of different filters. The convolution operation is equivalent to using a filter bank analysis, but is more convenient as the filter shape (as defined by Schroeder) is invariant across the bark scale, and no weighting has to be applied to account for changes in filter bandwidth.

The Bladon model differs slightly from Schroeder's in the calculation of the loudness density pattern, which is accomplished by conversion from critical band density to loudness level density in phons/bark followed by a conversion to loudness density in sones/bark. In this work, we have neglected to compute the loudness density pattern: justification for this omission in terms of resynthesis is that the phon curves are fairly flat in the region 200 - 4 kHz where the formants lie, and thus a displacement of the pattern along the bark scale would have little effect on the spectrum.

Once the excitation pattern has been calculated for the input segment, its position on the bark scale can be adjusted before resynthesis in order to investigate the perceptual effects of displacement.

Inverse Transformations

Since the filtering (convolution) operation has effected a data reduction on the original spectrum, it is impossible to recover the full spectrum for resynthesis. Some indirect evidence that the smoothed, auditory spectrum is a reasonable representation from which to resynthesise is given by certain other psychoacoustic findings. Using the relatively broad Schroeder filters, the physical formant pattern is smoothed to just two auditory peaks. This characteristic is consistent with the "centre of gravity" theory advocated by Chistovich and Lublinskaya [12] as well as with experiments in the matching of two-formant synthetic vowels to the full reference vowel - the so-called F-prime paradigm [13, 14]. Thus, the auditory spectrum should in principle be capable of retaining information concerning vowel identity. Confirmation of this notion is given in the work of Hermansky et al [15] who processed all-voiced sentences to show that a "reduced" spectrum produced by auditory filtering (18 critical band filters equispaced in the bark dimension) could yield "intelligible" speech.

The resynthesis operation involves conversion of the critical band density back to a spectral density by multiplication with the inverse density conversion factor, dz/df. However, the smearing operation removes much, if not all, of the voicing information. For the resynthesis process, therefore, two possibilities present themselves. Either the loss of voicing information could be ignored or appropriate voicing could be added. We intend to explore both of these approaches.

Finally, continuous speech output is obtained from the auditory spectra by inverse Fourier transformation using an overlap-add technique [16].

RESULTS

At this early stage, it is only possible to give initial results from some informal listening tests. The oral presentation will describe results of more extensive testing. Speech of telephone quality (low-pass filtered at 3.2 kHz and sampled at 8 kHz with 12-bits resolution) has been processed by the model. Two complete sentences have been studied: a male speaker saying "live wire should be kept covered" and a female saying "the kitten chased the dog down the street". At the resynthesis stage, no extra voicing has been added.

We wished first to examine the effect of smoothing but without shifting in the bark dimension. The speech was subjected to the forward transformations with zero bark shift and resynthesised. The speech output was slightly degraded but speaker identity was retained and the sentence was clearly intelligible. This observation lends weight to the belief that resynthesising is a valid technique for testing auditory models. If anything, the result extends the observation of Hermansky et al referred to above to speech consisting of voiced and unvoiced segments.

Subsequently, the effect of processing the male speech using the model, and including a shift of one bark, was investigated. Again, the speech was intelligible but more severely degraded. We speculate that this additional degradation is principally due to destroying the harmonic relation between voicing-frequency components when a linear bark shift follows the non-linear hertz-to-bark transformation. First impressions, however, were that speaker identity was markedly different. It was not easily possible to assign a perceived sex to the speaker with any confidence.

FUTURE WORK

The major priority is to conduct more formal matching experiments (perhaps using steady-state vowels) with a larger number of listeners.

Informal experimentation so far has not used added voicing. Further work is planned in which the speech spectrum will be deconvolved by cepstral techniques into excitation and envelope components. The envelope alone will be processed by the model and speech resynthesised with a variety of voicing components appropriate to different speakers (and including the natural voicing itself).

There are, of course, many specific details of the model which could be further tested by resynthesis. For instance, there is a good case to be made for employing auditory filters of much narrower bandwidth, such as the rounded-exponential (roex) filters described by Moore and Glasberg [8]. Arguably, in this case, equivalent rectangular bandwidth (ERB) would be a more appropriate frequency scale for shifting than the bark scale.

REFERENCES

[1] Bladon, R.A.W., Henton, C.G. & Pickering, J.B. (1984) 'Towards an auditory theory of speaker-sex normalisation', Language and Communication', 4, 59-69.

[2] Peterson, G.E. & Barney, H.L. (1952) 'Control methods used in the study of vowels', JASA, 24, 175-184.

[3] Potter, R.K. & Steinburg, J.C. (1950) 'Towards the specification of speech', JASA, 22, 807-820.

[4] Fletcher, H. (1940) 'Auditory patterns', Rev. Mod. Phys., 12, 47-65.

[5] Group Report (Fourcin, A.J. et al) (1977) 'Speech Processing by Man and Machine' in "Recognition of Complex Acoustic Signals", T.H. Bullock (Ed), Life Sciences Research Report 5, Abakon Verlag, Berlin.

[6] Zwicker, E. (1963) 'Uber die Lautheit von ungedrosselten und gedrosselten Schallen', Acustica, 13, 194-211.

[7] Patterson, R.D. (1976) 'Auditory filter shapes derived with noise stimuli', JASA, 59, 640-645.

[8] Moore, B.C.J. & Glasberg, B.R. (1983) 'Suggested formulae for calculating auditory filter bandwidths and excitation patterns', JASA, 74, 750-753.

[9] Holmes, J.N. (1985) 'Normalization in vowel perception' in "Invariance and Variability of Speech Processes", J.S. Perkell and D.H. Klatt (Eds), Lawrence Erlbaum Associates.

[10] Klatt, D.H. (1982) 'Speech processing strategies based on auditory models' in "The Representation of Speech in the Peripheral Auditory System", R. Carlson & B. Granstrom (Eds), Elsevier Biomedical.

[11] Traunmuller, H. (1983) 'Analytic expressions for the tonotopical sensory scale', Unpublished manuscript.

[12] Chistovich, L.A. & Lublinskaja, V.V. (1979) 'The "centre of gravity effect" in vowel spectra and critical distance between the formants: psycho-acoustical study of the perception of vowel-like stimuli', Hearing Research, 1, 185-195.

[13] Bladon, R.A.W. (1983) 'A study of two formant models for vowel identification', Speech Communication, 2, 295-303.

[14] Paliwal, K.K., Ainsworth, W.A. & Lindsay, D. (1983) 'A study of two-formant models for vowel identification', Speech Communication, 2, 295-303.

[15] Hermansky, H., Hanson, B.A. & Wakita, H. (1985) 'Perceptually based linear predictive analysis of speech', Proc. ICASSP 85, 509-512.

[16] Allen, J.B. & Rabiner, L.R. (1977) 'A unified approach to short-time Fourier analysis and synthesis', Proc. IEEE, 65, 1558-1564.

ИДЕНТИФИКАЦИЯ ДИКТОРА ПО ЧАСТОТАМ ФОРМАНТ, ИЗМЕРЕННЫМ СИНХРОННО С ОСНОВНЫМ ТОНОМ

ВАЛЕРИЙ ГИТЛИН

Кафедра "Вычислительная техника"
Устиновский механический институт
Устинов, Удмуртия, СССР, 426000

Идентификация диктора по голосу есть процесс идентификации некоторой физической системы, который требует достаточно точного определения параметров системы. В качестве таких параметров могут быть взяты частоты формант и их траектории. Однако процесс выделения формант достаточно труден. Эти трудности объясняются влиянием голосового источника на форму речевого сигнала и влиянием аппаратуры анализа. Учет этих влияний позволил снизить ошибку идентификации I04 дикторов мужчин по траекториям формант парольной фразы с I2,8% до 5,3%.

ВВЕДЕНИЕ

Человека можно представить как некоторую физическую систему. Речевой сигнал есть продукт этой системы и отражает ее конкретные физические особенности. В качестве меры параметров речевого тракта можно использовать частоты формант[I].

Структурная схема выделителя периода форманты показана на фиг. I. Речевой сигнал разделялся формантными фильтрами FF_n (n = I, 2, 3) на формантные полосы. В каждой полосе синхронно с основным тоном измерялся период форманты по методу Кампанеллы и Коултера [2]. Основной тон выделялся пиковым методом по речевому сигналу, ограниченному полосой 4,5 кГц [3]

В разделе I настоящей работы выполнен анализ погрешностей выделения частоты формант. В разделе П представлена методика оптимизации формантных фильтров. Результаты экспериментов по идентификации диктора даны в разделе Ш, раздел IУ - заключение.

I. ПОГРЕШНОСТИ ВЫДЕЛЕНИЯ ЧАСТОТ ФОРМАНТ

А. Влияние голосового источника

Был выполнен анализ реального речевого сигнала. Вычислялась огибающая спектра на интервале открытых (ОС) и закрытых (ЗС) голосовых связок и на периоде основного тона (ОТ) [4] при помощи методики линейного предсказания. Интервалы ОС, ЗС и ОТ выделялись ручным способом по осциллограммам речи. Огибающие спектров для гласной / i / из слова "электричество" показаны на фиг. 2 [5]. Аналогичные результаты получены Ларером и др. [6].

Из фиг. 2 видно, что в реальной речи открывание голосовых связок изменяет частоту и ширину формант, причем меньше всего подвержена изменениям резонансная частота первой форманты $F_{r1} = [F_1^2 - (B_1/2)^2]^{\frac{1}{2}}$. Спектр сигнала, вычисленный на периоде ОТ, был ближе к спектру, вычисленному на интервале ЗС, а не на интервале ОС.

В. Оценка взаимного влияния формант

Пусть сигнал искомой форманты записывается как $U_1 e^{-\sigma_1 t} \sin(\omega_1 t + \varphi_1)$, а сигнал мешающей форманты - $U_2 e^{-\sigma_2 t} \sin(\omega_2 t + \varphi_2)$. Для того, чтобы средняя частота пересечений нуля соответствовала измеряемой форманте, необходимо выполнить условие[9,5]:

$$U_1 > U_2 e^{-(\sigma_2 - \sigma_1)t} \,\Big|\, \omega_1 > \omega_2, \qquad (I)$$

$$U_1 > U_2 \frac{\omega_2}{\omega_1} e^{-(\sigma_2 - \sigma_1)t} \,\Big|\, \omega_1 < \omega_2 .$$

Из формулы (I) видно, что момент максимального подавления мешающей форманты зависит от амплитудных соотношений формант, частот формант и величин их затуханий.

С. Собственные колебания формантного фильтра

Представим речевой сигнал как [I] :

$$u(t) = \sum_{n=1}^{N} (-1)^n U_n^{in} e^{-\sigma_n t} \sin(\omega_n t + \varphi_n) \qquad (2)$$

Если сигнал $u(t)$, определяемый формулой (2) подать на формантные фильтры FF_n (фиг.I), то в каждом фильтре будут возникать собственные колебания с каждым новым возбуждением затухающей синусоиды [7], т.е. с каждым новым возбуждением речевого тракта.

Пусть FF составлен из L резонансных звеньев, включенных последовательно. Передаточная характеристика звена:

$$K_\ell(p) = -G_\ell (p + 2\alpha_\ell)/[(p+\alpha_\ell)^2 + \omega_{o\ell}^2], \qquad (3)$$

где ℓ = I, 2,..., L - номер звена, G_ℓ - коэффициент передачи, α_ℓ - затухание, $\omega_{0\ell} = (\omega_{o\ell}^2 - \alpha_\ell^2)^{\frac{1}{2}}$ - резонансная частота звена с учетом потерь.

Подадим на FF_n с передаточной характеристикой $K(p) = \prod_{\ell=1}^{L} K_\ell(p)$, сигнал вида (2). Выходной сигнал фильтра можно представить как [5] :

$$j\bar{u}^{out}(t) = U_i^{out} e^{-\sigma_i t + j(\omega_i t + \beta_i)} \bar{N}_i(t), \qquad (4)$$

где

$$\bar{N}_i(t) = 1 + \sum_{n=1, n\neq i}^{N} \frac{U_n^{out}}{U_i^{out}} e^{-(\sigma_n - \sigma_i)t + j[(\omega_n - \omega_i)t + \beta_n - \beta_i]} + \sum_{\ell=1}^{L} \frac{U_{n\ell}^{f}}{U_i^{out}} e^{-(\alpha_\ell - \sigma_i)t + j[(\omega_{o\ell} - \omega_i)t + \gamma_{n,\ell} - \beta_i]} \qquad (5)$$

- комплексный поправочный коэффициент, оценивающий влияние мешающих факторов на форму сигнала i-ой форманты, U_n^{out}, β_n - амплитуда и фаза вынужденного колебания на выходе FF_n : $U_{n,\ell}^{f}$, $\gamma_{n,\ell}$ - амплитуда и фаза свободного колебания в ℓ-ом звене FF_n. Формулы для расчета $U_n^{out}, \beta_n, U_{n,\ell}^{f}$ и $\gamma_{n,\ell}$ даны в работе [8] .

Примерный вид графика $|\bar{N}_i(t)|$ для $\sigma_n < \sigma_i$ показан на фиг. 3, где δ - допустимое отклонение $|\bar{N}_i(t)|$ от единицы, T_d - время задержки измерений, T_m - интервал времени, в течение которого возможны измерения параметров формант. Конкретный вид функции $|\bar{N}_i(t)|$ зависит от соотношения значений $\alpha_\ell, \sigma_n, \omega_{o\ell}, \omega_n$.

Анализ выражения (5) выполнялся путем численного моделирования на ЭВМ. Как показывают расчеты, если $\alpha_\ell \geq 3\sigma$, то скорость уменьшения T_d становится незначительной. Поэтому брать $\alpha_\ell > 3\sigma$ нецелесообразно.

Можно показать [7] , что при увеличении числа звеньев фильтра, скорость нарастания $|\bar{N}(t)|$ снижается, а T_d - растет. Для $\omega_{o\ell} = \omega_{oi}$, ℓ = I, 2,..., L ; $\alpha_\ell = \alpha_o$ для ℓ = I, 2,..., L и L = 3 имеем [7] :

$$\alpha_o T_d \cong 6. \qquad (6)$$

FF_n должен обеспечивать малое значение T_d, большое T_m и высокую избирательность по частоте. Эти требования противоречивы и необходима процедура оптимизации для получения наилучших характеристик фильтра.

П. ВЫБОР ПАРАМЕТРОВ ФОРМАНТНОГО ФИЛЬТРА

Было решено [8] остановиться на трехзвенной схеме формантного фильтра (L = 3). Если выбрать $T_d < T_{min}$ = 2 мс , то из формулы (6) получаем, что $\alpha \geqslant 3000$ и полоса пропускания отдельного звена FF_n $\Delta f_n \geqslant \alpha/\pi$ = 955 Гц.

Оптимизация параметров FF_n выполнялась [9] путем поиска минимума функции

$$Q(i,n,j) = U_{n,j}^{out} / U_{i,j}^{out}, \qquad (7)$$

где j = I, 2,..., 6 - индекс фонемы,для которой вычисляется функция Q, i = I, 2, 3 - индекс искомой форманты, n = I,2, 3, 4 - индекс мешающей форманты. При вычислении Q полагалось, что амплитуды формант одинаковы. С целью повышения избирательности диапазон F_2 был разбит на два поддиапазона. Параметры рассчитанных FF_n представлены в таблице, а их частотные характеристики показаны на фиг. 4.

Таблица

Форманта	Звено	Резонансная частота,Гц	Полоса,Гц
F_1	I 2 3	I70 240 280	720 740 740
F_{2-1}	I 2 3	950 I045 I2I5	800 800 800
F_{2-2}	I 2 3	I235 I365 I770	II25 800 800
F_3	I 2 3	24I0 2570 3260	720 925 I045

Ш. ЭКСПЕРИМЕНТЫ ПО ИДЕНТИФИКАЦИИ ДИКТОРА

На этапе предварительных экспериментов дикторы идентифицировались по средним частотам формант, выделенных на стационарных участках гласных. В качестве речевого материала использовались шесть русских гласных (/а/, /о/, /у/, /э/, /и/ и /ы/), произнесенных в составе слов семью дикторами (три женщины, четверо мужчин). В условиях машинного зала было сделано три сеанса записей по семь произнесений каждой фонемы в одном сеансе. Шесть произнесений использовались для обучения, седьмое - как контрольное. Результаты идентификации усреднялись по трем сеансам записи. В качестве меры сходства использовалось Эвклидово расстояние между частотами формант эталонной и контрольной выборки. В качестве формантных фильтров использовались либо рассчитанные формантные фильтры (РФФ), либо фильтры, параметры которых выбирались эмпирически (ЭФФ) [3] .

Число ошибок идентификации мужчин составило 19% и 44% , женщин - 27% и 50% при использовании РФФ и ЭФФ соответственно. РФФ обеспечили существенное снижение количества ошибок идентификации.

На этапе основных экспериментов дикторы идентифицировались по траекториям формант парольной фразы: "С[ода в луже убывала] слабо". В экспериментах участвовало I04 диктора мужчины. Фраза записывалась два раза в одном сеансе испытаний в условиях машинного зала. Первое произнесение использовалось для обучения, второе - как контрольное. Для описания траекторий формант применялся алгоритм Ламиса [I0] . В качестве меры сходства использовалось Эвклидово расстояние между коэффициентами разложения траекторий по ортогональным полиномам для трех формант плюс коэффициент корреляции между эталонной и контрольной выборками.

Ошибка идентификации составила для РФФ 5,3%, а для ЭФФ - I2,8%. РФФ и в этом случае обеспечил уменьшение количества ошибок идентификации.

IУ. ЗАКЛЮЧЕНИЕ

Снизить влияние источника и аппаратуры анализа на точность измерения параметров формант и, тем самым, повысить надежность идентификации диктора можно путем анализа временной функции речевого сигнала синхронно с основным тоном. Параметры формантного фильтра должны быть выбраны оптимальным образом с наиболее короткой импульсной реакцией и с максимально возможной избирательностью по частоте. Траектории формант обеспечивают повышение надежности идентификации диктора по сравнению со средними значениями частот формант стационарных участков гласных.

ЛИТЕРАТУРА

I. Фант Г. Акустическая теория речеобразования.- М.: Наука, I964.- С. 284.

2. Campanella S.J., Coulter D.C. Formant Period Tracker. Pat. USA.№ 3335225

3. Сапожков М.А. Речевой сигнал в кибернетике и связи.- М.: Связь, I963.- С. 422.

4. Гитлин В.Б., Сметанин А.М. О повышении точности измерения параметров формант // Проблемы построения систем понимания речи.- М.: Наука, I980.- С. I09-II5.

5. Гитлин В.Б. и др. Выбор интервалов измерений частоты и ширины формант. Тез. докл.АРСО-Х.- Тбилиси: Мецниереба, I978.- С. 20-22.

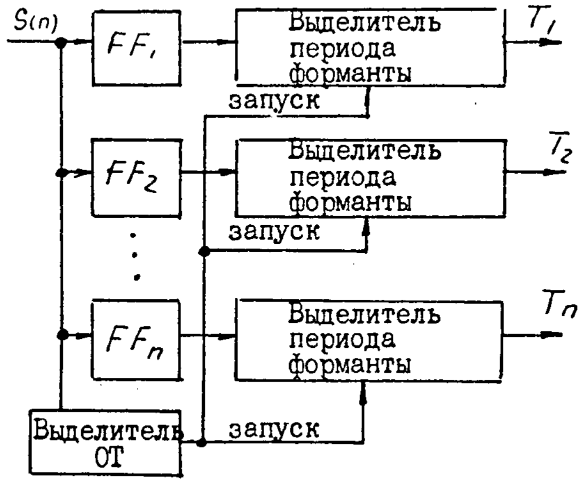
6. Larer J.N., Alsaka Y.A. Childers D.G. Variability in Closed Phase Analysis of Speech."ICASSP85, Tampa, Fla, March 26-29, 1985, Vol 3" New York, N. Y., 1985, 1089-1092.

7. Золотарев И.Д. Нестационарные процессы в резонансных усилителях фазово-импульсных измерительных систем.- Новосибирск: Наука. Сиб. отд.,I969.-С.I76.

8. Гитлин В.Б. К вопросу расчета формантных фильтров методом упрощенного преобразования Лапласа // Автоматические устройства учета и контроля.- Ижевск, I977.- Вып.П.- С. 83-9I.

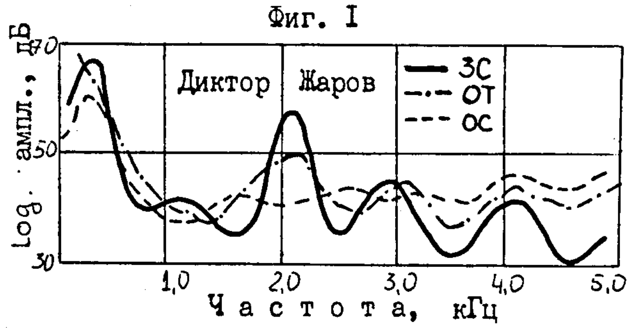
9. Сметанин А.М. Исследование и разработка методов повышения точности измерений параметров формант и голосового источника.- Диссертация на соискание ученой степени к.т.н. Ижевск, I980.

I0. Lummis R.C. Speaker verification by computer Using Speech Intensiti for Temporal Registration IEEE Trans. Audio and Electroacoust 1973, v21, №2, 80-89

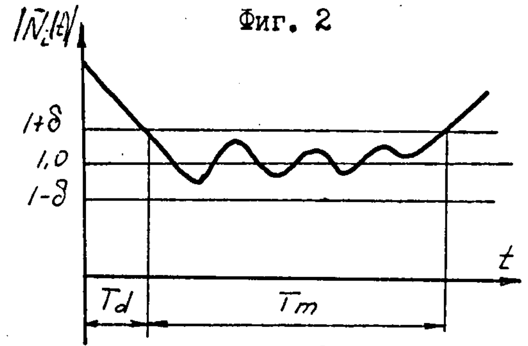


S(n)
FF1
FF2
FFn
Выделитель периода форманты
Выделитель периода форманты
Выделитель периода форманты
запуск
запуск
запуск
Выделитель ОТ
T1
T2
Tn

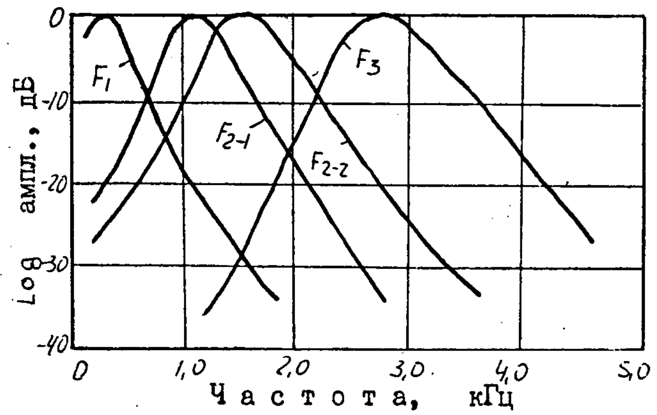
Фиг. I



Фиг. 2



Фиг. 3



Фиг. 4

A PRELIMINARY STUDY OF SPEECH RATE PERCEPTION

Pierre HALLE

Labo. de Psychologie experimentale.
CNRS, EHSS, Paris, FRANCE

ABSTRACT

Speech rate, whether physical or subjective has often been used as an experimental condition in speech perception studies, however, subjective speech rate itself has rarely been studied. One could simply assume that it is conveyed by the physical syllabic rate or, equivalently, by the periodicity of the vocalic cycle. This study is an attempt to examine the effect of intra-syllabic structure on perceived tempo. In particular, the vowel (acoustic) duration is shown to play a significant role, at least for producing a slow rate sensation.

INTRODUCTION

The importance of speaking rate in speech perception has been assessed in numerous studies ([1],[2],[3]) providing evidence that listeners identify speech segments in a rate dependent manner.

Although a large part of speaking rate variation is due to changes in the amount of pausing ([4]), it is rather the articulation rate, i.e. speaking rate in pause free stretches of speech, which could tune speech perception. The precise range which conveys rate information to the listener, as well as its extrinsic versus intrinsic nature, are still controversial issues. However, it seems clear that a given syllable of ambiguous phonetic identity with respect to some temporal cue, only needs the context of adjacent syllables ([2]), or even no context at all ([5]), for correct identification across different rates.

Since the smallest units conveying rate information examined so far are the size of the syllable, one may ask whether or not an even more detailed account of the intra-syllabic structure is required to explain the subjective rate of a given utterance. Indeed, the syllabic rate should be a prominent factor yielding the tempo sensation, but we could also take into account locally defined factors: the speed of acoustic changes in unsteady parts like release bursts and transitions, reflecting the underlying articulatory gesture velocity, and the duration of sustained sounds like fricatives and vowels.

The aim of this study was thus to examine more closely the possible effects of intra-syllabic rate manipulation. In short, we compared 2 ways of modifying the speech rate: in the 1st one, every portion of an original utterance was shortened or lengthened by the same factor ("uniform warping"); in the 2nd one, the warping factor was made dependent on the local spectral derivative ("non uniform warping"), ranging from 1, that is no warping at all, in the most unsteady parts, to a fixed warping target specification in the most steady parts. Thus, as a general rule, steady parts of vowels (or fricatives if any) were maximally distorted while transition from silence to release bursts, release bursts, onset of voicing, fast transitions and the like were preserved in as much as they were exhibiting fast acoustic change. Whether or not the latter warping scheme is closer to actual human speech production than the former goes beyond the scope of this study (anyhow, the existing data on speech production at different rates is quite conflicting, ranging from the observation that vowels are more elastic than consonants ([7]), as elastic ([6]), to less elastic ([2]). Also, speakers may adopt very different strategies when modifying their speech rate).

If our 2 schemes of rate manipulation, in the absence of any anchoring part which might yield a contrastive effect ([5]), do not influence differently subjective articulation rate, one could conclude that tempo sensation is conveyed mainly by the physical syllabic rate, or, equivalently, by the vocalic cycle tempo. Unfortunately, this is not apparent from our data.

METHOD

We used an AX experimental procedure to compare uniformly warped A stimuli with non uniformly warped X stimuli. All A and X stimuli were built from an original utterance of Japanese, /ikebukuro/, pronounced by a male speaker. This item was chosen because its syllables were homogeneous in structure, all one mora syllables with no geminates. This avoided problems of vocalic or consonantal quantity contrasts which might have interfered with tempo perception.

Four sets of AX pairs were prepared, using a modified version of the SOLA (Synchronized OverLap Add) technique ([8]). This technique produces very natural sounding time scaled speech. X stimuli were built from the original with a varying warping factor depending on the local value of the spectral derivative in the original speech and on a fixed warping target specification, as shown in Fig. 1. The latter was computed iteratively from the desired overall warping factor and the spectral derivative curve (see [9]). Overall time warping factors for A and X stimuli together with average syllabic and vocalic durations are reported in Table I. The overall warping factors for X stimuli were chosen on the basis of preliminary tests not reported here, so that extreme X stimuli of a given set would be close to the hesitation region. Also, since the processed speech included some silent portions beyond the region of interest, i.e. the acoustic word /ikebukuro/, we measured the acoustic length of all the different versions of /ikebukuro/ from spectrograms and energy curves. The 2 clear energy dips which consistently surrounded the word were chosen for its acoustic boundaries. Average syllabic duration was taken as the fifth of the duration defined by these boundaries. In addition, vocalic durations were estimated from the same spectrograms.

Each of the 4 sets consisted of a randomized sequence of 60 AX pairs containing 10 each of 6 AX pairs differing by their X stimulus only. Each pair consisted of a short 500 Hz beep, 1 second silence, A stimulus, 1 second silence, X stimulus and 4 seconds silence for written response. An extra silence was inserted every 10 pairs. The subjects, five male Japanese adults with normal hearing and phonetically naive, were required to select which stimulus of each pair "sounded faster" by circling a letter on an answer sheet. They sat for one session per set at 2 days interval.

The result of discrimination tests can be illustrated by Fig. 2 where the frequency of X being judged "slower" than A is plotted on a normal scale against its average syllabic duration. We assumed that the experimental data could best be approximated by cumulative normal distributions. The mean and standard deviation of such distributions were estimated by computing linear regression lines out of such graphs as in Fig. 2. This approximation held quite well for all individual data as well as for the pooled across subjects data. For each subject and each set, the estimated mean of the underlying normal distribution approximates the average syllabic duration of the X type stimulus which would sound just as fast as the A stimulus of the set. The standard deviation can be regarded as an index of the accuracy of listener's discrimination.

SET 1

stimulus	A	X1	X2	X3	X4	X5	X6
warping	0.6	0.53	0.56	0.59	0.61	0.64	0.67
syllabic length	155.6	143.2	150.8	158	163.2	170.4	177.6
vocalic length	103.8	70.6	79.8	84.4	88.6	95.2	103.4

SET 2

stimulus	A	X1	X2	X3	X4	X5	X6
warping	0.8	0.73	0.76	0.79	0.81	0.84	0.87
syllabic length	207.4	192	199.6	208.2	212.0	219.6	227.8
vocalic length	138.7	107.7	116.1	124.3	128.4	134.0	141.5

SET 3

stimulus	A	X1	X2	X3	X4	X5	X6
warping	1.2	1.13	1.16	1.19	1.21	1.24	1.27
syllabic length	317.8	296.2	305.0	312.6	316.2	324.2	331.2
vocalic length	213.7	204.9	213.4	222.2	226.6	235.5	240.3

SET 4

stimulus	A	X1	X2	X3	X4	X5	X6
warping	1.4	1.275	1.325	1.375	1.425	1.475	1.525
syllabic length	366	334.4	346.8	358	372	385.6	400.8
vocalic length	245.4	241.8	255.6	269.4	281.6	293	304.4

Table I. Stimuli used in the 4 sets (durations are given in ms).

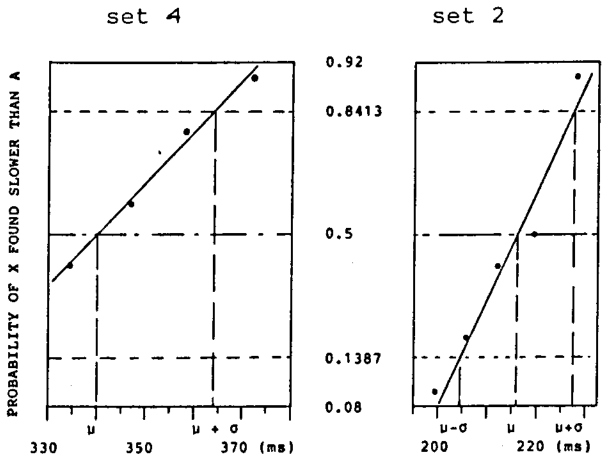


Fig. 2. Example of individual results for listener KH in the sets 2 and 4. The regression line yields μ and σ of the normal distribution best approximating the data.

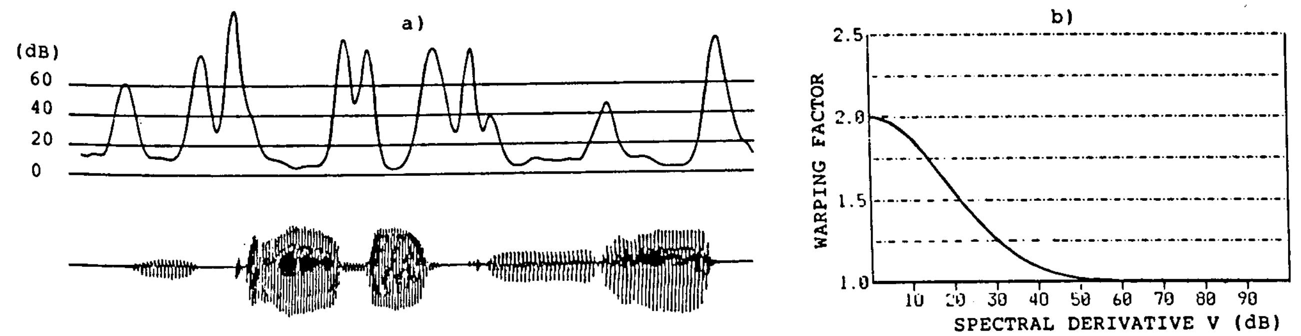


Fig. 1. a) the original utterance /ikebukuro/: audio signal and spectral derivative curve. b) the local warping factor as a function of the spectral derivative V and a fixed warping target specification U = 2.

RESULTS

Individual and pooled results are recorded in Table II.

SET 1

Subject	UH	MH	KH	WH	SR	pooled
μ	154.0	158.3	157.6	156.5	157.5	157.6
σ	9.9	8.7	11.2	6.	7.4	7.2

a	155.6
$\bar{\mu}$	156.74
SE	0.743

SET 2

Subject	UH	MH	KH	WH	SR	pooled
μ	207.9	208.2	215.2	207.8	209.5	210.9
σ	14.3	11.2	11.4	9.0	8.8	9.5

a	207.4
$\bar{\mu}$	209.86
SE	1.54

SET 3

Subject	UH	MH	KH	WH	SR	pooled
μ	303.1	313.3	312.7	312.1	310.4	311.6
σ	11.1	16.1	12.2	9.3	5.6	12.6

a	317.8
$\bar{\mu}$	310.32
SE	1.869

SET 4

Subject	UH	MH	KH	WH	SR	pooled
μ	353.5	361.5	340.1	356.1	343.8	353.4
σ	13.2	14.7	24	10.1	11.1	18.7

a	366.0
$\bar{\mu}$	351.0
SE	3.957

Table II. Individual and pooled results. The means μ represent the duration of the X type stimulus which would sound the same speed as the A stimulus for each set. Standard deviations σ are an index of subjects' discrimination accuracy.. The average syllabic of A stimulus, a, the mean and standard error of μ distribution, $\bar{\mu}$ and SE, yield the t of Student used for confidence estimation

Individual and pooled results are recorded in Table II.
In sets 1 and 2, which contain only shortened versions of the original utterance, the means exhibit a very weak tendency to be longer than A stimuli average syllabic duration ($t=1.54$, $p<0.2$ for set 1, $t = 1.6$, $p < 0.2$ for set 2). This might mean that shortened X stimuli are judged a little faster than A stimuli at the same syllabic rate, although this trend is very weak.
In sets 3 and 4, which contain only lengthened versions of the original, the means are significantly shorter than A stimuli average syllabic duration ($t = 4.002$, $p < 0.02$ for set 3, $t = 3.79$, $p < 0.02$ for set 4). Thus lengthened X stimuli sound slower than A stimuli for the same syllabic rate.
In order to check the possibility of a systematic bias introduced by the experimental procedure, a 5th experiment was conducted: A stimulus was replaced by the original utterance, and the 6 X stimuli overall warping factor were ranging from 0.93 to 1.07. The means, for all 5 subjects, were not found significantly different to the original utterance average syllabic duration ($t = 0.095$, $p > 0.5$), as shown in Table III. Thus the experimental procedure was considered as not introducing any systematic distortion.

SET 5

Subject	UH	MH	KH	WH	SR	pooled
μ	270.8	247.3	261.1	252.0	266.7	261.8
σ	15.2	12.1	12.6	21.5	10	11.8

a	259.8
$\bar{\mu}$	259.58
SE	4.4

Table III. Individual and pooled results for the 5th experiment.

DISCUSSION

From these results, we can hypothesize that the subjective articulation rate is affected, at least in the case of lengthening, by the vowel duration: for the same syllabic rate, X stimuli which have longer vocalic portions than A stimuli (see Table I), and thus shorter consonantal portions, sound slower. Does this rule out the possibility of a competing effect of relative consonantal shortening ? If yes, in the case of shortening, we should observe that the subjective rate is clearly affected by much shorter vowels in X stimuli than in A stimuli, but this is not the case. We then keep hypothesizing a competing effect arising mainly from the consonantal part of the syllable, or more precisely from the speed of fast acoustic changes which reflect consonantal gestures. The fact that the consonant effect is clearly dominated by the vowel effect in the case of lengthening, but not in the case of shortening might be explained by the ratio of consonantal to vocalic durations in the X type stimulus yielding the same speed sensation than A stimulus for each set. For each set, the means of the approximated normal distributions for pooled data were computed for both syllabic and vocalic durations. We assumed that both means corresponded approximately to the same ideal "A-equivalent" X stimulus whose consonant to vowel ratio was taken as representative of its set. These ratios, shown in Table IV, are in clear agreement with the assumption that vowel effect should dominate consonant effect in the case of lengthening.

SET	1	2	3	4
C/V ratio	87.4 %	67 %	41 %	34.5 %

Table IV. Average consonant duration to average vowel duration ratios for the "A-equivalent X stimuli" of the 4 sets.

We should keep in mind that all these stimuli were manipulated speech. In particular, shortened X stimuli were manipulated in such a way that the simulated consonantal gesture was essentially preserved. Somehow, this gave the feeling of a "careful" articulation which might have interfered with the required judgment of speed and could partly explain the asymmetry of our results.

Finally, if we turn to Nooteboom's research on "internal auditory representation of syllable nucleus durations" ([10]), it appears that our results are in good agreement with his finding that listeners are higly sensitive to vowel nucleus durations. The standard deviations of the normal distributions approximating our data (see Table II), give a quantitative indication of listeners' judgment accuracy: the order of magnitude is 10 ms. Nooteboom reports an even higher accuracy with which a syllable nucleus duration can be internally represented.

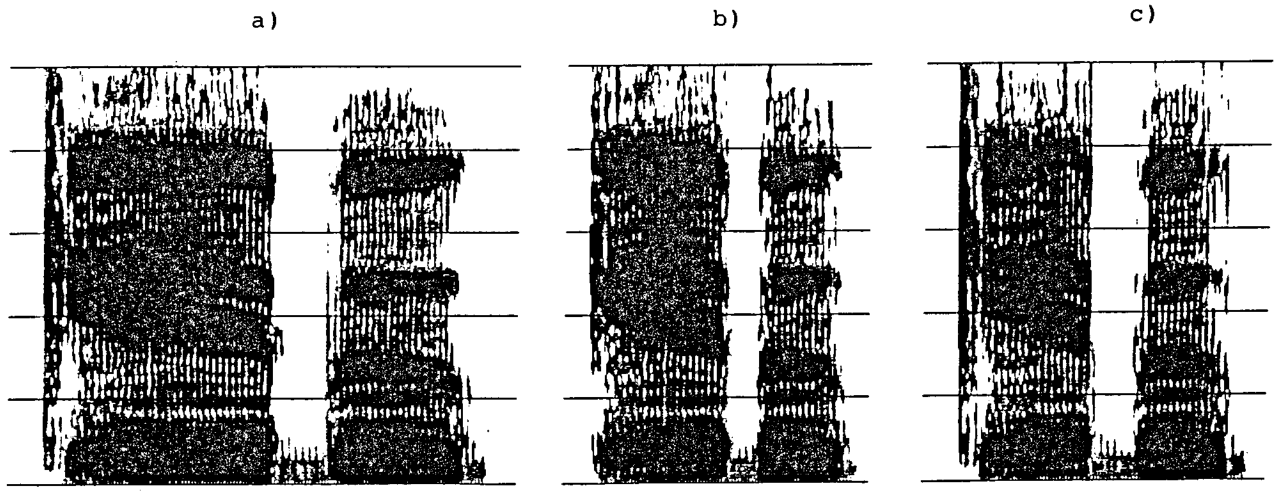


Fig. 3. Spectrograms of the portion /kebu/. a) original, b) overall uniform warping of 0.6, c) overall non uniform warping of 0.59.

CONCLUSION

To summarize, our experimental data indicates that intra-syllabic structure of speech may modify the tempo given by the syllabic rate. Namely, at least in the case of lengthened speech, it is not only the tempo given by such periodicity as the approximate one defined by temporal gaps between consecutive vowels (articulatory) onsets, that produces the speed sensation, but also, to a substantial extent, the duration of the steady parts of the vowels.

ACKNOWLEDGEMENTS

We wish to express our thanks to Dr. Fujisaki and the staff of his laboratory for the assistance received in the preparation of the perceptual tests.

REFERENCES

[1] Miller, J., L. (1981). Effects of speaking rate on segmental distinctions. In P.D. Eimas & J.L. Miller (Eds) Perspectives on the Study of Speech, Hillsdale, Erlbaum Associates.

[2] Johnson, J., L. & Strange, W. (1982). Perceptual constancy of vowels in rapid speech. JASA, 72 (6), 1761-1770.

[3] Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. J. Exp. Psychol.: Hum. Percept. Perform. 7, 1074-1095.

[4] Goldman-Eisler, F. (1968). Psycholinguistics: Experiments on Spontaneous Speech. London, Academic Press.

[5] Miller, J., L., Aibel, I., L. & Green, K. (1984). On the nature of rate dependent processing during phonetic perception. Perception & Psychophysics, 35 (1), 5-15.

[6] Kozhevnikov, V. A. & Chistovitch, L., A. (1965). Speech: Articulation and Perception. Joint Publication Research Service, 30543. (Washington D.C.).

[7] Shaffer, L. H. (1982). Rythm and timing in skill. Psychological Review, 89, 109-122.

[8] Roucos, S. & Wilgus, A., M. (1985). High Quality Time-Scale Modification for Speech. Proc. of ICASSP1985, 493-496.

[9] Halle, P. (1986). Non Uniform Time-Scale Modification for Speech. Proc. of ASJ 1986 Fall Meeting, 151-152.

[10] Nooteboom, S., G. (1975). On the Internal Auditory Representation of Syllable Nucleus Durations. In G. Fant & M.A.A. Tathan (Eds) Auditory Analysis and Perception of Speech, Academic Press, 413-430.

ARTICULATORY COMPLEXITY AND THE PERCEPTION OF SPEECH RATE

BERND POMPINO-MARSCHALL HANS G. TILLMANN WOLFGANG GROSSER KARL HUBMAYER

Institut für Phonetik und
Sprachliche Kommunikation
Universität München, FRG

Institut für Anglistik und
Amerikanistik
Universität Salzburg, Austria

ABSTRACT

It is shown that syllable sequences containing complex consonant clusters are perceived as faster than articulatorily less complex ones of the same duration; furthermore, that in AX-discrimination the second test item is perceived as faster.

INTRODUCTION

Although German is supposed to be stress timed, compression of complex stress feet to the duration of simple ones is known not to be complete [2], and thus complex syllables should be perceived as faster in contrast to simpler ones of the same duration [1]. With the following experiments we wanted to study this effect in more detail.

METHOD

Three five feet sequences identical with respect to number of syllables, trochaic foot structure, and vowels, but differing in foot initial consonance complexity were uttered by a native speaker of German with the stressed syllables in beat with a computer-generated metronome signal of variable frequency. These sequences are in accord with the phonotactic rules of German. The metronome frequency used to control the speech rate was varied in steps of 5 from 90 to 110 beats per minute. In this way we got five items in any of the following three sets:

(1) /'ʃe:te 'ʃa:te 'ʃi:te 'ʃo:te 'ʃu:te/
(2) /'ʃpe:te'ʃpa:te'ʃpi:te'ʃpo:te'ʃpu:te/
(3) /'ʃple:te'ʃpla:te'ʃpli:te'ʃplo:te
'ʃplu:te/.

Segment, syllable and foot durations were measured on sonagraphic displays (see below). For the perception experiment we removed the first and the last foot from these fifteen utterances. The utterances were combined in pairs in the following way (forming different subtests): The item with mean speech rate of set 1 first, i.e. /'ʃa:te 'ʃi:te 'ʃo:te/ at a rate of 100 feet per minute, followed by one item of set 2, or one item of set 1 followed by /'ʃpa:te 'ʃpi:te 'ʃpo:te/ at the mean rate of 100, and both combinations in reversed order, i.e. /'ʃpa:te'ʃpi:te' ʃpo:te/ in first position and /ʃa:te'ʃi:te'ʃo:te/ second. In the same way set 2 was combined with set 3 and set 1 with set 3, resulting in 54 stimulus pairs. The stimuli were presented to a group of 19 subjects in an AX-format (same/different rate of speech) six times each in randomized order.

RESULTS

Acoustical Analysis: Durational Measurements

The measurements of the relevant parts of the utterances used in the German perception experiments were made using broad band sonagraphic displays. The following segments were measured independently:
prestressed consonants from the beginning of the fricative noise till voicing onset of the vowel (/ʃ/, /ʃp/) or till the end of the /l/-obstruction marked by a clear increase of energy in the higher formants; stressed vowels till the /t/-occlusion; stressed syllables from /ʃ/-onset till offset of voicing of the vowel; unstressed syllables till onset of /ʃ/-frication; and single feet from one onset of /ʃ/-frication till the next.
Foot duration exhibits no difference between the different segmental compositions. It is always slightly longer than demanded by the presented metronome rate (mean 3.1%), i.e. the subject always is a little bit slower than the presented metronome pattern. Foot compression of 18.6% between the extreme metronome rates is in the range of the computed value (i.e. the relation 110/90: 18.2%). As expected, the different parts of the foot contribute differently to this overall compression: with 22.2% it is stronger in the unstressed than in the stressed syllable (16.5%), and in the stressed vowels with 18.8% it is stronger than in the prestressed consonants (13.5%).
Computed over all rates of speech, single two-factorial analyses of variance showed that the duration of the stressed syllable is significantly determined by the vowel (F(2,8) = 6.31; p < .05) and the initial consonants (F(2,8) = 38.22; p < .001): syllables containing /a:/ and /o:/ (with durations of 393.7 and 392.3 msec) are longer than those containing /i:/ (375.3 msec) and /ʃpl_/-syllables (408. msec) are significantly longer than /ʃp_/-syllables (394.3 msec) which in turn are significantly longer than /ʃ_/-syllables (359. msec)). These intrinsic differences at the foot level are compensated for in the given material by the reversed effects in the unstressed syllables (effect of preceeding vowel: F(2,8) = 11.54; p < .01; of preceeding consonants: F(2,8) = 71.3; p < .001): here the /te/ following a syllable containing /i:/ is longer (247.3 msec) than those containing /o:/ and /a:/ (230.3 and 229.7 msec), and /te/ following /ʃ_/ (261. msec) is longer than following /ʃp_/ (227.3 msec) which is longer than /te/ following /ʃpl_/ (219. msec).
At the segmental level one can see that even vowel duration is a complex function of the vowel itself (F(2,8) = 44.39; p < .001), the prestressed consonants (F(2,8) = 4.42; p ca. .05), and an interaction of both factors (F(4,16) = 3.74; p < .05). In general /a:/ (213.3 msec) is longer than /o:/ (196.7 msec; with the exception of /a:/ and /o:/ following /ʃ/, where both are not significantly different), which in turn is longer than /i:/ (171.7 msec).

Perception Experiments

The results of the different subtests (see above) are shown in Fig.1-3. For further analysis the median of the 'same'-response distribution was computed for every subject in all subtests. A two-factorial analysis of variance showed a significant effect of set combination (F(2,198) = 13.48; p < .001):
the results suggest that /'ʃpa:te .../ would have to be produced at a rate of 98.95 to be perceived as fast as /'ʃa:te .../ at a rate of 100 and, not significantly differing from this effect, that /'ʃpla:te .../ would have to be produced at a rate of 98.74 to be perceived as fast as /'ʃpa:te .../ at a rate of 100, but /'ʃpla:te .../ compared with /'ʃa:te .../ at a rate of 100 would have to be produced at the significantly slower rate of 97.5 to be perceived as equally fast. All computed rates are different from 100, the rate they are compared with (p < .01). An effect of the order of presentation within stimulus pairs, clearly visible in a pretest with real sentences is not to be seen in the results of the analysis of variance.

We replicated part of the experiment with American-English material and subjects (N = 9). The initial /ʃ/ in the material was replaced by /s/. Only the combinations of set 1 with set 2 and set 1 with set 3 were tested in the same way as before.
The results are shown in Fig.4 and 5. We can see an effect of order of presentation in Fig.5: stimulus pairs with the simpler /'sa:te .../-sequence in first position (open columns) result in very rare 'same'-responses that never reach 50%, whereas in the reversed order (filled columns) we have more 'same'-responses, exceeding 50% when the simpler sequence is maximally faster than the complex one, but both response functions are cut off at this stimulus pair (to be seen as well in Fig.4 and to a less degree also in the German results of Fig.3). This order effect means that the second part of the stimulus-pair is heard as faster than the identical one in first position probalbly due to a normally given slowing down at the end of utterances. Because the 'same'-response function is cut off at the 90-100, 110-100 pairs the order effect cannot become visible in the results of the analyses of variance based on the median measurements: these do not represent the actual point of perception of equal speech rates.
Parallel to the German results the analysis of variance shows a clear effect of set combination on the median of 'same'-responses (F(1,68) = 9.23; p < .01): /'spa:te../ would have to be produced at a rate of 98.24 to be perceived as fast as /'sa:te .../ at a rate of 100, whereas /'spla:te .../ would have to be produced at a rate of 96.12 to sound as fast as /'sa:te .../ at a rate of 100 (both computed rates differing from 100; p < .01). Because of the reasons mentioned above the analysis of variance again does not show a significant effect of order of presentation.

It should be metioned that it is not possible to correlate the data of the perception experiments with the measurements of acoustical segment durations since the median-based results of the

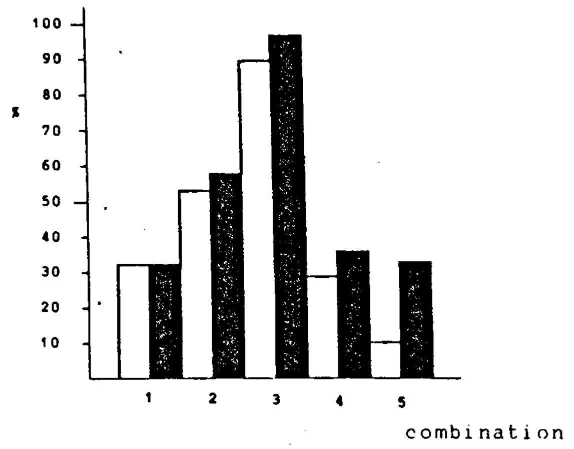


Fig. 1: /ʃ.../-/ʃp.../ (open columns) and /ʃp.../-/ʃ.../ pairs (filled columns)

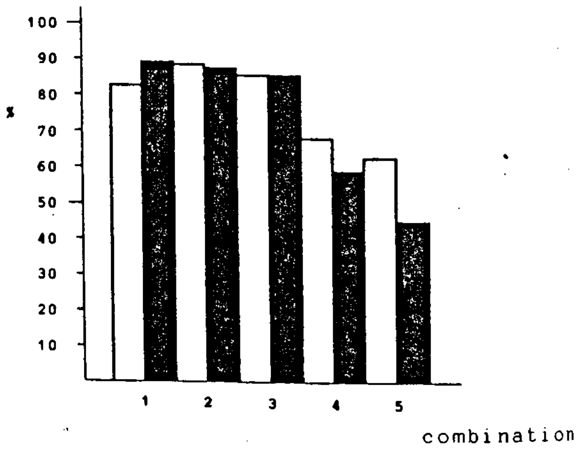


Fig. 4: /s.../-/sp.../ (open columns) and /sp.../-/s.../ pairs (filled columns)

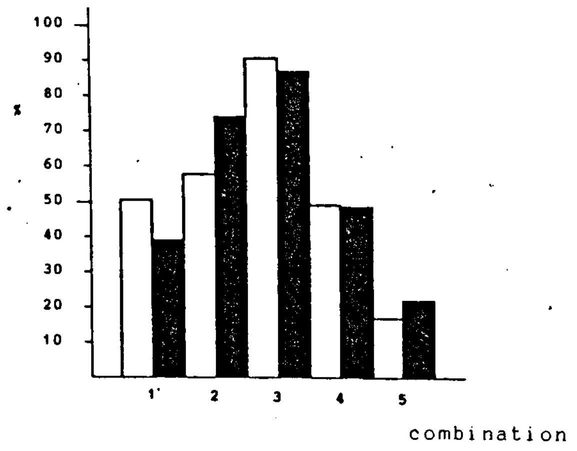


Fig. 2: /ʃp../-/ʃpl../ (open columns) and /ʃpl../-/ʃp../ pairs (filled columns)

Fig. 1-5: Percent 'same'-responses of German (1-3) and American-English subjects (4, 5) to stimulus pairs of the following rate combinations as shown on the abscissa: 1 = 100-90 and 110-100; 2 = 100-95 and 105-100; 3 = 100-100; 4 = 100-105 and 95-100; 5 = 100-110 and 90-100 (open columns) and in reversed order (filled columns)

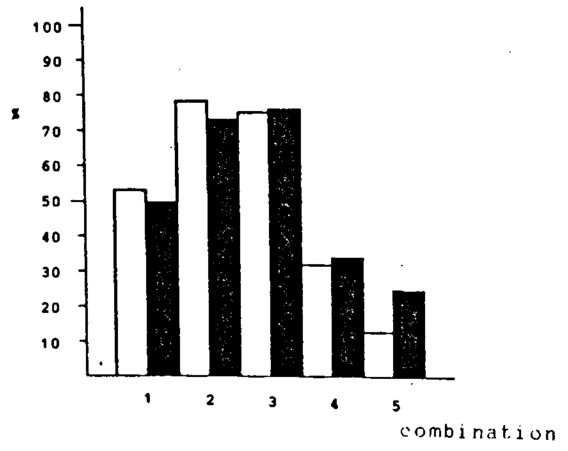


Fig. 3: /ʃ.../-/ʃpl../ (open columns) and /ʃpl../-/ʃ.../ pairs (filled columns)

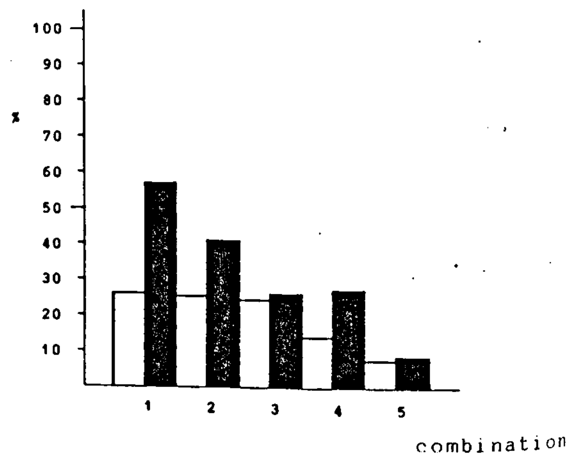


Fig. 5: /s.../-/spl../ (open columns) and /spl../-/s.../ pairs (filled columns)

perception experiments do not represent the actual measure of perceived equality of speech rate.

DISCUSSION

The durational measurements have shown that there was a good approximation of the metronome rate at the level of foot duration. But this durational compression is seen to work differently in different parts of the foot. Intrinsic durational differences of the stressed syllables for example are compensated for by the durational behaviour of the unstressed syllable. Although at the syllable level there are durational differences due to the complexity of the initial consonance up to 50 msec, the complex utterances are perceived as being uttered at a faster rate of speech. Clearly this judgement of the hearer must be based on a measure of articulatory movements per unit time.

REFERENCES

[1] Hoequist, C. 1983, Parameters of speech rate perception. Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK) 20, 99-138.

[2] Kohler, K. J. 1986, Invariance and variability in speech timing: from utterance to segment in German. In: Perkell, J. S. & Klatt, D. H. (ed.), Invariance and Variability in Speech Processes (Hillsdale, N. J.), 268-299.

P-CENTERS AND THE PERCEPTION OF 'MOMENTARY TEMPO'

BERND POMPINO-MARSCHALL HANS G. TILLMANN BARBARA KÜHNERT

Institut für Phonetik und
Sprachliche Kommunikation
Universität München, FRG

ABSTRACT

Variation in p-center location is shown to be dependent on the segmental composition of a syllable in a far more complex way than assumed by the model of Marcus [2] that represents a linear combination of two different linear effects: of the duration of the syllable-initial consonance and of the syllable rhyme. The mean distance of measured P-centers from one another in syllable sequences is shown to be a good indicator of perceived rate of speech.

INTRODUCTION

Alternating sequences of monosyllables, when presented with equal intervals between successive acoustical syllable onsets (so-called isochronous sequences) are not perceived as having a subjectively uniform rhythm because the perceived onset (P-center) of a syllable typically does not correspond to its acoustic onset. Generally, it is assumed that the location of the P-center of monosyllables is solely dependent on the duration of the initial consonant(s) and that of the syllable rhyme as represented by a linear equation proposed by Marcus [2].

In the following experiments we wanted to test this hypothesis with systematically varied material.

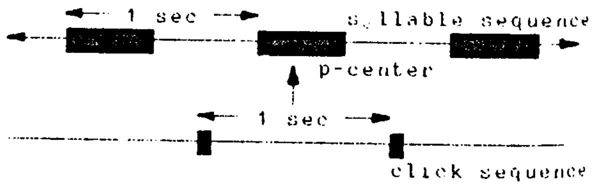
EXPERIMENT I

Method

In simple synthetic syllables composed of the consonant /m/ and the vowel /a/ segment durations were varied systematically:

25 /ma/-syllables with /m/ varying in duration from 40 to 200 msec in steps of 40 msec and /a/ varying from 100 to 260 msec also in steps of 40 msec and 25 /am/-syllables with /a/ varying as before and /m/ from 80 to 240 msec, again in steps of 40 msec. Finally 5 /mam/-syllables were synthesized with the following segment durations: 40 msec /m/, 260 msec /a/, 80 msec /m/; 80, 220, 120; 120, 180, 160; 160, 140, 200; and 180, 100, 240. Vowel duration included two symmetrical transitions (from and to /m/) of 40 msec each. Fundamental frequency was set at 100 Hz for the entire duration of the stimuli and amplitude was held constant over the steady-state parts. The syllables were synthesized on a PDP 11/50 with a program based on the Klatt software synthesizer [1].

In perception experiments the subjects had to adjust the timing of these syllables alternating with clicks (5-msec 1 kHz-tone bursts) in sequences of five signals with an overall tempo of 120 signals per minute to perceived isochrony by turning a potentiometer knob. We decided to use the time instant bisecting the duration between two successive clicks was used to determine the location of the p-center of the test syllable:



All stimuli were adjusted by two subjects in alternating sequences beginning with a click signal as well as in sequences beginning with the syllable itself six times in each session. The extreme adjustments were omitted from the analysis. There were five sessions for every stimulus, resulting in 40 adjustments (2 sequences ^ 4 adjustments ^ 5 sessions).

Results

The results of the adjudstments for the /ma/-syllables pooled over both subjects are seen in Figure 1. Two-factorial analyses of variance revealed highly significant effects of both duration of the initial /m/ and of the vowel /a/ as well as a significant interaction of the effects on P-center location for both subjects:

B.P.M. (male): F(4,975) = 1432.36; p < .001; F(4,975) = 70.67; p < .001; F(16,975) = 4.19; p < .001;
B.K. (female): F(4,975) = 7453.57; p < .001; F(4,975) = 769.23; p < .001; F(16,975) = 19.64; p < .001.

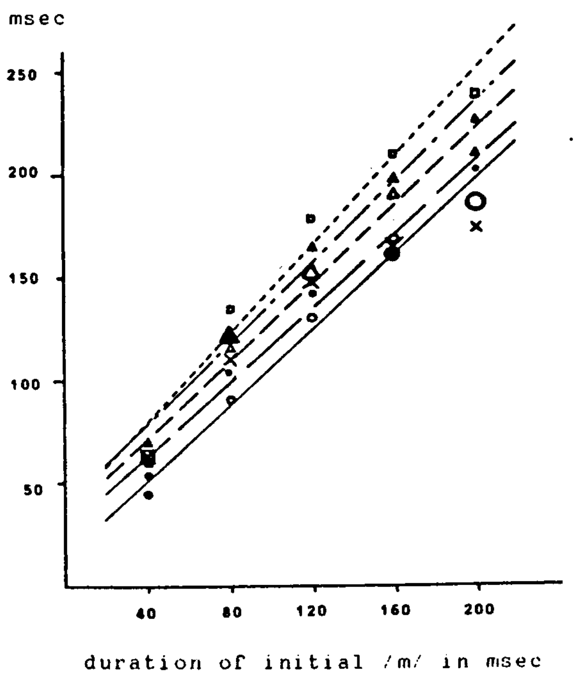


Fig.1: Variation in p-center position (ordinate) due to the duration of the initial /m/ (abscissa) with different /a/-durations: open circles, unbroken regression line: 100 msec /a/; filled circles, long dashed: 140 msec; open triangles, short dashed: 180 msec; filled triangles, dash-dotted: 220 msec; open rectangles, dotted: 260 msec; enlarged symbols represent those stimuli dB-paralleled in the psychoacoustic experiment which are represented as crosses.

The simple main effects of consonant and vowel duration are shown in Table I and II for both subjects individually along with the parameters of the regression lines and the levels of significance for linear and nonlinear components of trend. Às can be seen, quite generally there are significant nonlinear components speaking against the hypothesis of Marcus [2].

For the /am/-syllables, the results are shown in Figure 2 and Tables III and IV in the same way. Here too, we get significant effects of vowel duration, of final consonant duration and an interaction of both effects for both subjects:

B.P.M.: F(4,975) = 71.15; p < .001; F(4,975) = 24.24; p < .001; F(16,975) = 6.08; p < .001;

Table I:
Simple main effects of initial /m/-duration, parameters of the regression lines and significance of linear and nonlinear trend components

	level of	analysis of trend:				
duration of /a/ in msec	significance	r	a	b	lin.	non lin
100	^^^	.92	19.59	.97	^^^	^^^
	^^^	.97	9.62	.85	^^^	^^^
140	^^^	.89	32.92	.94	^^^	^
	^^^	.97	20.49	.84	^^^	^^^
180	^^^	.91	36.6	.97	^^^	^^^
	^^^	.97	30.6	.9	^^^	^^^
220	^^^	.91	39.74	1.04	^^^	^^
	^^^	.97	38.27	.91	^^^	^^^
260	^^^	.93	39.62	1.09	^^^	^^^
	^^^	.97	31.57	1.05	^^^	^^^

here and in the following tables: first line subject B.P.M., second line subject B.K.;
^^^: p < .001; ^^: p < .01; ^: p < .05;
- : n.s.

Table II:
Simple main effects of /a/-duration, parameters of the regression lines and significance of linear and nonlinear trend components

	level of	analysis of trend:				
duration of /m/ in msec	significance	r	a	b	lin.	non lin
40	^^^	.23	44.14	.11	^^^	^
	^^^	.65	25.34	.15	^^^	^^^
80	^^^	.47	83.28	.22	^^^	^
	^^^	.87	48.47	.31	^^^	^^^
120	^^^	.58	107.69	.29	^^^	-
	^^^	.87	94.07	.29	^^^	^
160	^^^	.48	157.69	.23	^^^	^^^
	^^^	.86	109.08	.35	^^^	^^^
200	^^^	.53	175.27	.28	^^^	-
	^^^	.92	135.15	.36	^^^	-

B.K.: F(4,975) = 459.47; p < .001; F(4,975) = 362.79; p < .001; F(16,975) = 2.22; p < .01.

As before we have quite a number of nonlinear effects. Furthermore the syllable rhyme seems not to be an integral part

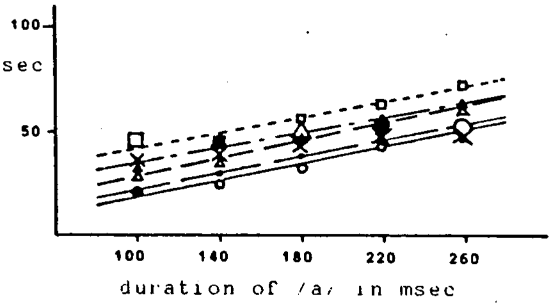


Fig. 2: Variation in p-center position due to the duration of /a/ (abscissa) with different final /m/-durations: open circles, unbroken regression line: 80 msec /m/; filled circles, long dashed: 120 msec; open triangles, short dashed: 160 msec; filled triangles, dash-dotted: 200 msec; open rectangles, dotted: 240 msec; enlarged symbols: dB-paralleled (crosses).

with respect to the determination of p-center location. As in open syllables, there are different influences of vowel and consonant durations interacting with one another in a complex fashion.
The results for the /mam/-syllables pooled over both subjects is shown in Figure 3.

EXPERIMENT II

Method

Some of the stimuli (those marked in Figure 1 and 2 by enlarged symbols and the /mam/-syllables) were paralleled with respect to dB-envelope by 100-Hz rectangular signals to test the influence of the envelope-parameter. Adjustments were done by the two subjects of Experiment I in the same fashion.

Results

The results for these nonspeech analogues are marked by crosses in Figure 1 - 3. Analyses of variance for both subjects individually show a speech-nonspeech effect for /ma/-syllables only for B.K. (p < .001: the nonspeech analogues always showing smaller p-center delays), for /am/-syllables for both subjects (p < .01 and < .001 respectively; again smaller delays for the nonspeech stimuli) and for the /mam/-syllables also for both subjects (p < .001) but here with different orientation (B.K. as before, B.P.M. with longer p-center delays for the nonspeech material). These latter results deserve further testing for a final interpretation.

Table III:
Simple main effects of /a/-duration, parameters of the regression lines and significance of linear and nonlinear trend components

duration of final /m/	level of significance	analysis of trend: r	a	b	lin.	non lin
80	***	.6	-4.32	.24	***	*
	***	.78	.01	.17	***	-
120	***	.52	3.58	.2	***	***
	***	.78	3.83	.18	***	*
160	***	.49	13.03	.19	***	**
	***	.85	2.92	.22	***	-
200	***	.33	21.33	.13	***	*
	***	.81	13.18	.22	***	-
240	***	.33	29.76	.14	***	**
	***	.81	17.77	.22	***	-

Table IV:
Simple main effects of final /m/-duration, parameters of the regression lines and significance of linear and nonlinear trend components

duration of /a/	level of significance	analysis of trend: r	a	b	lin.	non lin
100	***	.42	5.9	.16	***	***
	***	.78	3.51	.15	***	**
140	***	.34	18.29	.11	***	*
	***	.79	9.27	.16	***	-
180	***	.31	26.82	.11	***	***
	***	.77	16.2	.18	***	-
220	***	.09	48.23	.03	-	**
	***	.78	21.25	.19	***	-
260	***	.16	47.67	.06	**	***
	***	.75	26.03	.21	***	-

EXPERIMENT III

In a last experiment the /mam/-syllables and their nonspeech counterparts were concatenated to a five-item sequence paralleling those of Ventsov [6] and ours [3, 4] in the following form, yielding a sequence of open syllables of 300 msec and closed syllables of 340 msec duration:

40m260a80m220a120m180a160m140a200m100a240m

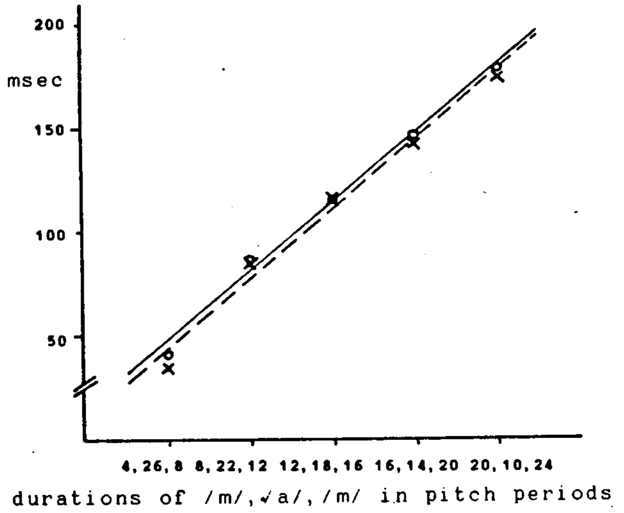


Fig. 3: P-center position (ordinate) due to the duration of syllable segments (abscissa): initial /m/-duration, /a/-duration, final /m/-duration (circles); dB-paralleled stimuli: crosses.

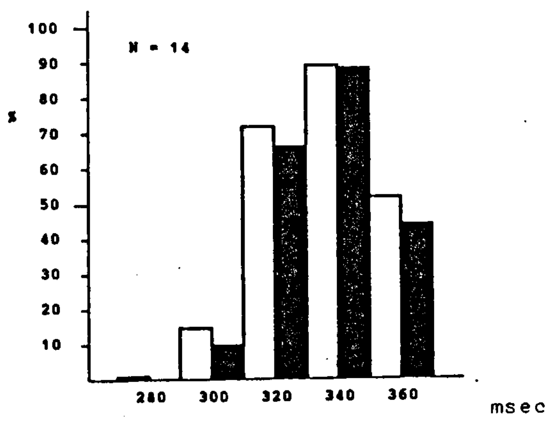


Fig. 4: Percent 'same'-responses to the /mam .../-sequence (see text; open columns) and dB-paralleled sequence (filled columns) in combination with click-sequences of variable onset intervals

This sequence was combined in pairs with a click sequence of varying click onset intervals (from 280 to 360 msec in steps of 20 msec). 14 subjects judged 10 randomized presentations of these five pairs as being same or different with respect to rate. The results are shown in Figure 4. For both stimulus sets the pooled median of the 'same'-response distribution lies near the duration of the closed syllable, but significantly differing from it (p < .001; speech: 336. msec, sd = 6.62; nonspeech: 336.3, sd = 6.02). These values resemble the mean p-center distances of the used stimuli (speech: 334.26, sd = 7.24; nonspeech: 334.95, sd = 9.55).

DISCUSSION

Our results clearly show that the location of the p-center is not simply dependent on segmental durations of the tested syllable. The psychoacoustic model of Schütte [5] which takes the rising auditorily filtered sound pressure envelope as the p-center determining parameter gives a better prediction of the p-center location for our /ma/-syllables. But it does not explain why there are different p-center locations in the set of /am/-syllables (because they all have the same rising envelope). And also the speech-nonspeech differences in our results lead us to the conclusion that the p-center location in speech material cannot be accounted for by a pure psychoacoustic modelling.

REFERENCES

[1] Klatt, D.H. 1980, Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America 67, 971-995.

[2] Marcus, S.M. 1981, Acoustic determinants of perceptual center (P-center) location. Perception & Psychophysics 30, 247-256.

[3] Pompino-Marschall B; Piroth, H.G.; Hoole, P.; Tilk, K; Tillmann, H.G. 1982, Does the closed syllable determine the perception of 'momentary tempo'? Phonetica 39, 358-367.

[4] Pompino-Marschall, B.; Piroth, H.G.; Hoole, P.; Tillmann, H.G. 1984, 'Koartikulation' and 'Steuerung' as factors influencing the perception of 'momentary tempo'. In: Van den Broecke, M.P.R.; Cohen, A. (ed.), Proceedings of the Tenth International Congress of Phonetic Sciences (Dordrecht, Cinnaminson - Foris), 537-540.

[5] Schütte, H. 1978, Ein Funktionsschema für die Wahrnehmung eines gleichmäßigen Rhythmus in Schallimpulsfolgen. Biological Cybernetics 29, 49-55.

[6] Ventsov, A.V. 1981, Temporal information processing in speech perception. Phonetica 38, 193-203.

PERCEPTION OF A PROSODIC BREAK: THE CASE OF INCIDENTAL PHRASES IN FRENCH

N. BACRI *
Lab. de Psychologie Expérimentale

A. NICAISE **
U. F. R. en Linguistique

* C.N.R.S. - E.H.E.S.S. 54 bd Raspail 75006 PARIS, France
** PARIS VII 2 pl Jussieu 75005 PARIS et Paris XII, France

An acoustic analysis of the production of incidental phrases provided a basis for a perceptual study of prosodic boundaries. A break is characterized by the introduction of a pause, a lengthening of the vowel preceding the pause and a F_0 contrast. The three types of modification were taken as potential cues and stimuli with 1, 2 or 3 cues for a break were synthesized. Subjects had to decide whether the utterance they heard contained an incidental phrase or not. Though the pause was the factor with the greatest weight, it was not in itself sufficient to induce a significant proportion of "break" responses. The conjunction of F_0 modification and pause was necessary to get that result. Interactions between the three parameters were significant. The notion of prosodic configuration based on the integration of several cues is proposed to explain these results.

Intonation has often been described as the succession of F_0 patterns and a F_0 contour as a "series of targets within an envelope specifying F_0 range" /7, p.985/. Pitch movements should be sufficient to convey information about the internal structure of the sentence and even about its linguistic meaning /2/. Some other studies have focused on the relevance of durational cues to the perception of phrase boundaries /4, 10/. However a number of recent studies have suggested that, at an auditory level, separate acoustic features, specifically durational and spectral cues, are evaluated and integrated to form a single percept /1, 3, 8, 9/. So, intonation has been studied from a multiparametric point of view and the question of the integration of several information sources into a prosodic pattern has been raised.

The first aim of the present experiment is to assess how the acoustic correlates of prosody are processed to produce a perceptual continuity or, on the contrary, a perceptual break. In order to determine the nature of integration processes, it was necessary to independently vary these acoustic features. Moreover, it is often taken for granted that the contrast between intonation contours is primarily based on the direction and amplitude of F_0 variations /5/. This hypothesis will be confronted with results derived from the study of the perception of a prosodic break in French. If the data pertaining to perceptual integration are really of importance, parsing should be accomplished either as the summation of separate evaluations or as the weighting of the outputs of temporal and pitch processors,- depending on the presence of significant interactions.

A second aim of the experiment is to analyse if acoustically different prosodic configurations can be judged, under certain conditions, as linguistically equivalent. The linguistic constraints, for a given language, can help the listener to construct the semantic interpretation for an utterance. The variability of prosodic organizations for a given sentence is nonetheless well known. Stimuli have, accordingly, been chosen from the productions of female and male speakers, in order to evaluate how listener's parsing strategies are matched with speaker's production strategies, and so to assess the degree of adjustment of the listeners to the speech signal.

Preliminary experiment

In a preliminary study of the production of 14 speakers, French utterances containing an incidental phrase as their second constituent were compared with utterances with the same segmental content but no intonational break. This study revealed that the following characteristics could be used to differentiate the two cases :

- Lengthening of the vowel preceding the break (present in 62% cases, it varies between 30% and 100% of the basic duration of the vowel in the utterance that does not include a break)

- Pause preceding the parenthetical clause or phrase (present in 76% cases, its duration can reach 325 ms)

- Pause following the parenthetical constituent (present in 76% cases, its duration can reach 210 ms)

	Continuity contour		Incidental Phrase	
	F	M	F	M
/o/ duration	80 ms	100 ms	125 ms	218 ms
left pause	0 ms	0 ms	290 ms	110 ms
right pause	110 ms	70 ms	120 ms	210 ms
/i/ F_0 value	263-238 Hz	159-152 Hz	192-182 Hz	149-143 Hz
/bo/ F_0 val.	222-200 Hz	145-182 Hz	204-323 Hz	130-170 Hz
/deka/F_0val.	189/250-260 Hz	120/147-156Hz	170/244-270 Hz	120/120-160 Hz
Speech rate	5.08 syl/s	4.60 syl/s	3.59 syl/s	3.57 syl/s
Artic. rate	6.18 syl/s	5.17 syl/s	5.26 syl/s	4.65 syl/s

Table I - Original stimuli: Continuity contour "Thibaut de Caen est là", Incidental Phrase contour "Thibaut, de Caen, est là", as produced by a female speaker, F, and a male speaker, M.

- Highest F_0-peak of the utterance on the final syllable of the parenthetical (86% cases).

Moreover, a comparison of the utterances containing a parenthetical clause with the corresponding "normal" utterances revealed that the strategy that was most frequently adopted by speakers was a succession Rise + Rise on the first two constituents of the utterance containing a parenthetical versus a Fall + Rise pattern in the other case (76% cases were of this type).

The experiment described attempts to :
1) establish the validity of those differences as potential cues
2) specify their role in the perception of an intonational break. Is one of the cues primary or even necessary ? When do integration processes take place ? The problem is to determine how the information from various dimensions of the signal is put together to produce an interpretation .
3) evaluate the stability of a prosodic configuration through two different realizations, and the extent of the adjustment of listeners to speakers.

Experiment: Stimuli

The stimuli used in the experiment were derived from two productions of the utterance: "Thibaut de Caen est là" read by two different speakers (one male, one female). Those stimuli were then sampled at 10 KHz (12 coefficients) and a LPC analysis was performed (10 ms window). The relevant characteristics of the original stimuli are summarized in Table I.

We have simulated the presence of a parenthetical on "de Caen" using the following modifications of the stimuli without a prosodic break (Table II):

	Stimulus Number											
	1	2	3	4	5	6	7	8	9	10	11	12
V.L.	50	Ø	50	50	25	25	25	Ø	Ø	50	Ø	25
Pause	100	100	0	100	100	0	100	0	100	0	0	0
F_0	RR	RR	RR	FR	RR	RR	FR	RR	FR	FR	FR	FR

Table II - Characteristics of the stimuli: vowel lengthening (ms) before the potential pause, pause (ms), F_0 contour (R: rise, F: fall).

- Duration of the final vowel of the constituent preceding the potential parenthetical. Vowel duration took 3 values : original value, 25 ms and 50 ms lengthening.

- Presence or absence of a pause before the potential parenthetical (duration : 100 ms). (Those values were deliberately chosen as minimal).

- F_0 contour on the first two constituents of the utterance. F_0 contour was modified so as to reproduce the F_0 contour of two utterances containing a parenthetical taken from the production of the speakers of the two original stimuli. It should be noted that the strategies used by the two speakers are different (Table I). The three factors were combined to produce 12 stimuli per speaker presenting 0,1,2 or 3 cues favoring the break interpretation.

Subjects and procedure

Four blocks of stimuli were composed (2 x 12-item blocks per original speaker, each stimulus being presented twice). The stimuli were randomized within each block. There was a 7-sec pause between items and a 30-sec pause between blocks. 16 subjects passed the experiment. The factorial design was as follows:

$$S_{16} * D_3 * P_2 * F_2 * O_2$$

(S: subjects, D: vocalic lengthening, P: pause, F: F_0 pattern, O: presentation order). Subjects had to decide whether the utterance they heard contained an incidental phrase or not. They were also asked for a confidence-rating on a 3-point scale. The experiment proper was preceded by 4 stimuli derived from the original productions of the speakers. The answers of one of the subjects were not analysed because he could not identify those stimuli correctly.

Results

All the stimuli were significantly identified either with one category or the other, except the stimulus with 50 ms lengthening, F_0 contour corresponding to the presence of a parenthetical but no pause. Taking the degree of certainty into account, the answers were coded on a 0-1 scale (0=Non-parenthetical, high certainty, 1=parenthetical, high certainty). The results are summarized in Table

III (the stimuli are classified according to the number of potential cues and numbered for easy reference). An analysis of variance revealed that the order of presentation and speaker factors had no significant effect, in spite of the fact that the F_0 strategies for the two speakers were different. A second analysis was performed, grouping the scores corresponding to the non significant factors. The presence/absence of a pause (F (1, 14) = 165.7, p<.001), duration of the final vowel (F (2, 28) = 34.35, p<.001) and F_0 contour (F (1, 14) = 41.3, p<.001) were highly significant factors. Interactions between vowel-duration and pause (F (2, 28) = 10.3, p<.01), vowel-duration and F_0 contour (F (2, 28) = 6, p<.01) were also significant.

	Stimuli											
	1	2	3	4	5	6	7	8	9	10	11	12
m	.83	.77	.61	.60	.76	.21	.55	.18	.50	.22	.12	.09
sd	.05	.06	.15	.10	.05	.08	.06	.06	.07	.07	.03	.03

Table III - Mean scores and standard deviations as a function of the number of continuity cues (0-1 scale, cf.text).

As expected, stimuli 1 and 11 were identified to an utterance containing a break or an utterance without a break (respectively) (Figure 1). An examination of the stimuli containing only one "break-cue" (stimuli 8-10 and 12) showed that the presence of a pause was the factor with the greatest weight, though it was not in itself sufficient to induce a significant proportion of "break responses" (st.9). The influence of vowel length is limited (that factor accounts for a variation of 20 to 25% in the scores). Comparisons on that factor reveal no significant difference between 0 and 25 ms lengthening but a difference between these and 50 ms lengthening. The original F_0 continuity contour is sufficient to counterbalance the influence of the other factors (st.4) but the contour corresponding to the incidental utterances has a smaller influence than expected (st.8).

Discussion

The parameters that varied in this experiment are traditionally considered as cueing a boundary. Though they were confronted with partially incompatible indices, the subjects of the experiment managed to reconstruct a linguistically pertinent prosodic organization : the distribution of their answers was not due to chance or determined solely by one of the factors manipulated. Though we noted that the pause was the factor with the greatest weight, the conjunction of F_0 modification and pause was necessary to induce a significant proportion of "break responses". The hypothesis of F_0 pattern as the dominant cue is therefore not confirmed. F_0 contours have an important effect but only when they are accompanied by another compatible characteristic. More, the results suggest that there is no dominant cue but rather that we are dealing with a prosodic configuration based on the integration of several cues. Though there was no global interaction between F_0 and pause, partial comparisons reveal a significant interaction between pause and sharp rise on the first constituent. We may interpret this in terms of contrast formation: a sharp rise and a pause are integrated in one percept and the resulting configuration introduces a contrast between two successive chunks of the signal. These results suggest a parallel processing of duration and pitch followed by a weighting of their outputs. The contrastive configuration sets off the decision process.

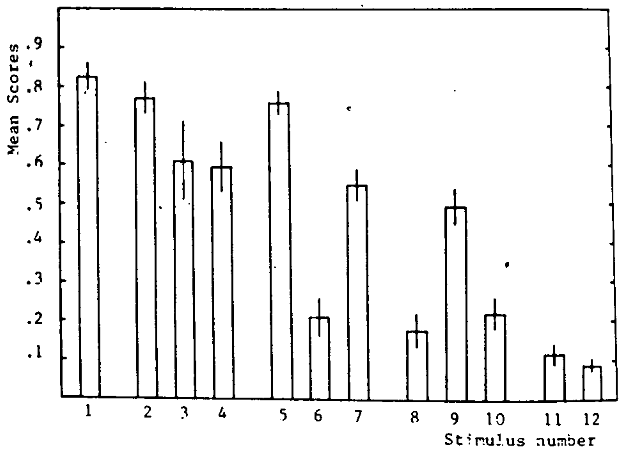


Figure 1 - Mean scores and confidence intervals as a function of the number of continuity cues. Stimulus 1: 0 continuity cue; stimulus 11: 3 continuity cues; 0-1 scale: cf. text.

As far as F_0 is concerned, the two speakers used two different strategies : the female speaker used a sharp rise resulting in a local contrast in the parenthetical case whereas the male speaker modified the hierarchy between two successive peaks (first peak higher than second peak in the non parenthetical case, two peaks of equivalent height in the parenthetical case). The effect of the two strategies is equivalent. But to account for the effect of the strategy used by the male speaker we must posit that a comparison is performed between two successive peaks, the two peaks being more than one second apart. Moreover, this comparison implies that a reference line (declination line) is taken into account : the parenthetical case would indicate that the break is accompanied by a resetting of the reference line.

The influence of vowel length is less easy to determine. The effect of this factor in the statistical analysis shows that it is taken into account but there is no linear effect of vowel lengthening. We propose that vowel length is not an independent cue but rather a property of the configuration : vowel lengthening causes the stimuli that include that characteristic to be judged as better tokens of the break configuration, provided this configuration is already established through other characteristics of the signal. Of course, the present conclusion applies only to rather small values for vowel lengthening.

CONCLUSIONS

The experiment showed that a set of characteristics including vowel length, pause and F_0 movement may be perceived as a configuration. It also showed that the F_0 contours that characterize a prosodic break in production are not always sufficient to cue that break in perception. The break is constructed primarily on the conjunction of F_0 pattern and pause. The significativity of the interaction between vocalic lengthening and pitch movements suggests that listeners do not proceed to a mere summation of information. Integration processes plausibly take place at the output of the duration and pitch processors, according to a weighting procedure /6/.

It's worth noting that the responses to the stimuli derived from the female and male utterances were on the whole the same. Though the two speakers used a different strategy, the F_0 patterns were characterized by a difference in F_0 level before and after the left boundary of the parenthetical constituent. The greater efficiency of F_0 movement in the presence of a pause might be explained in terms of the suppression of a retro-active masking /6/, induced by the pause.

The adjustment procedure seems to be automatic. The assumption is made that it is supported by the integration of the contrast between successive F_0 peaks.

The experiment reported here is also a confirmation that, in order to predict the effects of a prosodic configuration, a linguistic model of prosody should take into account, not only fundamental frequency movements, but also pauses and vowel length.

REFERENCES

/1/ A.S. Abramson, L. Lisker, "Relative power of cues: F_0 shift vs. voice timing", Haskins Lab., Status Report on Speech Research SR-77/78, 121-128, 1984.

/2/ W.A. Ainsworth, D. Lindsay, Perception of pitch movement on tonic syllables in British English, "J. of the Acoustical Soc. of America", 79(2), 472-480, 1986.

/3/ N. Bacri, "Fonctions de l'intonation dans l'organisation perceptive de la parole", Thèse de Doctorat d'Etat, Université Paris VIII, 1986.

/4/ I. Gee, F. Grosjean, Performance structures: a psycholinguistic and linguistic appraisal, "Cognitive Psychology", 15, 411-458, 1983.

/5/ Ph. Martin, Faits prosodiques et théorie de l'intonation, "Bulletin d'Audiophonologie", 1-2, 37-47, 1985.

/6/ D.W. Massaro, M.M. Cohen, Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction, "Perception and Psychophysics", 22, 373-382, 1977.

/7/ J. Pierrehumbert, Synthesizing intonation, "J. of the Acoustical Soc. of America", 70(4), 985-995, 1981.

/8/ B.H. Repp, A.M. Liberman, "Phonetic category boundaries are flexible", Haskins Lab., Status Report on Speech Research SR-77/78, 31-53, 1984.

/9/ M. Rossi, A. Di Cristo, D. Hirst, Ph. Martin, Y. Nishinuma, "L'intonation. De l'acoustique à la sémantique", Paris: Klincksieck, 1981.

/10/ N. Umeda, A.M. Quinn, Word duration as an acoustic measure of boundary perception, "Journal of Phonetics", 9, 19-28, 1981.

DISAMBIGUIERUNG GLEICHER WORTFOLGEN IM DEUTSCHEN DURCH PROSODISCHE VERÄNDERUNGEN

STEPHAN FORSTNER

Institut für Phonetik und Sprachliche Kommunikation
der Universität München
Schellingstr. 3, 8000 München 40, F. R. G.

ABSTRACT

Das Ausgangsmaterial für die Untersuchung bilden Äußerungspaare von vier männlichen Sprechern des Deutschen. diese Satzpaare weisen zum einen ein- und dieselbe Wortfolge. zum anderen unterschiedliche syntaktische Strukturen auf. Es wird untersucht. ob und in welcher Weise die phonetischen Parameter der Dauer. des F0-Bereichs und der Intensität einzeln und in Abhängigkeit voneinander eine disambiguierende Funktion bei oberflächenstrukturell ambigen Wortfolgen erfüllen.

EINFUHRUNG

Wenn man davon ausgeht. daß gleiche Wortfolgen in realen Kommunikationssituationen prosodisch verschieden realisiert werden. und daß diese unterschiedlichen Realisationen auch unterschiedliche kommunikative Funktionen erfüllen. so stellt sich die Frage. durch welche prosodischen Veränderungen diese Effekte zustande kommen. In der vorliegenden Untersuchung soll gezeigt werden. inwiefern die phonetischen Parameter F0, Dauer und Intensität bei der Disambiguierung solcher oberflächenstrukturell ambiger Sequenzen eine Rolle spielen.

Zu diesem Zweck wurden die Produktionsdaten mehrerer Sprecher erhoben und interpretiert. Sie sollen die Grundlage für experimentalphonetische Untersuchungen bilden. durch die geklärt werden soll. welche spezielle Funktion jeder Parameter für sich und in Relation zu den anderen Parametern erfüllt.

DAS AUSGANGSMATERIAL

Es wurde nach Material gesucht. das neben oberflächenstruktureller Ambiguität auch die Kriterien der Rekonstruierbarkeit natürlicher Kontexte. der entsprechenden Diskriminierbarkeit dieser Kontexte und der phonetischen Tauglichkeit zur akustischen Analyse und Manipulation natürlich produzierter Äußerungen erfüllt.

Aus einer Reihe von Satzpaaren verschiedenen Typs wurde gemäß der oben genannten Kriterien das folgende Satzpaar ausgewählt:

a) Ida meinte: "Nina hat angerufen".
b) "Ida" meinte Nina. "hat angerufen".

Für Äußerungen dieses Typs sind natürliche Kontexte durchaus vorstellbar. Die Disambiguierung erfüllt eine klare komunikative Funktion (etwa zur Beseitigung eines Mißverständnisses). Dementsprechend läßt sich auch die experimentalphonetische Aufgabenstellung formulieren (z.B. als Frage nach dem Namen des Sprechers, der etwas sagte oder meinte).

PRODUKTION DES AUSGANGSMATERIALS

Bei der Realisation des Ausgangsmaterials erhielten vier männliche Sprecher, deren normales Sprachverhalten keine dialektale Färbung aufweist, die Anweisung, das Satzpaar mindestens drei mal zu äußern. Der den Äußerungen zugrunde liegende Text war jeweils in eine Kontextbeschreibung eingebettet, aus der hervorging, daß die Sprecher eine Situation rekonstruieren sollten. in der ein Mißverständnis aufzuklären war. Ein imaginärer Gesprächspartner sollte davon überzeugt werden, daß Ida gemeint hat. Nina (und nicht Karin) habe angerufen. Analog dazu sind Kontextbeschreibung und Realisation von b) zu sehen. Zwangsläufig wurde dabei von allen Sprechern ein Kontrastakzent gesetzt.

ANALYSE DER DAUERVERHÄLTNISSE, DES F0-VERLAUFS UND -ONSETS UND DER INTENSITÄT

Das in einem schalltoten Raum auf Analogband aufgenommene Material wurde in eine digitale Kopie überführt. das resultierende digitale Oszillogramm am Bildschirm unter auditiver Kontrolle in zweifacher Weise segmentiert. Zum einen wurden die Segmente jeweils an den einzelnen Lautgrenzen definiert. und zwar bis zum Beginn des letzten Teils der Äußerung, also bis "hat angerufen". Die Dauern dieser Segmente wurden in Millisekunden gemessen. Zum anderen wurden die stimmhaften Passagen derselben Äußerungsteile periodenweise segmentiert. um die Hz- und db-Werte zu ermitteln.

In den folgenden acht Abbildungen sind die den ermittelten Zahlenwerten entsprechenden F0-Verläufe skizziert.

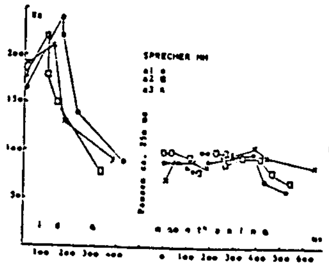


Abb. 1

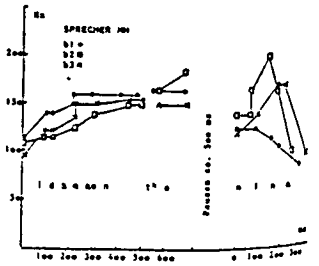


Abb. 2

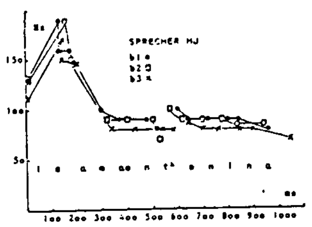


Abb. 3

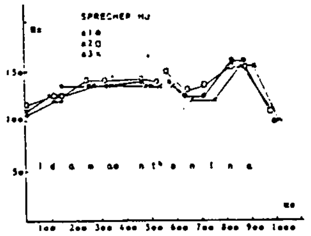


Abb. 4

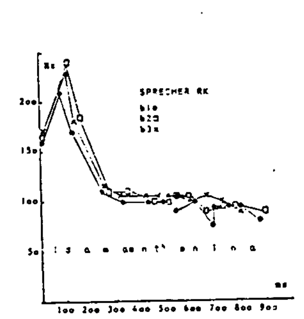


Abb. 5

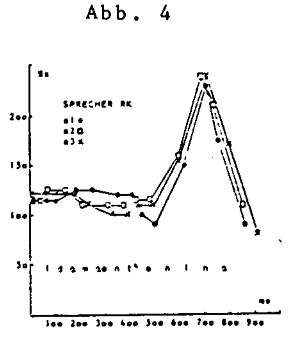


Abb. 6

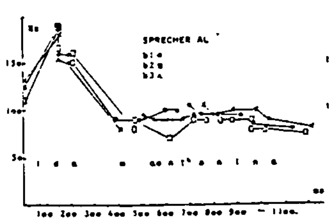


Abb. 7

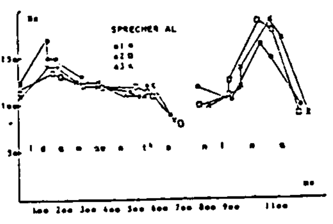


Abb. 8

1. Dauern

Für die Interpretation der Daten war hinsichtlich aller drei phonetischen Parameter in erster Linie die unmittelbare Umgebung der Phrasierungsgrenzen interessant. also der Schwalaut und die anschließende Sequenz /nina/ in den a-Versionen. die Sequenzen /ida/ und /nina/ in den b-Versionen.

Die vier Sprecher wiesen durchaus Unterschiede in der phonetischen Form ihres Disambiguierungsverhaltens auf, dennoch sind einige gemeinsame Regularitäten festzustellen.

Bei den Sprechern AL und MH ist zu sehen. daß an den Phrasierungsgrenzen zumindest parziell Pausen gesetzt werden. während die Sprecher HJ und RK in allen Fällen ohne den Einsatz von Pausen disambiguieren.
Es läßt sich also feststellen. daß die Pausensetzung als Disambiguierungsmittel zweifellos eine entscheidende Funktion übernehmen kann, zugleich stellt sie jedoch keinen zwingend notwendigen Parameter der Disambiguierung dar.

An den Stellen. an denen Pausen gesetzt werden. ist das sogenannte 'pre-pausal lengthening'. also eine längere Vokaldauer zu becbachten. In den a-Versionen ist hiervon der Schwalaut betroffen. Beim Sprecher AL liegen die Werte hier um durchschnittlich ca. 60 ms über denen der b-Versionen. beim Sprecher MH um durchschnittlich ca. 100 ms.

In den b-Versionen ist pre-pausal lengthening jeweils im a von "Ida" und "Nina" zu beobachten. Beim Sprecher MH liegen die Werte des a von "Ida" um durchschnittlich 110 ms über denen der a-Versionen. Nach dem a in "Nina" wird nur in einem Fall eine Pause gesetzt (b3). Der Wert liegt hier um mehr als 100 ms über den Werten von b1 und b2 und auch deutlich über den entsprechenden Werten der a-Versionen. Bei Sprecher AL wird in den b-Versionen nach "Ida" nur in einem Fall eine Pause gesetzt (b3), ohne daß sich der betreffende Wert wesentlich von denen der beiden anderen b-Realisationen unterscheidet. Bei diesem Sprecher zeigt sich jedoch im nachfolgenden Nasal von "meinte" (b1 und b2) ein deutlicher Glottalisierungseffekt, der sowohl auditiv als auch am digitalen Oszillogramm verifiziert werden kann. Die betreffenden Nasalsegmente weisen Dauern von 111 und 155 ms auf gegenüber 52 ms in b3. Diese Beobachtung legt die Vermutung nahe, daß Glottalisierung an Phrasierungsgrenzen anstelle von Pausensetzung und pre-pausal lengthening eine disambiguierende Funktion übernehmen kann. Dies wäre systematisch zu untersuchen. Auch nach dem finalen Vokal von "Nina" ist bei AL in den b-Versionen nur in einem Fall eine Pause vorhanden. ohne daß sich der entsprechende Wert von denen der beiden anderen b-Realisationen unterscheidet. Auch hier ist nachfolgende Glottalisierung zu erwarten.

Bei den Sprechern HJ und RK ist pre-pausal lengthening nicht festzustellen. Die betreffenden Werte der a- und b-Versionen unterscheiden sich nur in geringem Maße voneinander.

Eine weitere Auffälligkeit in den Dauerverhältnissen ist in Zusammenhang mit der Akzentsetzung zu sehen. Die Hauptakzente liegen durchwegs auf der jeweils ersten Silbe von "Nina" in den a-Versionen und von "Ida" in den b-Versionen. Es zeigt sich, daß die akzenttragenden Silben nahezu durchgehend höhere Werte aufweisen als die direkt vergleichbaren Werte der jeweils anderen Version. Die einzige Ausnahme bilden die a-Realisationen von "Nina" des Sprechers MH. Die Dauerwerte für die erste Silbe unterscheiden sich hier nicht wesentlich von den Werten der b-Versionen. Es ist zu vermuten. daß dies mit dem sprecherspezifischen Rhythmus und Tempo zusammenhängt, die vorhergehenden Pausen weisen immerhin eine relativ hohe Dauer von ca. 500 ms auf.

2. F0-Onset und -Verlauf

Bei den Sprechern HJ und AL weist der Schwa-Laut in den a-Versionen durchgehend ein Absinken der Tonhöhe zwischen 20 und 40 Hz auf, während die vergleichbaren F0-Werte in den b-Versionen progredient verlaufen oder gering abfallen. Zugleich liegt der F0-Onset der a-Versionen um bis zu 50 Hz höher als der der b-Versionen. Allgemein ist zu sagen daß die Schwa-Laute der b-Versionen aus einer nur sehr begrenzten Anzahl von Stimmtonperioden bestehen, an einigen Stellen war es überhaupt nicht möglich, periodenweise zu segmentieren.

Bei den Sprechern RK und MH verläuft der Schwa-Laut in den a-Versionen progredient oder zeigt einen Anstieg, während in den b-Versionen vor allem Progredienz (in einem Fall ein leichter Anstieg um 10 Hz) bei durchwegs niedrigerem Onset zu beobachten ist.

Somit läßt sich bereits ein erstes wichtiges Ergebnis aus der Erhebung der Produktionsdaten formulieren: Sowohl Level- als auch Fall- und Rise-Bewegungen können an gleicher Stelle als Mittel zur Disambiguierung eingesetzt werden. Als wesentlich konsistenterer Parameter erweist sich hier der F0-Onset.

Die akzenttragende Silbe von "Nina" weist im vokalischen Bereich bei allen Sprechern durchgehend einen deutlichen Anstieg um bis zu 90 Hz auf, während die vergleichbaren Silben in den b-Versionen bei ebenfalls deutlich niedrigerem Onset in allen Fällen progredient verlaufen.

Die zweite Silbe von "Nina" zeigt in den a-Versionen ein ebenso deutliches F0-Absinken um bis zu 100 Hz. In den b-Versionen ist dagegen nur ein leichtes Absinken um durchschnittlich ca. 10 Hz oder ein progredienter Verlauf festzustellen, auch hier bei durchgehend niedrigerem Onset.

Auch die akzenttragenden Silben von "Ida" in den b-Versionen zeigen eine deutliche Rise-Bewegung, während in den vergleichbaren a-Versionen Level- sowie leichte Rise- und leichte Rise-Fall-Bewegungen zu sehen sind. Ebenso weist die zweite Silbe von "Ida" in allen b-Versionen ein deutliches Absinken auf. In den a-Versionen findet sich dagegen in erster Linie Progredienz, aber auch ein leichtes Absinken und ein leichter Anstieg des Stimmtones.

3. Intensität

Hier sind in erster Linie die an den akzenttragenden Silben gemessenen db-Werte sowie jene an den damit direkt vergleichbaren Stellen der jeweils anderen Versionen interessant, also die jeweils ersten Silben von "Ida" und "Nina". Dabei war zu erwarten, daß die db-Werte der hauptakzenttragenden Silben wesentlich höher seien als die der nicht akzentuierten: (Ein Unterschied von ca. 6 db ist mit etwa doppelt so hoher bzw. niedriger Lautstärke gleichzusetzen.

Was die erste Silbe von "Nina" betrifft, so hat sich dies durch die erhobenen Daten voll bestätigt, während insbesondere bei den Sprechern MH und HJ in der ersten Silbe von "Ida" das erwartete Ergebnis nicht vorliegt. Dieser Umstand deutet darauf hin, daß bei diesen Sprechern im initialen Teil der Äußerungen wesentlich stärker mit F0- und Dauerverhältnissen disambiguiert wird.

ZUSAMMENFASSUNG

Die erhobenen Produktionsdaten von vier Sprechern zeigen in deutlicher Weise, daß Dauerverhältnisse, F0-Bewegungen und F0-Onset sowie die Intensität phonetische Parameter zur Disambiguierung strukturell ambiger Satzpaare sind. Daher ist auf der Basis der hier ermittelten Daten eine Grundlage für die Generierung von Teststimuli bzw Testkontinua gegeben, bei denen die genannten Parameterwerte so verändert werden sollen, daß es möglich wird, genauer zu untersuchen, unter welchen Bedingungen Ambiguität bzw. Disambuiguität resultiert.

Lexical effects in phoneme monitoring:
Facilitatory or inhibitory?

Uli H. Frauenfelder

Max-Planck-Institut für
Psycholinguistik
Nijmegen, The Netherlands

Juan Segui

Laboratoire de Psychologie Experimentale
Université René Descartes and EPHE
Paris, France

Ton Dijkstra

Max-Planck-Institut für
Psycholinguistik
Nijmegen, The Netherlands

Abstract

This paper addresses the questions of **how** and **when** lexical information influences phoneme detection in two phoneme monitoring experiments. In the first, the position of the stop consonant target (word initial, before uniqueness point, after uniqueness point, word final) and the lexical status of the target bearing item (word or nonword) were manipulated to pursue the temporal question. A contribution of the lexical level to phoneme detection (reflected by large RT differences between targets in words and nonwords derived from the words) was found only when the target came in the two positions after the uniqueness point. In a second experiment, the contribution of the lexicon was made incompatible with the bottom-up evidence for targets by placing them in words where they did not belong (p target substituted for l producing "stimupi"). No inhibitory effect of the lexical level was obtained even in cases where the target and substituted phonemes differed minimally. These results taken together indicate that the lexicon exerts its effect only **after** word recognition and as **positive** feedback suggesting strong limitations in the way in which lexical information can affect speech perception.

Introduction

No-one would dispute the claim that we recognize words on the basis of an analysis of the speech sounds of which they are composed. Controversial, at least in psycholinguistic circles, is the inverse claim that our perception of speech sounds depends upon the words they make up. In this paper, we will evaluate these claims about the relative importance of **bottom-up** and **top-down** processes mediating between the sublexical and lexical representations. To arrive at a proper description of the information flow between these two levels, we will address the questions of **how** and **when** the lexical and sublexical information sources are brought together.

To investigate these questions we used the phoneme monitoring task in which subjects are asked to detect as quickly as possible previously specified phoneme targets that appear in sentences or lists of words. Previous research [1,4,6,7] has shown that phoneme detection latencies are sensitive to lexical variables indicating an influence of the lexical level upon speech perception as reflected by the phoneme detection process. Our objective here is to examine empirically two opposed accounts of such lexical effects.

In **autonomous models** of language processing [3], it is assumed that bottom-up processes produce their output autonomously; top-down lexical information is not allowed to influence the bottom-up mechanisms responsible for phoneme perception. In order to account, nonetheless, for the presence of lexical effects in phoneme monitoring, "race models" [1,2] and "dual code models" [4] have been developed in which phoneme identification can be made on the basis of two different "outlets" or representations: a lexical and a sublexical level. In a race model account, there are two independent ways in which a phoneme target can be detected. The first target detection procedure depends upon the computation of a sublexical representation. In the second, target detection depends upon lexical access which makes available the phonological information associated with a particular accessed lexical entry. There is a race between these two processes with the one that reaches completion first providing the phoneme detection response. The presence or absence of lexical effects is explained in terms of the outcome of the race between these two independent and competing outlets.

Interactive activation models are designed to account for the integration of multiple sources of information or constraints in speech perception. The most explicit model constructed within this framework is TRACE [5]. In TRACE there are several levels of interconnected processing units corresponding to distinctive features, phonemes and words. The critical interactive aspect of this model is that word units can provide top-down feedback to phoneme units by increasing their level of activation. Hence, phoneme recognition (the moment a phoneme reaches a criterial level of activation with respect to the other phonemes) depends on **both** the amount of bottom-up activation from the distinctive feature level and the amount of top-down activation from the word level. Subjects responding in the phoneme monitoring task are assumed [8] to make direct and exclusive use of activated phoneme units. The presence or absence of lexical effects in phoneme monitoring is explained within the TRACE framework by varying lexical contributions to the phoneme's activation.

Although these two basic model types are radically different in nature, they make many of the same predictions and appear to be consistent with most of the data available in the phoneme monitoring literature. Given this state of affairs, it is critical to collect additional performance data that will allow us to further constrain these types of models. In particular, it is essential to determine how and when the lexical level contributes to the speech analysis as reflected by the phoneme monitoring task.

In order to trace the time-course of lexical effects, we selected targets in four different positions with respect to the uniqueness point (UP) of the word. The UP was defined as that point at which a word's initial part is shared by no other word listed in a phonetic dictionary. Nonwords were created from these target-bearing words by changing one or more phonemes, but keeping the target's local phonemic environment as constant as possible. The differences in detection times between phoneme targets in the same position in matched words and nonwords provided an approximate measure of the lexical contribution.

EXPERIMENT I

Subjects

Thirty-eight undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in the experiment.

Materials and procedure

The test stimuli consisted of 120 words and 120 matched nonwords. The target phonemes (p , t , or k) occured in four different positions within target-bearing words (word onset, before the UP, after the UP and word offset) and nonwords (nonword onset, before the nonword point, after the nonword point and nonword offset). The nonword point (NWP) is that point at which the item becomes a nonword moving from its beginning towards the end.

Target-bearing items were embedded in counterbalanced lists made up of other words and nonwords not containing the target phoneme. Twelve such lists, each containing 60 items, were created and divided into two blocks for counterbalanced presentation to the subjects. For each list subjects were asked to detect as quickly as possible one of the three targets (specified by means of a visual display).

Results

Mean reaction times (measured from the burst of item-initial targets and from closure for the targets in the remaining three positions) were computed for each subject and each experimental item. All responses less than 100 ms. or greater than 1000 ms. were not included in the computation of the means. Three subjects with more than 15 % errors were also excluded from the analysis. Figure 1 shows the results for words and nonwords broken down according to target position.

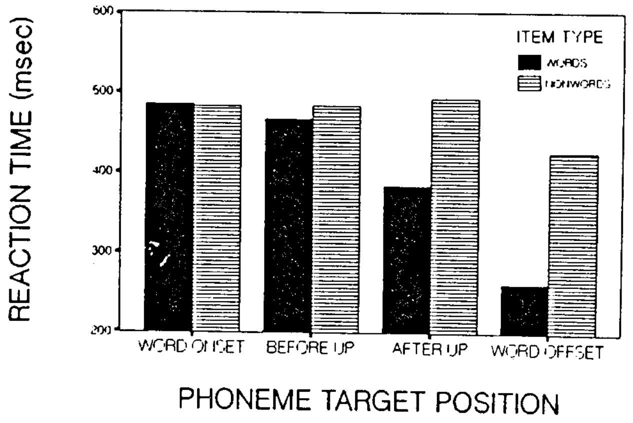


Fig. 1. Mean RTs for words and matched nonwords as a function of phoneme target position

An analysis of variance showed that the two main effects, lexical status and target position both were highly significant by subject ($F_1(1,34) = 371$; $p < .001$; $F_1(3,102) = 136$; $p < .001$) and by item ($F_2(1,9) = 48.2$, $p < .001$; $F_2(3,27)$, $p < .001$). The interaction between these two effects was also highly significant ($F(3,102) = 168$, $p < .001$). Post-hoc comparisons using the Scheffé test (S-method) revealed that the effect of lexical status was significant (at the .01 level) only for two target positions after UP.

Discussion

This experiment has shown a clear interaction between target position and lexical status of the target bearing item. The differences in detection latencies between targets in matched words and nonwords increased as the targets appeared later in the word. This increase is not linear; the differences for the two positions after the UP were significantly larger than those for the initial positions.

The global pattern of facilitation, in particular, the sharp increase in facilitation after the UP, is consistent with both the autonomous and interactive models. For the former, the lexical level can influence phoneme detection only after the phonological code associated with the target bearing word has been accessed. This code is normally assumed to become available at a discrete moment in time once the word has been recognized, that is, after the UP. In TRACE, the strength of lexical feed-back increases with the level of activation of the lexical units containing the target phoneme. Hence, the lexical level is involved from the beginning of the perceptual process, but its effects build up gradually and continuously as the lexical units themselves receive more activation. The activation of the target-bearing word should increase dramatically around the UP, and as a consequence so should the lexical influence on the phonemes.

In the following section we want to examine more closely **how** lexical information affects phoneme monitoring. In particular, we want to explore whether the lexical level can exert not only a **facilitatory** effect as in the first experiment but also an **inhibitory** effect upon the phoneme detection procedure. In order to test for this latter possibility, we created a situation in which the lexical information was made incompatible with the bottom-up evidence for the target. Subjects were asked to detect a target phoneme that appeared in the place of another phoneme situated after the UP (e g. replacing the t in "simplicité" by the target phoneme d giving "simplici*d*é").

The two models described above appear to make different claims concerning the existence of inhibition effects. The interactive activation model predicts that a phoneme arriving after the UP (such as the t in simplicité) receives excitatory feed-back from the lexical level. This phoneme in turn inhibits the other phonemes (such as the substituted d) that occur at this point in the sequence. The decreased activation level of the target phonemes (d) should translate into slower detection times as compared with those to detect the same phoneme d in another nonword such as "fimplicidé" where the target phoneme should submit to neither excitatory nor inhibitory influences from the lexical or phoneme levels respectively. As a consequence, any difference between the detection times for the identical targets d, in the same local phonemic environments " i **d** é " in the direction of slower detection times for the former type of nonword "simplicidé" would constitute evidence for mediated lexical inhibition effects.

The autonomous race model does not allow for inhibitory effects of this type. Since the two competing response outlets, lexical and pre-lexical, function independently, the lexical code cannot affect the elaboration of the pre-lexical code. Furthermore, the lexical code cannot contribute to the detection response for the target in "simplicidé" since this phoneme target is not contained in the lexical code for the word corresponding to the initial part of this nonword. As a consequence the target is always detected on the basis of the pre-lexical code for both nonwords leading the autonomous race model to predict no difference between the detection times in the two types of nonwords.

EXPERIMENT II

Subjects

Eighteen students of the University of Paris V, all native speakers of French, participated voluntarily in this experiment.

Materials and procedure

The test items consisted of 12 matched pairs of three- or four-syllable nonwords containing targets (4 pairs for each of the three targets: /d/, /t/, /k/). The inhibitory nonwords (INW) were constructed by replacing phonemes located after the UP in words by the target phoneme. Thus, for example, from the word "simplicité" whose UP lies well before the target, a INW item "simplicidé" was derived, in which the target-phoneme /d/ replaces the original phoneme /t/. Replaced and target phonemes differed only in the feature of voicing.

Matching neutral nonwords (NNW) were derived from each INW by replacing the initial phoneme of INW items by another phoneme of the same manner of articulation (e g., from the INW "simplicidé" the NNW item "fimplicidé" was created). All nonword-points for NNW items were located before the end of the second syllable, well before the target phoneme.

Eighteen target-bearing words (six for each target type) were also included to confirm the existence of lexical facilitation effects; in half of these the target-phoneme was located before the UP (for example, "ouverture" ("opening") with the target-phonerne /t/), the other half with the target after the UP (for example, "profitable" with the target-phoneme /t/). These target-bearing words and nonwords were embedded in one of three experimental lists of 64 items each (32 words and 32 nonwords).

Results

Less than 5 % of the responses were eliminated for the computation of the means (latencies smaller than 100 ms. or longer than 1000 ms.). Figure 2 summarizes the means for both the words and the nonwords.

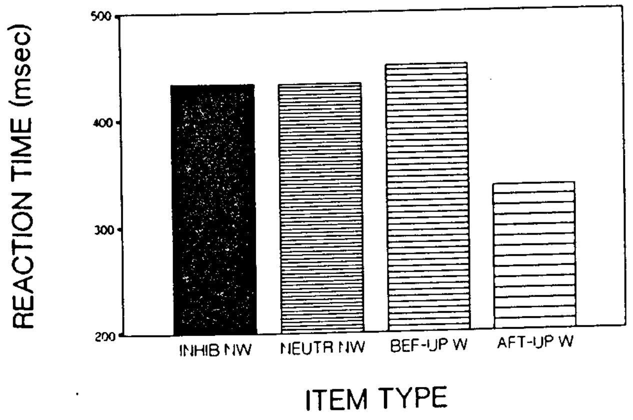


Fig. 2. Mean RTs for inhibitory and neutral nonwords and for words with target before and after UP

An analysis of variance performed on the nonword data indicated that neither the factor nonword type nor the type of target-phoneme (/d/, /t/, /k/) introduced significant effects ($F < 1$ in both cases). The interaction between the factors was also not significant ($F(2,34) = 2.16$, $p > .10$). Furthermore, there was no difference in the percentage of errors or omissions for the two nonword conditions. An analysis of variance for the words, however, showed that the difference between the two conditions (before and after UP) was highly significant ($F_1(1,17) = 55.40$, $p < .0005$).

Discussion

In this experiment, we have investigated whether the lexical influence, observed to be highly facilitatory after the UP in the first experiment, can also be inhibitory when the lexical information is incompatible with the target to be detected. The results provided no evidence for such inhibitory lexical effects, but did replicate facilitatory lexical effects. Predictions of inhibition are implicitly made in the interactive activation framework, but are excluded by the autonomous model. As a consequence these results appear to be more compatible with the autonomous model. The absence of inhibition, although problematic for interactive models can, nonetheless, be explained within this framework. According to such an account, the bottom-up activation of the target phoneme is so effective that the inhibitory influence of the replaced (appropriate) phoneme does not show up. The strength of the bottom-up activation dominates and hides the ephemeral inhibitory lexical effect. If this account is correct, then one might expect inhibitory effects to vary as a function of the strength of bottom-up activation. We are now in the process of exploring the possibility that inhibitory effects will emerge with acoustically less clear targets.

The two experiments taken together suggest an asymmetry in the way lexical information can contribute to the bottom-up analysis underlying phoneme detection. The results presented here indicate that the lexicon exerts a **facilitatory** but not an **inhibitory** influence upon bottom-up processing **after** word recognition. Thus, these results show strong limitations in the way in which lexical information can affect phoneme processing that must be taken into account by both interactive and autonomous models.

REFERENCES

1) Cutler, A., Mehler, J., Norris, D. & Segui, J. Phoneme identification and the lexicon. *Cognitive Psychology*, 1987, (in press)

2) Cutler, A. & Norris, D. Monitoring sentence comprehension. In W.E. Cooper & E.T.C. Walker (Eds.) *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, Hillsdale, N.J.:Lawrence Erlbaum Associates, 1979.

3) Forster, K.I. Levels of processing and the structure of the language processor. In W.E. Cooper & E.C.T. Walker (Eds.) *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

4) Foss, D.A. & Blank, M.A. Identifying the speech codes. *Cognitive Psychology*, 1980, 12, 1-31.

5) McClelland, J.L. & Elman, J.L. The TRACE model of speech perception. *Cognitive Psychology*, 1986, 18, 1-86.

6) Marslen-Wilson, W.D. Function and process in spoken word recognition. In H. Bouma & D.G. Bouwhuis (Eds.) *Attention and Performance X: Control of Language Processes*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1984.

7) Segui, J., & Frauenfelder, U. The effect of lexical constraints upon speech perception. In F. Klix & H. Hagendorf (Eds.) *Human Memory and Cognitive Abilities: Symposium in Memoriam Hermann Ebbinghaus*. Amsterdam: North Holland, 1986.

8) Stemberger, J.P., Elman, J.L. & Haden, P. Interference between phonemes during phoneme monitoring: Evidence for an interactive activation model of speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1985, 11(4), 475-489.

INITIAL SPEECH SOUND PROCESSING IN SPOKEN WORD RECOGNITION

PHILLIP DERMODY KERRIE MACKIE RICHARD KATSCH

Speech Communication Research Section
National Acoustic Laboratories
126 Greville St., Chatswood. N.S.W. 2067. AUSTRALIA.

ABSTRACT - The present study uses the gating paradigm to investigate the initial speech sound (ISS) in spoken word recognition. Results are presented for spoken words and consonant-vowel (CV) syllables, which both show recognition of the ISS in the first 30 msec. Acoustic analyses of the ISS show similarities between the words and syllables and are consistent with the templates proposed by Stevens & Blumstein (1). It is suggested that the time course of ISS perception indicates the need to change present models of spoken word recognition.

INTRODUCTION

In a previous investigation of initial speech sound processing we confirmed that listeners could identify the initial speech sound in CV syllables, consisting of a stop consonant plus the vowel /a/ in the first 30 msec after onset (2). This was paralleled by similar performance for initial consonant recognition during the first 30 msec in words. The present study investigates the same effects but examines in more detail the time course for speech sound recognition during the initial 30 msec for individual sounds.

STIMULI

A set of 12 English consonants (including the six stops) plus the vowel /a/ were recorded by a male speaker onto a computer based speech storage/ editing system. In addition, a set of words beginning with similar consonants were recorded. A subset of these words included words which began with a stop consonant plus the vowlel /a/ (eg. tartan; gardener; particle). The stimuli were digitised at 36KHz sampling rate with filters for input/output set at 12KHz. Each syllable and word was visually displayed and segments in increments of 10 msec, were marked and labelled. The endpoint of each gated stimulus was made to the nearest zero crossing point to avoid audible clicks. The gated stimuli were output to audio tape in sequential order (ie. 10, 20, 30 etc msec presented in order) with 4 seconds between each presentation. For example, the first 10 msec stimulus was produced (gate 1) followed 4 seconds later by the 20 msec stimulus (gate 2) etc. For the CV syllables, gates in increments of 10 msec from 10 to 100 msec were recorded. For the spoken words, 10 msec gate increments were used for the first 6 gates (ie. to 60 msec of the word after onset) and then incremented in 30 msec gates to the completion of the word. That is, for the words, gate 7 was 90 msec in duration from onset while gate 8 was 120 msec etc.

PROCEDURE

All subjects were given minimal practice to familiarize them with the task. The words and syllables were given in separate test sessions using different sets of subjects. The subjects were University undergraduates with normal hearing (N= 23 per test) who were paid for their participation. Subjects were instructed that they would hear short bits of a speech sound which would increase on succeeding trials and that they should write down their response after each presentation. Subjects were instructed to guess if unsure since recording a response per trial was mandatory. They were also told that the syllable or word could begin with any consonant sound in English and their task was to identify the initial consonant for the CV syllables and the initial consonant and the whole word for the word stimuli.

RESULTS

While the full set of CV syllables and words were presented to the subjects only the results for the stimuli beginning with the stop consonants are presented here. Figure 1 shows the results for the 6 stop consonants plus the vowel /a/. The cumulative percentage of subjects who obtained correct initial consonant recognition without subsequent errors are shown for each gate duration. For comparison, the results of our previous experiment (2) for tha same CV gates presented as part of a closed set of 6 alternative responses are also presented. The figure shows similar performance patterns for open and closed set responses. The open set performance reaches an asymptope slightly later (at about 40 msec)than the closed set performance which reaches an asymptope at the 30 msec gate.

Figure 2 shows the contribution of two of the CV syllables to the curve including the best result (from /ta/) and the worst result (from /ga/). The figure presents the percentage of subjects who obtain correct recognition as a function of gate duration. In the case of /ta/ all subjects have correctly recognised the initial sound correctly at the 20 msec gate, while for the /ga/ only 4% have recognised the sound at the 20 msec gate and at the 40 msec gate 80% of subjects have identified the initial /g/.

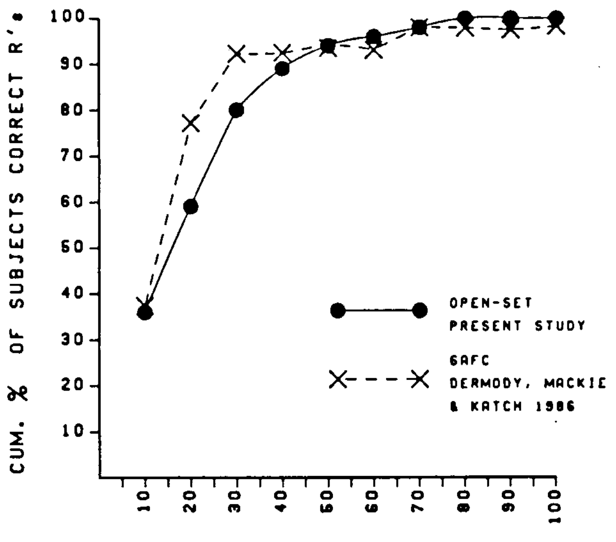


Figure 1
Initial consonant recognition point for 6 stops

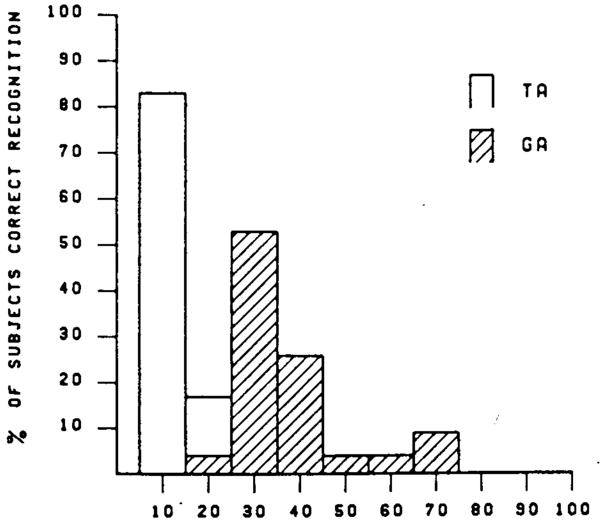


Figure 2
Initial consonant recognition point for /ta/,/ga/.

The results for the identification of the ISS in words were similar to those for the CV syllable. Figures 3 and 4 present the results for "tartan" and "gardener" for comparison with /ta/ and /ga/ in figure 2. In figure 3 -"tartan"-, the initial /t/ is correctly recognised by the majority (84%) of subjects at the first gate (10 msec), while the /g/ in "gardener" requires about 60 msec for equivalent performance levels. This shows a similar pattern as shown for gate recognition in syllables. In both /ta/ and "tartan", the initial /t/ is correctly recognised by most subjects at 10 msec while the /g/ requires about 60 msec for "gardener" and about 40 msec for /ga/.

Figures 3 and 4 also show the vowel and word recognition points for the words. The overall results for the word gates is similar in pattern to the word gates reported by Grossjean (3). Figure 3 shows the /a/ sound in "tartan" is recognised by only 4% of sujects by the 60 msec gate while in figure 4 , about 30% of subjects recognise the vowel at 60 msec. Both words however,require 90 to 120 msec for the majority of subjects to correctly identify the /a/. Despite similar word durations "tartan" is recognised by the majority of subjects by about 570 msec, while recognition for gardener is delayed until 690 to 720 msec.

DISCUSSION

In a previous study (2) we reported that recognition of the initial speech sound occurred for both syllables and words containing initial stops in about 30 msec. Acoustic analysis revealed the patterns suggested by Stevens & Blumstein (1) and Blumstein & Stevens (4) for these durations. It was also noted that the same acoustic patterns were also present for the 10 and 20 msec gate stimuli where subjects were able to recognise the initial sounds at better than chance level. Similar templates were also found for the stimuli in the present study which is consistent with the reports by Blumstein & Stevens (4) and by Kewley- Port, Pisoni, & Studdert-Kennedy (5) that the information is sufficient to recognise the initial speech sound. The present study extends these findings to the case of word recognition and indicates that the initial speech sound could play and important role in lexical access. These results suggest that the word recognition process begins much earlier than the 100 to 200 msec after word onset as suggested by Marslen-Wilson (6). That is, at least for stops in the word initial position, lexical access may begin around 30 msec after word onset. This finding probably needs to be incorporated into theories of the time course of lexical access and word recognition and would certainly provide a time allowance sufficient for elaborate search procedures in these processes.

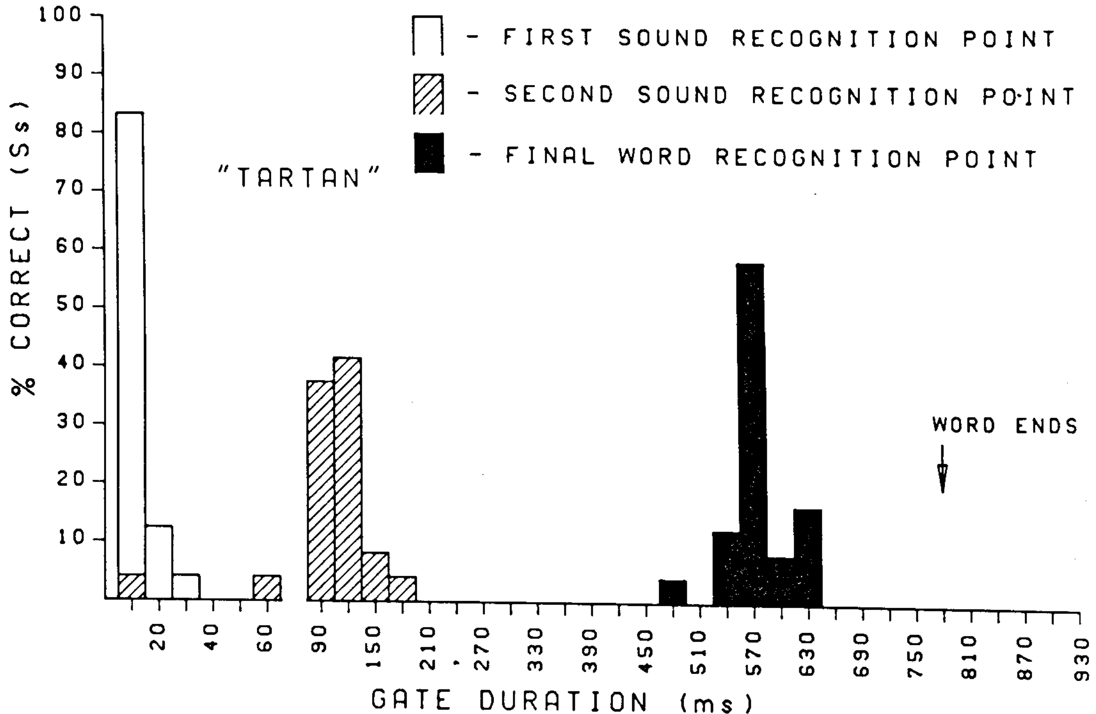


Figure 3 Percentage of subjects obtaining correct identification at each gate.

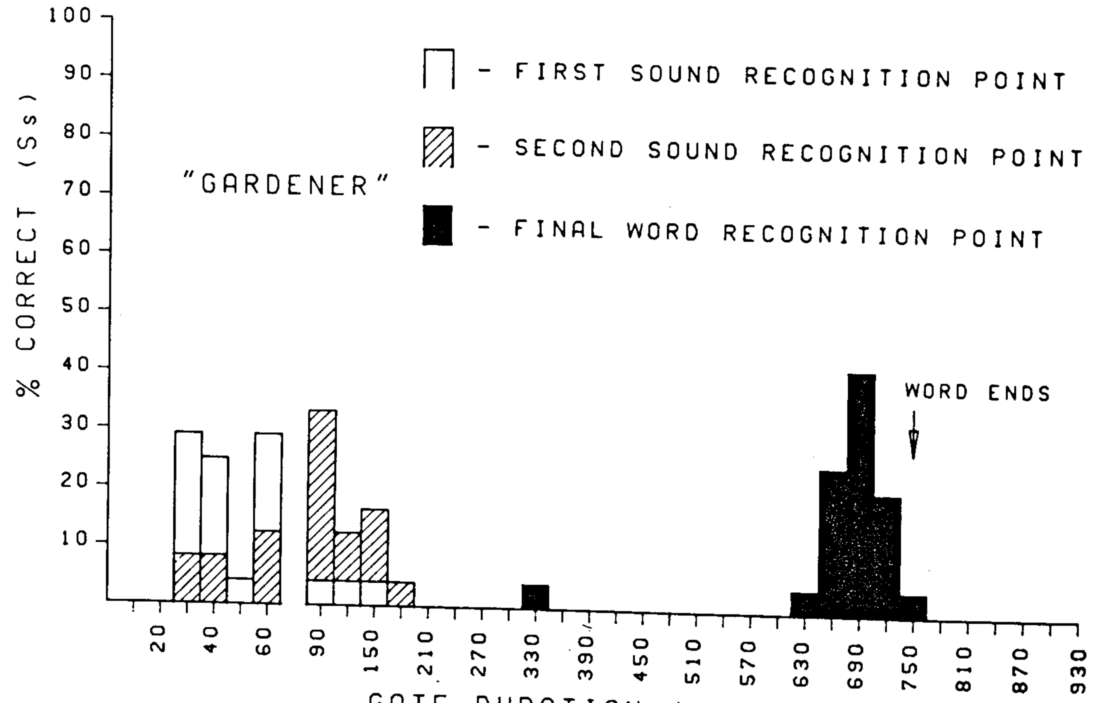


Figure 4 Percentage of subjects obtaining correct identification at each gate.

REFERENCES

(1)Stevens, K. & Blumstein, S. (1978) "Invariant cues for place of articulation in stop consonants". J. Acoustical Society of America, 64, 1358-1368.

(2)Dermody, P., Mackie, K. and Katsch, R. (1986) "Initial speech sound processing in spoken word recognition." Proceedings of the 1st Australian Conference on Speech Science & Technology, Canberra: Australian National Univ.

(3)Grossjean, F. (1980) "Spoken word recognition processes and the gating paradigm", Perception & Psychophysics, 28, 267-283.

(4)Blumstein, S. & Stevens, K. (1980) "Perceptual invariance and onset spectra for stop consonants in different vowel environments" J. Acoustical Society of America, 67, 648-662.

(5)Kewley-Port,D., Pisoni,D. & Studdert-Kennedy, M. (1983) "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants", J. Acoustical Society of America, 73, 1779-1793.

(6)Marslen-Wilson, W. & Tyler, L. (1980) "The temporal structure of spoken language understanding". Cognition, 8, 1-71.

Speech perception in predictable and non-predictable contexts.

EVA MAGNUSSON KERSTIN NAUCLÉR
Department of Linguistics
Lund University
Helgonabacken 12
S-223 62 Lund, Sweden

ABSTRACT

The ability of language disordered pre-school children to perceive speech masked by noise and to use contextual cues is investigated and compared with the same ability shown by two other groups: a matched group of normally speaking pre-school children and a group of normally speaking adults. The scores on the perceptual tasks are correlated with the pre-school subjects' performance on different language tasks such as syntactic production, comprehension and awareness as well as on verbal short term memory.

INTRODUCTION

In the perception and interpretation of linguistic messages two types of processes are considered: data driven, or bottom-up processes, and concept driven, or top-down processes. The processes work simultaneously, but disturbances in the signal may cause one of the processes to dominate occasionally. This compensatory mechanism was described for reading in an interactive model by Stanovich /1/.
When interpreting speech (listening), compensation for the constantly present but varying noise results in a more or less heavy reliance on top-down strategies. When interpreting writing (reading), heavier reliance on top-down strategies occurs when e.g. misprinting makes word decoding difficult. However, poor readers with poor decoding ability (poor bottom-up strategies) are said not always to compensate by a more efficient use of contextual cues (top-down strategies). Therefore, an interesting question is whether poor readers and children running the risk of becoming poor readers, i.e. language disordered pre-school children, show the same reluctance or inability to use top-down strategies in spoken language when bottom-up strategies are insufficient.

PURPOSE

This study is part of an investigation on language disordered and normally speaking pre-school children's reading and spelling acquisition (cf Magnusson & Nauclér, /2/). The aim of the study reported here is to examine language disordered pre-school children's ability to use contextual cues when the signal-to-noise ratio in speech is too high to permit data processing alone.
The questions to be answered are the following:
- Are language disordered pre-school children able to understand speech masked by noise to the same extent as normally speaking children of the same age in the absence of contextual cues?
- Do language disordered pre-school children benefit from contextual cues to the same extent as normally speaking children do?
- If not - what specific linguistic deficiencies are mainly preventing their use of contextual cues?

TEST ITEMS

Ten words, masked by white noise, occurred in the final position of 20 sentences designed so as to make the masked words non-predictable in 10 cases and highly predictable in the other 10 cases. The test items were selected from a material that has been developed by Axelsson /3/, based on Kalikow et al. /4/. All the words were familiar to Swedish pre-school children and differed in length and structure as is shown below:

1- svans (tail)
2- säng (bed)
3- ljus (candle)
4- katten (the cat)
5- fötterna (the feet)
6- sågen (the saw)
7- vas (vase)
8- bordet (the table)
9- saxen (the scissors)
10- tråd (thread)

SUBJECTS

The subjects were 39 language disordered and 39 normally speaking pre-school children and eleven normally speaking Swedish adults.
The two groups of pre-school children were matched on an individual basis as to sex, age, and non-verbal cognitive level. In each group there were 27 boys and 12 girls.
The mean age in the language disordered group was 6 years, 3 months, and in the normally speaking group 6 years, 4 months.
The mean of the non-verbal cognitive level, measured by Raven's coloured matrices, was 16.74 (S.D. 3.8) in the language disordered group, and 16.87 (S.D. 3.6) in the normally speaking group.

PROCEDURE

The 20 test sentences were randomized and tape recorded and presented individually to the children and as a group task to the adults. The task was to identify the masked word, but in most cases the subjects responded by repeating the whole sentence, which made it possible to register whether the context preceeding the masked word had been correctly perceived.

RESULTS

The scores on the two identification tasks , i.e. the identification of predictable or non-predictable masked words were calculated separately for the language disordered group (LD group), the normally speaking group (LN group) and the adult group (AN group). The scores differed both within the groups and between the groups, as can be seen in table 1. In all the groups the predictable words showed significantly higher scores (M = 8.8, 9.4 and 9.8 resp.) than the non-predictable words (M = 5.4, 5.8 and 6.9 resp.). This indicates that all subjects, regardless of group, take advantage of the contextual cues.

Table 1. Identification of non-predictable and predictable words in the language disordered (LD), linguistically normal (LN), and adult (AN) groups.

	Un-predict. word			Predict. word		
	LD	LN	AN	LD	LN	AN
Mean	5.4	5.8	6.9	8.8	9.4	9.8
S.D.	1.9	1.9	1.8	1.4	1.0	0.4
Var.	3.6	3.7	3.9	1.8	1.0	1.2
Min.	1	1	2	4	6	9
Max.	9	9	9	10	10	10

Comparing the results for the non-predictable words in the three groups it is evident that the LD group scores lower than the LN group and the AN group. Although the difference between the pre-school groups is not significant, it illustrates the negative effect of language disorders on interpreting speech in noise.
The LD group also scores lower than the LN group, which in turn scores lower than the AN group when the task is to identify predictable words. The difference is significant between the LD group and the LN group and between the LD group and the AN group but not between the LN group and the AN group. Thus, the normally speaking subjects, whether children or adults, benefit significantly more from contextual cues than the language disordered subjects do.
Figures 1-3 show the frequency distributions for each of the test items of the two identification tasks in the LD group (fig.1), the

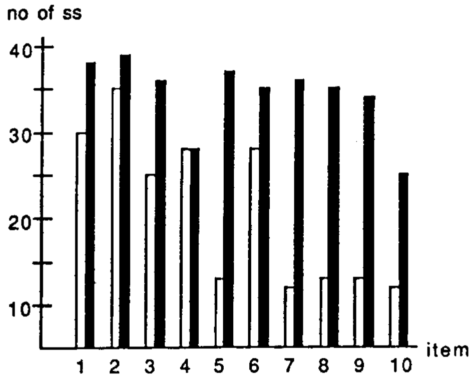


Fig.1. Number of subjects in the language disordered group (N=39) who identify non-predictable (white columns) and predictable (black columns) words correctly.

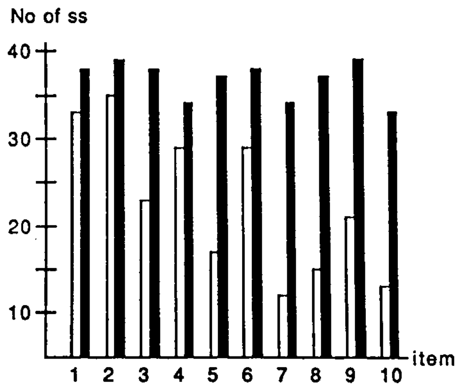


Fig. 2. Number of subjects in the normal language group (N=39) who identify non-predictable (white columns) and predictable (black columns) words correctly.

LN group (fig. 2), and the AN group (fig. 3). It is apparent from any of the figures 1, 2, or 3, that more subjects identify the test words correctly when contextual cues are provided.

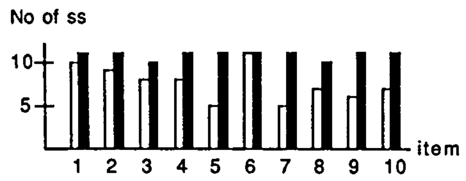


Fig. 3. Number of subjects in the adult group (N=11) who identify non-predictable (white columns) and predictable (black columns) words correctly.

Some of the words are more difficult to identify without contextual cues than others, as e.g. words no 5 (fötterna) and no 7 (vas), which are hard for all the groups, and words no 8 (bordet) and no 10 (tråd), which are penalising for the two pre-school groups. On the other hand, word no 6 (sågen) is correctly identified by all the adults in the non-predictable task.

Apart from the fact that some of the words are harder to identify than the others in a non-predictable context it is also to be seen in figures 1 and 2 that the influence of the contextual cues on the words to be identified varies in the pre-school groups. On one of the test items, no 4 (katten), the LD group does not benefit from the contextual cues at all, as their scores are the same as in the non-predictable task. Not even the AN group makes full use of the context, as is shown for items no 3 and no 8 in figure 3. However, the majority of the pre-school children, i.e. 25 of the subjects (64%) in the LN group and 16 (41%) in the LD group make full use of the contextual cues. This can be seen in table 2, in which it is also shown that the top scores in the predictable task cannot be inferred from the scores on the non-predictable task. Only very low scores (1 and 2) on the non-predictable task seem to imply poor or incomplete use of contextual cues on the predictable task. There is, nevertheless, a correlation between the performance on the predictable and non-predictable tasks in both the LD group (.60) and the AN group (.60), but not in the LN group (.37).

Table 2. Number of children identifying predictable and non-predictable words.

Number of non-predictable words	Number of predictable words 4	5	6	7	8	9	10	
1	1				1			2
2			1	1	1			3
3			1		2		3	6
4				3	1	2	5	11
5					3	3	7	13
6						5	9	14
7				1	2	8	7	18
8							7	7
9						1	3	4
	1		2	5	10	19	41	

Since the LD group performs worse than the matched LN group on both the predictable and non-predictable tasks it seems appropriate to look for the cause in the LD group's deficient linguistic competence. However, there are no correlations found between the performance of the two pre-school groups' performance on the predictable or the non-predictable tasks and their scores on linguistic tasks such as syntactic production, comprehension and awareness. As the scores of the LD group and the LN group overlap considerably, the scores of the worst performers (N=8) and the best performers (N=11) were calculated separately. There were no correlations between language tasks and the predictable or the non-predictable identification tasks among the worst performers, but there was a very high correlation between the predictable identification task and two of the syntactic measures (.97 and .91) among the best performers, as well as a moderate correlation with short term memory (.63).

DISCUSSION

The data in the present study allow us to make the following comments on the questions which were posed in the beginning of the paper:

Language disordered pre-school children are not able to understand speech masked by noise to the same extent as normally speaking children of the same age in the absence of contextual cues. This is in accordance with the results obtained by Brady et al. /5/ in a study where good and poor readers listened to speech in noise. The authors were able to conclude that the poor readers' low performance was due not to vocabulary but to a perceptual difficulty, and that "poor readers require more complete stimulus information than good readers in order to apprehend the phonetic shape of spoken words." (p. 21)

Language disordered pre-school children do not benefit from contextual cues to the same extent as normally speaking children do. This does not necessarily imply a general inability to use contextual cues. Rabbitt /6/, in a study of adults, found that recall of items presented without noise was inpeded if subsequent items were presented in noise. Thus, it might have been the case in our study, that the sentences that preceded the masked words were hard to perceive for some of the subjects and therefore did not facilitate the identification of the test words.

Syntactic ability and short term memory seem to be important factors for the use of contextual cues. The scores obtained by the subjects who performed best on identification of predictable words (i.e. subjects from both the LD and the LN groups) correlat with syntactic ability and short term memory. No such correlation is found for the subjects who performed low on the identification of predictable words. This implies that the low performing subjects are a heterogenous group as regards syntax and memory. Some of them do not have sufficient syntactic ability and memory to perform the tasks successfully, while others have these abilities but do not use them.

The variation of syntactic ability and memory in the low performing group shows that these factors are not the only determinants for the ability to use contextual cues in identification tasks such as reported in this study.

REFERENCES

/1/ Stanovich, Keith E. 1980. "Toward an interactive-compensatory model of individual differences in the development of reading fluency". Reading Research Quarterly XVI:1, 32-71

/2/ Magnusson, E. & Nauclér, K. 1987. "Language disordered and normally speaking children's development of spoken and written language. Preliminary results from a longitudinal study". Reports from Uppsala University Department of Linguistics, RUUL no 16

/3/ Axelsson, U. fortcoming. "Tal i brus". Department of Linguistics, Lund University

/4/ Kalikow, D.N., Stevens, K.N. & Elliot, L.L. 1977. "Development of a test of speech intelligibility in noise using sentence material with controlled word predictability". The Journal of Acoustical Society of America 61, 1337-1351

/5/ Brady, S., Shankweiler, D. & Mann, V. 1982. "Speech perception and memory coding in relation to reading ability". Haskins Laboratories: Status Report on Speech Research SR-69

/6/ Rabbitt, P.M.A. 1968. "Channel-capacity, intelligibility and immediate memory". Quarterly Journal of Experimental Psychology, 20, 241-248

This research was supported by a grant from HSFR, Swedish Council for Research in the Humanities and Social Sciences.

The use of the gating paradigm for studying the long - short vowel opposition

Laurens W.M. Damen Louis C.W. Pols

Institute of Phonetic Sciences
University of Amsterdam
The Netherlands

ABSTRACT

A gating task was used to study the long-short vowel oppositions in Dutch. It is shown that the duration of the vowel segment is an important cue to identify the long vowels. Also, an effect of the spectral differences between the vowels of an opposition pair was found.

INTRODUCTION

One of the striking characteristics of speech is its variability; the acoustic fine structure of a given speech segment varies considerably depending on many different factors. One such factor is the rate at which speech is produced. Yet listeners appear to have little difficulty in understanding speech correctly accross a wide variety of speech rates. Several theoretical explanations have been given of this fact. The theory of invariability states that the variability in the speech wave is only a surface phenomenon, and that invariants can be found. On the other hand, the theory of normalization postulates that listeners perceive speech sounds correctly by taking into account the acoustic-phonetic context.

Several studies have shown that the interpretation of speech segments depends on the acoustic-phonetic context. For example, Johnson and Strange (1982) asked subjects to identify vowels in tVt syllables. The syllables were spoken in a neutral carrier sentence at a normal and at a fast speech rate. The subjects heard the syllables in isolation, in a rate-appropriate carrier sentence, or in a rate-inappropriate carrier sentence. Vowels spoken at a normal speech rate were always identified accurately. This was not the case for the vowels in the fast-spoken syllables. The context in which the fast-spoken syllables were presented had a significant effect on the number of errors made. Analysis of the errors revealed that most of the errors were confusions of intrinsically long vowels with their spectrally similar short counterparts.

Van Bergem and Drullman (1985) tried to replicate this effect for the Dutch language. The replication failed; no confusions of long and short vowels were found. A possible explanation for this 'null' result is that the effect of context is totally obscured by the high performance of the listeners.

Both Johnson and Strange (1982) and van Bergem and Drullman (1985) used an identification task in their experiments. This task does not give detailed information about the process of vowel identification. Therefore, we have chosen a different method. The task used in the present experiment employs an extension of the gating paradigm (Grosjean, 1980), somewhat similar to that of Salasoo and Pisoni (1985).

The purpose of our study was to investigate the suitability of the gating task for the study of perceptual confusions between spectrally similar long and short vowels.

METHOD

Subjects

Subjects were 18 students, who were paid for their services. All subjects were native speakers of Dutch, with no reported hearing loss.

Materials and Design

The Dutch language has four spectrally similar long-short vowel pairs. Twelve monosyllabic pairs of nouns were selected for each vowel pair (e.g. /tak/-/tɑk/, /bot/-/bɔt/, /pen/-/pIn/ and /pøl/-/poel/). The words were uttered as the last word of a nonsense sentence. The sentences were comparable to those used by Nakatani and Dukes (1973). In addition, 60 bisyllabic and 20 monosyllabic nouns were selected to serve as filler words. None of the filler words contained any vowel used in the experimental words. The filler words were spoken as the last word of a neutral carrier sentence.

All the sentences were read by a male speaker and recorded on audio tape. The sentences were low-pass filtered at 4.5 kHz, sampled at 10 kHz and stored on computer disk. The experimental and filler words were excised from the stored sentences with the aid of a digital waveform editor and four points were marked in each experimental word. The first point marked the end of the initial consonant(cluster), the second point marked the end of the vocalic (CV) transition, the third point the begin of the final vocalic (VC) transition and finally the last point marked the begin of the final consonant(cluster). Relative to these four markers, additional segment markers for the gates were specified. The excised parts of the word were replaced by envelope-shaped speech noise. This procedure maintained amplitude and durational cues of the word, while removing all spectral cues of the replaced speech segment.

The gate durations of the experimental words are phonetically motivated: The first gate consisted only of the initial consonant(cluster). The initial consonant(cluster) and the CV transition were presented in the second gate. Two periodes of vowel signal were added in the third gate, etc. In the penultimate gate of a word, the complete vowel, including the VC transition, was presented. In this gate only the final consonant(cluster) was replaced by noise. The last gate of a word consisted of the original speech waveform.

It is difficult for the listeners to give a word as a response to gates in which only a very short segment of speech is presented. This was one of the reasons to add filler words. The speech segment in all filler words was at least 60 ms long. It is fairly easy to give a word reponse to such gates in comparison to the many very short speech segments of the experimental words. Thus, we tried to make the task easier for the subjects and we hoped that it would motivate the subjects to give word responses to all the gates presented. The filler words had another important function: it prevented the subjects from recognizing the main purpose of the experiment.

The presentation sequence used in the present experiment differs from that used in other studies that employ the gating paradigm. Words were entered into the presentation sequence in twelve 'steps'; four experimental and four filler words at a step. The order of these eight words was randomized. At the next step eight new words were introduced in the sequence. The set of words in the presentation sequence then totals sixteen. These sixteen words were again randomized. This process was repeated until all words had been introduced. This presentation format is halfway between a successive presentation sequence (e.g. Grosjean,1980) and a single presentation format (e.g. Cotton & Grosjean, 1984).

Six stimulus tapes were prepared from the digitally stored stimuli using a 12-bit D/A converter and a Revox A77 tape recorder. Each tape contained half of the experimental words: two series of twelve words containing only long vowels (e.g. /a/ and /o/) and two series of twelve words containing short vowels (e.g. /I/ and /oe/). The filler words were the same for all six stimulus tapes. Each tape contained about 815 stimuli.

Procedure

Subjects were tested in groups of three in a quiet room. Each group heard one stimulus tape only at a comforting listening level over Sennheiser HD424 earphones. Each experimental word was recorded on three different tapes and was heard by nine subjects. Thus, the design was a between-subjects design.

Subjects were told that they would hear one word at a time and that a large part of the word was made unintelligible by noise.

The subjects were instructed to write down the word they thought they had heard after each presentation of a stimulus. They were required to guess if they were not certain of a particular word. The tape ran without interruption; the interstimulus interval was 4.5 seconds long.

Every page of the response form contained 70 numbered lines, for each response a line. Two cue tones were used to signal the subjects to turn to the next page. A short pause was held after presentation of 280 and 560 stimuli. An entire listening session took about one hour and a half.

RESULTS

Two types of dependent measures were obtained. First, we computed the 'Vowel Isolation Time' (henceforth VIT) for all experimental words. The VIT is defined as the mean amount of vowel signal in ms needed by seven out of nine subjects to guess the identity of the vowel correctly without subsequently changing that guess. Second, we divided all reponses for each word and for every gate into three categories: (a) correctly identified vowels, (b) confusions with the spectrally similar long/short vowel, and (c) all other responses. The data are collapsed over words and over subjects and are given as Cumulative Response Curves.

Vowel Isolation Times (VITs)

The mean VITs for the eight different vowels are depicted in Figure 1.

The VITs were entered in an ANOVA with Long vs Short vowels (/a/,/o/,/e/ and /ø/ vs. /ɑ/,/ɔ/, /I/ and /oe/) and Vowel as fixed effects. Words were considered as the random effect.

A main effect of Long versus Short vowels ($F(1,88) = 169.78$; $p < 0.01$), as well as a main effect of Vowel ($F(3,88) = 7.59$; $p < 0.01$) was found. The interaction of both factors was not significant ($F < 1$).

Cumulative Response Curves

The Cumulative Response Curves of the vowels /o/ and /ɔ/ are given in Figures 2a and 2b. The results of the vowels /e/ and /I/ are plotted in the Figures 3a and 3b. The graphs of the other vowels have the same general appearance but are not presented here, because of lack of space.

Looking at Figures 2a and 3a, it is clear that the long vowels are often confused with their spectrally similar short vowel counterparts, especially at the first four gates. This effect is seen for all four long vowels.

The pattern of results of the short vowels (Figures 2b and 3b) is completely different from that of the long vowels. The short vowels are identified quite accurately, even if only a very short vowel segment is presented. Another interesting point can be seen clearly in these two graphs. Short vowels hardly ever give rise to confusions with their spectrally similar long

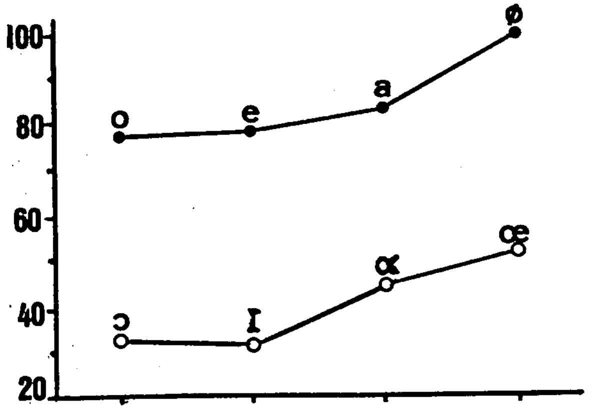


FIG 1. Mean Vowel Isolation Times of the 4 vowel oppostion pairs. Each mean is based on 12 words and 7 ss. ● long vowels; 0 short vowels

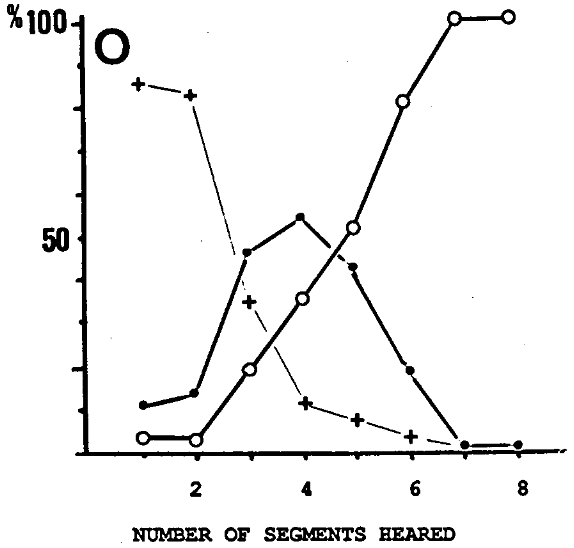


FIG. 2a. Presented vowel: /o/.
Response curves: O—O long vowel responses; ●—● short vowel responses; +—+ other responses.

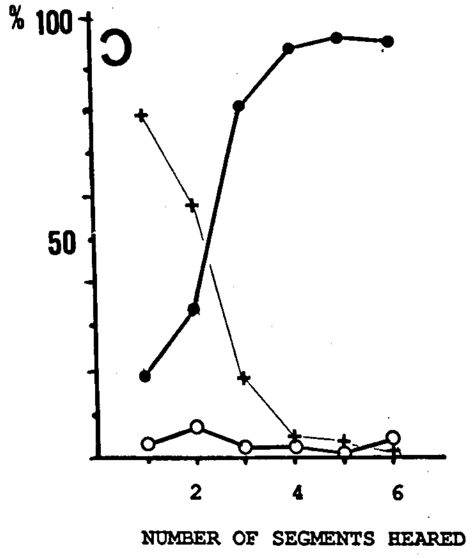


FIG. 2b. Presented vowel: /ɔ/.
Response curves: O—O long vowel responses; ●—● short vowel responses; +—+ other responses.

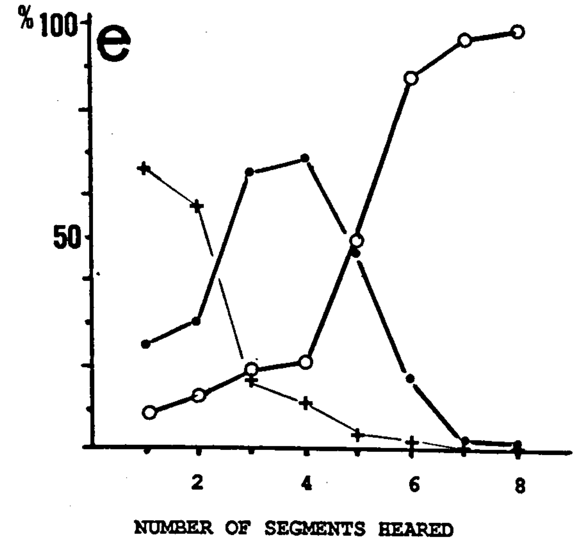


FIG. 3a. Presented vowel: /e/.
Response curves: O—O long vowel responses; ●—● short vowel responses; +—+ other responses.

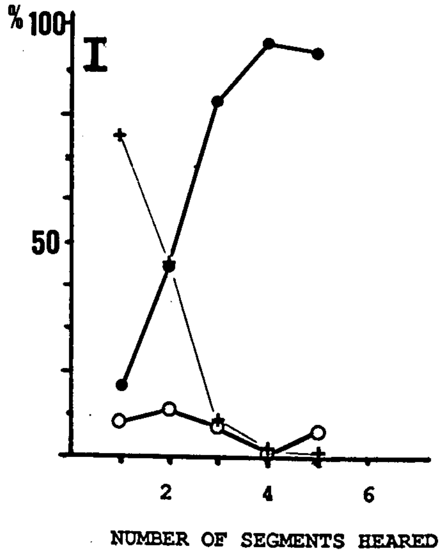


FIG. 3b. Presented vowel: /I/.
Response curves: O—O long vowel responses; ●—● short vowel responses; +—+ other responses.

counterparts. This effect is observed for all four vowels, though to a lesser extent for the short vowel /oe/. This finding points strongly to spectral differences between similar short and long vowels. In an attempt to evaluate these spectral differences, we compared the percentage of incorrect short responses to long vowels (e.g. /I/ responses with the vowel /e/ presented) to the percentage of correct short responses to the short vowels (e.g. /I/ responses to the vowel /I/ presented). The data of only the first four gates were analyzed. The percentages of responses were first square-root transformed, as recommended by Winer (1971, p. 399) and subsequently entered into an ANOVA. Long vs. Short vowel and Gate were fixed effects. Vowels were the random effect.

The results are very clear. A main effect of Long vs. Short vowel (F(1,24) = 27.40; p< 0.01). As expected, there was also a main effect of Gate (F(3,24) = 44.00; p < 0.01). The interaction between both factors was not significant (F(3,24) = 1.77). The outcome of this analysis indicates that the listeners used the fine spectral differences between the long and the short vowel of a pair, and furthermore that they did so independently of segment duration.

Another interesting result that becomes apparent from Figures 2b and 3b is the relatively high percentage of correct vowel responses to the first gate. Only the initial consonant (cluster) was presented in this gate (see Method section). Nevertheless, the identification of the intended vowel is above chance level. The validity of this finding is dependent on the defined boundary between initial consonant(cluster) and vowel. Still, the data indicate that the consonant contains coarticulatory information about the immediately following vowel.

DISCUSSION

This experiment was set up to ask a specific empirical question: can confusions between spectrally similar vowels, differing mainly in duration, fruitfully be studied using a gating task. The results clearly show that this can be done. Both dependent measures (Vowel Isolation Times and percentage correct responses) lead towards the same conclusion, i.e. duration of the presented vowel segment is an important cue to distinguish long vowels. It has also become clear from the analysis on percentage of short vowel responses on the long versus the short vowels, that fine spectral differences between the members of a vowel pair were used by the listeners to identify the vowels. Some caution must be taken with this interpretation, because an alternative explanation can not be ruled out completely. It is likely that the gating task biased the listeners to give more short than long vowel responses. This effect of bias can be expected to be strongest for the shortest speech segments and only the data for these shortest segments were analyzed. Thus, parts of the observed differences in response percentages might be due to bias.

One aspect of the results still needs to be discussed. In the analysis of the Vowel Isolation Times a main effect of Vowel was found. This effect was mainly due to the vowels /ø/ and /oe/. This result can readily be interpreted. These two vowels have a low frequency of occurrence in Dutch and it is likely that this will lengthen the Vowel Isolation Time.

Recently, a follow-up experiment has been carried out to evaluate the presence or absence of noise in the gates. This time the gates were not filled with noise but left silent, all other things being equal. The global pattern of results of this experiment was similar to the results of the experiment reported above. This indicates that the observed effects are robust effects. Furthermore, a small effect of the presence or absence of noise was found. We take this as evidence that the gating task is a very sensitive measuring method.

Presently, an experiment is being carried out, in which acoustic-phonetic context is manipulated in a manner analoguous to the experiment of Johnson and Strange (1982), mentioned in the Introduction of this paper. Fast-spoken words are presented to subjects in a rate-appropriate or in a rate-inappropriate carrier sentence, as well as in isolation. Assuming that the gating task is a sentitive method, it is hypothized that these context manipulations will have a distinct effect on both the Vowel Isolation Times and the response curves.

The advantage of using the gating task in such an experiment is twofold. First, the task provides a detailed picture of the vowel indentification process, as has been shown by the present experiment. It is reasonable to expect that even very small context-induced differences can be found. Furthermore, using the gating task the 'dependent measure' is the moment at which the vowel is identified correctly, and not whether it is identified correctly or not. This allows for the use of high quality speech material, and this reduces the chance that the results of the experiment depend on the specific speaker and speaking situation.

REFERENCES

Bergem,D.R.van,& Drullman,R.(1985) The influence of contextual speech rate on the identification of Dutch vowels in normally and rapidly spoken tVt utterances. IFA-report nr.77 (in Dutch).

Cotton,S.,& Grosjean,F.(1984) The gating paradigm: A comparison of succcessive and individual presentation formats.
Perception and Psychophisics,35,41-48.

Grosjean,F.(1980) Spoken word recognition and the gating paradigm.
Perception and Psychophysics,28,267-283.

Johnson,T.L.,& Strange,W.(1982) Perceptual constancy of vowels in rapid speech. Journal of the Acoustical Society of America,72,1761-1770.

Nakatani,L.H.,& Dukes,K.D.(1973) A sensitive test of speech communication quality. Journal of the Acoustical Society of America,53,1083-1091.

Salasoo,A.,& Pisoni,D.B.(1985) Interaction of knowledge sources in spoken word identification Journal of Memory and Language,24,210-231.

Winer,B.J.(1971) Statistical principles in experimental design. New York:McGraw-Hill.

LISTENERS' IDENTIFICATION OF SPEECH SOUNDS IS INFLUENCED BY ADJACENT "RESTORED" PHONEMES

JOHN J. OHALA and DEBORAH FEDER

Department of Linguistics
University of California
Berkeley, California 94720 (USA)

ABSTRACT

When listeners' identifications of speech sounds are influenced by adjacent sounds is it only the quantitative phonetic characteristics of these neighboring sounds that matter or could their qualitative linguistic identity play a role? We tested this by leading subjects to restore or induce the noise-obliterated medial consonant in V_1CV_2 utterances by first presenting them with several prior utterances where this medial consonant could be heard clearly and was consistently the same, either a /b/ or a /d/. Included as V_1 were synthetic vowels from the /i - u/ continuum. More /u/'s were identified out of this continuum in the environment of physically present /d/'s than /b/'s. Restored /d/'s had the same effect (vis-a-vis restored /b/'s), thus indicating that the influence of context need not operate only via physical phonetic features. These results challenge the 'direct realist' theories of speech perception as well as claims that 'invariant' features of speech sounds are to be found by normalizing these features with respect to the physical phonetic characteristics of their surroundings.

INTRODUCTION

There is abundant evidence that listeners identify speech sounds in part by normalizing them with respect to their phonetic context [1, 2, 3, 4, 5]. How is this done? Are the physical phonetic parameters of the context used to adjust recognition threshholds or is it enough for the listener just to know the (categorized) linguistic identity of the context? We investigated this questions through a series of perceptual tests involving listeners' identification of synthetic vowel stimuli in isolation and in consonantal contexts. (In what follows, we collapse descriptions of two of these tests--a pilot study and a main test, which differ in some details. The description is kept general and details and differences given only where essential.)

IDENTIFICATION OF VOWELS IN ISOLATION

First, we constructed a 17-step linear stimulus continuum between the vowels /i/.and /u/; see Fig. 1. The continuum endpoints were modeled on the first 100 msec of natural /i/ and /u/ pronounced in isolation by an adult male native speaker of American English. (Since the 'crossover' from /i/ to /u/ was expected to happen in the middle of this continuum, some stimuli near the end points were omitted from the study, those steps showing absence of 'x''s on the formant parameters.) In a forced-choice identification task, listeners gave the response function shown in Fig. 2, where the ordinate shows percent identification of tokens as /u/ and the abscissa, the /i/ -- /u/ continuum (/i/ at the left and /u/ on the right). These results were obtained from 28 native American English-speaking listeners, each responding twice to each stimulus for a total of 56 responses per data point.

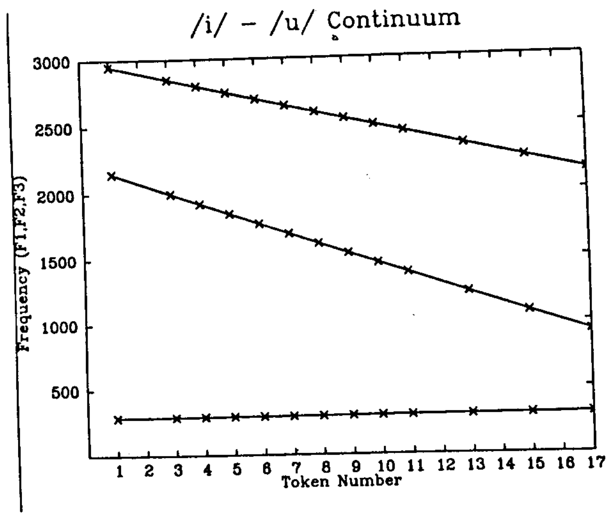


FIGURE 1

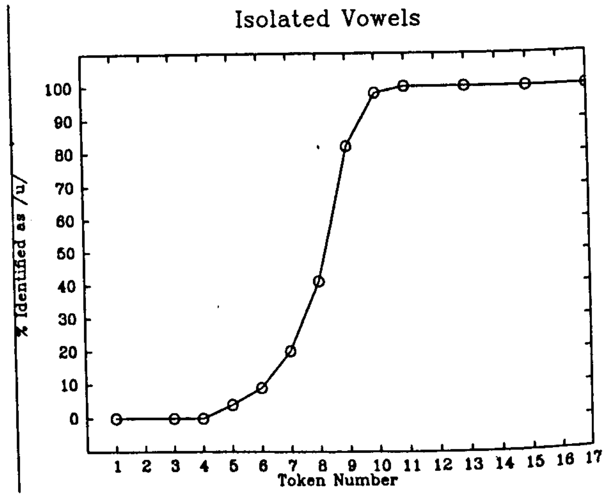


FIGURE 2

IDENTIFICATION OF VOWELS IN CONSONANTAL CONTEXT

Second, we sought to replicate the finding that this function shifts when the vowels are put in certain consonantal contexts [6, 7]. Using digital splicing, we embedded our vowels in nonsense words of the form /ibə/, /ubə/, /idə/, /udə/--where the /bə/ and /də/ were excised from the same speaker's natural utterances of /abə/ and /adə/. In another forced-choice task, listeners exhibited a shift in the earlier response so that more /u/'s were heard in the context of a following /d/--presumably because listeners allowed for and factored out the elevated F2 that alveolar consonants produce on back vowels [8, 9]. Fig. 3 shows the results from the pilot test which had 14 listeners and a total of 28 responses per stimulus and Fig. 4, the results from the main test with 28 listeners and 8 judgements per stimulus per listener for a total of 224 responses per data point.

IDENTIFICATION OF VOWELS IN CONTEXT OF RESTORED CONSONANTS

Third, we asked whether this same shift in the function due to consonantal context would appear even if the consonants were not physically present in the signal but if the listeners instead just imagined that they were. We attempted to make listeners believe that the /b/ in the /ibə/, /ubə/ tokens or the /d/ in the /idə/, /udə/ tokens were present when they weren't, by using the technique called "phoneme restoration" [10] where high redundancy of the message induces the listener to "fill in" missing elements. The redundancy in our case was provided by presenting all our stimuli in two major blocks, one in which the medial consonant was or seemed to be a /b/, and the other in which /d/ was or appeared to be the consistent medial consonant. To enhance this priming, we also began each block with a number of tokens in which the consonant was clearly present. In approximately 15 to 20% of the stimuli in each block we completely replaced the medial consonant by white noise (always equal in intensity to the average intensity of the voicing during the consonantal closure).

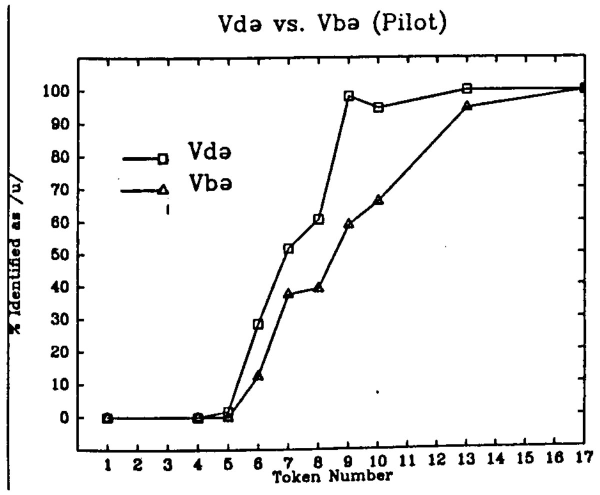


FIGURE 3

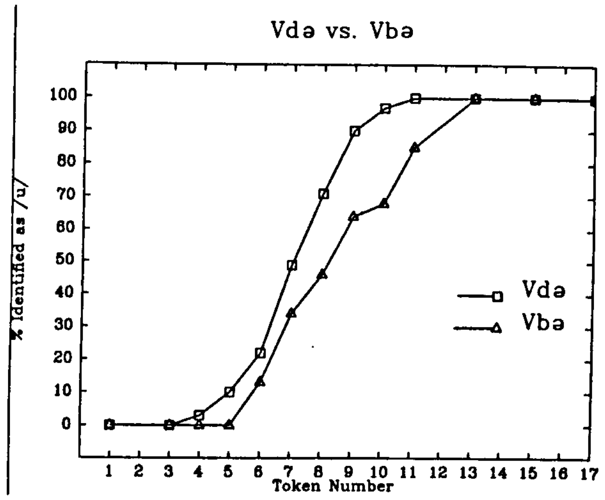


FIGURE 4

The schwa portion of these latter stimuli consisted of several periods from the center of the naturally spoken schwa, i.e., a portion with minimal, if any consonantal 'coloring'. To insure that restoration of the consonants did not derive from any residual cues remaining in the schwa, we used the schwa that had originally followed a /b/ in the stimuli where we wanted a /d/ to be restored by our listeners, and, similarly, a schwa that had originally followed a /d/ where we wanted them to restore a /b/. As foils to prepare listeners for hearing noise, another 15% of the tokens also contained noise bursts at various locations, such as during the intervals between stimuli, over the vowel, or superimposed on (but not replacing) the consonantal closure. Listeners were told that the noise bursts served as distractor to the identification task.

The results from the pilot test are shown in Fig. 5 (where only a fraction of the entire vowel continuum was studied). There was an important difference between the pilot and the main test. In the former, subjects wrote down the entire VCV utterance they heard. The results shown in Fig. 5

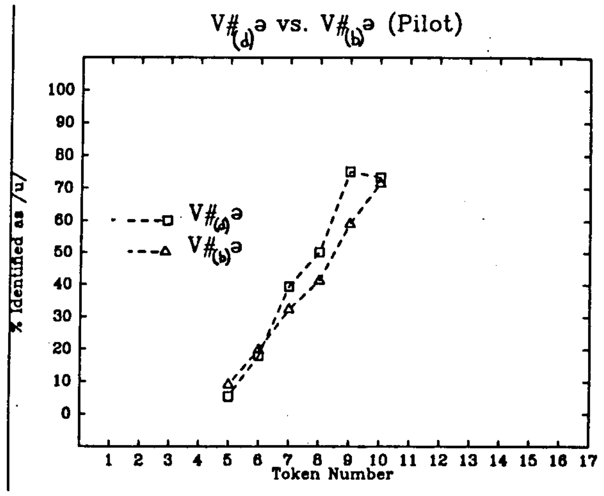


FIGURE 5

are those for which the consonant reported was the one we were trying to get subjects to restore. The shift for these tokens is in the same direction as that for the physically present consonants. Each data point represents the average of 28 responses. For the main test we gave subjects answer sheets with the "b"'s and "d"'s already present; they were only required to fill in the vowel. These results are shown in Fig. 6. Here there were 56 judgements per data point: 28 subjects times 2 judgements per stimulus.

The difference between the two consonantal contexts has been shown to be significant by several preliminary curve-fitting analyses. However, our statistical findings cannot yet be considered conclusive since the irregular shapes of some of our curves have made it difficult to fit enough of them to any single statistical model to make comparisons among them possible and meaningful.

DISCUSSION AND CONCLUSION

We conclude that since the magnitude of these shifts under the restored or imagined phonemes is not as great as the case with physically present phonemes, listeners adjust their identification threshholds in part due to processing of the actual acoustic parameters of the signal and in part--perhaps as a default case--on the basis of the linguistic identity of contextual segments which may in some cases be provided through non-phonetic channels. In the latter case, one imagines that the listener knows from experience the typical effects of one segment on another and uses this information to adjust recognition threshholds.

This result should not be surprising: it is well recognized, for example, that in the visual domain we achieve a high degree of color constancy in part by factoring out the distorting influence of the hue of ambient illumination but also by our knowledge of what the colors of typical objects are and how these colors are modified in various situations. It would be remarkable if something similar did not apply in the case of speech perception.

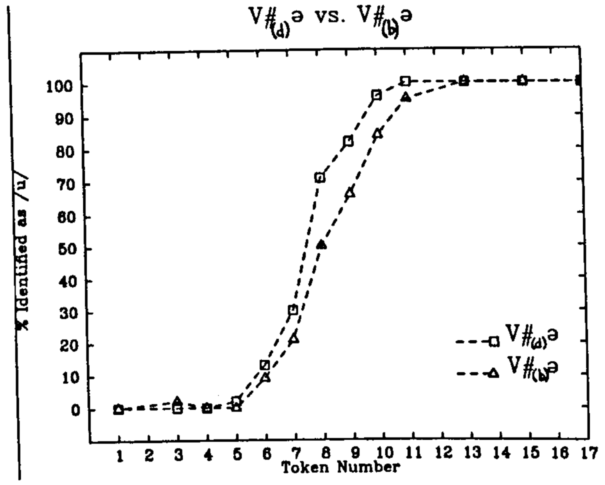


FIGURE 6

We believe that these results present a challenge to the direct realist view of speech perception which holds that all the information needed to identify the intended message elements of speech are present in the acoustic signal and can be discovered by the listeners without the need for inferences and the like; for example, Fowler [11] claims that

> "...[speech] perception must be direct and, in particular, unmediated by cognitive processes of inference or hypothesis testing, which introduce the possibility of error."

Here, listeners showed that their speech sound identification was influenced by entities not physically present in the signal. Specifically, the identity of an ambiguous stimulus was resolved by reference to predicted effects of an assumed, i.e., hypothesized, environment. In this latter respect, our results are compatible with those of Mann and Repp [4], who showed that listeners' identification of one variable stimulus shows a discontinous shift as a function of their identification of an another adjacent ambiguous segment.

These results do not actually refute the direct realist view, though, since in one recent formulation of it [11, 12], it has been allowed that listeners can sometimes operate on what might be called "automatic pilot"--that is, by making assumptions, even unwarranted assumptions, about what is present in the signal. Nevertheless, direct realists would maintain that, in principle, if listeners just paid closer attention to the speech signal, speech perception would be accomplished "directly" and they wouldn't make the kinds of perceptual "mistakes" as they did in our study. Our results challenge this view, too, though, by raising the following question: if listeners are capable of integrating non-phonetic information into their recognition task, isn't it likely that the speaker knows this and only puts enough energy, precision, and detail into the generation of the speech signal as the listener requires? It is our impression that the speaker often does not in fact put sufficient phonetic details into the speech signal to permit decoding of the message in a direct way.

These results also bear on the question of whether there are or should be acoustic invariants of phonemes (or other message units) in speech [13]. Clearly, these results add to the evidence that absolute invariance is not necessary; the listener has ways of accomodating variation. Stevens [14] has suggested that relative invariance may be more likely than absolute invariance, i.e., a given unit or a distinctive feature characterizing it may be invariant with reference to the phonetic environment it appears in--in his view they physical phonetic environment. The notion of "relative invariance" is compatible with our results but only if the linguistic identity of the context, not exclusively its physical properties, are admitted as figuring in the normalizing process.

Finally, we think we have demonstrated a potentially quite useful way of inducing listeners to restore missing elements in speech which does not require construction of semantic, syntactic, or other higher-order redundancies.

ACKNOWLEDGEMENTS

We benefited from the comments of Klaus Kohler, Bruno Repp, Mary Smith, and Richard Warren on earlier versions of this paper. Responsibility for the contents of the paper, including any errors, is ours, though. This research was supported by a grant from the Sloan Foundation to the Cognitive Science Program at the University of California, Berkeley.

REFERENCES

[1] Ladefoged, P. & Broadbent, D. E. 1957. Information conveyed by vowels. J. Acous. Soc. Am. 29.98-104.

[2] Pickett, J. M. & Decker, L. 1963. Time factors in perception of a double consonant. Lang. & Speech 3.11-17.

[3] Mann, V. A. & Repp, B. H. 1980. Influence of vocalic context on perception of the [š] vs [s] distinction. Perception & Psychophysics 28.213-228.

[4] Mann, V. A. & Repp, B. H. 1981. Influence of preceding fricative on stop consonant perception. J. Acous. Soc. Am. 69.548-558.

[5] Fowler, C. A. 1981. Production and perception of coarticulation among stressed and unstressed vowels. J. Speech & Hearing Res. 46.127-149.

[6] Ohala, J.J, Riordan, C. J., & Kawasaki, H. 1978. The influence of consonant environment upon identification of transitionless vowels. J. Acous. Soc. Am. 64.S18.

[7] Ohala, J. J. 1981. The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), Papers from the Parasession on Language and Behavior. Chicago: Chicago Linguistic Society. 178-203.

[8] Lindblom, B. 1963. Spectrographic study of vowel reduction. J. Acous. Soc. Am. 35.1773-1781.

[9] Stevens, K. N. & House, A. S. 1963. Perturbations of vowel articulations by consonantal context: An acoustical study. J. Speech & Hearing Res. 6.111-128.

[10] Warren, R. M. 1970. Perceptual restoration of missing speech sounds. Science 167.392-393.

[11] Fowler, C. A. 1986a. An event approach to the study of speech perception from a direct realist perspective. J. Phonetics 14.3-28.

[12] Fowler, C. A. 1986b. Reply to commentators. J. Phonetics 14.149-170.

[13] Stevens, K. N. & Blumstein, S. E. 1981. The search for invariant acoustic correlates of phonetic features. In P. Eimas & J. Miller (eds.), Perspectives on the study of speech. Hillsdale, NJ: Erlbaum.

[14] Stevens, K. N. 1986. Models of phonetic recognition II: An approach to feature-based recognition. Proc., Montreal Symposium on Speech Recognition, July 21-22, 1986, McGill University, Montreal. Canadian Acoustical Association. 67-68.

CONSONANT COMBINATORICS IN GERMAN

MART RANNUT

Dept. of Computational Linguistics
Institute of Language and Literature
Tallinn, Estonia, USSR 200106

ABSTRACT

In the paper we present a list of possible (existing as well as typologically relevant non-existing) consonant sequences in German and their unified phonological distributional description. The works of L.Hirsch-Wierzbicka /7/ and V.Tarамets's group /26/ were used as source material for sketching out phonotactic rules, methodologically the work is based on Rannut /17/. The initial version was prepared as a computer program which was reprocessed and improved to obtain a better final result. The aim of the work was to find a set S' maximally similar to the set S of German consonant sequences, fulfilling the condition $|S-S'|<\epsilon$, where ϵ is minimal. The final version of phonotactic restrictions is presented in the form of a context-sensitive grammar.

1. INTRODUCTION

Consonant sequences in German and regularities of their formation have previously been studied and described by Twaddell /24/, Menzerath /11,12/, Moulton /13, 14/, Trubetzkoy /23/, Seiler /19,20/, Tanaka /22/, Hirsch-Wierzbicka /17/ and Taranets's group /26/. Bierwisch /1/, Wurzel /25/ and Copeland /13/ have described them by means of generative grammars. In addition, we have used the data provided by Kelz /9/, Stock /21/ and Kästner /10/.

The main research object of Twaddell was the structure of German words consisting of one syllable. He provided the rules of word-initial and word-final consonant clusters. When studying the structure of consonant sequences in the middle of the word (he found 394 of them) he came to the conclusion that their structure follows certain regularities. Menzerath followed the work of Twaddell in studying monosyllabic words, Moulton and Trubetzkoy studied certain aspects of them (Moulton studied the correlation between the structure of a consonant cluster and the length of the preceding vowel, Trubetzkoy formulated the distributional rules of the word-initial consonant cluster), Seiler constructed its structural formula and the work was completed by Hirsch-Wierzbicka, who applied a computer and added a functional load to every case. After Twaddell the work in studying word-internal consonant sequences was continued by Tanaka. He examined the structure of intermediate clusters, which he defined as clusters between two vowels, not taking into account word boundary signals. From this material he drew 5 statistically significant conclusions.

Taranets has studied the same problem with an aim of finding articulatory correlates to the syllable boundary. Bierwisch and Wurzel have studied certain morphonological aspects of the problem while Copeland has given separate distributional rules for coda and onset.

In our paper we deal with the word-internal consonant sequence as a unit with its own structure. Word-initial and word-final consonant clusters will not be thoroughly discussed as there are above-mentioned fundamental works available on this topic. Neither are those clusters difficult to generate when applying a few additional rules to the word-internal consonant sequences (see p.9).

2. PRINCIPLES OF DESCRIPTION

The principles of the description of consonant sequences are as follows:
1) The composition and occurrence of German consonant sequences are determined by regularities, part of which are general linguistic while the rest are characteristic of the phonological system of the German language, the latter being primary with regard to the general linguistic ones.
2) Syllable structure is regarded as primary with respect to phonemes and capable of dictating phoneme sequences in it (see /5/). The consonant sequences containing a syllable boundary are restricted by the syllable structure as well as by consonant sequence constraints. Such a model can be considered as a submodel of the model of the rhythmic pattern of speech (see /18/). As the structure of the syllable is influenced by the morphological component, the deviations resulting from this are regarded as a unnatural heterogenous set Q_n(see p.7).

3) In working out an integral model of phonotactic restrictions we have proceeded from word-internal consonant sequences as they are more regular than word-initial and word-final consonant clusters. By extending the restrictions of word-internal consonant sequences to word-initial and word-final consonant clusters and applying a few additional rules we have obtained a unified system of phonotactic restrictions of German consonant sequences (cf./15,16/).
4) Unlike the above studies which have dealt with the analysis of consonant sequences with a view to discovering the rules lying as the basis for the structure, this investigation aims at a synthesis of consonants by means of a generative model, the purpose of which is to obtain an adequate final result with the help of a minimal number of rules.
5) Considered in the study are only consonant sequences occurring in native German words while the difference between a German word and a foreign one has been made on the basis structure (see /6/).
6) The structure model of consonant sequences contains only predominating rules. As the phonological component is constantly affected by the environment through the adoption of foreign words and the phonetic peculiarities of other languages, added to its grammar are more and more rules which can be opposed to the existing ones. At the same time the language dislodges part of the rules to retain the level of homogeneity necessary for its existence. For that reason from the synchronic point of view a language is a complex of rules opposed to each other and some of them dominate over the others. In the case of opposing rules, considered in constructing the phonological generator are the rules which dominate over the others. The choice of rules in this case has been made with the help of functional load.
7) The present phonotactic description is based on Duden's /4/ transcription, containing the following consonant phonemes:

/p	t		k
b	d		g
pf	ts	tʃ	
		dʒ	
f	s		ç x h
v	z	ʒ	
m	n		ŋ
	l		
	r		
	j /		

As the consonants dʒ, tʃ and ʒ do not appear in native German consonant clusters, they are not taken into account in the inventory of our generative grammar.
8) The present work is not concerned with consonant sequences generated at compound component junctions, regarding only sequences in the words consisting of stem and affix(es), The latter are taken from Duden/4/

3. DEFINITIONS

A consonant sequence is an arbitrary sequence of different consonants within the boundaries of one simple word.

A consonant cluster is a syllable-internal consonant sequence. A syllable is a sequence of phonemes between two successive syllable boundaries within one word. Syllable boundary is fixed according to traditional German grammar.

A fortis structure is a consonant sequence containing stop consonants and/or affricates.

A lenis structure is a consonant sequence containing neither stop consonants nor affricates.

A base structure is a sequence of non-terminal symbols.

A surface structure is a sequence of German consonant phonemes.

A pronunciation strength structure is a sequence of non-terminal symbols marking pronunciation strength classes (see /8/).

4. GENERAL PRINCIPLE OF THE GENERATOR

A generator of consonant sequences represents a formal grammar which determines one phonological subsystem of a language - the phonotactics of consonant sequences - and gives its exact description. According to Chomsky's /2/ grammar the consonant sequences generator can be described by the formula D=(I,E,T,P), where D is the consonant sequences generator, E - the ultimate number of non-terminal symbols (the auxiliary symbols of the base structure), T - the number of terminal symbols (the consonant phonemes of the surface structure) different from E, P - the ultimate number of restriction (X) and derivation rules (Y), which describe the process of generating consonant sequences and restrict sequences not characteristic of the language, and I - the initial symbol.

E comprises the auxiliary symbols of the first- and second-level base structures (M1,K2,M3,K4,M5, 1,2,3,4,5, G,L,N,V,F, K,A,B,R,Q_1,Q_2). T comprises the consonant phonemes occurring in sequences in the surface structure(the 3rd level) of the German language (k,p,t,g,b,d,pf,ts,s,f,ʃ,ç,x,rm, h,v,z,m,n,ŋ,l,r,j,rn) and a blank or empty string (ϵ). Different variants are separated in the description by the symbol (,). The length of a consonant sequence, i.e. the number of phonemes in the sequence between two vertical strokes is marked by a number following the mark of equation. The symbol C stands for any consonant phoneme, * is a string of consonants comprising from 0 to 5 phonemes and $*_1$ is a string of consonants comprising from 1 to 5 consonant phonemes.

The generator has been compiled on the basis of the combinatorical regularities of German consonant sequences and it constructs the existing consonant sequences whose structure is acceptable to the phonological system of the German language as well as the non-existing consonant sequences which are typologically relevant to the existing ones. It also determines the position of syllable boundary. The work of the generator is based on the application of the hierarchical character of the phonological system of the language where restrictions are applied on all levels. This makes it possible with few restrictions to obtain from 5.5 million potential sequences ($\prod_{k=1}^{5} C_k = 23*23*23*22*21 = 5621154$) a result that in number approximately corresponds to the one in reality. The generator does not pretend to psychological reality or "natural" processes in the language, but represents a black-box-type model in which the application of phonological rules provides a result close to linguistic realities.

5. FORMAL DESCRIPTION OF THE GENERATOR

The formal description of the generator of German consonant sequences is given in the form of a grammar where Y marks the derivation rules and X the restriction rules, and the number following them denotes the hierarchical level. The number in brackets refers to the subsection where the respective operation is presented in more detail. The rules with the number 0 (e.g. X-1.0) point out those consonant strings which do not correspond to the definition of consonant sequences.

Y-1 $I \to 1\ 2\ 3\ 4\ 5$
(6) $1 \to M_1, \epsilon$
$2 \to K_2, \epsilon$
$3 \to M_3, \epsilon$
$4 \to K_4, \epsilon$
$5 \to M_5, \epsilon$

X-1.0 $C \to \epsilon$

Y-2 $M_1 \to L,N,Q_1$
(7) $K_2 \to K,A$
$M_3 \to F,R$
$K_4 \to K,A,B$
$M_5 \to G,L,N,V,Q_2$

X-2.1.1 $Q_1*\ Q_2 \to \epsilon$
2 $*A*\ A* \to \epsilon$
3 $*BQ_2 \to \epsilon$
$*BV \to \epsilon$

X-2.1.4 $*FB* \to \epsilon$
$*AB* \to \epsilon$
$*KB* \to \epsilon$

X-2.2.1 $Q_1*R* \to \epsilon$
$N*R* \to \epsilon$
$*KR* \to \epsilon$
$*AR* \to \epsilon$
2 $*RK* \to \epsilon$
$*RB* \to \epsilon$
$*RA* \to \epsilon$

X-2.3.1 $G* \to \epsilon$
2 $*CCG \to \epsilon$
3 $BG \to \epsilon$
$AG \to \epsilon$
$RG \to \epsilon$
$FG \to \epsilon$

X-2.4 $*FA* \to \epsilon$

X-2.5 $Q_1AK* \to \epsilon$
$Q_1KA* \to \epsilon$
$Q_1KK* \to \epsilon$

X-2.6.1 $C_1C_2C_3C_4* \to \epsilon$, if $C_n = B,A$
2 $C_1C_2C_3C_4C_5 \to \epsilon$ if $C_5 = N,Q_2,V$-
if $C_1 = Q_1$

Y-3.1 $G \to j$
(8) $L \to l,r$
$N \to n,m,\text{ŋ},rn,rm$
$V \to v,z,h$
$R \to l,m,n$
$F \to s,f,\text{ʃ},x$
$B \to b,g,d$
$A \to pf,ts$
$K \to k,p,t,\text{ç}$

X-3.0 $*C_iC_j* \to \epsilon$, if $i=j$

X-3.1 $*C\text{ŋ} \to \epsilon$
$*Cx* \to \epsilon$

X-3.2.1 $\text{ŋ}C \to \epsilon$, if $C \neq v,z,h,l,n,k$
2 $\text{ŋ}C_1C_2* \to \epsilon$, if $C_1 \neq k,s$

X-3.3 $xC* \to \epsilon$, if $C \neq t,ts$

X-3.4 $C\text{ç}* \to \epsilon$, if $C \neq l,r,n$

X-3.5 $*kf* \to \epsilon$
$*pf* \to \epsilon$
$*\text{ç}f* \to \epsilon$
$*tf* \to \epsilon$
$*tsf* \to \epsilon$
$*pff* \to \epsilon$
$*sf* \to \epsilon$
$*mf* \to \epsilon$

X-3.6 $Q_1fCC \to \epsilon$

X-3.7 $*gm \to \epsilon$
$*bm \to \epsilon$
$*dm \to \epsilon$
$*km \to \epsilon$
$*pm \to \epsilon$
$*fm \to \epsilon$
$*fm \to \epsilon$
$*tsm \to \epsilon$
$*pfm \to \epsilon$

X-3.8 $m\text{ʃ}C* \to \epsilon$
$*Csm \to \epsilon$
$*sp* \to \epsilon$, if $* \neq \epsilon$

X-3.9 $*CCpn \to \epsilon$
$*CCkn \to \epsilon$

X-3.10 $*tsp* \to \epsilon$
$*pfp* \to \epsilon$
$*kp* \to \epsilon$
$*tp* \to \epsilon$
$*\text{ç}p* \to \epsilon$
$*fp* \to \epsilon$

X-3.11 $mpC_1C_2 \to \epsilon$, if $C_2 \neq t$
$lpC_1C_2 \to \epsilon$, if $C_2 \neq t$
$rpC_1C_2 \to \epsilon$, if $C_2 \neq t$

X-3.12 $mtCC \to \epsilon$
$Q_1tCC \to \epsilon$

X-3.13 $Cpt* \to \epsilon$
$C\text{ç}s* \to \epsilon$

X-3.14 $*tk* \to \epsilon$
$*pk* \to \epsilon$
$*pfk* \to \epsilon$
$*tsk* \to \epsilon$
$*\text{ʃ}k* \to \epsilon$
$*fk* \to \epsilon$

X-3.15 $*sk* \to \epsilon$, if $* \neq e$
$*\text{ç}k* \to \epsilon$, if $* \neq e$

X-3.16 $Cpf* \to \epsilon$, if $C \neq m,Q_1,r,n$

X-3.17 $rpfC \to \epsilon$, if $C \neq l,r$
$npfC \to \epsilon$, if $C \neq l,r$
$Q_1pfC \to \epsilon$, if $C \neq l,r$

X-3.18 $CCs* \to \epsilon$
$CCts* \to \epsilon$

X-3.19 $C_1tsCC \to \epsilon$, if $C_1 \neq n,r,l$

X-3.20 $*tts* \to \epsilon$
$*kts* \to \epsilon$
$mts* \to \epsilon$

X-3.21 $*\text{ç}r \to \epsilon$
$*sr \to \epsilon$
$*tsr \to \epsilon$

X-3.22 $C_1C_2C_3r \to \epsilon$ if $C_2 \neq \text{ʃ}$

X-3.23 $*k\text{ʃ}* \to \epsilon$
$*p\text{ʃ}* \to \epsilon$
$*t\text{ʃ}* \to \epsilon$
$*\text{çʃ}* \to \epsilon$
$*ts\text{ʃ}* \to \epsilon$
$*pf\text{ʃ}* \to \epsilon$

X-3.24 $Q_1s* \to \epsilon$
$ms* \to \epsilon$
$*tss* \to \epsilon$

X-3.25 $rsC* \to \epsilon$, if $C \neq t$

X-3.26 $mg* \to \epsilon$

X-3.27 $*pfz \to \epsilon$
$*tsz \to \epsilon$
$*\text{ʃ}z \to \epsilon$
$*fz \to \epsilon$
$*sz \to \epsilon$
$*pz \to \epsilon$
$*tz \to \epsilon$
$*\text{ç}z \to \epsilon$

X-3.28 $*pfv \to \epsilon$
$*fv \to \epsilon$
$*pv \to \epsilon$
$*tv \to \epsilon$
$*\text{ç}v \to \epsilon$

X-3.29 $*\text{ʃ}Cl \to \epsilon$
$*\text{ʃ}Cn \to \epsilon$
$*\text{ʃ}CQ \to \epsilon$

X-3.30 $*pfsC \to \epsilon$, if $C \neq t$
$*tsC \to \epsilon$, if $C \neq t$
$rnC \to \epsilon$, if $C \neq t$

X-3.31 $ktC \to \epsilon$
$ptsC \to \epsilon$
$pftC \to \epsilon$

X-3.32 $npC* \to \epsilon$, if $C=s,n,Q_2$

X-3.33 $Q_1Cn \to \epsilon$, if $C = t,p,ts,f,b,d$

X-3.34 $Q_1Cl \to \epsilon$, if $C=t,d$

X-3.35 $npn \to \epsilon$

X-3.36 $rmCCC \to \epsilon$
$rmC \to \epsilon$

X-3.37 $rmC_1C_2 \to \epsilon$, if $C_1 \neq s$
$rnC_1C_2 \to \epsilon$, if $C_1 \neq s$

X-3.38 $C_1C_2C_3C_4 \to \epsilon$, if $C_1 = k,t$

X-3.39 $C_1C_2C_3C_4C_5 \to \epsilon$,
if $1.C_n = n,x,\text{ç},f,\text{ŋ},m,v,Q_n,k$
v $2.C_1 = s$
v $3.C_2 = t$

X-3.40 $C_1C_2C_3C_4C_5 \to \epsilon$, if $C_5 \neq l$

Y-3.2 $Q_1 \to s,f,nt$ etc.
$Q_2 \to \text{ç},b,h,t,f,$.etc.

VOCALIC CLUSTERS AND THE NATIVIZATION OF LOANWORDS IN POLISH

BARBARA NYKIEL-HERBERT

Program in Linguistics, State University of New York
Binghamton, New York 13901 U.S.A.

One significant fact about Polish phonotactics is that no vocalic clusters are permitted within a single native morpheme -- either at the underlying or phonetic levels of representation. Borrowed morphemes, on the other hand, commonly contain sequences of two vowels: teatr 'theatre', oaza 'oasis', jubileusz 'anniversary', aorta 'aorta'. Both phonological and phonetics vowel sequences occur at prefixal boundaries in words of all categories (e.g. pootwierać 'open', nieostry 'blurred', nausznik 'earmuff'), but at suffixal boundaries the occurrence of vocalic clusters is heavily restricted and limited almost entirely to words formed from foreign lexical material (both stems and suffixes).

In this paper, I attempt to analyze and explain the constraints on the occurrence of vocalic clusters as well as the strategies that the language employs to prohibit the violation of these constraints by loanwords.

The existing analyses of the morphophonological system of Polish convincingly argue that vowel sequences can appear in the underlying representations of complex words of all major lexical categories as a result of morpheme concatenation [3,4]. Rubach [4] claims that Polish permits, on the level of phonetic representation, the occurrence of vowel clusters in nouns but not in verbs; in his analysis, any underlying sequence of vowels in a verb (except those in the prefixal position) is reduced to a single vowel by means of Vowel Deletion Rule, which reads as follows:

(1) $[+\text{syll}] \rightarrow \varnothing \;/\; \underline{\quad\quad}[+\text{syll}]]_V$

The above rule explains how a surface form such as [xrapjõ] 'they snore' derives from the underlying /xrap-a-om/, but it fails to explain why the vocalic clusters are not simplified in verbs such as:

(2)		
	ewakuować	'evacuate'
	sytuować	'situate'
	konstytuować	'constitute'
	substytuować	'substitute'
	instruować	'instruct'
	tatuować	'tatoo'

and a handful of others. In a footnote, Rubach suggests that these verbs contain the labiovelar glide /w/ in their underlying representations and that the glide also appears in pronunciation, e.g. ewak[uwo]wać, syt[uwo]wać. This solution is not workable because it fails to explain why the glide does not appear in some of the corresponding derived nouns: there are no forms such as *konstyt[uw]cja, *substut[uw]cja. A subsequent rule of glide deletion before consonants cannot be postulated in the presence of forms such as [bawvan] 'snowman', [kawtsja] 'deposit', [kuwko] 'wheel', [awktsja] 'auction', etc.

Since all the verbs listed in (2) are obviously loanwords supplied with verbalizing and inflectional morphemes in order to adapt to the native morphological system of Polish, it seems simpler to restrict the application of the Vowel Deletion rule to etymologically native verbs. The labiovelar glide that breaks up the vowel cluster in the verbs of (2) in some pronunciation styles (but not all) has then to be treated as inserted by an optional rule formulated as follows:

(3)
$$\varnothing \rightarrow \begin{bmatrix}-\text{cons}\\-\text{voc}\\+\text{round}\end{bmatrix} \;/\; \left\{\begin{array}{l} \overset{\text{V}}{\begin{bmatrix}+\text{high}\\+\text{back}\\+\text{round}\end{bmatrix}} \;\underline{\quad\quad}\; \text{V} \\ \text{V} \;\underline{\quad\quad}\; \overset{\text{V}}{\begin{bmatrix}+\text{high}\\+\text{back}\\+\text{round}\end{bmatrix}} \end{array}\right\}$$

i.e., the labiovelar glide is inserted between two vowels if at least one of then is /u/. Rule (3) is responsible for labialization not only in verbs such as those in (2) but also in nouns and adjectives, for example: jubil[ewu]sz 'anniversary', akt[uwa]lny 'current', seks[uwo]log 'sexuologist'. Note that the above rule also explains why the glide does not break up the vocalic cluster in kreować 'create'.

Etymologically native nouns and adjectives never exhibit vowel sequences on the level of phonetic representation in what is considered 'Standard Polish'. This is due to the interaction of several morphological and phonological factors:

(a) all nominal stems in Polish are consonant-final
(b) all nominal and adjectival suffixes are vowel-initial
(c) any sequences of vowels that appear in the process of morphological derivation are eliminated by the independently motivated phonological rules (Gliding and yer-Deletion [3]).

The Polish lexicon contains a number of loanwords which violate the constraint requiring that all nominal stems be consonant-final, e.g.:

(4)	(a)		(b)	
	kakao	'cocoa'	boa	'boa'
	makao	'card game'	rodeo	'rodeo'
	kamea	'cameo'	video	'video'
	idea	'idea'	stereo	'stereo'
	orchidea	'orchid'	Mao	'Mao'
	gwinea	'guinea'		
	statua	'statue'		

The nouns in column (a) are pronounced with or without the glide between the two vowels, but there seems to be no such option for the words in (b), in which the vowels are always pronounced as a sequence. The examples in (b) are also undeclinable, as opposed to those in (a), which do take appropriate case endings, at least in informal speech styles. These differences point to a greater degree of nativization of the words in the (a) group. The words in (b) behave as genuinely foreign words, both from the phonological and morphological points of view. They still have the status of monomorphemic words and therefore their vocalic clusters remain intact (cf. teatr 'theatre', Beata (personal name), toast 'toast'). Words in column (a), however, have been reanalyzed as bimorphemic with the final vowel fulfilling the function of inflection. This operation leaves the stems with a final vowel, thus creating impermissible (from the point of view of the native vocabulary) sequences of vowels across the stem-suffix boundary. The glide that appears, at least in some pronunciation styles, at the end of the stems, makes these stems conform to the phonotactic constraints of the language. The nature of the glide seems to be determined by the quality of the stem-final vowel: the labiovelar glide [w] occurs after back vowels ([kakawo], [statuwa]) and the palatal glide [j] after front vowels ([kameja], [ideja]).

The above reasoning inevitably leads to the conclusion that the glides in focus are not inserted by a phonological rule, but are present in the stems in their underlying representations, i.e., stems are consonant-final on both the underlying and phonetic levels. Consider further the data below, which offer support for this claim:

(5)			
	kakao	-	kaka[w]ko (dim.)
	'cocoa'		kaka[w]ek (dim. gen.pl.)
	kamea	-	kame[j]ka (dim.)
	'cameo'		kame[j]ek (dim.gen.pl.)
	statua	-	statu[w]ka (dim.)
	'statue'		stat[w]ek (dim.gen.pl.)
	orchidea	-	orchide[j]ka (dim.)
	'orchid'		orchide[j]ek (dim.gen.pl.)

as well as the word kafejka - kafejek 'coffee shop', which is a nativized form of French café. The suffix employed in these derivations is the diminutive morpheme -ek, represented as /-ɨ̆k/ on the underlying level. Note that the glide occurs not only in intervocalic position but also preconsonantally. If the glide were inserted by a phonological rule to break up a vocalic cluster, then it would have no reason to before consonants as well. Obviously, sequences such as -ak-, -uk-,-ek-, are permissible in Polish, as is demonstrated by such words as brakować 'to lack', stukać 'to knock' and those data in (5) as well.

The glide following vowel-final stems appears also in adjectives formed with the suffix -ski (phonologically /-ĭsk-i/, as the examples below demonstrate:

(6)		
	Dante	dantejski
	Galileo	galilejski
	Prometeusz	prometejski
	Gwinea	gwinejski

and the name of the famous street in Paris, the Champs Elysées, is translated into Polish as Pola Elizejskie.

Yet another piece of evidence comes from the case forms of some of the words under discussion. The locative case of the words kakao, statua is [kakale], [statule], respectively: the segment that appears intervocalically is the lateral [l] rather than the labiovelar glide. The alternation between [w] and [l] is regular in Polish phonology and is considered part of the anterior palatalization process; the alternation kakao - kakale, statua-statule parallels that of szkoła - szkole 'school', koło - kole 'wheel'. It has been argued [2,4] that the segment underlying the [w] - [l] alternation is the velarized lateral /ł/; it changes to [l] when followed by a front non-low vowel, and otherwise it surfaces as the labiovelar glide [w]. Thus, the stem-final consonant (the velarized lateral) has to be present in the underlying representations of loanwords discussed here. The underlying representations of the words kaka[w]ko, statu[w]ka, ide[j]ka are /kakał-ɨ̆k-o/,

/statuł-ɨ̆k-a/, /idej-ɨ̆k-a/, respectively.

It has been demonstrated that the stems which came into Polish as vowel-final have adjusted to the native phonotactic constraints by adding a non-vocalic segment at the end, which segment appears on the surface of some derivatives and some case forms. There are also, however, complex words containing the stems in focus which fail to show the glide in the crucial environment on the surface, or where the glide appears in some pronunciation styles but not all (no such alternative exists for the forms in (5) and (6)). Some of these derivatives are:

(7)	kakaowy	'cocoa' (adj.)
	ideowy	'ideological'
	ideologia	'ideology'
	kameowy	'of a cameo'
	prometeizm	'Prometeism'
	prometeiczny	'Prometeic'

The comparison of the data in (5) and (6) with (7) points to a morphological factor in the distribution of the intervocalic glides: some suffixes (the diminutive -*ek* /-ɨ̆k/, the adjectivizing -*ski* /-ĭsk-i/, and the inflectional endings) do not attach to vowel-final stems that end in a vowel, and etymologically foreign stems ending in a vowel must adjust phonologically to comply with the constraint. What these suffixes have in common, and what distinguishes them from the suffixes used in the formation of the items in (7) is that they are all etymologically native. It comes as no surprise then that etymologically native suffixes attach only to bases whose shape does not violate the phonotactic constraints of the language. The suffixes represented in (7) are all etymologically foreign except for -*owy*, which nevertheless functions with the others and attaches freely to foreign bases (cf. *żakardowy* 'made on a Jacquard loom', *finansowy* 'financial', etc.). It can thus be concluded that the class of foreign suffixes (and -*owy*) do not require consonant-final bases, and so they can attach to stems ending in a vowel.

To account for the variation in the shape of the stem exhibited by the forms in (5) and (6) as opposed to (7), one is forced to consider an allomorphy rule that adds a consonantal extension to etymologically foreign stems when these stems are followed by etymologically native suffixes. Such a rule, even though phonological in form, is actually morphological in nature and applies prior to phonological rules. (The exact status, form, and function of allomorphy rules are extensively discussed in [1].)

As expected, the process of adaptation of borrowings to the native phonomorphological system is not uniform for all items, the time of borrowing and the frequency of usage being the most crucial variables. Another important facts that appears to influence the rate of nativization of the loanwords in focus is the level of education of speakers. The nativizations (phonological and grammatical adaptation) are usually met with long and bitter opposition from the dedicated linguistic prescriptivists who insist on preserving the foreign status of loanwords by adhering to their original pronunciations (which are frequently simply spelling pronunciations) and prohibiting their declension. The here much-discussed example *kakao*, for example, is still considered undeclinable by *Słownik Poprawnej Polszczyzny* [5] and the pronunciation with the labiovelar glide breaking up the final cluster is banned [6]. The diminutive form *kakałko* is not even listed in any of the Polish dictionaries, perhaps because, as one of the prescriptive informants pointed out, "one would not even know how to spell it". However, since *kakao* has a relatively high frequency of usage, the diminutive form *kakałko* is gaining popularity among those speakers who do not necessarily consult a dictionary before uttering a word.

The word *kakao* is probably the most interesting of the loanwords discussed here because it displays the greatest amount of variation in its forms and pronunciations. This variation correlates with the extent to which the word is perceived as "foreign" or "native" by speakers. On the one end of the spectrum, there is the prescriptive pronunciation that does not permit an intervocalic glide in any form; speakers with this pronunciation do not generally decline or diminutivize the word. This pattern predominates among educated, language-conscious speakers who have an awareness of the foreign source of the work, but it is not representative of the majority of Polish speakers. At the other end, there are speakers who have the glide (or, necessarily, [l] in some inflectional forms) throughout the whole pattern: [kakawo], [kakawko], [kakawovɨ], [kakawem], etc. This pronunciation style occurs among speakers with little education and it is considered nonstandard by prescriptivists. At the same time, this pattern shows full nativization to the phono-morphological system; for these speakers, *kakao* has the same status as *koło* 'wheel', which is a native word. In the intermediate group are mixed-pattern pronunciations: with the glide in diminutive and inflectional forms, but without the glide in the citation form and the adjective *kakaowy*. The reverse distribution (with the glide in citation and adjective forms, but absent in inflectional and diminutive forms) does not occur. Such distribution is predicted by the analysis presented earlier: if the glide occurs, it is only in forms with native suffixes immediately following the stem. Prescriptive speakers do not exhibit the allomorphy rule mentioned above, and they have vowel-final representations in their lexicons, which makes diminutives and any inflectional forms impossible. "Nonstandard" speakers do not have the allomorphy rule either, but their representations for the stems under discussion are consonant-final. It thus appears that the allomorphy rule which adjusts foreign stems so that they comply with the demands of the native system is a transitional device in the process of loanword nativization.

The analysis presented in this paper suggests the following conclusions:

(1) restrictions on the occurrence of vocalic clusters in Polish are of phonological, morphological, and lexical natures. Crucial factors include the native/foreign status of stems and suffixes, and phonotactic constraints on the phonological shape of morphemes,

(2) although Polish permits vocalic clusters across prefixal junctures, it tends not to permit them across suffixal boundaries,

(3) there is a correlation between the degree of loanword integration and the occurrence of vocalic clusters; words which are more fully nativized will exhibit fewer vocalic clusters in their various derivational and inflectional formations.

References

[1] M. Aronoff. *Word Formation in Generative Grammar*. MIT Press, 1976.

[2] E. Gussmann. *Contrastive Polish-English Consonantal Phonology*. PWN (Warsaw), 1978.

[3] E. Gussmann. *Explorations in Abstract Phonology*. UMCS Press (Lublin), 1978.

[4] J. Rubach. *Cyclic Phonology and Palatalization in Polish and English*. Warsaw University Press, 1981.

[5] *Słownik Poprawnej Polszczyzny* (W. Doroszewski, ed.). PWN, 1977.

[6] *Słownik Wymowy Polskiej* (M. Karas and M. Madejowa, eds.). PWN, 1977.

[7] A. Zajda. Problemy wymowy polskiej. In [5], pp. xxvii-xxxviii.

A STATISTICAL APPROACH TO SPANISH AMERICAN PHONOLOGICAL UNITS

MIGUELINA GUIRAO MARIA A. GARCIA JURADO

Laboratorio de Investigaciones Sensoriales
CONICET, Universidad de Buenos Aires
CC 53 1453 Buenos Aires, Argentina

ABSTRACT

In this paper we report an account of Spanish American phonological units. The sample consisted of 74460 syllables distributed in 43306 words. The frequency of occurrence of phonemes are presented each one labeled according to articulatory features. Dentals have an incidence of almost four times more than velars. Palatals have a low frequency. Voiced phonemes gave higher figures than unvoiced ones. A table is presented with a ranking order of the first 50 syllables. Thirty one of these syllables are also words. Included is the percent incidence of each of the first 50 syllables in initial and final word position. Type CV is equally distributed in both positions and CVC tend to terminate words. An observation is made on the most frequent articulatory combinations encountered at both extremes of words.

INTRODUCTION

In this work we intent to address the problem of segmentation of morphemic units in continuous speech on the basis of statistical data. Other aims were to provide useful information for spectrogram reading and for cross language comparison.

Our system consists of five vowels and seventeen consonants. All but four phonemes /ʒ/ (spelled y and ll), /tʃ/ (ch), /x/ (j) and /ɲ/ (ñ) can be represent with orthographic symbols.

In spoken Spanish the syllabic structure predominate over the morphemic one. It is also known that CV constitutes more than 50% of all syllabic types. The expansion of this pair produces CV+C, CV+V, CV+V+C which add over to 90% of all syllabic types /1/. It is also true that in Spanish there is a tendency for syllables between words to be fused. This and other phonological changes take place according to principles that could be determined. Taking this facts into account we centered our attention into the syllabic segments that are encountered at both margins of words. Before entering into this problem we examined a previous inventory of phonemes and classified the sounds adopting broad conventional articulatory terms /2/. A more detailed articulatory with allophonic variants is presented elsewhere /3/.

The corpus covered 163861 phonemes and 74460 syllables distributed in 43306 words. The text was extracted from five modern plays written in conversational style, by contemporary argentine authors.

DISTRIBUTION OF VOWELS AND CONSONANTS

In Figure 1 we present the five vowels ordered according to tongue height.

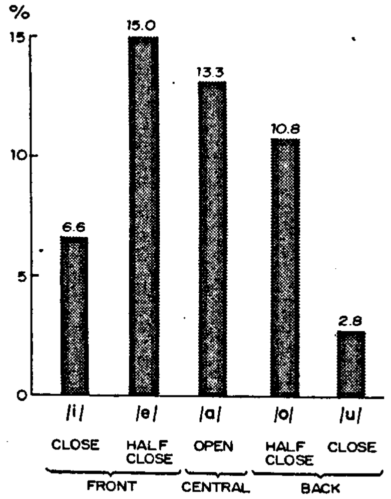


Figure 1. Distribution of Spanish vowels by articulatory configurations.

Taking as a reference frontal half closed /e/, which is the most used sound, scoring 15%, central /a/ and back closed /o/ follow with relativately small differences. Closed vowels are less used. Frontal /i/ by a factor of two and half and back /u/ by a factor of five.

In Figure 2 consonants are distributed according with place, voicing and manner of articulation. Dentals are produced approximately four time more than labials, six times more than velars and twenty six times more than palatals. Voiced phonemes are more frequent than unvoiced ones.

As for manner of articulation nasal /n/ is more than twice more recurrent than labial /m/ while palatal /ɲ/ is infrequently produced. Within the group of liquids vibrant /r/ is more recurrent than lateral /l/ but /rr/ is much less pronounced.

The more recurrent stops are dentals /t/ and /d/, velar /k/, followed by labials /b/ and /p/. Velar /g/ is less used.

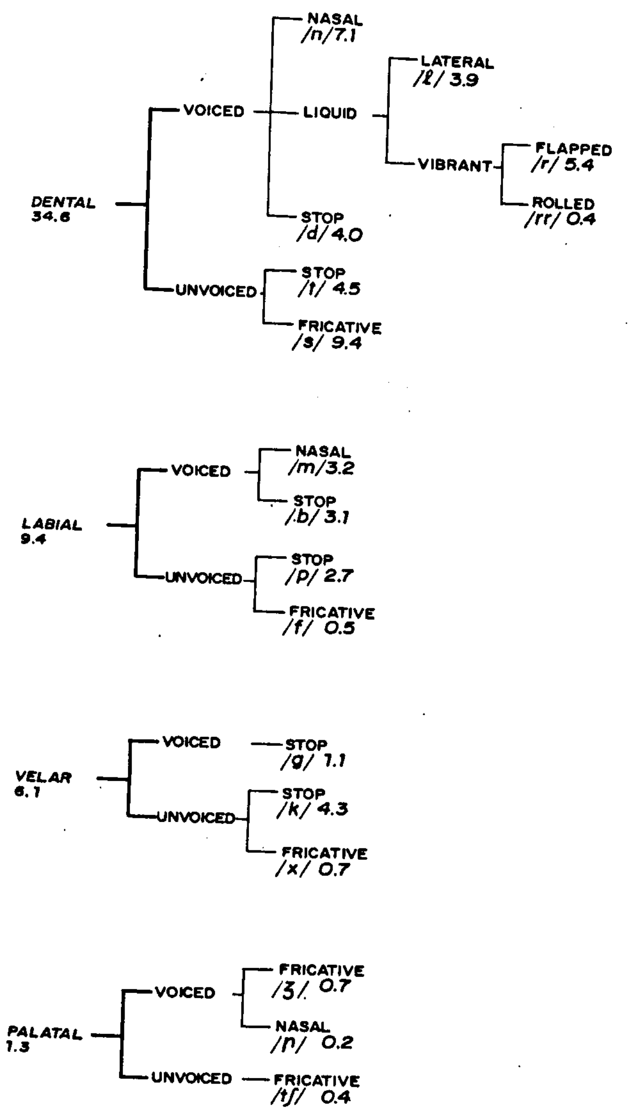


Figure 2. Distribution of Spanish consonants by articulatory configurations.

In general we observe that a large number of sounds are produced at the front of the mouth. Two frontal vowels plus dentals and labials summate over three thirds of all phonemes occurrences.

Voicing is another apparent feature vowels and voiced consonants added together make more than three fourths of the total occurrences.

A comparison with the Peninsular Spanish pronuntiation gave minor deviations which are correlated with some differences in the phonological system /2/.

FREQUENTLY REPEATED SYLLABLES

It is noted that the high incidence of certain phonemes result from a group of frequently repeated syllables.

Table 1. The first fifty syllables (with asterisk are also words)

Syllables	Percent	Syllables	Percent
/a/*	5.529	/el/*	1.050
/ke/*	3.920	/ro/	1.043
/no/*	3.826	/pe/	1.019
/de/*	2.484	/en/*	0.992
/se/*	2.460	/so/	0.957
/es/*	2.375	/le/*	0.870
/i/*	2.080	/por/*	0.864
/te/*	2.042	/di/*	0.835
/si/*	1.775	/mo/	0.819
/do/*	1.676	/u/*	0.768
/ta/	1.649	/ma/	0.742
/to/	1.598	/be/*	0.687
/la/*	1.518	/mi/*	0.679
/me/*	1.489	/ne/	0.640
/ra/	1.449	/bi/*	0.620
/sa/	1.407	/po/	0.608
/ko/	1.321	/ʒa/*	0.596
/na/	1.304	/ʒo/*	0.593
/pa/	1.286	/kon/*	0.589
/e/*	1.249	/go/	0.585
/ba/*	1.200	/un/*	0.549
/lo/*	1.193	/mos/	0.531
/da/*	1.136	/mas/*	0.529
/o/*	1.075	/des/	0.486
/ka/	1.073	/ga/	0.474

In Table 1 we present the first 50 items which represent 66.3% of the total sample but only 7% of the different syllables. Thirty seven are formed by a CV pair, five by CVC, five by V and four by VC. The most frequently pronounced consonants are dentals /d t s n l/ which summate 27%, labials /m p b/ 10% and velar /k/ 7%. These consonantal components are mainly combined with the three called strong vowels /e a o/. Thirty one of the units listed in Table 1 are monosyllabic words. These units appear with a relative much higher incidence 48.7% than bisyllabic 34.7%, trisyllabic 13.0% or tetrasyllabic words 3.1%. However when looking at the number of different words which in the total sample are 3993, we found that those formed by one syllable are only 3.3% while those of two are 34.8%, of three 41.2% and of four 17.1%.

SYLLABLES AT BOTH MARGINS OF WORDS

To count syllabic sounds located either at the onset or at the offset of words, we took again a sample formed by the first 50 syllables. Most of these items resulted from bisyllabic words. The data are displayed in Table 2 and 3.
The units in Table 2 represent 74% of the total corpus and 1.7% of the different syllables initiating words. The range of scores is in the order of 1 to 20. To facilitate the inspection of our data we have divided the table in two sets of 25 syllables. The first from 9.8 to 1.0% and the second from 0.97 to 0.49%. In the first set we see that vowel alone or in VC pair are the most frequently produced sounds initiating syllables. Open vowel /a/ 9.8% is first. Vowel /e/ alone and combined with a fricative in /es/ and a nasal in /en/ summate 10.4%. Back closed vowel /u/ alone and /us/ make 3.8%. With lower scores isolated vowels /o/ and /i/ appeared also in the first set. The rest of the list is completed by three labials /p b m/, four dentals /t d s n / and one velar /k/.

Table 2. Syllables in initial word position

CV: 35.838		CVV: 4.229
/pa/ 3.697	/rre/ 0.720	/bue/ 1.307
/ko/ 3.070	/rra/ 0.559	/tie/ 1.051
/pe/ 2.878	/mo/ 0.501	/pue/ 0.716
/to/ 2.143	/ti/ 0.492	/kie/ 0.635
/di/ 2.072	/ʒe/ 0.492	/bie/ 0.591
/ka/ 1.987		
/de/ 1.969	V: 18.566	CVC: 3.997
/sa/ 1.964		
/se/ 1.575	/a/ 9.802	/des/ 0.971
/po/ 1.315	/e/ 3.334	/kon/ 0.725
/na/ 1.293	/u/ 2.542	/tam/ 0.649
/ma/ 1.172	/o/ 1.602	/ten/ 0.604
/me/ 1.114	/i/ 1.284	/por/ 0.595
/te/ 1.033		/pen/ 0.452
/ba/ 1.002		
/mi/ 0.917	VC: 8.965	CCV: 1.597
/bi/ 0.890	/es/ 4.735	/kla/ 0.819
/ke/ 0.877	/en/ 1.378	/tra/ 0.778
/mu/ 0.872	/us/ 1.307	
/be/ 0.774	/al/ 0.868	CVVC: 0.693
/no/ 0.747	/an/ 0.675	/kuan/0.693

Among the labials stop /p/ articulated in pair with strong vowels constitute the favourite combination. Follows a voiced stop in /bue ba/ and a nasal in /ma me/. In the category of dentals stops /t d/ in /to te ti/, /di de/ and fricative /s/ in /sa se/ ranked first. Velar /k/ appeared relatively often in /ko ka/. Subtotals are vowels 26.9, labials 12.2 dentals 12.1 and velars 5%.
Most of the consonantal sounds listed in the first set appeared again in the second but in pair with vowels of lower incidence or in CVC or CCV context. The distribution of all items is quite even. This time initial vowels-with exception of /al/ and /an/-were not registered. Fricative /s/ was also missing but another dental in /rra rre/ and a palatal in /ʒe/ entered in the list. Subtotals are: vowel 1.5, labials 5.8, dentals 5.6 and velars 3.7%.
Syllables in Table 3 represent 70% of the total sample and 1.8% of those terminating words. In this list scores are ranging from 5.3 to 0.49%. This proportion, close over 1 to 10, is practically half of the range observed in Table 2. In this table we observed a clair predominance of dentals. This category covers about two thirds of the total percentage listed in the table. The first eleven sounds in CV pair are dentals. Also, in CVC all but two, belong to this category. Stops /t/ and /d/ fricative /s/ nasal /n/ and vibrant /r/ combined mainly with the strong vowels scored each an average close to 10%. Labials /m/ in /mo mos ma mas/ and /b/ in /ba be bien bre/ follow with 4.7 and 3% respectively. Among the velars stop /k/ and /g/ ranked in /ko ka ke/ 2.8% and in /go ga/ 2.4%. Fricative/x/ in /xa/ 1.3%. Palatals in /ʒa/ and /tʃo/ scored only 0.6 and 0.5% respectively. Two isolated vowels are found in this list /a/ 2.4% and /i/ 0.6%.
In comparing Tables 1, 2 and 3 we observe that eighteen of the items are common to the three lists. Nine are words /a i ke de te se ba be no/. Forty percent of the items in Table 1 are not found either in Table 2 or 3. Among these there are six words /la lo el le o un/ which because of their grammatical function (articles and pronouns) frequently occur as isolated one syllable words. It was also noticed than when forming part of larger segments eleven words /es me e o en por di u mi bi kon/ appeared in initial position and five /si do a mas da/ were located at the end of words.

Table 3. Syllables in final word position

CV: 55.064		
/do/ 5.308	/xo/ 0.863	/nas/ 0.559
/ra/ 4.435	/ga/ 0.738	/tas/ 0.550
/ta/ 4.122	/ke/ 0.734	/res/ 0.541
/ro/ 3.450	/ʒa/ 0.671	/nos/ 0.510
/to/ 2.989	/ma/ 0.671	/sır/ υ.510
/na/ 2.801	/po/ 0.514	/mas/ 0.492
/te/ 2.788	/xa/ 0.505	
/da/ 2.743	/tʃo/ 0.505	V: 3.083
/no/ 2.667		
/so/ 2.327	CVC: 11.144	/a/ 2.475
/se/ 2.193		/i/ 0.608
/mo/ 2.112	/mos/ 1.696	
/sa/ 2.027	/ted/ 0.944	
/go/ 1.785	/nes/ 0.801	CVVC: 1.145
/de/ 1.383	/ses/ 0.796	/pues/0.598
/ne/ 1.351	/tar/ 0.734	/bien/0.555
/ko/ 1.239	/des/ 0.626	
/ba/ 1.159	/dos/ 0.622	
/be/ 1.025	/ser/ 0.617	CCV: 0.496
/ka/ 0.989	/sas/ 0.577	
/si/ 0.957	/ron/ 0.564	/bre/ 0.496

FINAL REMARKS

The range of frequently repeated syllabic sounds at both margins of words is twice as big for those initiating words.
Combinations of CV type are almost equally distributed while CVC is more frequent terminating words. Types V, VC, CVV and CCV are mainly in initial position.
As for distribution of articulatory combinations we differentiate four significant categories. Two are prevalent at the onset of words: a) Vowels alone and in VC pair /es/ and /en/; b) Labials /p b m/ plus strong vowels. The third category is formed by dentals combined with strong vowels which are limiting both sides being three times more frequent at the end of words. Velar /k/ is similarly distributed at both extremes.
Monosyllabic words that are part of larger lexical units are more frequently occurring in initial than in final position.

REFERENCES

/1/ M. Guirao, M.A. García Jurado, "Frequency of acoustical patterns in Spanish syllables". Journal of the Acoustical Society of America Suppl.1,vol.78:S55, 1985.

/2/ M. Guirao, M.A. García Jurado, "Frequency of occurrence of phonemes in American Spanish" (to be published).

/3/ M. Guirao, A.M.Borzone de Manrique, "Fonemas, sílabas y palabras del español de Buenos Aires" Filología 16: 136-165, 1972.

DATEN DER PHONOLOGISCHEN STATISTIK ALS INDIZ FÜR DEN GRAD DER SPRACHVERWANDTSCHAFT (AM BEISPIEL URALISCHER SPRACHEN)

WOLFGANG VEENKER

Finnisch-Ugrisches Seminar
der Universität Hamburg

ZUSAMMENFASSUNG

Überblick über ein Projekt zur einheitlichen Erforschung, Beschreibung und Vergleichung der Lautstände in den uralischen Sprachen; Hinweise auf den Grad der wechselseitigen Verwandtschaftsverhältnisse dieser Sprachen im Spiegel der phonologischen Statistik.

*

AUSFÜHRUNGEN

(1) Um eine Vergleichbarkeit der Phonemsysteme der einzelnen uralischen (finnougrischen) Sprachen vornehmen zu können, habe ich vor einigen Jahren ein Beschreibungsmodell /1/ entworfen, das nach einheitlichen Prinzipien aufgrund der phonetischen Eigenheiten der einzelnen Laute unter Berücksichtigung der phonologischen Relevanz eine Klassifizierung ermöglicht; ich bediene mich dabei eines sechsziffrigen Codes für jedes Phonem. Haben zwei Phoneme in verschiedenen Sprachen den gleichen Code, so sind sie phonologisch identisch, wenngleich es phonetisch unterschiedliche Nuancen geben kann, die jedoch im Hinblick auf das phonologische System irrelevant sind.
Mein Beschreibungsmodell ermöglicht des weiteren Vergleiche zwischen den einzelnen uralischen Sprachen resp. Dialekten im Hinblick auf die Ausgestaltung und Belastung innerhalb der verschiedenen phonologischen Korrelationen. Dies kann für die synchron-konfrontierende Betrachtungsweise wie auch für die diachrone Entwicklung interessante Ergebnisse zeitigen.

(2) In einem langfristigen Projekt habe ich für eine Reihe uralischer Sprachen die Frequenzdaten der einzelnen Phoneme ermittelt; der Ausgangspunkt war ein hinreichend umfangreicher Text /2/ von jeweils ca. 2500 Phonemen, der in allen finnougrischen Sprachen vorliegt. Dieser Text ist zwar insofern ein wenig manipuliert, als er bewußt bevorzugt genuin finnougrischen Wortschatz und keine Fremdwörter enthält, er vermittelt in dieser Hinsicht jedoch ein gutes Bild über die Verbreitung der eigenständigen Phoneme der jeweiligen Sprache (die in der UdSSR beheimateten finnougrischen Sprachen unterliegen auch im Bereich der Phonetik/Phonologie einem starken Einfluß des Russischen, da mit der Übernahme von russischen Fremdwörtern heutzutage im Gegensatz zu der Übernahme von Lehnwörtern in früheren Zeiten auch "Fremdphoneme" allmählich in diesen Sprachen heimisch werden). Eine vorgenommene Analyse der Häufigkeit der einzelnen Phoneme in elf finnougrischen Sprachen (UNG, VOG-N, OST-Š, KPM, SYR, UDM, ČRW, ČRB, MOK, ERZ, FIN) liefert ein hinreichend umfangreiches Material, um daraus weitere Rückschlüsse ziehen zu können: alle Zweige der finnougrischen Sprachen (mit Ausnahme des Lappischen) sind vertreten, wobei besonderes Gewicht auf die Analyse der "kleineren" Sprachen der volgaischen und permischen Gruppe gelegt wurde.

(3) Ermittelt wurden die Daten nicht nur der Häufigkeit des Vorkommens der Phoneme, sondern auch in bezug auf ihre Position im Wort und ggf. in der Silbe. Der Umfang des Materials kann nur angedeutet werden. Die Daten im einzelnen sind in einer Reihe von Arbeiten /3/ zugänglich gemacht worden. Eine Beschränkung ergibt sich bislang noch insofern, als die äußerst wichtigen phonotaktischen Daten noch nicht ermittelt werden konnten: für die Fragen der Distribution und vor allem der Kombinatorik ist das von mir zugrunde gelegte Corpus zu gering. In dieser Hinsicht habe ich auf der Basis umfangreicherer Corpora Untersuchungen fürs Syrjänische /4/ und fürs Tundrajukagirische (i. e. eine paläosibirische Sprache!) /5/ durchgeführt und vorgelegt.

(4) Die zeitliche Distanz der heutigen uralischen Sprachen zu der angenommenen uralischen Protosprache dürfte ca. 6000 Jahre betragen, die Distanz der heutigen finnougrischen Sprachen zu der finnougrischen Protosprache sicherlich auch ca. 4500 Jahre oder mehr. Sprachdenkmäler liegen - und auch das nur für einige Sprachen - erst seit wenigen Jahrhunderten vor, zudem oftmals in einer Form, die eine phonologische Analyse stets mit vielen Fragezeichen versehen lassen muß, soweit es die älteren Sprachzustände betrifft. Phonetisch zuverlässige Aufzeichnungen, die dann auch eine phonologische Analyse ermöglichen, liegen erst seit Ende des vorigen Jahrhunderts vor. In dieser Hinsicht ist die Situation in der Uralistik/Finnougristik beträchtlich ungünstiger als in anderen Sprachfamilien.

(5) In der Uralistik haben seit Beginn dieser wissenschaftlichen Disziplin die Lautforschung und die Etymologie eine tragende Rolle gespielt, wobei beide Bereiche in wechselseitiger Abhängigkeit zueinander stehen. Bei meinen Untersuchungen zur Frequenz der Phoneme ist mir daher der Gedanke gekommen, daß sich in den ermittelten Ergebnissen - wenn sie nur entsprechend interpretiert werden - beim Vergleich zwischen den einzelnen Sprachen auch Rückschlüsse auf den Grad der wechselseitigen Verwandtschaft ziehen lassen müßten, oder anders ausgedrückt: in den Daten müßte sich der Grad der Verwandtschaft widerspiegeln:
Wenn es zu einem "Lautwandel" kommt, so geschieht dies ja nicht durch die Ersetzung eines Phonems durch ein beliebiges anderes, sondern das dem Lautwandel unterworfene Phonem verändert sich nur in einem oder im Verlaufe der Zeit in einigen wenigen seiner Merkmale, d.h. ein Lautwandel wie etwa /t/ > /s/ beinhaltet eine Spirantisierung, Artikulationsstelle und andere Eigenheiten bleiben aber (zunächst) erhalten, können sich sehr wohl aber im Verlaufe von weiteren Wandeln verändern.
Im Bereich der uralischen Sprachen sind diese Wandel schwer nachzuvollziehen, da es - wie schon erwähnt - an älteren Sprachdenkmälern mangelt; lediglich die Möglichkeit einer relativen Chronologie, wodurch ersichtlich wird, daß der eine Lautwandel dem anderen vorausgegangen sein muß, läßt gewisse Rückschlüsse zu.

(6) Meine Idee ist die folgende: da bei einem Lautwandel nicht alle Charakteristika eines Phonems betroffen sind, weil die einzelnen Sprachen sich unterschiedlich verhalten, müßte es möglich sein, bei der Aufsplitterung der einzelnen Charakteristika der Phoneme deren statistische Frequenzdaten zu vergleichen: Sprachen, die gemäß dem Stammbaummodell enger verwandt sind, müßten dann größere Ähnlichkeit in den statistischen Daten aufweisen et vice versa.

(7) Der knappe zur Verfügung stehende Raum läßt nur einige wenige Beispiele zu. Betrachten wir zum Beispiel das Verhältnis von Vokalen (VOC) zu Konsonanten (CNS): Nach dem üblichen, den Verwandtschaftsgrad widerspiegelnden Stammbaummodell ergibt sich folgendes Bild:

Sprache	VOC	CNS	(in %)
UNG	41,72	58,30	
VOG-N	39,65	60,38	
OST-Š	39,82	60,18	
KPM	42,00	58,00	
SYR	41,24	58,77	
UDM	42,26	57,72	
ČRW	41,33	58,67	
ČRB	42,21	57,80	
MOK	39,31	60,69	
ERZ	39,00	61,01	
FIN	43,25	56,73	

im Vergleich dazu zwei nicht-uralische Sprachen:

*DEU	37,85	62,17
*JUK-T	42,61	57,39

Auf der Basis von elf finnougrischen Sprachen erhalten wir einen Durchschnittswert von

41,07 % VOC ./. 58,93 % CNS,

wozu beispielsweise die Nähe oder Ferne jeder der untersuchten finnougrischen Sprachen in Relation gesetzt werden kann. Ordnet man diese Daten jetzt nach den Frequenzwerten, ergibt sich folgende Reihung:
FIN - *JUK-T - UDM - ČRB - KPM -
UNG - ČRW - SYR - OST-Š - VOG-N -
MOK - ERZ - *DEU

(8) Es ist natürlich evident, daß ein einziges Kriterium nicht ausreicht. Es sollen daher einige weitere Ergänzungen gegeben werden, um die Methode zu verdeutlichen. Bei einer Klassifizierung der Vokale nach der Zungenstellung (Öffnungsgrad) ergibt sich für den Anteil der Vokale mit hoher Zungenstellung folgendes Bild:

(a) Auflistung gem. Stammbaum	
UNG	12,98
VOG-N	22,90
OST-Š	19,25
KPM	39;65
SYR	43,47
UDM	51,60
ČRW	16,75
ČRB	8,45
MOK	18,30
ERZ	24,36
FIN	34,68
*DEU	42,35
*JUK-T	28,95

(b) Auflistung gem. Frequenz	
51,60	UDM
43,47	SYR
42,35	*DEU
39,65	KPM
34,68	FIN
28,95	*JUK-T
24,36	ERZ
22,90	VOG-N
19,25	OST-Š
18,30	MOK
16,75	ČRW
12,98	UNG
8,45	ČRB

Die "aus dem Rahmen fallenden" Daten einzelner finnougrischer Sprachen lassen sich bei Kenntnis des Phoneminventars und der

Lautgeschichte teilweise recht leicht interpretieren.

(9) Es sei des weiteren ein Beispiel angeführt aus dem Bereich des Konsonantismus: Analyse nach der Artikulationsart, in diesem Fall der Anteil der Fricativae (in % aller Konsonanten):

(a) Auflistung gem. Stammbaum		(b) Auflistung gem. Frequenz	
UNG	24,28	37,33	SYR
VOG-N	32,53	35,64	UDM
OST-Š	30,77	34,18	ČRB
KPM	33,45	33,89	ČRW
SYR	37,33	33,45	KPM
UDM	35,64	33,21	ERZ
ČRW	33,89	33,15	MOK
ČRB	34,18	32,53	VOG-N
MOK	33,15	30,77	OST-Š
ERZ	33,21	28;66	*DEU
FIN	26,26	26,26	FIN
*DEU	28,86	24,28	UNG
*JUK-T	15,29	15,29	*JUK-T

(10) Aus der Fülle des Materials kann ich hier nur einen kleinen Ausschnitt vorlegen, die Untersuchung und Analyse sind auch noch nicht abgeschlossen. Gleichwohl scheint mir schon jetzt deutlich zu sein, daß mit dieser Methode in Verbindung mit den Verfahren, die ich an anderer Stelle /4, 6/ beschrieben habe, ein neuer Ansatz geschaffen wird, um zur Aufklärung strittiger Probleme in der Lautforschung beizutragen. Das von mir eingangs erwähnte Beschreibungsmodell /1/ mit Verwendung des Code-Systems ermöglicht eine rasche Orientierung, die bei der weiteren Untersuchung besonders phonotaktischer Fragen nützlich sein kann.

(11) Abschließend sei die vorläufige Aussage gemacht, daß sich die in der Forschung einhellig angenommene genetische Klassifizierung der finnougrischen Sprachen auch in der Frequenzstatistik widerspiegelt.

ABKÜRZUNGEN

ČRB	bergčeremissisch
ČRW	wiesenčeremissisch
DEU	deutsch
ERZ	eržamordvinisch
FIN	finnisch
JUK-T	tundrajukagirisch
KPM	komipermjakisch
MOK	mokšamordvinisch
OST-Š	ostjakisch (Dialekt von Šerkaly)
SYR	syrjänisch
UDM	udmurtisch, votjakisch
UNG	ungarisch
VOG-N	vogulisch (nördl. Dialekt)

LITERATURHINWEISE

/1/ W. Veenker: "Vorschlag für ein phonologisches Beschreibungsmodell der uralischen Sprachen und Dialekte". - Dialectologia Uralica, Wiesbaden 1985, 33-47.

/2/ "Tekst dlja perevoda. Lisa." - Osnovy finno-ugorskogo jazykoznanija (1). Moskva 1974, 439-481.

/3/ W. Veenker: "Zur phonologischen Statistik der komipermjakischen Sprache." - Finnisch-Ugrische Mitteilungen 3 (Hamburg 1979), 13-27. -- "Zur phonologischen Statistik der vogulischen Sprache." - Festschrift für Wolfgang Schlachter. Wiesbaden 1979, 305-346. -- "Bemerkungen zur Verteilung der Vokale im Vogulischen." - Finnisch-Ugrische Mitteilungen 4 (Hamburg 1980), 75-83. -- "Zur phonologischen Statistik der čeremissischen (marischen) Schriftsprachen." - Sovetskoe Finno-ugrovedenie 16 (Tallinn 1980), 106-134. -- "Problemy fonologičeskoj statistiki chantyjskogo jazyka." -- Teoretičeskie voprosy fonetiki i grammatiki jazykov narodov SSSR. Novosibirsk 1981, 84-96. -- "Zur phonologischen Statistik der votjakischen Sprache." - Lakó-Emlékkönyv - nyelvészeti tanulmányok. Budapest 1981, 196-213. -- "Zur phonologischen Statistik der mordvinischen Schriftsprachen." - Ural-Altaische Jahrbücher NF 1 (Wiesbaden 1981), 33-72. -- "Zur phonologischen Statistik der syrjänischen Sprache." - Études Finno-ougriennes 15 (Budapest/Paris 1982), 435-445. -- "Konfrontierende Darstellung zur phonologischen Statistik der ungarischen und finnischen Schriftsprache." - Nyelvtudományi Közlemények 84 (Budapest 1982), 305-348. --

/4/ W. Veenker: "Zur Architektonik der syrjänischen Sprache." - Ural-Altaische Jahrbücher NF 5 (Wiesbaden 1985), 30-44. -- "Architektonika komi-zyrjanskogo jazyka." - Sovetskoe Finno-ugrovedenie 22 (Tallinn 1986), 39-49.

/5/ W. Veenker: "Architektonik der jukagirischen Sprache." - Lingua Posnaniensis 28 (Poznań 1985), 79-107.

/6/ W. Veenker: "Beschreibung der Wortstruktur." - Wiener Linguistische Gazette, Beiheft 3 (Wien 1984), 266-271.

PHONOTACTICAL CONSTRAINTS STRENGHT CHANGES AS FUNCTION OF INSIDE WORD SYLLABLE POSITION

ALESSANDRO FALASCHI

University of Rome "La Sapienza"
Information and Communication Department
Via Eudossiana 18, 00184 Roma, Italy

ABSTRACT

The aim of the work is to show how much and where the spoken chain is constrained by paradigmatic neighbouring effects. The constraints strength is evaluated in information theory terms by the ratio between the average mutual information of phonetic symbols pairs and the entropy of the phonetic sequence. Positional variation of the symbols predictability is evidenced by evaluating the strength measure along a probabilistic phonetic word model, which takes into account the phonological knowledge about the syllable structure. The parameters of the model are estimated from a phonetic sequence training corpus obtained by automatic translation and syllabic segmentation of a training text data base.

I. INTRODUCTION

Recent advances in automatic speech recognition technology [1][2] have lead to a statistical point of view of linguistic message analysis. Such approach can easily bring to new tools to be utilized in verifying linguistical hypothesis.

Along this study the phonetic message is regarded as a sequence generated by a symbolic information source whose alphabet consists of the italian language phonemes set. As the symbols emitted by a source are not completely predictable they convey information; such quantity will be so much greater as the symbols have little probability. The time average of the information is called entropy, and represents a global information about the source predictability and is measured in bits/symbol. This number represents in fact the average number of binary choices necessary to "guess" the next symbol once the past ones and the statistical behaviour of the source are completely known. It is clear that more constrained are the sequences emitted by the source and smaller will result the entropy.

The phonetic sequences are constrained by many low and high level factors like syllable structure, sintax and semantic: a perfect knowlege of them will bring us to the true entropy of the language. Ignoring some of them will produce higher entropy values, and the difference indicates how much the omitted knowledges conditionates the allowable symbol sequences.

Since our interest is to evaluate the neighbouring effects strength, a comparison between the entropy values obtained supposing known or not the last symbol emitted is done. In order to analyze if and how the phonotactical constraints varies with the position inside a generalized word model, the entropies will be evaluated for every morphological state of the model, giving rise to non-stationary entropy functions. It is to be highligthed that this approach results innovative with respect to a stationary source model [3].

The following section illustrates how the morphological knowledge is organized in a data structure representing the phonetic information source as a markov source. Section III will give the analitical formulas for computing the constraints strenght measure and section IV will illustrate how the phonetic conditional probabilities to be utilized in such formulas are estimated. Finally, the results of the analysis are presented in section V in form of istograms of the computed quantities.

II. THE WORD MODEL

The word source model (WSM) that will be now described is just an economic way to represent the phonetic non stationary conditional probabilities which are needed for evaluating the entropy functions briefly describer above. In oder to take into account the neighbouring effects and the morphological word structure, the phoneme probabilities will be treated as functions of the following four events: 1) Previous phoneme (pp); 2) Syllable number (sn); 3) Syllabic morphological state (ss) 4) Syllable position whit respect to the lexical stress (a).

The dependece upon the syllable number and position with respect to the stress can be modelled by the transition diagram of Fig.1, where $S^0{}_i$ represents a syllabic source model (SSM) for syllable number i which precedes or contain the stressed vowel of a word (a=0), $S^1{}_i$ represents SSMs (a=1) for syllables following the stressed vowel and the arcs between the models individuate the allowable transition between syllables. Note that transitions from $S^0{}_i$ to $S^1{}_{i+1}$ or to the lower bar (representing end-of-word symbol emission) may occur

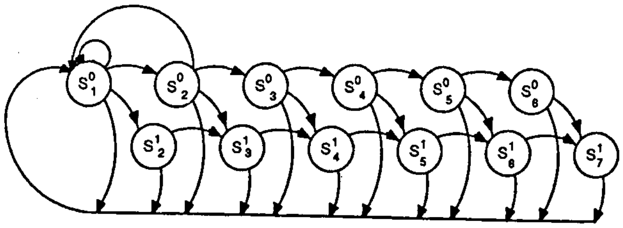


Fig. 1 - Word Source Model main structure

only after a stressed vowel, thus assuming the existence of only one stressed vowel for word. The only exception is for apostrophed words, which are treated as prefixes and modeled by means of the upper backward transition of Fig.1.

In order to represent the syllabic morphology every SSM is then substituted with another transition diagram which exploites the allowable phoneme class sequences as described in [4]; such graphs are drawn in Figg. 2 and 3 for the S^0 and S^1 type SSM respectively. These diagrams identify the italian syllabic structure by means of an outlined syllabe state number (indicated as ss before) and phoneme class label superimposed over the transitions; back transitions for apostrophed words are not shown.

The previous phoneme conditioning is also not represented in Figg. 2-3; the SSM adopted will then have, for any ss shown, as many states as the number of different phonemes (pp) which can be emitted by the WSM while it executes a transition ending in such position.

At this point it should be clear how the four conditioning events have been encoded inside the WSM. It remains to emphasize that the set (a,ns,ss,pp) is in fact a markov source state identifier, whose outgoing transitions are described not only by the phonetic symbol identity but also by the destination state to be reached. In the following the transition probabilities will be indicated as $P(t_i/a,ns,ss,pp)$ where i spaws from 1 to the state outgoing number of transition.

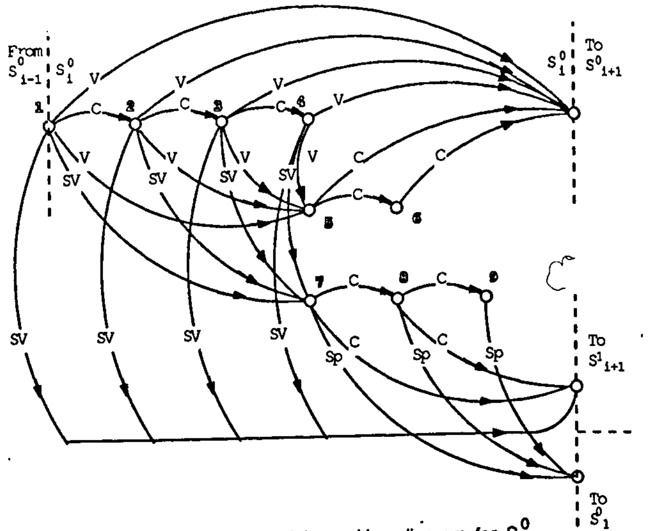


Fig. 2 - Syllabic Source Model transition diagram for S^0_i

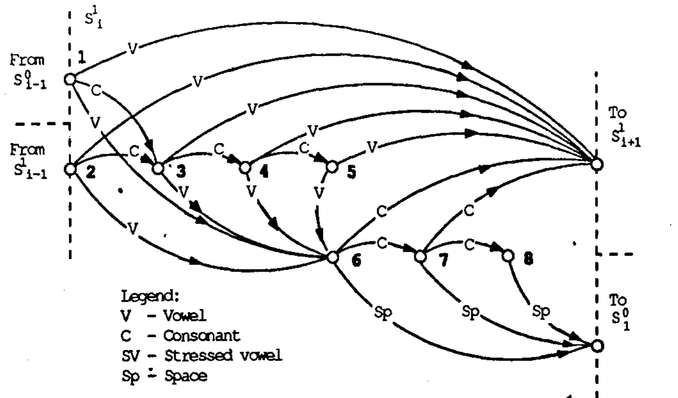


Fig. 3 - Syllabic Source Model transition diagram for S^1_i

III. CONSTRAINTS STRENGHT MEASURE

As anticipated in Sect. I, the amount of symbol predictability due to the previous one is evaluated by means of the difference between the source entropy H_0 and the conditional source entropy H_1. This quantity is called average mutual information (I_m) of the symbols pairs emitted by the source and it also will be a function of the word morphological state, thus having

(1) $I_m(a,ns,ss) = H_0(a,ns,ss) - H_1(a,ns,ss)$

Let us now considerate the behaviour of I_m as a function of the entropy values. If the knowledge of the last symbol emitted by the source do not increase the sequence pre'ictability the conditional entropy value H_1 results equal to H_0, and the zero value of the mutual information can be regarded as no constraints on the symbols sequence. On the other hand, I_m assumes the maximum value H_0 when H_1 results zero, indicating that the previous symbol knowledge impose an unique choiche for the next one, i.e. a maximum phontactical constraint strength condition. In oder to derive an omogeneous strenght measure along the WSM states, the I_m is normalized by the entropy H_0 thus having the normalized average mutual information

(2) $I_N(a,ns,ss) = I_n(a,ns,ss)/H_0(a,ns,ss)$

whose values spawn between zero and one for the cases of null or absolute sequence predictability.

As the entropy is the expected value of the information emitted by the source and having defined the information received after a transition in the model has taken place as the base 2 logarithm of the inverse of the transition probability, the entropies computation formulas once given the parameters of the model are

(3) $H_0(a,ns,ss) = \Sigma_i Q(p(t_i/a,ns,ss))$

(4) $H_1(a,ns,ss) = \Sigma_j p(pp_j/a,ns,ss)\Sigma_i Q(p(t_i/pp_j,a,ns,ss))$

where $Q(a) = aLg_2(1/a)$.

The non stationary functions (1)-(4) can be statistically averaged over their arguments, in order to point out the dependencies from syllable number or syllable state or stress relative position only, or even a global value. Here below follows the formulas for obtain such values. for a generic non stationary function R(a,ns,ss):

$$R(a,ns) = \Sigma_{ss} p(ss/a,ns)R(a,ns,ss)$$

$$R(a,ss) = \Sigma_{ns} p(ns/a,ss)R(a,ns,ss)$$

$$R(a) = \Sigma_{ss}\Sigma_{ns} p(ss,ns,/a)R(a,ns,ss)$$

$$R = \Sigma_a \Sigma_{ns} \Sigma_{ss} p(a,ns,ss)R(a,ns,ss)$$

IV. MODEL ESTIMATION

This section illustrates the estimation method for the probabilities values that appear in the above formulas by using a knowledge of the world.

A set of 14 text files were selected from newspapers, novels, textbooks, giving a total of 4981 words occurencies and 2073 different lexical items. These files were phonetically transcribed and then syllabized in agreement to the rules described in [4] by a computer program, thus obtaining a set of phonetic word sequences (referred as DB1) whose composition reflects the function and content words frequency of occurence which is proper of the language.

The WSM is then utilized as a parser, individuating the paths inside the model that generate the phonetic sequences belonging to the training corpus. This is a straightforward operation since every word in the lexicon corresponds to an unique path in the model; a simple count of the number of times that each transition has been crossed during the training corpus analysis allows us a maximum likelihood estimation of the probabilities values. In fact the ML estimate can be expressed by the ratio of the counts of the joint event and the conditioning one; the number of times count that a state (or a set of states) has been visited is easily derived by summation over the counts of the outgoing transitions.

As a first result, Fig.4 reports the stationary source entropies H_0 and H_1 variations while the source is being trained with the phonetic sequences. As knowledge is added, the new created transitions makes the WSM a more informative source. But larger is the data base already analysed and lesser will be the probability of observing new events; this consideration motivates the saturation effect visible in Fig.4. This entropies behaviour validates the results that will be given below as significative for our investigation purposes, even if obtained from a relatively modest data base size.

As a side experiment, a secondary data base DB2 was derived from DB1 by elimination of duplicate words, thus obtainig a list of equiprobable lexical items. This list is then used in a separete model estimation for comparison purposes.

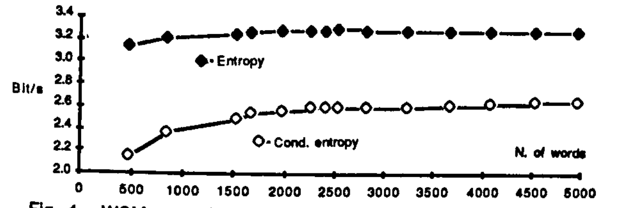


Fig. 4 - WSM entropies as function of the training corpus size

	DB1			DB2		
	a = 0	a = 1	all	a = 0	a = 1	all
H_0	3.86	2.14	3.29	4.05	2.17	3.36
H_1	3.26	1.43	2.66	3.35	1.49	2.67
I_m	0.59	0.7	0.63	0.7	0.68	0.69
I_N	0.15	0.33	0.19	0.17	0.31	0.20
P(a)	0.67	0.33		0.63	0.37	

Tab. I - Stress position dependent and global information values

V. RESULTS

Let us begin to examinate results derived from DB1 analysis. Formulas (1)-(4) were evaluated and the information matricies statistically averaged over their dimensions in order to deal with more readable series.

The first two columms of table I give the entropy and mutual information values as function of the syllable type only; by using the syllable type probability in the last row of table 1 as wheigths in a further averaging the stationary values of column three are obtained. The comparison of the entropy values reveal the syllables which follows the stress are less informative with respect to the type 0 ones (the difference is more than one bit/symbol) and exhibit a phonotactical constraints strenght nearly doubled. This fact is easily explained remembering that in italian the stress is often placed near the end of the lexical root, so type 1 syllables mainly belongs to a closed set of suffixes. Columns four to six of Table I report the values obtained by using DB2 as training corpus. The lexical items equiprobability produces only a slight entropy increase, mainly due the 0 syllable type.

A more accurate feeling on the symbols predictability changes is given in Fig. 5, where the non-stationary functions (1)-(4) expected values with respect to the syllable state number are reported in form of istograms, allowing an easier visual evaluation of the results. Black bars refer to type 0 SSM and grey bars to type 1 ones; Table I values are computed from Fig. 5 ones by means of the SSM probability wheighting reported in Fig 5.e. As an example consider the a) plot, which refers to the WSM entropy $H_0(a,ns)$. Also if the first three type 0 SSMs exhibit nearly equal entropy values, their contribution to the value reported in Tab.I are unequal because of their different frequency of occurence.

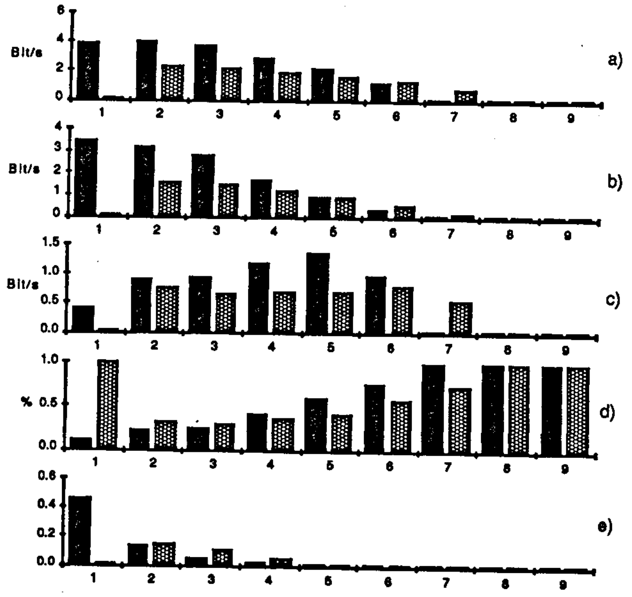


Fig. 5 - Information values as function of the syllable number and type:

a) - Entropy H_0 (a,ns)
b) - Conditional Entropy H_1 (a,ns)
c) - Average Mutual Information I_m(a,ns)
d) - Normalized Mutual Information I_N(a,ns)
e) - Syllable Probability P(a,ns)

By looking to Fig 5.a, it can be noted that the type 1 SSMs entropy is quite constant, while type 0 SSMs entropy decrease more quickly after syllable number three; about the same effect can be noted also for H_1 (Fig 5.b). This result can be explained by thinking that the statistical composition of syllables which follow the stress is nearly the same for every syllable number because it originates from the same suffixes joined with word roots of different lengths. For what regars type 0 SSMs entropy decrease, a motivation can be the word roots progressive fading in SSM statitical composition with the syllable number increase, making room for the influence of a closed set of inflectional morphemes.

Figg. 5.c and 5.d shows the average mutual information values, and it is possible to appreciate how the normalization of I_m whith respect to H_0 is important in giving the exact measure I_N of the constraints strenght variations. The neighbouring effects become strongher in a linear fashion as function of the syllable number for both the syllable types. The unitary value means an absolute predictability, and this is obvious for syllable numbers to which no word in the training corpus has given contribution!

The use of DB2 as training corpus do not give much additional information; just an entropy increase can be noted, mainly for the initial syllables whose statistic is heavily influenced by the high frequency short words.

As a final analysis it is interesting investigate on the constraining power of the SSM states as they are evidenciated in Figg. 2-3; for this purpose the expected values with respect to the syllable number of the functions (1)-(4) were calculated, obtaining the plots of Figg. 6-7 for syllable types 0 and 1 respectively.

Fig 6.a shows the values of the conditional entropy: it is possible to note an information decrease among the state numbers which are destinations of the consonants belonging to the initial consonant cluster (states number 2-4), while their constraining power shown in Fig 6.b by means of I_N(a=1,ss) is nearly equal. The state following an unstressed vowel (#5) is sligthly more informative than the one following a stressed vowel (#7), and exhibits a predictability just a little bit weaker than the latter. For what regards the after-vowel consonant cluster, its information contribute is very low. Fig 6.c reports about the probability of occupancy of the SSM states, which is in agreement with the high frequency of occurence of short syllables.

The states numbering of Fig. 7 reflects the type 1 SSM morphology of Fig. 3. The entropy and predictability distributions are given in Figg. 7.a and 7.b; the states probability values of Fig. 7.c evidenciate the great prevalence of CV syllable structure for type 1 SSMs.

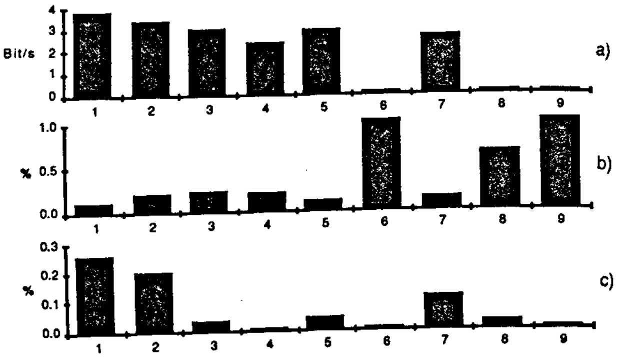


Fig 6 - Information values as function of the state number for syllable type 0
a) - Conditional entropy H_1 (0,ss)
b) - Normalized mutual information I_N (0,ss)
c) - State probability P(0,ss)

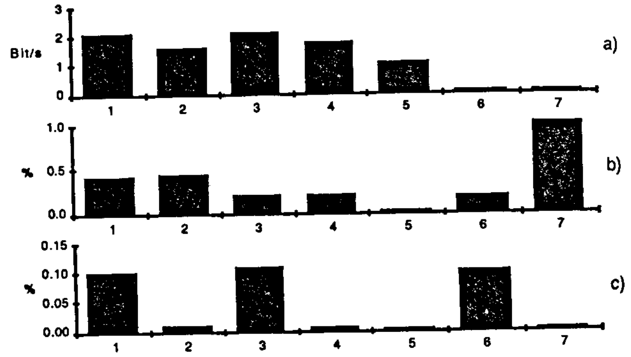


Fig 7 - Information values as function of the state number for syllable type 1
a) - Conditional entropy H_1 (1,ss)
b) - Normalized mutual information I_N (1,ss)
c) - State probability P(1,ss)

VI. CONCLUSIONS

The word and syllable structures driven non stationary statistical analysis of phonetic chains has highlighted a heavy disuniformity in the phonotactical constraints strenght as a function of the inside word syllable position. Although this effect could be foreseen from morphology, its quantification can result very useful in the area of automatic speech recognition for very large lexicon systems. Future work will be addressed towards the use of the WSM as a language model for an automatic phonetic recognizer.

REFERENCES

[1] - L.R. Bahl, F. Jelinek, R.L .Mercer, "A Maximum Likelihood Approach to Continuos Speech Recognition", IEEE Trans. PAMI-5, No.5, March 1983

[2] - S.E.Levinson, "Structural Methods in Automatic Speech Recognition", Proc. IEEE, Nov. 1985

[3] - J.P.Tubach, L.J.Boe, "Quantitative Knowkedge on Word Structure, from a Phonetic Corpus, with Applications to Large Vocabularies Recognition Systems", Proc. ICASSP 86, April 1986, Tokyo

[4] - R.A.Hall Jr., "La Struttura dell'Italiano", Armando Armando Ed., 1971 Roma

THE FEATURE [FLAT] IN CROSS-LANGUAGE PERCEPTION

MEL GREENLEE CHARLES A. FERGUSON DOROTHY HUNTINGTON
Linguistics Department, Stanford University, Stanford, California

JOHN J. OHALA DEBORAH FEDER
Phonology Laboratory, Linguistics Department, University of California
Berkeley, California

ABSTRACT

The legitimacy of [+/- flat] has been repeatedly discussed (Jakobson et al, 1969; McCawley, 1972; J. Ohala, 1985). One argument offered in justification consists of unified acoustic-perceptual correlates in spite of distinctive articulatory characteristics. Following an earlier suggestion and an empirical test, this paper examines the extent to which [+/- flat] of one language is heard as [+/- flat] in another language when the articulatory correlates in the two languages are different. Languages chosen were Arabic (pharyngealization) and Bengali (retroflexion); each language has a traditional orthography which indicates the [+/- flat] distinction. Speakers of the two languages listened to both Arabic and Bengali nonce words contrasting [+/- flat] consonants between vowels and transcribed these according to the orthography of their language. Subjects accurately perceived [+/- flat] in their own respective languages, but [+ flat] consonants of one language were rarely heard as [+ flat] in the other. Also, Bengali listeners often identified Arabic [-flat] as the corresponding Bengali [+flat] consonants. Thus, the unity of perceptual correlates for [+/- flat] appears to be questionable.

INTRODUCTION

Among the set of distinctive features proposed in Preliminaries [1], was the distinction flat verses plain: [+flat] segments manifested "a downward shift of a set of formants or even of all the formants in the spectrum," as compared to plain segments. The proposed feature [+flat] encompassed labialization, pharyngealization, and retroflexion, which were held to be similar in acoustic/auditory effect and never phonologically contrastive in the same language. The utility of the proposed feature has been challenged, both on formal [2], and substantive grounds [3]. McCawley argued against the feature as requiring as many descriptive and interpretive levels as taxonomic phonemics, while other authors noted that all three articulatory manifestations of [+ flat] are not in strict complementary distribution [4]. In Chomsky & Halle's feature set, [+/- flat] was discarded. J. Ohala [5], however, noted several reasons for the usefulness of the feature, including distributional similarities, effects on neighboring segments, and phenomena of borrowing and sound change.

Ferguson [6] proposed an empirical test of the perceptual unity of [+ flat] consonants, investigating the perceptual judgements of Arabic pharyngealized consonants and South Asian retroflex consonants, since speakers were readily available and the respective orthographies afforded representation for the hypothesized [+/- flat] distinction.

Feder [7] conducted a cross-linguistic perception test of [+/- flat] using Arabic and Hindi words and niñe Arab listeners. Stimulus words were recorded by a number of native speakers of each language. In order to reduce the influence of Arabic vowels which co-vary with the [+/- flat] consonant distinction, all CV stimulus words were edited to include only a very short /i/ or /a/. Arab listeners usually responded correctly on Arabic words, but they rarely heard Hindi retroflex stops as [+ flat], although more such identifications occurred when the following vowel was /a/ (16%) than when the syllable contained a high vowel (2%). Feder concluded that a more refined cross-linguistic test of [+/- flat] was needed.

METHODS

The present experiment was designed to further test the perceptual unity of the feature [flat] with speakers and listeners from the same language areas tapped by Feder. However, there were several methodological differences.

First, rather than reducing influence of the vowels surrounding [+/- flat] consonants, we sought to include as much naturally-occurring information as possible. It has often been noted that the auditory effects of retroflexion are more striking on the vowel preceding [+ flat] consonants, while for pharyngealized consonants, although both preceding and following vowels may be affected, the more prominent auditory effects typically occur on the following vowel. In order to give listeners from both language groups equal opportunity to perceive these effects, the [+/- flat] contrast was placed in a medial position between two similar vowels in a CV_V format. Duration of the adjacent vowels was not manipulated.

Second, we used only a single, male speaker of each language in recording the stimulus tape, but relied on a number of native speakers of each language as listeners (11 Bengalis and 13 Arabs), who transcribed recorded tokens in their entirety according to the conventions of their respective orthographies. By asking listeners to transcribe the whole "word", we hoped to obtain information about potential vowel effects of the [+/- flat] distinction, as well as data on consonant perception per se. Finally, in the present experiement, all stimulus items were nonce words in both languages. By excluding real words, we intended to avoid potential semantic effects and focus listeners' attention on the phonetic correlates of [+/- flat].

Items contrasting [+/- flat] consonants

consisted of three minimal pairs in each language (/kiṭi/-/kiti/, /niḍi/-/nidi/, and /šaṭa/-šata/) and one non-minimal pairing, because of real-word constraints, Arabic /naḍa/-/rada/ and Bengali /kada/-/rada/. The stimuli were created by digitizing and splicing the target "words" which were spoken in a frame sentence. Arabic tokens were recorded by a male speaker of the urban Palestinian dialect, while the Bengali speaker was a native of Calcutta, India. Each token was repeated ten times, then randomized onto a stimulus tape containing items in both languages.

Spectrographic analysis of the stimuli showed that the acoustic correlates of [+ flat] were generally more pronounced in the Arabic than in the Bengali tokens. This was particularly true when Arabic [flat] consonants occurred between two high vowels. For example, Arabic /niḍi/ showed a steep rise in the second format of both the first and second instances of /i/, a consequence of pharyngealization described by earlier acoustic analyses [8,9]. Bengali /niḍi/ differed from its plain counterpart less dramatically in the location of vowel formants, but manifested a rather salient difference in locus and amplitude of the release burst for /ḍ/ in comparison to /d/.

Listeners were not told that they would be hearing two languages, but were instructed to write as closely as possible, in their own language, the speech on tape. They were told that they would hear possible but non-occurring words in their own language, and that the recorded words had been processed by a computer.

RESULTS

Table 1 shows each group of listeners' responses to flat and plain (dental) stops in either language. In tallying responses, only the value of the feature [+/- flat] was considered, disregarding other misperceptions (e.g., of consonant voicing). For both [+ flat] and [- flat] consonants, listeners were much more accurate when judging their own language. Each group correctly perceived [+ flat] consonants in their own language over 90% of the time. Yet only rarely were [+ flat] stops of one language identified as the corresponding [+ flat] stops in the other. Arabs heard Bengali retroflex stops as pharyngealized Arabic /ṭ/ or /ḍ/ less than 8% of the time. As Table 1 shows, except for one subject's unscorable responses, Bengalis never heard the Arabic emphatic /ṭ/ or /ḍ/ as retroflex.

When listening to [- flat] stops in their own language, neither group of subjects erred more than 3% of the time. Arab listeners also had a relatively low error rate on Bengali plain stops, misidentifying them as [+ flat] only 10% of the time. On the other hand, Bengalis perceived nearly half of the Arabic [- flat] stops as retroflex. A t-test revealed that Bengali listeners had a significantly higher rate of false-positive responses (i.e., misidentification of [- flat] consonants as [+ flat]) in this cross-linguistic task (t=5.17(df 22), p< 0.001, 2-tailed).

One possible source of the Bengali listeners' bias for hearing plain stops as retroflex might lie in the pronunciation of Arabic by the speaker we recorded. It may be that he sometimes pronounced the Arabic plain stops with an alveolar place of articulation, since no distinction between dental and alveolar stops exists in Arabic, allowing for free-variation. If our speaker pronounced the Arabic plain stops as alveolar, then Bengali listeners' frequent misidentification of these stops as retroflex is not surprising, given that Hindi speakers perceive American English alveolars as retroflex 91% of the time [10].

We also examined the influence of vowel context on listeners' errors. Results are shown in Table 2, which lists percent errors for each set of stops and each group of listeners. As can be seen, for [+ flat] consonants, neither group's errors were much affected by the surrounding vowels. But, for [-flat] consonants, each group of listeners was affected by vowel context, but only when perceiving their own language. As in Feder's earlier experiment, Arabs more often mistook Bengali [-flat] stops for their pharyngealized /ṭ/ or /ḍ/ when the Bengali plain stops were presented between low vowels (t=3.72(df 12), p< 0.01, 2-tailed). Since in Arabic, /a/ is fronted to [æ] in the context of [- flat] stops, Arab listeners, hearing a [- flat] consonant surrounded by low back vowels, transcribed the consonant as [+ flat].

Bengali listeners, in contrast, made more errors in judging Arabic [- flat] stops when these were surrounded by high vowels (t=4.17 (df 10), p< 0.001, 2-tailed). Thus, Arabic plain stops were most often heard as retroflex when in the context of /i/. This pattern of errors is somewhat surprising, given the more pronounced acoustic effects of retroflexion on the high second formant of /i/ rather than on the already low formant structure of /a/. At present, we have no explanation for this paradoxical result.

DISCUSSION

Our findings have demonstrated that while both Arabic and Bengali speakers accurately perceive the phonological feature [+ flat] in their own languages, they rarely identify [+ flat] in the other language. This result argues against the proposed unitary acoustic/perceptual correlates of [+/- flat] as a phonological feature.

In particular, the large percentage of false-positive responses by Bengali listeners would seem to challenge the distinctiveness of the proposed acoustic correlates of [+ flat]. Since Bengali listeners most often heard Arabic plain stops as retroflex when in the context of a high vowel, one might infer that for a Bengali listener, the proposed acoustic correlate of lowered formant structure is not a necessary cue for judging a consonant as [+ flat] (Cf. [10]).

While our findings as a whole did not confirm the proposed cross-linguistic identification, we are not yet ready to discard the notion of [+ flat] as a class of perceptually similar sounds. Our reservation is based not only on the conflict of these experimental results with earlier explanations of sound change [5], but also on limitations in the scope of the present study. We tested only a small set of stimuli in two languages, which included other phonetic distinctions as well as the consonant contrasts. Perhaps a more thoroughly controlled test, including other [+ flat]

segments (such as labialized consonants) and minimizing co-varying phenomena would yield a more unified picture of flat perception, with greater cross-linguistic agreement.

REFERENCES

[1] Jakobson, R., G.M. Fant, & M. Halle. 1969. Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, MA: MIT Press.

[2] McCawley, J.D. 1972. The role of a phonological feature system in a theory of language. In V.B. Makkai (Ed.) Phonological theory: Evolution and current practice, 522-8. Lake Bluff, IL: Jupiter Press.

[3] Chomsky, N. & M. Halle. 1968. The sound pattern of English. New York: Harper & Row.

[4] Catford, J. 1977. Fundamental problems in phonetics. Bloomington, IN: Indiana U. Press.

[5] Ohala, J.J. 1985. Around flat. In V. Fromkin (Ed.) Phonetic linguistics, 223-241. New York: Academic Press.

[6] Ferguson, C.A. 1966. Linguistic theory as behavioral theory. In E.C. Carterette (Ed.) Brain function. Berkeley, CA: U. California Press.

[7] Feder, D. 1984. Pharyngealization, retroflexion, and the distinctive feature flat: A perceptual experiment. Ms. U. California-Berkeley Phonology Laboratory.

[8] Obrecht, D.H. 1968. Effects of the second formant on the perception of velarization consonants in Arabic. The Hague: Mouton.

[9] Al-Ani, S. 1970. Arabic phonology: An acoustical and physiological investigation. The Hague: Mouton.

[10] Ohala, M. 1972. Conflicting expectations for the direction of sound change. POLA Reports, Series 2, no. 16: 58-62. U. California-Berkeley Phonology Laboratory.

Table 1: Perceptual Confusions According to Native Language of Listeners

	STIMULUS			
	Arabic [+flat]	Bengali [+flat]	Arabic [-flat]	Bengali [-flat]
RESPONSE				
1. Arabic-speaking listeners (2076 responses)				
[+flat]	90.2%	7.9%	2.3%	9.5%
2. Bengali-speaking listeners (1756 responses)				
[+flat]	0.0%[a]	92.0%	47.3%	2.3%

[a] One Bengali subject consistently added an extra syllable for 3/4 of the Arabic [+ flat] stimuli. These unscorable responses are omitted from the table.

Table 2: Percent Listener Errors in Identification of [+/- flat] Consonants According to Vowel Context

	Arabic [+flat]		Bengali [+flat]		Arabic [-flat]		Bengali [-flat]	
	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/
1. Arabic-speaking listeners								
	0.0	19.6	99.6	84.6	0.4	4.2	0.4	18.5
2. Bengali-speaking listeners								
	89.0	95.4	7.3	8.6	62.7	31.8	4.1	0.4

UNIVERSELLE WECHSELBEZIEHUNGEN ZWISCHEN DEN EINHEITEN MIT SONANTISCHEN MERKMALEN

IRINA MELIKISCHVILI

Institut für Orientalistik
Tbilissi, Georgien, UdSSR, 380062

ZUSAMMENFASSUNG

In Beziehung zu entsprechenden Phonemen (vom Typ w,y,n,r,?/q,h̲,h) haben die resonanten differenziellen Merkmale einen sekundären, markierten Charakter. Diese Beziehungen lassen sich in implikative Universalien formulieren.

Die Untersuchung der Wechselbeziehungen von sonoren Phonemen und entsprechenden Resonanzmerkmalen (Bezüglich des Zusammenhanges von verschiedenen Realisationen von sonantischen Einheiten vgl. [1]) offenbart eine besondere Rolle von Sonanten im Aufbau der Phonemsystemen. Die bekannte Universalie von R. Jakobson: Sprachen, die die aspirierte - nichtaspirierte Phoneme unterscheiden, enthalten auch das Phonem /h/ [2], hat keinen isolierten Charakter und stellt einen Sonderfall der Realisation eines umfassenden Prinzips dar. Dieses Prinzip bestimmt noch eine Reihe von anderen Sonderuniversalien, die folgenderweise formuliert werden können:

1. Die Sprache, die labialisierte. Phoneme besitzt, enthält auch das bilabiale sonore Phonem vom Typ /w/. Wir wollen betonen, dass die Voraussetzung der Labialisation der labiale Sonant, nicht aber das dento-labiale Phonem vom Typ /v/ bildet.

2. Die Sprache, die palatalisierte Phoneme besitzt, enthält auch das palatale sonore Phonem vom Typ /y/.

3. Die Sprache, die nasale Phoneme besitzt, enthält auch das sonore nasale Phonem vom Typ /n/ [3] .

4. Die Sprache, die retroflexe Phoneme besitzt, enthält auch den sonoren Vibrant vom Typ /r/. Auf die Möglichkeit dieser Generalisation hat uns D.I. Edelmann hingewiesen. Wir denken, dass das Vorhandensein des retroflexen /ṛ/ in den Sprachen mit der Korrelation der Retroflexion (wie z. B. im Beludschischen, Paratschi, Ormuri, Bengali) Spricht nicht

gegen diese Generalisation. Allem Anschein nach kann das Resonanzmerkmal auch mit entsprechendem Grundphonem kombinieren. Im Jakutischen, zum Beispiel, existiert das palatalisierte Phonem /y'/ neben dem nichtpalatalisierten /y/.

Vermutlich kann das Phonem /l/ die Voraussetzung der Velarisation sein, obwohl hier der Zusammenhang nicht so anschaulich ausgeprägt ist.

Wir nehmen den Gesichtspunkt an, der die laryngale und pharyngale Phoneme zu der Sonantenklasse zählt und die entsprechenden Merkmale der Pharyngalisation, Aspiration, Glottalisation als ihre Metamorphismen betrachtet. Im faucalem Raum, der die pharyngalen und laryngalen Höhlen vereinigt, kann man drei Typen der Artikulation unterscheiden: Verschluss, Verengung und Ausdehnung [4] . Es gibt Sprachen mit sehr reichem System faucaler Konsonanten, wo diese Artikulationsarten in allen drei faucalen Hauptzonen auftreten: das sind oberpharyngale, unterpharyngale und laryngale Phoneme mit Verschluss, Verengung und Ausdehnung. Als differenzielle Merkmale aber können in einer Sprache zugleich höchstmöglich nur drei faucale Merkmale nebeneinander bestehen. Die Merkmale der Verengung: Uvularisation, Pharyngalisation, Emphatisation erscheinen in einer Sprache niemals gleichzeitig und so können wir sie in einem Merkmal vereinigen. Die Bezeichnung "Emphatisation" betrachten wir als Beste für dieses Merkmal, weil sie keinen Hinweis auf die Stelle der Artikulation enthält. So können wir nach der Art der Artikulation drei faucale Resonanzmerkmale unterscheiden: Glottalisation (faucaler Verschluss), Aspiration (faucale Ausdehnung), Emphatisation (faucale Verengung). Die entsprechenden faucalen sonoren Phoneme bezeichnen wir als faucale Phoneme mit Verschluss, Ausdehnung und Verengung ohne Artikulationsstelle zu präzisieren. Bei Solcher Lösung können wir mehrere Schwierigkeiten vermeiden, die bei der Präzisierung der Artikulationsstelle entstehen. Das Phonem /h/ wird in einigen Sprachen als laryngaler, in anderen als pharyngaler Laut bezeichnet (z. B. in mehreren iranischen Sprachen wird /h/ als pharyngal qualifiziert [5]).

Die meisten Sprachen mit dem Merkmal der Glottalisation haben den laryngalen Verschlusslaut /?/, aber einige wie z.B. das Georgische, Swanische, Ossetische enthalten keinen laryngalen Verschluss, haben aber den pharyngalen Verschlusslaut / q̣/, den wir zum Vertreter des faucalen Verschlusses zählen. Also können wir bezüglich des faucalen Raumes folgende Generalisationen formulieren:

5. Die Sprache, die aspirierte Phoneme besitzt, enthält auch das Phonem vom Typ /h/ (mit faucaler Ausdehnung - Universalie von R. Jakobson).

6. Die Sprache, die glottalisierte Phoneme besitzt, enthält auch das Phonem vom Typ /ʔ/ oder /q̣/ (mit faucalem Verchluss).

7. Die Sprache, die emphatisierte oder pharingalisierte Phoneme besitzt, enthält auch das Phonem vom Typ /h̲/ (mit faucaler Verengung).

Alle diese Universalien sind vom selben Typus und stellen die Realisation eines allgemeinen Prinzips dar: Resonante differenzielle Merkmale sind sekundär, markiert in Bezug zu entsprechenden linearen sonoren Phonemen. Sonantische Phoneme stellen eine Quelle von vielen differenziellen Merkmalen dar, die den minimalen Konsonantismus und Vokalismus modifizieren und erweitern. In der Tat haben alle Merkmale, die mit dem minimalen Konsonantismus und Vokalismus kombinieren, den sonantischen Ursprung. Es ist erwähnungswert, dass diese Einteilung der Merkmale- in primäre (die am Aufbau des Minimalsystems teilnehmen) und sekundäre oder sonantische ist im Einklang mit den Ergebnissen der experimentellen Untersuchung der Lautperzeption: Auf der sensoren Ebene werden die primären Merkmale nicht wahrgenommen, in besonderen Verhältnissen aber können die sonantischen Merkmale wahrgenommen werden [6].

Also hat die Ganzheit, die zusammengesetzte Einheit in der Wechselbeziehung von sonoren Phonemen und entsprechenden Merkmalen einen primären Charakter. Das muss wiederum die Aüsserung einer allgemeinen Gesetzmässigkeit sein, die sich auf allen Ebenen der Sprache offenbart. So haben, zum Beispiel, die morphologischen Merkmale stäts lexikalische Entsprechungen. Darauf beruht die kontextuelle semantische Analyse der grammatischen Kategorien - da wird mittels der entsprechenden lexikalischen Umgebung die Bedeutung der grammatischen Kategorien festgestellt. Man kann die Aüsserung dieser Gesetzmässigkeit auch in anderen Gebienten der Sprache suchen.

Universalien von diesem Typus haben auch diachronische Implikationen. Ohne der Rekonstruktion der entsprechenden linearen Sonanten können in den Sprachen keine Resonanzmerkmale rekonstruiert werden. Rekonstruktion der Merkmale der Aspiration und Glottalisation, zum Beispiel, fordert die Rekonstruktion der entspechenden faucalen Phoneme. So bilden diese Generalisationen noch zusätzliche Argumente für die Rekonstruktion der faucalen Phoneme im Gemeinindoeuropäischen.

Literaturverzeichnis

[1] D.I. Edelmann, J.S. Stepanov, Beschreibung des Ausdrucksplanes der Sprache auf semiologischer Grundlage, Prinzipien der Sprachbeschreibung, Moskau, 1976, S. 267.

[2] R. Jakobson, Typological studies and their contribution to historical comparative linguistics, Proceedings of the Eigth international congress of linguists, Oslo, 1958, S. 17-25.

[3] C.A. Ferguson, Assumptions about nasals, a sample study in phonological universals, Universals of Language, ed. J.H. Greenberg, Cambridge, 1963, S. 53-60.

[4] S.W. Kodsasov, Phonetik der artschinischen Sprache, Versuch der strukturellen Beschreibung der artschinischen Sprache, Moskau, 1977.

[5] L.A. Pireiko, D.I. Edelmann,Die nordwestliche Gruppe der neuiranischen Sprachen; W.A. Efimov, D.I. Edelmann, Die östliche Gruppe der neuiranischen Sprachen, Sprachen von Asien und Afrika, II, Moskau, 1978, S. 112 u. 203.

[6] Z.N. Japaridse, On the perseption of distinctive features, Abstracts of the 10th international congress of phonetic sciences, Utrecht, 1983.

Модификации сонорных согласных, связанные с различиями фонетических систем по признаку "вокальность — консонантность"

С.В.Бромлей

Институт русского языка
АН СССР

Звуки класса сонорных дают специфическую информацию об уровне "вокальности — консонантности" сопоставляемых по этому признаку систем, особенно существенную для микротипологии.

При изучении различия фонетических систем по признаку "вокальность — консонантность" в говорах русского языка, где по этому признаку·противопоставлены русские центральные (Ц.), более консонантные, и периферийные (П.), более вокальные говоры [1], обнаружилось, что звуки класса сонорных способны давать о признаке "вокальность — консонантность" специфическую информацию [2]. Она оставалась неучтенной при обычном дихотомическом делении звуков на два класса — гласные и согласные — в работах, посвященных сравнительно-типологическому сопоставлению систем по этому признаку , ср.[3; 4; 5].

Специфичность этой информации состоит в том, что звуки класса сонорных, в отличие от гласных и шумных согласных, варьируются не столько по числу единиц (фонем), различающихся в сопоставляемых системах, сколько по соотношению у каждой из фонем класса сонорных голоса (тонального элемента) и шума, — т.е. не по количеству, а по качеству. Таким образом, варьируется степень "сонорности сонорных", или потенциал их звучности, отражая различия систем по признаку "вокальность — консонантность".

Разность потенциала звучности сонорных не в малой степени определяет общий характер языка как более звучного или менее звучного в чисто эмпирических оценках при аудитивном его восприятии. На уровне системы различия в степени вокализованности сонорных эксплицируются в их позиционном поведении. В более вокальных системах в состав позиционных модификаций сонорных фонем входят звуки более вокального характера: слоговые сонорные, неслоговые гласные и даже слоговые гласные; в других, более консонантных, — в их состав входят шумные согласные и нуль звука. Каждая отдельная сонорная фонема имеет свой специфический набор модификаций, которые манифестируют ее вокально-консонантный потенциал.

Губной спирант в более консонантных (Ц.) системах реализуется лабиодентальным [в], проявляющим себя в одних положениях как шумный согласный (ср. на конце слова и перед глухими наличие парного глухого [ф]), в других (перед гласными и сонорными) — как сонорный, перед которым звонкие и глухие шумные различаются: ср. [с во]до́й — бе[з во]ды́, [с вр]ачо́м — бе[з вр]ача́. В более вокальных (П.) системах губной спирант представлен сильно вокализованным билабиальным сонорным [w] при неслоговом [ў] в конце слова и слога (ро[ў], вере́[ў]ка), а в начале слова перед согласным эта фонема реализуется слоговым [у], ср. [у]ну́к, [у]дова́. В соответствии с [w] может появляться [в] лабиодентальный, но он проявляет себя здесь как сонорный в любых условиях, не оглушаясь перед глухими согласными, ср. ла́[вк]а, [вп]ере́д...[6].

Твердому смычно-проходному боковому сонорному [л] в более вокальных (П.) системах в позиции конца слова и слога соответствует неслоговой гласный [ў] или [w] — звук более сонорный, чем [л], ср. да[ў], па́[ў]ка или да[w], па́[w]ка. Эти же замены реже можно встретить и в начале слова перед гласными ([ўо́]шадь, [wо́]шадь).

Фонема <j> . В некоторых более консонантных (Ц.) говорах [j] может выступать на месте [γ'] перед гласными переднего ряда ([jи́]бель, [jе́]на = Гена). Совпадение в этом положении [j] и [γ'] свидетельствует о значительной силе трения как компоненте артикуляции [j] , что приближает его к шумным согласным, ср. различение перед таким [j] звонкости-глухости шумных : бе[з jи́]ри, но с[jи́]рей (без гири, с гирей). В этих системах [j] выступает в ряду других шумных мягких среднеязычных согласных: [г'']—[к'']—[j] —[х'']. Звук [х'']— "безголосое j" [7]— звучит во многих Ц. русских говорах, а также в речи части носителей литературного произношения в конце слова и фразы: откро́[х''] , скоре́[х''] . В этом проявляется такая степень шумности [j], которая выражается в оглушении. В говорах же более вокальных (П.) фонема <j> реализуется в основном в виде неслогового гласного [й]. Слабая выраженность консонантных свойств приводит в части севернорусских П. говоров к его утрате в интервокальном положении (ду́м[аэ]т, бо́льш[ы́э]...).

Известна способность сонорных в пределах одной системы менять свое качество в довольно широком диапазоне. Это объясняется тем, что, обладая нефонологической звонкостью, сонорные могут вести себя относительно этого признака довольно свободно, не совпадая при этом с другими звуками [8]. Существенно, однако, что рамки этой свободы различны для Ц. и П. говоров. В системах более консонантного типа (Ц.) сонорные в соседстве с шумными теряют свою звучность, спирантизируются; в позиции конца слова после согласного может происходить полная утрата их, ср. ру[п'] (рубль), жи[с'] (жизнь). В более вокальных (П.) системах модель их поведения другая. Здесь имеет место усиление звучности, что приводит к развитию побочной слоговости, ср. ру́[бл'^{ъ}]:ру́[бъл'] :ру́[б^{ъ}л']; жи́[з'^{ъ}н']: жи́[з'ьн'], а также примеры типа [ал'н]у́ : [ил'н]у́ (льну); [арж]и́ : [ирж]и́ (ржи), широко известные в П. западных говорах.

Для более вокальных (П.) говоров характерны разнообразные преобразования, связанные с сонорными согласными:

1) Специфическое поведение сонорных в интервокальном положении: чередование их с нулем звука. Ср. выше об утрате [й] в этой позиции. Широко представлено также отсутствие интервокального звука [w] ([w']), ср. кор[о́у]шка, д[е́у]шка, сам[оа́]р, пр[а́и]льно, де́р[ео]. Эллипс согласных в П. говорах захватывает и не сонорные согласные, но сонорные в этом процессе, безусловно, на первом месте. Причину этого следует видеть в слабой выраженности здесь у сонорных их консонантных свойств.

2) В сочетании шумных с сонорными и сочетаниях двух сонорных широко представлены результаты процессов ассимиляции, ср. долгие сонорные на месте сочетания звонких зубных с сонорными, объединенных общим местом образования: [дн]>[нн] (ла́[нн]о), [бм]>[мм] (о[мм]а́н); разнонаправленные преобразования сочетания [л'н'] (ср. примеры типа [л'л']яно́й, пра́[н'н']ик; а также бо́[л'л']о, бо́[н'н']о вм. бо́льно (с предшествующей ассимиляцией по мягкости). Во всех подобных случаях активизатором процессов являются сонорные.

3) Широкая взаимная мена сонорных. Обычно она осуществляется в соседстве с сонорными и имеет определенную упорядоченность. Чаще всего наблюдается взаимная

мена сонорных [н’], [л’], [j]. Примеры такой мены перед [н’] многочисленны, ср. пе́се[л’]ник, моше́[л’]ник, племя́[л’]ник , пра́[й]ник, шве́[л’]ня, са́[й]ник... Эти же сонорные могут заменять друг друга между гласными и на конце слов, ср. рассто[л’]я́ние, ма́[н’]енькая, карто́фе[й], горноста́[н’], при́ста[л’]; только в этих положениях осуществляется мена [р’] и [л’]: косты́[р’], моты́[р’], ва́[л’]ежки, а также мена [j] и [w’]: сыро[w’]ежки, кро́[j]. Реже — в соседстве с несонорными, ср. кры́[мк]а, ко[с’л’]е́вище, брю́[кл]а, [жв]и́твина.

Явления 2) и 3) групп также непосредственно связаны с высокой степенью вокализованности сонорных. Известно, что сонорный перед следующим сонорным произносится звучно, так как не теряет в этой позиции своей связи с центром слога — гласным: при произнесении слога он как бы вливается в следующий сонант. В диалектных системах П. типа, где сонорные обладают значительным потенциалом звучности и связанной с ним длительностью тонального элемента [9], легко происходит объединение двух сонорных в одном звучании с утратой специфических тембровых характеристик одного из сонорных, составляющих сочетания. Те же качества сонорных — сильная вокализованность и длительность вокального элемента при слабой выраженности их тембровой специфики — способствуют их смешению в аудитивном восприятии, что в свою очередь объясняет и их высокую вариабельность в речи. Все эти особенности сильно вокализованных сонорных стоят в прямой связи с двуединой вокально-консонантной природой сонорных: при сильной вокализованности их вокальная общность берет верх над консонантной спецификой, связанной с характером преграды.

4) В сочетаниях с последующими твердыми зубными, губными и задненебными на месте мягких [л’], [р’], [н’] выступают твердые [л], [р], [н], ср. бо́[лн]о , бо́[лш]е..., пе́[рв]ый, ве́[рб]а, ве[рх], четве[рг]а́..., ме́[нш]е, ра́[нш]е... Отвердение сонорных — также свидетельство их сильной вокализованности, поскольку установлено, что чем звучнее сонорный, тем легче он утрачивает мягкость [IO; 6].

Соотношение голоса и шума в звуковых характеристиках сонорных, совмещающих эти признаки, является параметром очень мобильным. Именно поэтому сонорные способны тонко реагировать на уровень вокальности-консонантности фонетической системы. В системах более вокальных, где представлено большее число гласных фонем и они проявляют бо́льшую фонологическую независимость, сонорные приближаются по своим свойствам к чистым гласным; в системах же более консонантных, с бо́льшим числом согласных фонем и более развитой категорией твердости-мягкости, сонорные, наоборот, по своим признакам сближаются с согласными. Кроме того, разное соотношение голоса и шума у разных звуков этого класса достаточно индивидуально проявляется у разных сонорных в рамках их общей повышенной вокальности в одних системах или, наоборот, пониженной — в других. Так, при общей значительной вокализованности звуков, представляющих фонемы <л> и <в> в П. говорах, высокое сонорное качество <л> проявляется в большей степени на северо-востоке, тогда как наибольшая вокализованность <в> характерна преимущественно для говоров юго-запада. Актуализация тех же свойств <j> в этих же группах П. говоров проявляется в различных формах ее поведения, что, по-видимому, также отражает не одинаковую степень их выраженности. Индивидуальность проявления свойств повышенной вокальности приводит к неполному совпадению по говорам порядка следования сонорных по признаку нарастания (либо убывания) звучности. С этой точки зрения различия в реализации сонорных в разных системах требуют дальнейшего пристального изучения.

Учитывая всю совокупность приведенных данных, следует считать целесообразным при типологическом изучении языковых систем по признаку "вокальность — консонантность" особо выделять показания звуков класса сонорных. Способные в специфической форме отражать вокально-консонантный уровень фонетической системы в целом, сонорные могут достаточно отчетливо сигнализировать о различиях систем по этому признаку, даже при относительно незначительной их выраженности в пределах общей дихотомии: гласные — согласные. Поэтому показания этого класса звуков дают возможность сопоставлять близкородственные, в том числе диалектные системы, где эти типологического характера различия выражены обычно не столь отчетливо, относясь к так называемой микротипологии. Сопоставления подобного рода обычно делаются на уровне языков, с использованием для этой цели литературных систем. Но эти системы не всегда бывают достаточно представительными для языка в целом, мыслимого как совокупность его диалектов, поскольку отдельные диалекты могут значительно отклоняться от литературного. Так, по известной шкале вокальности-консонантности славянских языковых систем [3] русский литературный язык стоит ближе всех к наиболее консонантному — польскому. Однако степень вокальности говоров Вологодской группы северного наречия, наиболее выделяющихся по своей вокальности среди П. говоров русского языка, позволила бы расположить их в непосредственном соседстве с украинским языком, занимающим на этой шкале срединное место.

Литература.

1. Захарова К.Ф., Орлова В.Г. Диалектное членение русского языка, М.,Просвещение, I970.

2. Бромлей С.В. Различия в степени вокализова́нности сонорных и их роль в противопоставлении центральных и периферийных говоров. — В кн.: Диалектография русского языка, М.,Наука, I985.

3. Исаченко А. Опыт типологического анализа славянских языков. — Новое в лингвистике. Вып.Ⅲ, М.,I963.

4. Левков И. Насоки в развоя на фонологичните системи на езици. София, I960

5. Скаличка В. Типология славянских языков и в особенности русского. - См. **Skalicka V. Vyvoy yazyka. Soubor statí. Statni pedagogicke nakladatelstvi. Praha, 1960.**

6. Васильева А.К. О закономерностях возникновения в русских говорах юго-западной зоны спирантной пары [в] — [ф]. — В кн.: Диалектологические исследования по русскому языку. М.,Наука, I977.

7. Аванесов Р.И. О качестве задненебной фрикативной согласной перед гласными переднего ряда в русском языке. — Доклады и сообщения Института языкознания АН СССР, М.,I952, № 2.

8. Барановская С.А. О согласных фонемах русского языка, не включенных в корреляцию по глухости-звонкости и твердости-мягкости. — В кн.: Проблемы теоретической и прикладной лингвистики и обучение произношению. М.,I973.

9. Трахтеров А.Л. Основные вопросы теории слога и его определение. — ВЯ, I956, № 5.

IO. Брок О. Говоры к западу от Мосальска. Пг.,I9I6.

UNIVERSAUX DES NASALES

IRENE JGUENTI

Docteur ès Sciences,Professeur, Chef de la Chaire de Phonétique

Institut des Langues Etrangères
Tbilissi, Géorgie, URSS, 380068

RESUME

En nous fondant sur les données de notre étude expérimentale des phonèmes nasaux et de notre enquète effectuée en 1972 à Paris, nous pouvons conclure que la nasalisation en français et les tendences modernes de l'évolution continue des phonèmes nasaux concordent de façon univoque avec les universaux des nasales proposés par Ch.Ferguson et que la langue française est une des langues qui peut servir d'illustration à ces universaux.

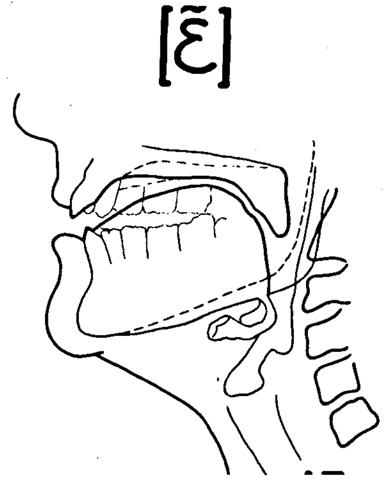
Après avoir mené à bien des recherches sur le système des voyelles et des semi-voyelles de la langue française à l'aide des méthodes de phonétique expérimentale /1/, nous avons décidé d'approfondir l'étude des phonèmes nasaux du français contemporain et d établir certains rapports entre les données de nos recherches concernant les nasaux et les Universaux des nasales de Ch.Ferguson/2/. Dans ce but, en 1972, nous avons fait enregistrer le même texte à 40 personnes d'origine française, habitant Paris, d'âge et de profession différents..
Des quinze universaux edmis par Ch.Ferguson, nous nous contenterons d'examiner les universaux des consonnes nasales primaires,"CNP" et des voyelles nasales "VN".
D'après Ch.Ferguson, la "CNP" est le phonème dont l'allophone est représenté par la nasale sonore occlusive, c'est à dire le son dont l'articulation est le résultat de l'occlusion complète dans la cavité buccal (appical, labiale), quand la voie de la cavité nasale est libre et les cordes vocales vibrent.
Les résultats de nos recherches, obtenus grace aux méthodes de l'analyse spectrale, à la radiographie et à la tomographie, nous permettent de constater le fait que les consonnes nasales primaires, "CNP" du français sont justement les memes que celles décrites par Ch.Ferguson (voir radiogrammes et tomogrammes).
Les autres universaux des consonnes nasales ne seront pas examinés dans ce rapport et nous passerons directement auxxvoyelles nasales "VN".
Ch.Ferguson considère que la voyelle nasale est un phonème dont l'allophone caractéristique provient du résultat du passage de l'air expulsé par les deux cavités - buccale et nasale avec vibration des cordes vocales. Si dans la langue il n'y a pas d'autres voyelles nasales avec des signes contraires, cette voyelle peut avoir un allophone avec occlusion dans la voie buccale ou nasale sans vibration des cordes vocales.
Il faut ajouter à la définition de l'articulation des "VN" que le mouvement de l'abaissement du voile du palais est toujours accompagné de l'élargissement de la région palatopharyngale, de l'augmentation du volume général de la chambre de résonnance et du relâchement des cordes vocales et que les fentes entre les cordes vocales s'élargissent. L'effet acoustique de ces mouvements articulatoires est l'abaissement des fréquences du spectre acoustique, conditionné par l'augmentation du volume des résonateurs.Ainsi la comparaison des universaux de Ch.Ferguson avec ceux de nos recherches nous fait considérer que le résultat de nos expériences sur les nasales coîncide avec celui de Ch.Ferguson.
D'après le Xe universel de Ch.Ferguson, si dans la langue il existe des voyelles nasales, l'existence des consonnes nasales est indispensable. La langue française est l'illustration de ce fait; dans le système de la langue française, il y a 4 voyelles nasales et 2 consonnes nasales.
Selon le XIe universel de Ch.Ferguson, le nombre des voyelles buccales ne peut être inférieur au nombre des voyelles nasales, le fait que nous rencontrons dans la langue française, où les voyelles buccales sont au nombre de 11 et les voyelles nasales de 4. De plus, après la disparition de l'opposition des phonèmes /œ̃ - ɛ̃ /, les voyelles nasales ne sont plus que 3. La déphonologisation de la nasale /œ̃/ a été démontrée par nos recherches effectuées, en 1972 sur 40 personnes habitant Paris, mais originaires du Nord de la France/3/.
Nos expériences ont montré que malgré la confusion complète des phonèmes /œ̃ - ɛ̃/, une différence minimale physique subsiste toujours, mais la fonction distinctive s'efface ou passe d'un phonème à l'autre: de /œ̃/ à /ɛ̃/. Nous considérons qu'une des causes de la déphonologisation de/œ̃/ est la surcharge des fonctions grammaticales. Nous supposons que ce fait provoque une répétition fréquente d'un même phonème dans différents contextes, ce qui entraîne leur confusion et puis la disparition complète de l'opposition. Nous avons noté la confusion des phonèmes /œ̃ - ɛ̃/ ce qui indique également une tendance générale vers une articulation ouverte, une délabialisation et une réduction du nombre des voyelles nasales.
L'essentiel du XIIe universel de Ch.Ferguson est que la fréquence d'emploi des voyelles buccales doit être toujours plus élevée que celle des voyelles nasales. La langue française illustre clairement ce fait.
L'importance du XXVe universel est que l'origine des voyelles nasales est toujours le résultat de la disparition des "CNP" (consonne nasale primaire).
Dans la langue française, les "VN", sont justement de cette origine. Au moment de la disparition de la "CNP", le trait de nasalité de "CNP" passe sur la voyelle buccale précédente et s'ajoute aux composantes du phonème buccal, ce qui nous donne un ensemble nouveau de traits distinctifs de la voyelle nasale. L'analyse acoustique des voyelles et des consonnes nasales, a montré que les indices acoustiques de "VN" et "CNP" - les formants bas, sont identiques. L'identité des indices acoustiques des "VN" et "CNP" confirme bien que les voyelles nasales proviennent du résultat de la disparition des "CNP" et de l'assimilation régressive, ce qui correspond à l'universel de Ch.Ferguson au sujet de la genèse des voyelles nasales.
Conformément aux Universaux de Ch.Ferguson, il convient de dégager que: 1) la formation des "VN" de la langue française est le résultat de la disparition des "CNP"; 2) le nombre des "VN" est inférieur au nombre des "CN"; 3) la fréquence de l'emploi des voyelles buccales est toujours plus élevée que celle des voyelles nasales, car les voyelles buccales sont beaucoup plus nombreuses que les voyelles nasales.
En nous fondant sur les données de notre étude des phonèmes nasaux, nous pouvons conclure que la nasalisation en français et les tendances modernes de l'évolution continue des phonèmes nasaux (la disparition du phonème /œ̃/) concordent de façon univoque avec les Universaux des nasales proposés par Ch.Ferguson. Ainsi la langue française est une des langues, qui peut servir d'illustration des universaux des nasales.

LITTERATURE :

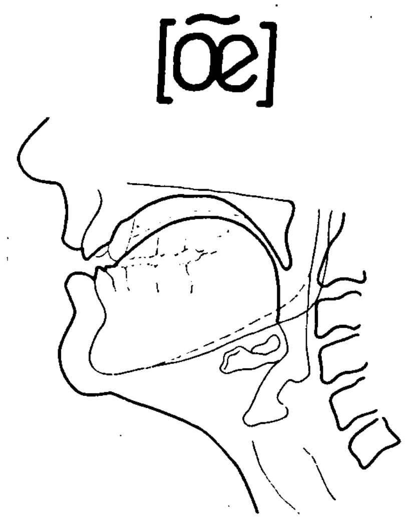
1. Jguénti I. Schéma des correspondances de niveau dans le système des voyelles et des semi-voyelles de la langue française. Abstracts of papers Eighth International Congress of phonetics. Lids. 1975.
2. Ferguson Ch. Implications abaut the nasals. Universels of language. Ed. Greenberg. Cambridge (Mass). 1963.
3. Jguénti I. Dephonologisation des voyelles /ɑ/ postérieure et /œ̃/ nasale dans le système vocalique du français contemporain. Ed. Université de Tbilissi. (174, Linguistique).1976.

ILLUSTRATIONS :

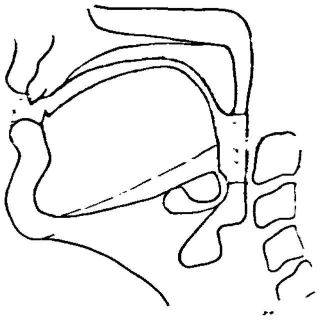
1. Radiogramme de / ɛ̃ / (Schéma)



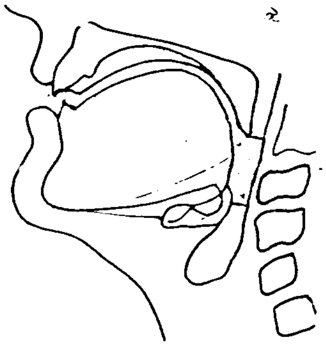
2. Radiogramme de /œ̃ / (Schéma)



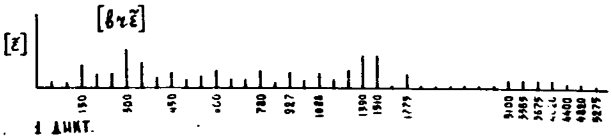
3. Radiogramme de / ɛ̃ /   (Schéma)



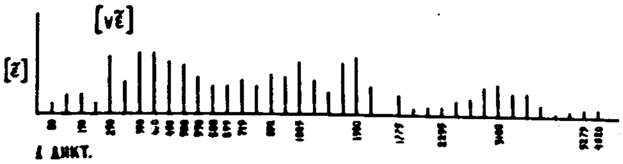
4. Radiogramme de / õe /   (Schéma)



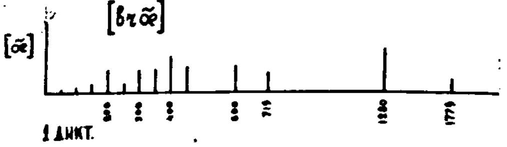
5. Spectrogramme de / ɛ̃ / dans /br ɛ̃ /
(Schéma)



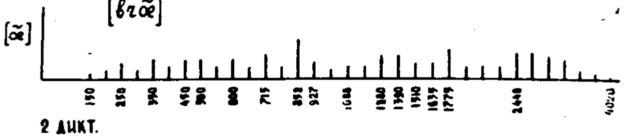
6. Spectrogramme de / ɛ̃ / dans /v ɛ̃ /
(Schéma)



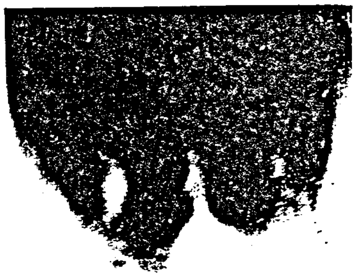
7. Spectrogramme de /õe/ dans /brõe/
(Schéma)



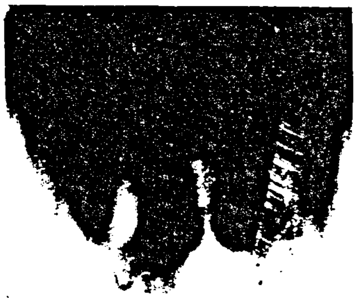
8. Spectrogramme de /õe/ dans /brõe/
(Schéma)



9. Tomogramme de / ɛ̃ /



10. Tomogramme de /õe/



ACOUSTIC STUDIES OF VOWEL REDUCTION IN SWEDISH

LENNART NORD

Department of Speech Communication and Music Acoustics
Royal Institute of Technology (KTH), Box 70014
S-100 44, Stockholm, Sweden

ABSTRACT

An acoustic analysis has been performed on a number of Swedish vowels, spoken in varying context. A complementary matching experiment using synthetic speech was made to see whether the analyzed differences in vowel quality were perceptually significant. Subjects had to adjust the vowel quality in words produced by an interactive rule-synthesis computer program. The purpose with these investigations was to describe and quantify the relations between vowel quality and influences of stress and position.

INTRODUCTION

In this study we focus on the concept of vowel reduction, here taken to mean the reduction in phonetic contrast between vowels. For a number of languages it has been found that the vowel space is reduced as the level of stress placed upon the vowels is reduced. Acoustic studies by Tiffany /1/, Shearme & Holmes /2/, Delattre /3/ Stålhammar, Karlsson & Fant /4/, Koopmans-van Beinum /5/ and others have shown that vowels in unstressed positions are displaced towards a more central (neutral) position in the vowel plane. A number of factors contribute to obscure vowel color in speech, see for example the study by Delattre (ref. /3/) who lists factors such as stress, rhythm, duration and contextual assimilation.

PRESENT STUDY

The aim with the present study was to study in detail some of the factors that contribute to vowel reduction in Swedish. We need a better understanding of these problems to improve the quality of synthetic speech, a typical impression being that synthesizers often over-articulate unstressed syllables.

TEST HYPOTHESIS

The phonetic context that will influence the formant pattern of a given vowel in a two-syllable word is: i) surrounding consonants, ii) the neutral position of the vocal tract, and iii) the second syllable, especially its vowel nucleus.

We focus on one aspect of the reduction phenomenon; is there a difference in formant pattern between two vowel samples of the same duration, one stressed and the other unstressed but of equal duration due to final lengthening? If there is a difference, could it be accounted for in terms of varying degrees of contextual influence?

Four types of two-syllable words were chosen with the following structure: The lexical stress on the initial or the final syllable, with dental consonants surrounding the analyzed vowel: CVC´S2, ´S1CVC, ´CVCS2 and S1´CVC, with V = the short Swedish vowels /a,i,e,ʉ/ and C-C = /s-l, l-s, s-s/, S1 and S2 =first and second syllable. This means that each analyzed vowel was placed either in initial or final position or in a stressed or unstressed syllable in an invariant consonantal frame. The words were read in isolation with no carrier phrase.

MATCHING EXPERIMENT

We were also interested in testing the perceptual significance of the results from the acoustic analysis by means of an interactive matching paradigm. That is, how reliably would subjects be able to adjust F1F2-values for given synthetic words in order to match to some internal criteria? A number of phonetic details can be tested with this type of interactive matching paradigm, using the specially developed rule synthesis program (Carlson & Granström, /6/). As long as the quality of the speech is acceptable to the subjects, segmental as well as suprasegmental cues can be evaluated. One could, for instance, let the subjects manipulate duration, pitch, intensity, etc. Few matching experiments of this type have been reported /7//8/ on segment duration. Öster /9/ also used the Carlson & Granström rule-synthesis program to systematically map typical features of the speech of deaf children.

EXPERIMENT: ANALYSIS OF NATURAL SPEECH

A high-quality recording was made of four Swedish male speakers from the Stockholm area reading twice a list of 38 lexically meaningful words with no carrier phrase in an anechoic chamber. The words contained the short vowels /a,i,e,ʉ/ and were constructed as described above. Formant frequencies and vowel durations were measured manually. The sample point for the formant measure was chosen by means of broad-band spectrograms in the middle of the vowel segment. The actual measurements were made from narrow-band spectral sections. In a few cases of uncertainty, the measurements were adjusted after comparison with selective inverse filtering which was used to display one single formant ringing at a time and enable measurements in the time domain. Based on systematic comparisons with measurements on synthetic vowels, we estimated the accuracy of the formant measurements to be +/- 20 Hz.

As it was impossible to always find content words of the right format, a few proper names were used. Also the demand on invariant CVC-syllable forced us to modify the consonantal frame for the different vowels, but still only use dentals. C-C were for the most words /s-l, l-s, s-s/. Accordingly, a comparison across vowels has to be taken into account the difference in consonantal coarticulations that occurred.

RESULTS AND DISCUSSION. VOWEL ANALYSIS

To find out the sensitivity of formant perturbations to changes in word material, a set of words were tried with variation of consonantal frame: /s,l,d,t/ as well as a change of vowel nuclei in the other syllable. The spread turned out to be small and the tendencies the same. Therefore, it was decided to consider the influence of the different dental consonantal frame negligible and base mean values of the entire word list material.

By placing the vowel in stressed or unstressed, initial or final CVC-syllables, we obtained vowel durations ranging from 70 to 190 msec. Comparing vowels in final unstressed syllables with vowels in initial stressed syllables, we were also able to study the influence of stress in those cases where the duration did not differ between vowels, i.e., word categories 2 and 3, which both got duration values around 125 msec.

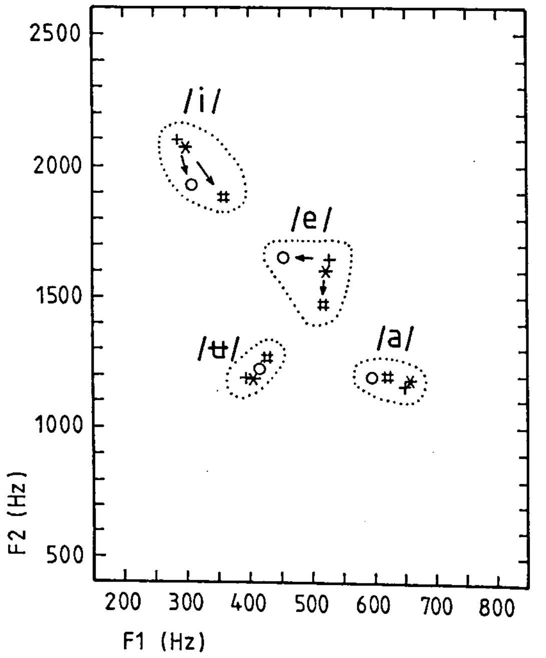


Fig. 1. Results from the analysis of real speech. Mean values of first and second formant of the short Swedish vowels /a,i,e,ʉ/. 4 male speakers, 8-40 samples/point.

	initial	final (position)
stressed	o	⌗
unstressed	+	*
(syllable)		

In Fig. 1 the mean values of first and second formants are shown. As can be seen, the unstressed initial and final vowels are displaced away from the target values of the stressed vowels. For the short /e/ and /i/, it is evident that the unstressed initial samples (o) are displaced differently compared to the unstressed final samples (⌗) as the arrows indicate. This difference could be expressed as a difference in coarticulation: the unstressed vowels coarticulate with the consonantal frame, i.e., they move towards the dental locus of approximately 350/1650 Hz for Fl/F2, while the unstressed final vowels are reduced towards a more neutral place in the vowel plane (500/1500 Hz). Formant values for the initial stressed vowels (+), that are of the same duration as the finally lengthened unstressed vowels (⌗) are thus not identical. Duration is thus not the sole determinant of the degree of reduction. These tendencies are not as evident for all the vowels, probably depending on the relative position of the target, the consonantal locus and the neutral vowel. For the /a/ vowel it is thus not possible to distinguish a perturbation caused by neutralization or by coarticulation as both effects will lower Fl and raise F2.

Another way of showing this effect for /e/ is to plot F2 as a function of vowel duration, see Fig. 2. As can be seen, the duration alone cannot predict the formant value. The stressed initial vowel (+) has approximately the same duration value as the unstressed final vowel (⌗), but speakers choose different formant values depending on the syllable context (in terms of stress and position).

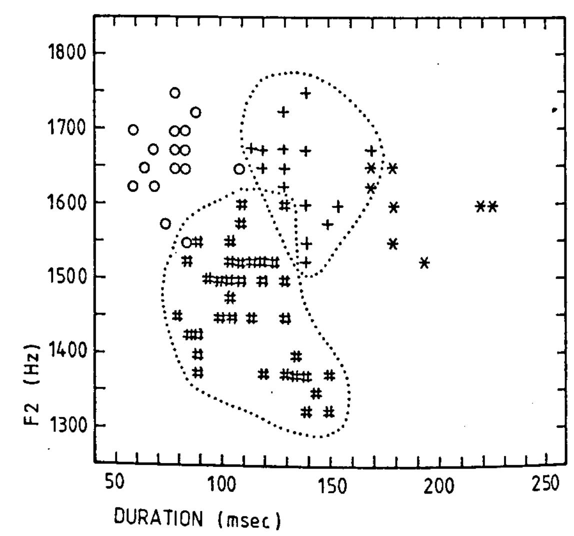


Fig. 2. Second formant value as a function of vowel duration for the short vowel /e/. Each point represents the mean value of the two readings of each speaker.

	initial	final (position)
stressed	o	⌗
unstressed	+	*
(syllable)		

The intention was to maintain an invariant C-C frame for each vowel (for the four-word categories). Due to the demands for meaningful words, the consonantal frame differs somewhat for the different vowels, but still being dental. Thus, for /a/ and /e/: /s-l/, for /i/: /l-s/ and for /ʉ/: /s-s/. This causes the differences in vowel duration. /ʉ/ and /i/ become shorter than /a/ and /e/ as they are followed by a voiceless consonant.

One might also wonder whether all speakers behave in the same way. An analysis of individual performances for the four speakers shows that the tendencies vary. Two of the speakers, one of which was used in an earlier study show clear tendencies /10/. The other two speakers perform a bit differently. One shows less Fl-perturbations and the other has small vowel areas in general. The vowel /ʉ/ differs appreciably between the speakers, probably due to sociolect differences.

MATCHING EXPERIMENT

As a complement to the acoustic analysis, an interactive rule-synthesis program (see ref. /6/) including an OVE III formant synthesizer, was used in a matching experiment. The task of the subject was to listen to synthetic words taken from the list of previously analyzed material, and by means of a joystick connected to the computer, adjust the quality of one vowel at a time in a word to make it sound as natural as possible. This interactive method has been used earlier for duration studies (ref. /8/). For this experimental set-up the x- and y-coordinates of the joystick were programmed to give the Fl- and F2-values of the synthetic vowel that was tested. The quality of the vowel could thus, instantly, be changed by moving a cursor around in a grid pattern on the terminal screen. Different scaling and offset values were used for each test word in order to avoid learning effects. A minor modification of the duration rules made the unstressed finally lengthened vowel of equal duration as the initially stressed vowel.

With this paradigm it is possible to get valuable information about the perceptual importance of acoustic parameters. Here, where one aim is to improve the synthetic speech with regard to the vowel dynamics, it is especially interesting trying to optimize the setting of the synthesis parameters directly, using the rule synthesis.

The test was run in the following way: The subject had a list of test words with one vowel in each word marked. The task was to listen to a synthetic version of one word at a time and adjust the phonetic quality of the marked vowel to sound as natural as possible. The subjects were first instructed on the task of moving the joystick and listen to the result. The test demanded some effort in terms of concentration by the subjects so it was felt necessary to limit the word list. The same type of test as for the reading list was made to evaluate the influence due to different dental C-C frames, comparing /s-l/, /l-s/ etc. The variation in matching did not change systematically with the different frames. Therefore, the mean values are pooled over the entire word list.

Preliminary tests with phonetically untrained subjects showed that they could manage the task quite well. However, in order to keep the variability as low as possible, it was decided to use phonetically non-naive subjects.

The matched formant values were automatically stored by the program and could be analyzed immediately after each session.

Eight subjects participated, among them the four speakers in the previous test. Each subject performed two matchings on a list of 25 words; one to three words for each vowel and word category. The amount of words were limited to a selected part of the reading list as the test was rather exhausting.

MATCHING EXPERIMENT. RESULTS AND DISCUSSION

The results from the matching experiment are shown in Fig. 3 where the mean values of the first and second formants are plotted. The vowel areas are smaller than for the the spoken samples, cf., Fig. 1, but the same tendencies can be seen, although to a lesser extent. Thus, the unstressed final /e/ is matched differently than the initial stressed one, the former moving towards schwa, the latter towards the dental locus.

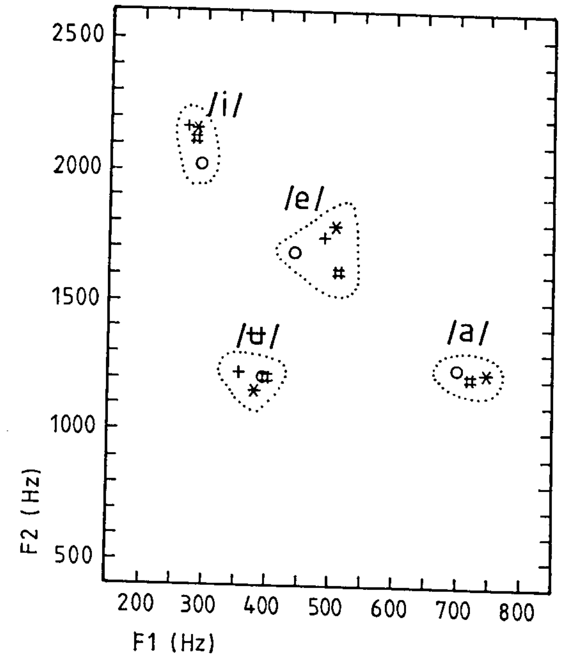


Fig. 3. Results from the matching test with synthetic words. Mean values of first and second formant of the short Swedish vowels /a,i,e,ʉ/. 8 subjects, 16-48 matchings/point.

	initial	final (position)
stressed	o	⌗
unstressed	+	*

A number of reasons can account for the discrepancy between the two tests. The synthetic quality will probably affect the matchings depending on the subject's acceptance of the voice quality. As only F1 and F2 were manipulated while higher formants were kept constant, especially the F2 of high, front vowels will differ from F2 of natural vowels. The matching session was experienced as a difficult but manageable task by the subjects. Also the spread between subjects was small. In conclusion, the method seems to be useful for this type of optimizations.

SUMMARY

The first and second formants were measured for four Swedish short vowels /a,i,e,ʉ/ in varying context, the purpose being to investigate factors of vowel reduction, such as stress, position and duration. The vowels were placed in stressed and unstressed, initial and final syllables in two-syllable words.

The result supports the findings in the previous pilot study by Nord (see ref. /10/). A tentative explanation to the distribution of formant data is that the perturbations are caused by contextual influence of surrounding consonants and in unstressed final position by a neutralization gesture, which in this word list material with no carrier phrase also belongs to the immediate context. If we do not reach for a phonological rule to explain the observations, specifically regarding the unstressed short /e/ in final syllable position, we could formulate the vowel reduction process in the following manner: irrespective of their duration, unstressed vowels coarticulate strongly with context: in non-final syllable position with surrounding phonemes and in final syllable position with a neutral position corresponding to a centralized schwa vowel. These tendencies were seen in varying degrees, probably depending on the relative locations of vowel targets, schwa and consonantal loci.

A supplementary study was performed using synthetic speech in order to evaluate the perceptual importance of formant perturbations in the realization of vowels in varying contexts. During the experiment, subjects were exposed to synthetic words of the same structure as in the previous experiment. The task was to adjust the quality of one vowel in each word by means of a joystick, connected to the rule-synthesis program, controlling the first and second formant of that particular vowel.

The results from this test were compared with the previous analysis. The same tendencies were seen, although to a lesser extent. This was probably due to the design of the experiment. As only two formants were manipulated, there were some difficulties in finding suitable vowel qualities during the matching procedure. Also the synthetic quality of the stimuli might have had some influence on the subjects' matching strategies. Although the task was rather difficult, subjects performed well with small deviations. One conclusion from this test is that the matching procedure using synthetic stimuli is an efficient way of evaluating perceptual cues and testing theories of speech dynamics.

REFERENCES

/1/ W.R. Tiffany, "Non-random sources of variation in vowel quality", J. Speech Hearing Res. 2, pp. 305-317, 1959.

/2/ J.N. Shearme, J.N. Holmes, "An experimental study of the classification of sounds in continuous speech according to their distribution in the Formant 1 - Formant 2 plane", pp. 234-240 in A. Sovijärvi & P. Aalto, eds., Proc. 4th Int.Congr.Phon.Sci., Mouton & Co., The Hague, 1962

/3/ P. Delattre, "The general phonetic characteristics of languages. An acoustic and articulatory study of vowel reduction in four languages", Final Report, Univ. of California, Santa Barbara, CA, USA. 1969

/4/ U. Stålhammar, I. Karlsson, G. Fant, "Contextual effects on vowel nuclei", STL-QPSR 4/1973, pp. 1-18 (KTH, Stockholm).

/5/ F.J. Koopmans-van Beinum, "Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions", Doct. thesis, Univ. of Amsterdam, 1980

/6/ R. Carlson, B. Granström, "A text-to-speech system based entirely on rules", pp. 686-689 in Conf. Record, 1976 IEEE Int.Conf. on Acoustics, Speech and Signal Processing, April, Philadelphia, PA, 1976

/7/ S.B. Nooteboom, "Production and perception of vowel duration. A study of durational properties of vowels in Dutch", Doct. thesis, Univ. of Utrecht, 1972.

/8/ R. Carlson, B. Granström, "Perception of segmental duration", STL-QPSR 1/1975, (KTH, Stockholm), pp. 1-16.

/9/ A-M. Öster, "The use of a synthesis-by-rule system in a study of deaf speech", STL-QPSR 1/1985 pp. 95-107 (KTH, Stockholm).

/10/ L. Nord, "Vowel reduction - centralization or contextual assimilation?", pp. 149-154 in G. Fant, ed., Speech Communication, Vol. 2, Almqvist & Wiksell Int., Stockholm. 1975

REGIONAL DIFFERENCES IN THE REALIZATION OF STANDARD GERMAN VOWELS

ANTTI IIVONEN

University of Helsinki
Helsinki, Finland

ABSTRACT

The socially high standard regional realization of German vowels shows systematical differences in quality of monophthongs and diphthongs, in vowel duration and height of F0. Especially the speech varieties spoken in Vienna and in East Middle German area are compared with each other.

METHODS

For the comparison of the vowel qualities a F1/F2-plane has been used in which the scales are presented linearly between 200 and 510 Hz, logarithmically above 510 Hz (practically the mel-scale is simulated). Instead of plottig the vowel means on the F1/F1-plane as dots a vowel quality is described as a circle of 1 Bark size around the mean value of the vowel type (see Fig. 1 and /1/). The Bark-circle, based on the critical band concept /2/, and considered as a mobile (not as a fixed entity), freely moving on the F1/F2-plane, is used assuming that it is capable to show the psychoacoustic vowel distances. An IBM and MSX compatible computer program BARKF1F2 is used for plotting. The formant measurements were as follows:

1) SPS-method developed by M. Karjalainen. FFT-spectrum analysis has been carried out with a time window comprising either one glottal period or 30 ms. The monophthongal vowel formants have been

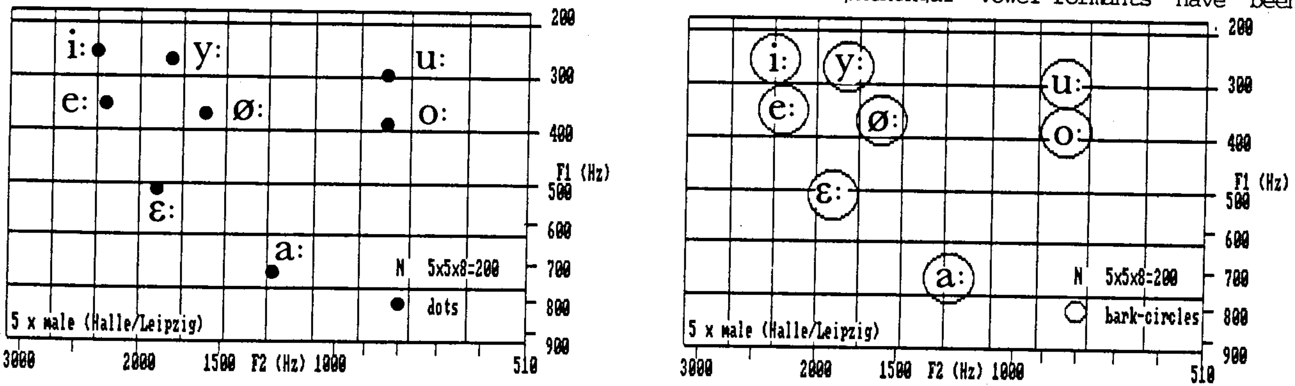
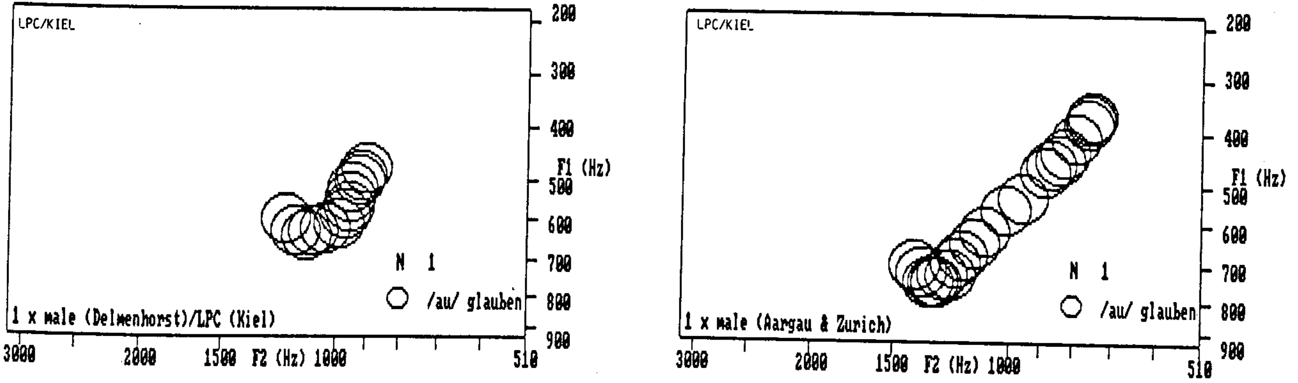


Fig. 1. The means of German long vowels on the F1/F1-plane. The representation of vowels as dots (1a) does not show whether their distances are audible or not. The representation as freely moving 1 Bark size circle (1b) shows that the distance between the close and mid vowels is audible. The lines indicate the boundaries of critical bands (cf. Zwicker 1961).



a male speaker/Delmenhorst/39 years old/computer scientist

GERMAN DIPHTHONGS (NORTH GERMAN)

a male speaker/Aargau and Zurich/26 years old/univ. teacher

GERMAN DIPHTHONGS (SWISS GERMAN)

Fig. 2. A diphthong can be presented as Bark-circles on the F1/F2-plane in 10 ms time intervals /au/ in glauben in North German (2a) shows a shorter gliding than that in Swiss German (2b).

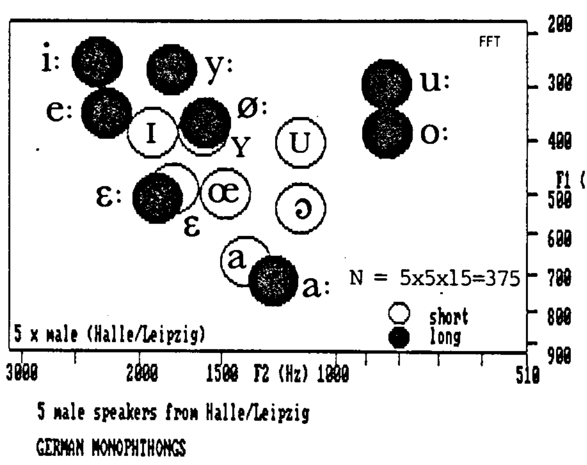


Fig. 3a. The system of 15 German monophthongs of East Middle German on the basis of 5 male speakers.

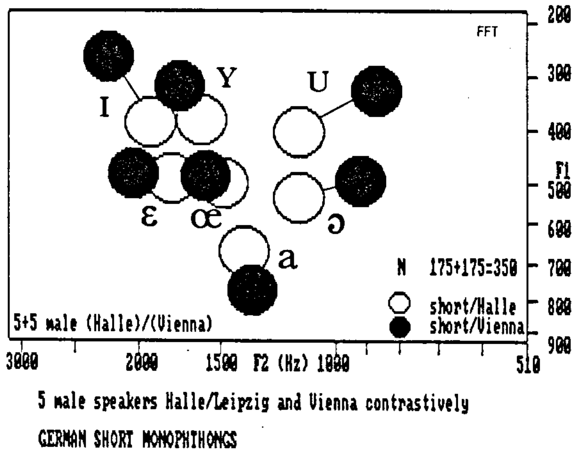


Fig. 4a. Comparison of the short monophthongs between East Middle German and in Viennese German.

measured at the temporal mid point of the vowels.
2) Mainly diphthongs were analyzed by means of the LPC-method (SSP-method) available at the University of Kiel (Institut für Phonetik und digitale Sprachverarbeitung).

QUALITY OF MONOPHTHONGS

The vowels of five male informants from East Middle German area (Halle/Leipzig, EMG) and from Vienna were compared with each other. A material consisting of (5+5)x90 isolated one syllable words was analyzed. Each vowel class of the 15 primary stressed monophthongs and 3 diphthongs comprised 5 word examples. All the informants had university background as students or teachers. The five speakers from EMG area were students or teachers of speech science and therefore they can be assumed to be good representatives of Standard German.
The means of the EMG and Viennese monopthongs (Fig. 3a and 3b) and their comparisons (Fig. 4a and 4b) show following main differences.
The long monophthongs have almost the same quality, but Viennese German (=ViennG) has a more back quality in /u:/ and /o:/ (or possibly they are more

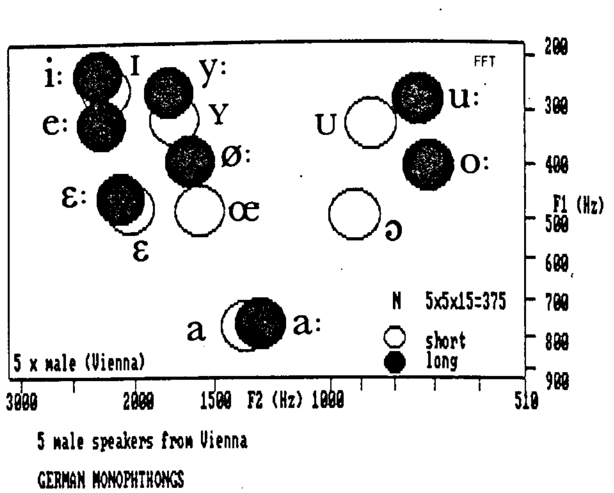


Fig. 3b. The system of 15 monophthongs of Viennese German on the basis of means of 5 male speakers.

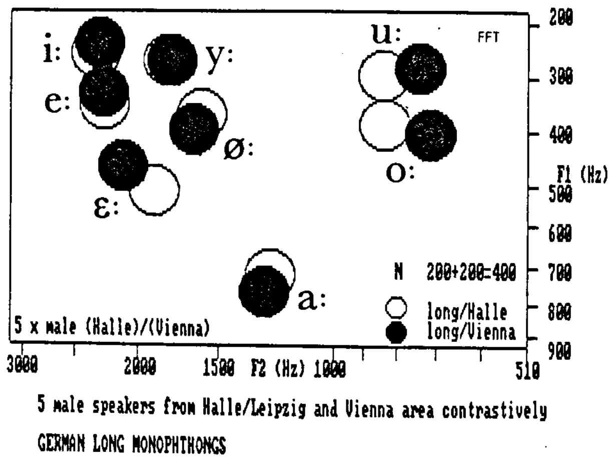


Fig. 4b. Comparison of the long monophthongs between East Middle German and in Viennese German.

round because of their longer duration; cf. the chapter on duration). /i:/ and /e:/ have a slightly closer quality in ViennG than in EMG. The open / :/ tends to be closer in ViennG than in EMG. /a:/ is more open in ViennG than in EMG.
The short monophthongs are considerably more centralized in EMG than in ViennG. This concerns also the vowel /a/. The differences are in the cases of /I/, /U/ and / / more than 1 Bark. In the cases of /a/, / / and /Y/ almost 1 Bark. Only in /ɛ/ the difference is smaller.
The vowels /a/ and /a:/ of ViennG show almost the same quality /a/ being slightly more fronted (but not centralized). In EMG the difference between the series /i:, y:, u:/ and /e:, ø:, o:/ comprises practically only 1 Bark which means psycho-acoustically the smallest possible distance.
When the formant data from Rausch /3/ (see Fig. 5) are plotted on the Bark F1/F2-plane it can be seen that they show very similar monophthongal structure with the EMG data obtained here. The background of the informants in Rausch was probably West Middle German (WMG). Also the data concernig the means of three informants from Berlin in Jørgensen /4/ show similar structure in spite of the

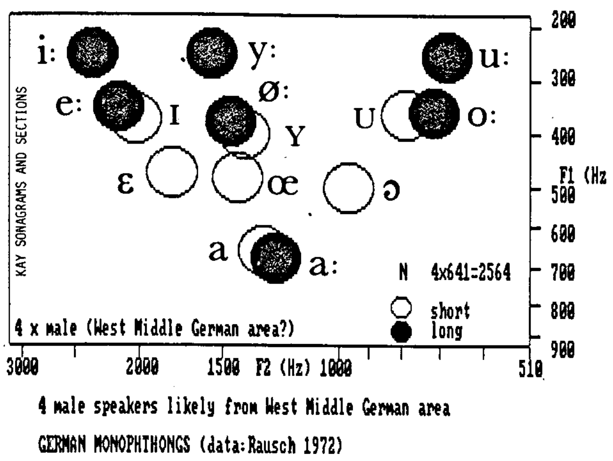


Fig. 5. The system of 14 German monophthongs on the basis of data from Rausch (1972).

considerable differences in the test arrangements. It is remàrkable that the comparison with the results obtained by Delattre /5/ shows that the best monophthongal vowel structure on the basis of listening synthetic F1&F2 stimuli corresponds more to that of ViennG than to that of EMD or WMG (= results by Rausch). /a/ is not centralized but slightly fronted compared with /a:/. /I, Y, U/ are only slightly centralized like in ViennG. Delattre does not tell, whether his listeners come from the South. It can also be argued that the synthesis of lax vowels is no easy task. If the 14 vowel qualities obtained by Delattre represent some kind of deep structure of German monophthongs in general we can say that the centralization of the Middle and North German vowels depends on the rapid speech delivery, i.e. on the performance (cf. durations).
Fig. 7a shows that the centralization of the short monophthongs is even stronger in the vowels of a male speaker from Hamburg (North German). The series /I, Y, U/ is more open than the series /i:, ø:, o:/. The series /ɛ, œ, ɔ/ comes close to /a:/. Fig. 7b shows that the monophtongal system of a Swiss male speaker resembles much that of ViennG, but his /u:/ has a lower F2 on an average.

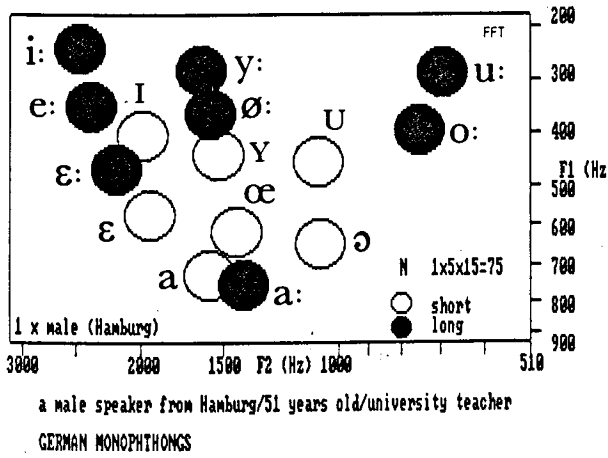


Fig. 7a. The idiolectal system of monophtongs of a North German male speaker from Hamburg.

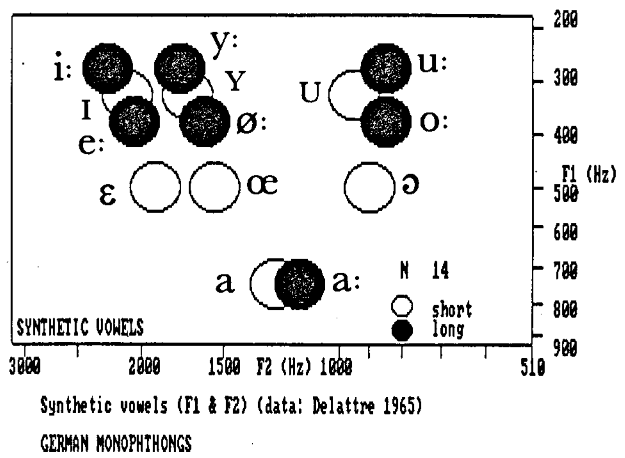


Fig. 6. The system of German 14 monophthongs on the basis of synthetic stimuli (Delattre 1965).

QUALITY OF DIPHTHONGS

The glide of diphthongal quality is described in two idiolects by means of starting and end point of the glide (Fig. 8). The means of 5 words for each class of diphthongs were analyzed. The diphthongs of a male East Middle German speaker show minor span of gliding than that of a male Viennese speaker. The same feature distinguishes also the North German and the Swiss speaker in Fig. 2. A more explicit quality seems to be a common feature for the monophthongs and diphthongs in ViennG. In both varieties of German the first element of /ai/ is a front vowel, whereas that of /au/ is a back vowel. In /ai/ of EMG that vowel is ә, in /ai/ of ViennG æ. In ViennG the first element of /oi/ is more a central than a back vowel.

RELATIVE AND ABSOLUTE DURATION OF VOWELS

Fig. 9 shows that the relative duration of the 15 monophthongs in EMG and in ViennG is similar, but the absolute durations are much shorter in EMG. The quantity quotient V:/V was 2.3 in EMG and 2.1 in ViennG.

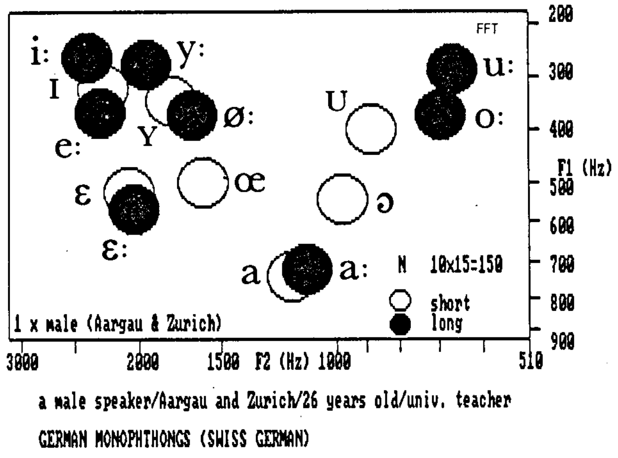


Fig. 7b. The idiolectal system of the monophthongs of a male Swiss speaker from Aargau and Zurich.

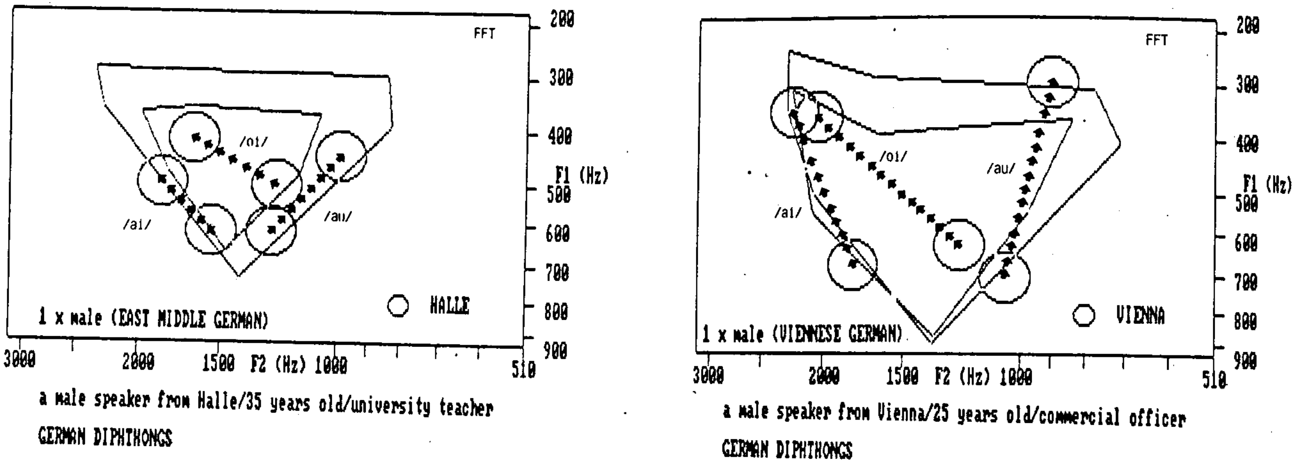


Fig. 8. The idiolectal qualities of the diphthongs of an East Middle German male speaker (8a) and a male speaker of Viennese German (8b). The lines combine the mean values of their short/long vowels. The circles show the starting and end points of the diphthongal glides.

Comparison of two idiolects showed that the diphthongs were ca. 10 % longer than the long monophthongs in EMG and ViennG.

FUNDAMENTAL FREQUENCY OF VOWELS

Fig. 10 shows that the fundamental frequency of the monophthongs in ViennG is systematically higher than in EMG. The short and long close vowels are higher than the other monophthongs in ViennG. This might be in connection with the 5 degree openness system of vowels in the Viennese German Dialect. The variation of the F0 height is smaller in EMG.

CONCLUDING REMARKS

The following features are characteristic for East Middle German (and probably also for North German): centralization of the short monophthongs, smaller dispersion of the total vowel area, shorter gliding of the diphthongs, shorter duration of all types of vowels, lower level of F0. The Viennese German vowels (and probably those of South German in general) are more explicit in quality. Following explanations can be considered: 1) more careful pronounciation of ViennG because of the dialect/Standard diglossia; 2) more rapid speech delivery of East Middle German and North German; 3) difference in the articulation base; 4) 5 degree vowel openness system in Viennese Dialect and its interference in the socially higher speech form. Further details in /6/.

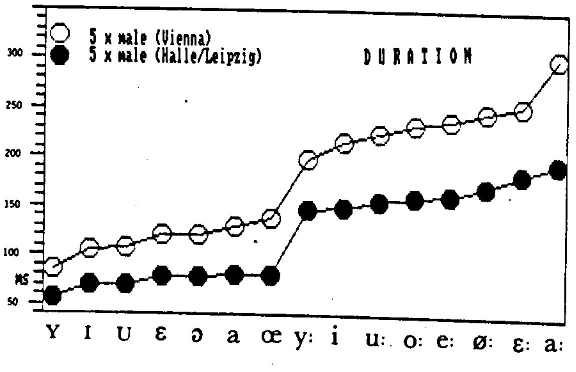


Fig. 9. Durations of the 15 monophthongs in East Middle German and Viennese German contrastively.

REFERENCES

/1/ A. Iivonen, "The critical band in the explanation of the number of possible vowels and psychoacoustical vowel distances", Mimeogr. Series of the Dept. of Phonetics, University of Helsinki 12,1987.

/2/ E. Zwicker, "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)", JASA 33(2),248,1961.

/3/ H.P. Jørgensen, "Die gespannten und ungespannten Vokale in der Norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantenfrequenzen", Phonetica 19,217-245,1969.

/4/ A. Rausch, "Untersuchungen zur Vokalartikulation im Deutschen", Beiträge zur Phonetik, IPK-Forschungsberichte 30,35-82,1972.

/5/ P. Delattre, Comparing the Phonetic Features of English, German, Spanish, and French, Heidelberg,1965.

/6/ A. Iivonen, "Zur regionalen Variation der betonten Vokale im gehobenen Deutsch", Festschrift für das 100-jährige Bestehen des Neuphilologischen Vereins, Helsinki (im Druck).

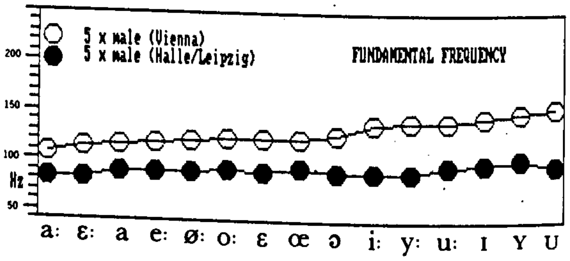


Fig. 10. The F0 height of the 15 monophthongs in East Middle German and Vienn. German contrastively.

SOME ACOUSTIC OBSERVATIONS ON HALF NASALS IN SINHALESE

MASATAKE DANTSUJI

Dept. of Linguistics
Kyoto University
Kyoto 606 JAPAN

ABSTRACT

Sinhalese is one of the Indic languages and is spoken in Sri Lanka. The present study is intended to explore the phonetic characteristics of half nasals in Sinhalese. From the examinations of the acoustic analysis making use of a minicomputer and a linear prediction algorithm, some phonetic properties can be clarified as follows. It appears that a half nasal consists of a nasal murmur portion, a voiced oral murmur portion, a burst and a transition portion to the following vowel. From the spectrum analysis, it can be examined that frequencies of the first formant (F1) of the nasal murmur portion are slightly higher than those of the oral murmur portion. It has been able to distinguish places of articulation by making use of spectrum features of the nasal murmur portion.

INTRODUCTION

Sinhalese is a national language of Sri Lanka, and is spoken by about 11 million people, or 75 percent of the population, living mainly in the southern and western two-thirds of the Island of Ceylon [1]. It is said that Sinhalese is an Indo-European language descended from Sanskrit, and this language was brought to the island by settlers from northern India in the 5th century B.C.[1]. There have been pointed out several problems connected with Sinhalese which arouse the interest of the linguists, e.g. Sinhalese is notable among the major Indo-Aryan languages of the past and present in having no aspirate stop phonemes nor clusters [2], literary Sinhalese is very different from spoken Sinhalese [3], etc. One of them is the phenomenon of "half nasals". In the intervocalic positions, there occur medial clusters composed of nasal plus voiced stop. There are two types of the first nasal element. One is often referred to as the single nasal and the other is referred to as the doubled nasal. The two types present a contrast, e.g., /kandə / 'trunk' : /kanndə / 'mountain'. In regard to voiceless stops, there are no such oppositions in the same position, only the normal type of cluster with doubled nasal occurs. In such cases, however, the doubled nasal is written with a single letter according to the convention. The length of the nasal in a cluster of single nasal plus voiced stop varies from normal to very short [2]. It has been customary in Sinhalese studies to treat the single nasal in these clusters as a special class of sounds to which was given the name "half nasal" and this is in accord with the traditional Sinhalese orthography, which uses special signs for the "half nasals" and the regular nasal letters for the "full (doubled) nasals" in the same position. From a synchronic point of view, there can be two different phonological interpretations. Some linguists regard them as independent phonemes. For example, Jones [4] treats them as separate independent phonemes. On the other hand, others regard them as consonant clusters. For example, Coates and De Silva [2] criticize Jones' view as it is an unnecessary complication of the phonemic system, increasing the number of consonant phoneme by nearly 20 percent, and treat them as consonant clusters with a single nasal, contrasting with a doubled nasal in similar clusters. However, the literature on the phonetic detail of half nasals is quite limited in quantity and quality. The present study is intended to explore the phonetic characteristics of half nasals through acoustic investigation. One of our main concerns is to confirm if the articulator is already ready for the place of articulation during the first nasal element.

ACOUSTIC ANALYSIS

Material

Lists of words were prepared, which contained four types of half nasals [mb, n̪d̪, ṇḍ, ŋg] in the intervocalic position. Each half nasal was preceded and followed by 5 vowels / i, e, a, o, u / leading to five different V_1CV_2 combination, where V_1 is the same vowel as V_2. The list includes meaningless words. Each word was written in Sinhala scripts.

Subjects

One subject (M,30) who was born in Kurunegala and brought up in Kegalla, the northeast of Colombo which is the capital of Sri Lanka, served as informant for this investigation. The subject came to Japan as foreign research student on Japanese government scholarships and were at Osaka University of Foreign Studies for six months for Japanese language training. Recording was made five months after he had arrived in Japan.

Procedure

Four lists were prepared, each containing a subset of the words including half nasals at the intervocalic position, and some of them were nonsense words. Each word was appeared four times in the recording. Therefore, 80 tokens in all (4 types of half nasals x 5 different combinations x 4 times) were analyzed. The lists were read by the informant in the soundproof room of the phonetic studio of Osaka University of Foreign Studies, Osaka, Japan. The material was taped on a TEAC A-6700 tape recorder, using a highly sensitive microphone. The distance to the microphone was set at about 30 cm. The informant was instructed to read the material at a natural tempo with a pause after every word, and with the same loudness.

Acoustic Analysis

Acoustic analyses were made at Doshita Laboratory, Department of Information Science of Kyoto University. Speech waveforms were digitized from the output of the tape recorder. A JEIC 3118 low-pass filter with 70 dB/oct for anti-aliasing (the cut-off frequency was set to 8.9 kHz) and a DATEL DAS-250 16-channel 12-bit A/D converter of 4-microsecond sampling period were used for digitization. Through the A/D converter which was connected to a FACOM U-200 minicomputer at the common bus with direct access mode, speech data samples at every 54 microsecond (18.5 kHz sampling) were stored into a cartridge-disc of 2-megabyte in real-time for about 1 minute continuously. Individual utterances were input from a digital magnetic tape to a minicomputer and the waveforms were drawn on an X-Y plotter. It was found out that each half nasal includes two types of murmur portions from the expanded waveform. From the 3500 samples of data (190 msec), waveforms including both murmur portions were sampled manually using a cursor. Each murmur portion was excised with 27 msec Hamming windows for fast Fourier transform (FFT), and linear prediction analysis was applied to obtain a smoothed spectrum. The optimal prediction order was estimated in a preliminary experiment to be 26. These spectra were analyzed and displayed on the plotter. Formants corresponding to spectral peaks were computed from the solution of higher order polynomial equation by the Muller method. We adopted the lowest three prominent peaks of murmur spectra as the first formant (F1), the second formant (F2) and the third formant (F3).

RESULTS AND DISCUSSION

From the waveforms, it was observed that a half nasal is sub-segmented into four portions; a portion of periodic repeated waveforms adjacent to the preceding vowel, a portion of quasi periodic waveforms following to the former portion, a small protrusion upon a running smooth waveforms and fluctuated waveforms after the protrusion. It is assumed that the small protrusion is a burst of the plosive. The fluctuation portion after the protrusion is assumed to be a transition portion to the next vowel. From the precise observation on earlier two portions, it was found out that the former portion has rather steady repeated waveforms in comparison with the latter portion. This implies the presence of continuous tone in this portion, and this is one of the characteristics of the nasal murmur portion. The duration of this portion is 36 ~ 141 msec. On the other hand, the intensity of the latter portion is less than that of the former portion. The latter portion has quasi repeating waveforms, though it shows the continuation of periodicity, that is to say, each waveform is quite similar but differs each other in some ways, especially in the energy. The quasi periodic waveforms of this portion show gradual decreasing energy with the aim of the burst for a plosion. This indicates that this portion is the portion of vibration of the vocal folds of voiced plosives, which precedes the release burst of the voiced plosives. The duration of this portion is 13 ~ 43 msec. In addition to these observations, fast Fourier transform (FFT) and linear prediction analysis were applied in order to obtain spectrum structures. It was noticed that the latter portion has only very low-frequency energy. The energy falls off rapidly after the first formant and is very weak in the middle- and high-frequency range. The mean values and standard deviations (SDs) of the first formant of both portions are presented in Table 1. The mean F1 value of the latter portion is 215 Hz and that of the former portion is 290 Hz. The latter portion has lower F1 than the former portion. This is statistically significant ($p < 0.01$). This indicates that the latter portion has only very low frequency energy. This corresponds with the fact that the low frequency value of F1 in comparison with the adjacent segment is one of the characteristics of voiced plosives. On the other hand, it was remarked that the former portion has several resonances below 3500 Hz and anti-formant. It was assumed that the former portion is the nasal murmur portion. Therefore, a half nasal is sub-segmented into a nasal murmur portion, anoral murmur portion, a burst and a transition portion to the next vowel. These observations indicate that a half nasal is a kind of prenasalized voiced plosive.

Table 1. Analysis frequency values (in Hz) and standard deviations of the first formant for the former portion and for the latter portion (means and standard deviations).

	Former Portion		Latter Portion	
	Mean	(S.D.)	Mean	(S.D.)
F1	290	(37)	215	(21)

Table 2. Analysis frequency values (in Hz) for the nasal murmur portion of half nasals for each place of articulation (means and standard deviations).

	Bilabial		Dental		Velar	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
F1	256	(29)	294	(43)	339	(35)
F2	1009	(139)	1338	(169)	1051	(92)
F3	2607	(225)	2787	(129)	2700	(104)
B1	112	(42)	151	(51)	175	(39)

In the next analysis, we examined if the nasal murmur portion of a half nasal has a cue for places of articulation. The mean values and standard deviations of formants and band-widths of the nasal murmur portion for each place of articulation (bilabial, dental and velar) are presented in Table 2. The properties of murmur of ordinary nasals are reported as follows (> = higher frequency than).

F1 frequency values: [ŋ] > [n] > [m]

F1 bandwidth values: [ŋ] > [n], [m]

It is assumed that the differences in F1 values are presumably related to differences in the size of the coupling section at the velo-pharyngeal passage and of the pharyngo-nasal tract [5],[6]. Although these are characteristics of ordinary nasals, it has been also reported that the murmur portion of voiceless nasals in Burmese shows the same tendency [7]. The mean values of F1 of bilabial half nasals, dental half nasals and velar half nasals are 256 Hz, 294 Hz and 339 Hz, respectively. In this analysis, retroflex half nasals are excluded, as they are considered to make the matter complicated. It can be said that there is a tendency for F1 of bilabial half nasals to be slightly lower than those of dental and velar half nasals. [m] < [n̪], [n̪] < [ŋ] are statistically significant ($p < 0.01$).

The mean value of F2 of dental half nasals is 1338 Hz. Those of bilabial half nasals, velar half nasals are 1010 Hz and 1054 Hz. There is a tendency for F2 of dental nasals to be slightly higher than those of bilabial and velar half nasals. Both [n̪] > [m] and [n̪] > [ŋ] are statistically significant ($p < 0.01$). In the framework of Jakobson, Fant and Halle [8], distinctive features were also defined by acoustic aspects. Dental nasals have an "acute" feature and, when the second formant "is closer to the third and higher formants, it is acute." Velar nasals have a "compact" feature and when the first formant "is higher (i.e. closer to the third and higher formants), the phoneme is more compact". Therefore, we tried to distinguish dental half nasals by means of the higher frequencies of the second formant, and separate velar half nasals from bilabial half nasals by means of the higher frequencies of the first formant.

Table 3 represents classification matrix of the group classification and the percent of correct classifications of discrimination analysis. The left column represents source, and the uppermost row represents judged groups after analysis. The percent of correct classifications in total are 73 %. This shows tolerably high performance. In order to obtain better resolution, all of the first, the second, the third formants and bandwidth of the first formant were used as variables for the step-wise discrimination analysis. Results are presented in Table 4. The percent of correct classifications in total are 93 %. It can be said that each

Table 3. Classification matrix (Variables: F1,F2,F3,B1)

	[mb]	[n̪d̪]	[ŋg]	
[mb]	19	0	1	95 %
[n̪d̪]	0	19	1	95 %
[ŋg]	1	1	18	90 %
TOTAL	20	20	20	93 %

group of the place of articulation is effectively discriminated. These results indicate that the nasal murmur portion of the half nasal includes necessary information to distinguish place of articulation. This confirms the view that the articulator is already ready for the place of articulation during the murmur portion of the first nasal element.

SUMMARY AND CONCLUSIONS

So far as our informant is concerned, the properties of half nasals in Sinhalese have been clarified as follows. A half nasal is sub-segmented into a nasal murmur portion, a voiced oral murmur portion, a burst and a transition portion to the next vowel. The plosive element keeps voicing before release of the consonantal constriction. These results indicate that a half nasal is a kind of prenasalized voiced plosives. Bilabial half nasals, dental nasals and velar half nasals could be distinguished by means of a step-wise discriminant analysis utilizing the value of F1, F2, F3 and B1 of the nasal murmur portion, and this confirms the view that articulator is ready for the place of articulation during the first nasal element.

ACKNOWLEDGMENT

I would like to thank Prof. Tatsuo Nishida of Kyoto University, who have encouraged my study in many ways. I also would like to thank Prof. Syuuji Doshita of Kyoto University and Assoc. Prof. Shigeyoshi Kitazawa of Shizuoka University, for making their facilities available. I also wish to thank Mr. N. S. J. Nawaratne for his assistance as my informant.

REFERENCES

[1] K. Katzner,<<The Languages of the World>>, Routledge & Kegan Paul, revised edition 1986.

[2] W. A. Coates, M. W. S. De Silva, "The Segmental Phonemes of Sinhalese", University of Ceylon Review 18, 163-175. 1960.

[3] M. W. S. De Silva, "Sinhalese", in T. A. Sebeok (ed.)<<Current Trends in Linguistics>>,Mouton, 1969.

[4] D. Jones,<<The Phoneme: Its Nature and Use>>, W. Heffer & Sons, 1950.

[5] O. Fujimura, "Analysis of Nasal Consonants", <<J. Acoust. Soc. Am>> 34, 1865-1875, 1962.

[6] D. Recasens, "Place Cues for Nasal Consonants with Special Reference to Catalan", <<J. Acoust. Soc. Am.>> 73, 1346-1353, 1983.

[7] M. Dantsuji, "An Acoustic Study on the Distinction of Place of Articulation for Voiceless Nasals in Burmese", <<Folia Linguistica>> 21(forthcoming).

[8] R. Jakobson, G. Fant, M. Halle, "Preliminaries to Speech Analysis", <<Technical Report>>13, Acoustic Laboratory MIT, 1952.

THE ANCIENT INITIAL "VOICED" CONSONANTS IN MODERN WU DIALECTS

CAO JIANFEN

Institute of Linguistics
Chinese Academy of Social Sciences
5 Jianguomennei Dajie, Beijing, CHINA

ABSTRACT

The Ancient Initial 'Voiced' Consonants in Modern Chinese Wu Dialects have phonetically become completely voiceless cosonants when they are in isolated monosyllabic words or in the first syllable of polysyllabic words. However, in running speech, they conditionally alternate between typical voice and complete voicelessness.

INTRODUCTION

Phonologically, the most important feature of the Modern Chinese Wu Dialects is the retention of the Ancient Initial Voiced Stop Consonants(i.e. 并 ,定,群,etc.) in its consonant system. The stop consonants of Wu can be divided into three categories: voiced, voiceless unaspirated, and voiceless aspirated /1/. This division is well known in the linguistic circle both in China and overseas. However, what is the phonetic value of the I.V.C. in Wu is a controversial issue in the linguistic circle.
Before the 1920's, linguists did not seem to distinguish phonological category from phonetic(or physical) nature of the I.V.C. They tended to believe that there must exist vocal cords vibration during the phonation of these consonants. In the 1920's, the famous linguists Liu Fu and Zhao Yuan Ren /2/ suggested that the voiced consonants of Wu are not really voiced, but begin with a voiceless sound and finish with a voiced glide. They are marked with the diacritic /ɦ/, such as /$p^{ɦ}$,$t^{ɦ}$,$k^{ɦ}$ /. It means that there is not vocal cords vibration during the closure, but a whiff of voiced breath between stop release and following vowel(s). Bernhard Karlgren /3/ regarded this breath as a slight 'voiced aspiration' and directly transcribed it with the diacritic /ʻ/, such as /bʻ ,dʻ ,gʻ /. Sinse then, the above statements have become the dominant theory.Early in the 1980's, in order to exam the real nature of this so-called 'voiced glide', the present author investigated these consonants in the Chang Yinsha dialect of Wu with spectrographic analysis. The result /4/ showed that there is not any 'voiced glide' or 'voiced aspiration' between the stop release and following vowel(s) when these consonants are in isolated monosyllabic words or in the first syllable of polysyllabic words, but the tone pattern of these syllables is quite different from that of those syllables beginning with corresponding voiceless consonants.In running speech, on the other hand, the phonetic value.of the I.V.C. conditionally alternate between typical voice and complete voicelessness. The condition of this alternation is the syllables tone pattern. When it is pronounced with a 'Yang' tone (阳 调) /*/, the real value of the I.V.C. is complete voicelessness; when it is pronounced with a 'Yin' tone (阴 调) /*/, or neutral tone, the value of the consonat is typical voice. In spite of above evidence, linguists of the traditional school are still sketical to what the spetrograms showed. Later, the above result was verified by the measurement of the air flow through the glottis, supraglottal presure and subglottal presure /**/. Fig.1 (see p.2) shows a few recordings of these air flow and air presure with the related spectrograms. There the word /dao/稻 begins with /d/ of the I.V.C. Both the air flow and the air presure recordings and the spectrograms show clearly, that in the isolated word or in the first syllable of the polysyllabic word /dao za/ 稻 柴 , the value of /d/ is the same as the corresponding voiceless /t/ in word /tao/ 岛 , but the syllable keeps its 'Yang' tone. While in the polysyllabic word /ɕie dao/ 籼 稻 , the 'Yang' tone of the syllable changes to neutral tone which is the same as that of syllable /tao/ 岛 in /ɕie tao/ 仙 岛 , and the value of /d/ becomes typical voice. Sinse 1983, the present author has investigated about 15 subdialects of Wu, recorded about 30 speakers' utterances, including isolated words, short sentences and conversations. The spectrograms of these utterances show that the situation of the I.V.C. in these dialects is almost

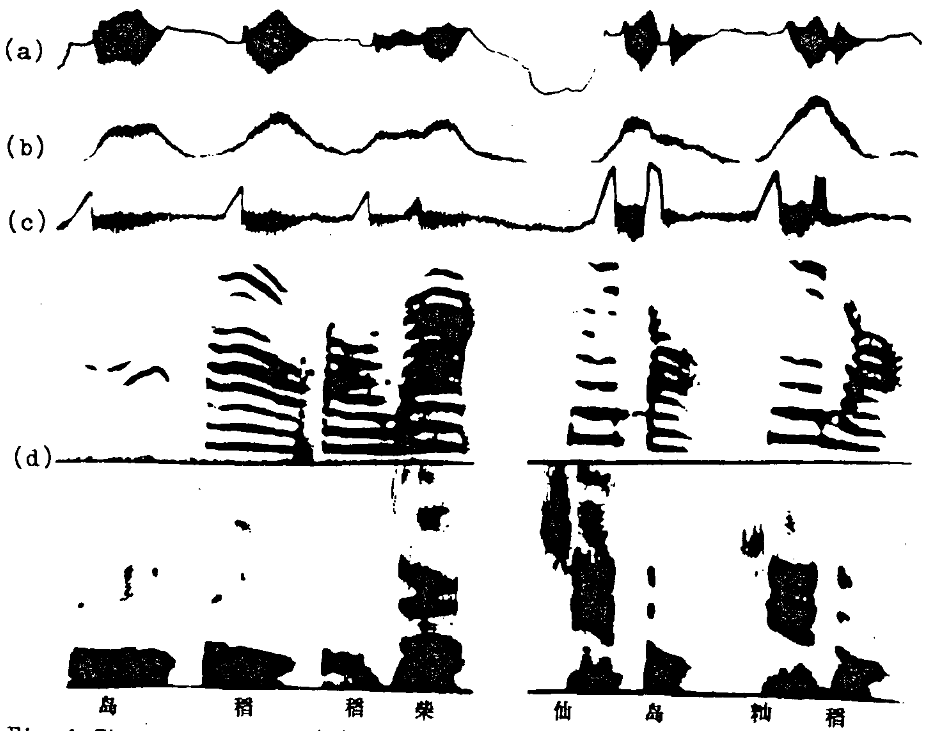


Fig.1 The records of (a)air flow through glottis,(b) subglottal presure,(c) supraglottal presure with (d) the rėlated spectrograms.

the same as it is in Chang Yinsha dialect. On the bases of the above investigations, we posit two hypotheses. First, if we view the situation from isolated words, the phonetic value of the I.V.C. in Wu has become completely voiceless, its abstract distinctive function in a word has been taken over by the `Yang' tone pattern of the syllabie. Secondly, if we view them from running speech, the value of the I.V.C. alternate between voice and voicelessness. In this case, distinctive role of the I.V.C. in a word is played by the phonetic value of the I.V.C, or by the `Yang' tone pattern of the syllable, and this regular alternation seems to be dependent on the stress type of relavant syllable in speech.

EXPERIMENTS

In order to test above hypotheses, a perception test and a synthesized speech test, as well as a spectrographic analysis were carried out.

Perception Test

In this test, one female speaker of a Wu dialect pronounced 64 pairs of monosyllabic words which begin with I.V.C. and the corresponding voiceless consonants respectively; another female native speaker of the Beijing dialect pronounced 126 monosyllabic Beijing words with the tone pattern similar to the `Yang'tone pattern of Wu. Their recorded utterances were mixed up in random order on the same tape. Then, 13 subjects were asked to listen to the tape recording and determine which words begin with voiced consonants and which ones the voiceless consonants. The result is very interesting. Our subjects have different background and their responses are quite different, but showed the same rule operates in different aspects.

The first type of the subjects, who are phonologists and familiar with the Wu Dialects, have a great degree of agreement in their responses. They judge both of those words beginning with the I.V.C. and most of the Beijing words as beginning with the voiced consonants. It seems that what is significant for these subjects is the sense of voicing, but not the presence of vocal cords vibration. It also seems to be the case, that they tend to perceive that there is vocal cords vibrátion during the consonant of a syllable with a `Yang' tone, even though its spectrographic correlate is the same as that of the corresponding voiceless consonant.

The second type of the subjects are both phonologist and phonetician. Their responses show inconstantly. The typical subject is a phonetician, but he has the Wu dialect background. He said he had tried to judge these words by means of physiological or physical cues. However, he still judged about 16 % of consonants of the Beijing words as voiced consonants by mistake. So it seems that he still could not avoid being motivated by the so-called voicing sense which likely to be caused by the `Yang-like' tone pattern of these words.

The third type of the subjects is a student in phonetics. At that time, he just graduated from an English department, knew nothing about Chinese Phonology and the Wu Dialects, so his judgement could only be in terms of the acoustic or physiological characteristics of these consonants. The accuracy of his judgement reaches as high as 98%. He commented that he thought almost none of these initial consonants were voiced. That is true.

Actually, the value of these so-called voiced consonants here is the same as that of the corresponding voiceless consonants according to their spectrograms, and those consonants of Beijing words here are doubtlessly complete voicelessness. Why did the majority of the subjects consider that there was vocal cords vibration during the phonation of these consonants? The only explanation seems to come from the `Yang' tone or `Yang-like' tone patterns of these syllables. Because this tone pattern generally has either a lower pitch or a lower beginning, it is easy to give a sense of voicing. Moreover, this pattern is always accompanied by the I.V.C. in the monosyllabic words. This is a phonological rule of Wu. Consequently, the subjects who are familiar with the sounds of Wu are used to this rule. Ones they hear a syllable with a `Yańg' tone, they will naturally identify its initial consonant as being voiced, as if there exists vocal cords vibration during the phonation of the consonant.

Synthesized Speech Test

The data in this test come from the spectrographic measurement of the voiceless consonants, vowels and the tones of Wu. A procedure of synthesis by-rule /5/ is used to produce the test samples. First, the consonant and vowel data and the `Yin' tone data were input to a computer, consequently, a syllable with a `Yin' tone was produced and it sounded like a word beginning with a voiceless consonant in Wu. Secondly, while the consonant and vowel data remained constant, but a `Yang' tone data replaced the `Yin' tone data. As a result, a syllable with a `Yang' tone was produced and it sounded just like a word beginning with the I.V.C. in Wu. Followed this procedue, we got a series of word pairs. In 1985, the tape recording of these words

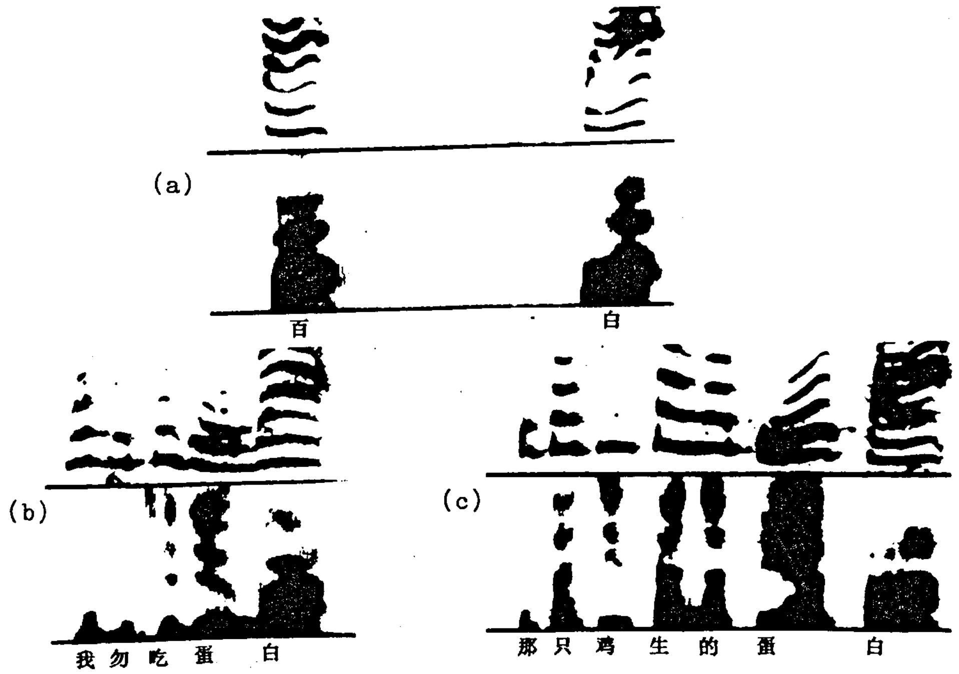


Fig.2 The spectrograms of (a) the monosyllybic word 白 (white) and 百 (handred), (b) the sentence 我勿吃蛋白 (I dislike to eat albumen), (c) the sentence 那只鸡生的蛋白 (the agg laid by that hen is more white).

was played at the Third Meeting of the Linguistic Society of China, and the audiance was asked to judge them without any knowledge of the source of these sounds. As the result, those linguists did judge those words with a ‘Yang’ tone as beginning with voiced consonants. This result indicates that the sense of voicing which perceived by these audiance is caused by the ‘Yang’ tone of the syllable, instead of the value of the initial consonant.

Experiment of Running Speech

In this experiment, a series of syllables beginning with the I.V.C. were put in comparative context of different sentences, to observe how the variation of phonetic value of the I.V.C. depend on the changes of the syllable’s tone and stress type. This test involved several subdialects of Wu, Fig.2 (see p.3) is a sample of this variation. Here the /b/ of /ba/白is an I.V.C., and the /p/ of /pa/百is the corresponding voiceless consonant. When they are pronounced in isolation, the spectrographic correlates of the /b/ and the /p/ are the same. it means that the I.V.C. /b/ is voicelessness. However, when the word is in running speech, the value of the /b/ varies. In sentence (b), as you might have noticeor there is obvious voice bar during the closure of /b/ in /diø ba/ 蛋 白 as shown in its spectrogram, it is typical voice. But in sentence (c), the spectrographic correlate of /b/ in /diø/ 蛋 /ba/ 白 is the same as that of /b/ or /p/ in (a), it is completely voiceless. Compare (b) and (c), we can see that what causes this variation is the different changes of tone and stress in this two sentences. In sentensce (b), /diø ba/ 蛋 白 is a bisyllabic word. the syllable /ba/ is unstressed and its ‘Yang’ tone has changed to be ‘Yin’ tone. In this case, the distinctive function of the ‘Yang’ tone has lost, and it must be replaced by something else. Consequently, the feature of voice becomes neccessary; In contrast, the /diø/蛋/ba/白in sentence (c) is a subject-predicate structure, the syllable /ba/ 白 is the predicate here and has to be stressed, so the ‘Yang’ tone pattern remains unchanged and keeps its distinctive function, therefore, the value of /b/ need not be changed in this case. This example supports our second hypothesis and clearly indicates that the alternation of the I.V.C.’s value in running speech regulary matches the changes in tone and stress type of those syllables where they located. This alternation can be formulized by the following rules:

$$\text{I.V.C.} \longrightarrow \overset{C}{[-\text{voice}]}\ /\ \#\text{——}V+(CV)0\ \#$$

$$\text{I.V.C.} \longrightarrow \overset{C}{[+\text{voice}]}\ /\ V+\text{——}\overset{V}{[-\text{stress}]}$$

Here the zero in (cv)0 means the numbers from zero to any other integer, so the $\#\text{——}V+(CV)0\ \#$ means any isolated monosyllabic or polysyllabic words. The $V+\text{——}\overset{V}{[-\text{stress}]}$ means the I.V.C. is in running speech and must be in an unstressed syllable.

CONCLUSION

This paper has tried to clarify two points about the I.V.C. in Wu. First, spectrographic analysis indicates that the I.V.C. in the Modern Wu Dialects has become completely voiceless. They are just the retention of the abstract distinctive category of the original phonological ‘voice’ rather than real physiological vocal cords vibration. Our perception test and the synthsized speech test show that the voicing sense of these syllables is caused by the ‘Yang’ tone pattern and motivated by the perceptual mechanism. Secondly, in running speech, there are alternate variants between typical voice and complete voicelessness. The condition of this alternation is the rule of the tone sandhi and the stress type in particular subdialects.

NOTES

/*/ The ‘Yang’ tone is called 阳调 in Chinese and the ‘Yin’ tone is阴调 . Generally, the ‘Yang’ tone has either a lower pitch or a lower beginning, while the ‘Yin’ tone has a higher pitch or a higher beginning in Wu.

/**/ This test was taken under Prof. P. Ladefoged’s guidance in 1983.

REFERENCES

/1/,/2/ Zhao Yuan Ren, Studies in the Modern Wu Dialects, Qing Hua College Res. Inst. Monogr. 4, Beijing, 1928.

/3/ Karlgren, Bernhard, Etudes Sur La Phonologie Chinoise, Stockholm, 1915-1926. See Zhao Yuan Ren et,al., Chinese Transl.

/4/ Cao Jianfen, The Characteristics of the Ancient Initial Voiced Consonants in Chang Yinsha Dialect, Zhongguo Yuwen,No.4, 1982, pp. 273-279.

/5/ Yang Shunan and Xu Yi, A software system for synthesizing Chinese speech, Proc. 1987 Inter. Conf. on Chinese Information Processing, Aug. 4-6, Beijing, China.

JAW KINEMATICS IN HEARING-IMPAIRED SPEAKERS

Nancy S. McGarr, Anders Löfqvist, Robin Seider Story
Haskins Laboratories
270 Crown Street
New Haven, CT 06511, U. S. A.

ABSTRACT

This study examines jaw movements as a function of vowel height and stress in real-word phrases produced by deaf and hearing speakers. There were statistically significant main effects for vowel height in nearly all measures for all speakers. However, there were no statistically significant main effects for stress. The hearing speaker distinguished stressed and unstressed segments by maintaining the jaw in a lowered position for a longer period in the stressed vowels. With few exceptions, kinematic values for the hearing and the deaf were comparable.

INTRODUCTION

Persons who sustain severe-profound congenital hearing loss learn to produce speech using limited residual hearing as well as information derived from visual and kinesthetic sources. Acoustic cues accompanying changes in stress are perceptible to even the most profoundly impaired [1]. With respect to visual and kinesthetic information, deaf speakers frequently place their articulators fairly accurately especially for places of articulation that are highly visible, but fail to coordinate articulatory movements [2]. However, the overall timing of the articulatory event is longer than normal. Unfortunately, the locus of this timing difference at the articulatory level cannot be recovered. Existing cineradiographic data on hearing-impaired speech production [5,6,7] is equivocal having examined either a limited set of utterances (owing to the methodology) or having averaged across speakers with different etiologies. Also, some [8,9] argue that differential vowel productions by deaf speakers are made with extreme movements of the jaw, a visible articulator, as a substitute for the more appropriate but less visible tongue configuration. Thus, the present study examines kinematics of jaw movements in deaf subjects with particular focus on vowel and stress effects.

METHOD

Subjects The subjects were three congenitally, severely-profoundly deaf adult females (mean pure tone average for .5, 1, and 2 kHz 90 dB+ ISO in the better ear), and a hearing adult female who served as a control. All of the deaf subjects received their early training in oral schools for the deaf; two of the subjects, D1 and D2, were mainstreamed in hearing schools. No subject had any additional handicaps. Using a rating scale for intelligibility [10] deaf speakers 1 and 2 could be characterized as difficult to understand although the content could be understood. Deaf speaker 3 was difficult to understand with only isolated words or phrases intelligible.

Procedure Articulatory movements in the vertical dimension of the jaw, the lower lip, and the upper lip were recorded tape using an optical tracking system. Acoustic recordings were obtained simultaneously and all signals were digitized. Velocity records for the different articulators were obtained and a number of measurements made: amplitude, duration, and peak velocity of raising and lowering movements. In some cases, jaw lowering and raising was not executed as a single uninterrupted movement. Rather, the jaw maintained a lowered position for a short period of time. In these instances, the lowering and raising movement was taken as the interval of uninterrupted movement and the "hold" phase was analyzed separately.

Linguistic material The stimuli were short phrases of English words containing a labial medial consonant [p, b, f, v, m, w] flanked by one high (or close) [i] and one low (or open) [a] vowel, respectively. The noun "Pa" was paired with the verbs "peal, beep, meet, weed, feel, veto"; the noun "Bea" was paired with the verbs "pop, bop, mop, want, farm, varnish". The words were produced in the carrier phrase "And ... it". (e.g., "And Pa peals it") with sentence stress occurring on the noun or on the verb. Five repetitions of each utterance type were recorded, giving a total of 120 tokens for each speaker (6 consonants * 2 vowels * 2 stress patterns * 5 repetitions).

*This work was supported by NIH Grants NS-13870 and NS-13617 to Haskins Laboratories.

RESULTS

Movement duration The effects of vowel and stress on kinematic measures of jaw movements were tested using a 2-way analysis of variance for repeated measures for each subject. Significant differences in main effects reported below are≤0.01.

Movement durations of the jaw lowering gesture for the first vowel, the raising gesture into the medial consonant, and the lowering gesture for the second vowel were made. Overall, the means and standard deviations for the deaf speakers are similar to the hearing speaker. The duration of the jaw lowering gesture for the first vowel was significantly longer for both hearing and deaf subjects when the vowel was open. There was also a strong vowel effect on the duration of the raising gesture for the medial consonant and the lowering gesture for the second vowel for all speakers. Again, duration was longer when the vowel was open. Thus, all movement durations were significantly affected by the vowel for the speakers.

Turning to the effects of stress, the results are much more variable. In particular, it was not always the case that stressed segments were produced with movements of longer duration than unstressed ones. For the lowering gesture of the first vowel, there were no significant stress effects for any subject. The raising gesture for the medial consonant was significantly effected by stress for Deaf speakers 1 and 3; for the hearing speaker and Deaf speaker 2, the effect was not significant. Finally, the lowering gesture for the second vowel was significantly increased in the stressed condition for the hearing speaker but there were no statistically significant differences for any of the deaf speakers.

Thus, most of the movement durations were not sensitive to stress variations for any of the speakers. Since this finding was somewhat unexpected, at least for the hearing speaker, we analyzed the "hold phase". Deaf speaker 3 differs from the other subjects in this measure as her durations are significantly longer. Statistically significant differences in "hold phase" for the hearing speaker were noted i.e., longer for the open than for the closed vowel and also longer for the stressed than for the unstressed condition. Only for Deaf speaker 1 was there a significant effect of vowel;for Deaf speakers 2 and 3, the effect was not significant. Stress was not significant for the hold phase for all three deaf subjects.

Movement displacement These measurements show that the values for the hearing and the deaf subjects again do not differ in any systematic manner as to the absolute values; this is also true for the standard deviations. Vowel quality had a significant effect on all displacements for all subjects. Displacements were greater for the open than for the close vowel. The effect of stress on jaw displacement was more variable and not consistent. The lowering movement for the first vowel was significantly greater in stressed syllables than in unstressed ones for Deaf speakers 1 and 3. The difference between stressed and unstressed segments was not significant for the hearing subject and Deaf speaker 2. The raising gesture for the medial consonant was unaffected by stress in all speakers. Similarly, the lowering movement for the second vowel was significantly longer in the stressed condition for the hearing subject and Deaf speaker 3; for the other two Deaf subjects, 1 and 2, no difference was found. The results for displacement thus indicate that movement amplitude was always larger for the open than for the close vowel for all subjects. The stress effects were more variable and mostly non-significant.

Movement velocity For peak velocity, it is, again, not necessarily the case that the hearing subject and the deaf subjects differ in any systematic way. Noteworthy is that Deaf speaker 3 executed the raising gesture of the jaw for the medial consonant with a very low peak velocity. Vowel quality had a significant effect on peak velocity of jaw lowering for the first vowel for all speakers. Peak velocity was higher for the low vowel. The difference in peak velocity between stressed and unstressed segments was significant for the hearing subject and Deaf speakers 1 and 3. The effect of stress was not significant for Deaf speaker 2. Also the closing velocity for the medial consonant was higher for the open vowel for all subjects; However, stress had no significant effect for any speaker. Finally, peak velocity during jaw lowering for the second vowel was significantly faster for the open vowel compared to the close vowel for all subjects. As for stress, stressed segments were produced with a faster lowering gesture for the hearing subject and Deaf speaker 3. For Deaf speakers 1 and 2, this effect was not significant. Again, there was a clear and consistent vowel effect on peak velocity of jaw movements; the open vowel [a] was produced with higher velocity than the close vowel [i]. Stress mostly affected the lowering movements of the jaw which were produced with higher peak velocity in stressed segments.

DISCUSSION

The results of this study indicate that vowel quality has a strong and reliable effect on jaw movements in speech. That is, the open vowel [a] was consistently produced with jaw movements of greater amplitude, longer duration and higher peak velocity than those associated with the vowel [i]. This, was true, in general, for both the hearing and the deaf speakers.

The effect of stress on jaw movements was, on the other hand, much less reliable and consistent. In many cases, there was no discernible difference between stressed and unstressed segments for either group of talkers. However, the hearing speaker reliably differentiated stressed and unstressed vowels by the "hold phase" of jaw lowering. This phase was longer for stressed than for unstressed vowels. Peak velocity of jaw lowering was also reliably higher for the hearing subject and Deaf subject 3; also, Deaf speaker 1 had a higher peak velocity of jaw lowering for the first vowel in the stressed productions.

The hearing subject in the present investigation did not show stress effects on jaw displacement. This differs from previous results for normals [11,12,13,14] and may be due to methodological differences. This study did not use reiterant speech or nonsense syllables. Further, the present results suggest that speakers can choose among different strategies in producing stressed and unstressed segments. Thus, at the articulatory level, the hearing speaker differentiated stressed and unstressed segments by a longer "hold phase" of the jaw gesture in the stressed condition while holding movement times and displacement constant. By inference, vowel duration was longer in the stressed syllables. While the hearing subject differentiated stressed and unstressed segments, the deaf speakers did not do so. At the same time, the deaf speakers showed reliable vowel effects. Overall, the kinematic measures of jaw movements did not differ between the hearing and the hearing-impaired speakers. Only Deaf speaker 3, the least intelligible, differed in that some measures were significantly longer than those for the other subjects.

These data refine some of the notions frequently reported to characterize deaf speech. In fact, deaf speakers do coordinate fairly accurately articulatory movements as evidenced by the results of this study for jaw and lip control. This is not too surprising since movements of the lips and jaw are visible. Moreover, we found no evidence that these hearing-impaired subjects distinguished vowel height by exaggerated jaw displacement. Durations, displacements and peak velocities did not differ remarkably among the subjects in the present study. Thus, the slow speaking rate of the deaf is not necessarily due to the fact that they move their articulators more slowly than hearing speakers. Deaf speaker 3, the least intelligible, had speech that was characterized by pauses between words reflected in the measures of the jaw hold phase and the interval from onset of jaw lowering for the first vowel to offset of jaw raising for the medial consonant. However, it is significant to note that this speaker is not distinguished from the other deaf talkers, or the hearing speaker, in any of the other measures.

Variability has often been reported as one of the hallmarks of the speech of the deaf [15]. The results of the present study do not show any good evidence of such variability in kinematic measures of a highly visible articulator. Moreover, we obtained similar results in our study of laryngeal-oral coordination [16]. We argue that this is the result of examining articulators which inherently have few degrees of freedom such as the jaw or laryngeal abduction/adduction. Measures for an articulator such as the tongue may show more variability. However, the nature of articulatory variability in normal speakers is far from understood and is most likely substantial. Results of stress implementation indicate that several strategies are available. We argue that an understanding of normal articulatory variation is a necessary prerequisite before we can hope to understand speech in disordered groups.

REFERENCES

[1] Rubin-Spitz, J., McGarr, N. S., and Youdelman, K. (1986). "Perception of stress contrasts by the hearing-impaired," J. Acoust. Soc. Am. 79, S10.

[2] Levitt, H., Smith, C., and Stromberg, H. (1974). "Acoustic, articulatory, and perceptual characteristics of the speech of deaf children," in Speech communication, edited by G. Fant (Stockholm: Almqvist and Wiksell), Vol 2: Speech Production and Synthesis by Rules, pp. 126-139.

[3] Osberger, M. J., and Levitt, H. (1979). "The effect of timing errors on the intelligibility of deaf children's speech," J. Acoust. Soc. Am. 66, 1316-1324.

[4] Maassen, B., and Povel, D. (1984). "The effect of correcting temporal structure on the intelligibility of deaf speech," Speech Communication 3, 123-135.

[5] Stein, D. (1980). "A study of articulatory characteristics of deaf talkers," unpublished doctoral dissertation, University of Iowa.

[6] Zimmermann, G., and Rettaliata, P. (1981). "Articulatory patterns of an adventitiously deaf speaker: Implications for the role of auditory information in speech production," J. Speech and Hear. Res. 24, 169-178.

[7] Tye-Murray, N. (1984). "The articulatory behavior of deaf and hearing speakers over changes in rate and stress: A cinefluorographic study," unpublished doctoral dissertation, University of Iowa.

[8] Ling, D. (1976). Speech in the Hearing Impaired Child: Theory and Practice (A. G. Bell Association, Washington, D. C.).

[9] Martony, J. (1968). "On correction of voice pitch level for severely hard-of-hearing subjects," Am. Annals of the Deaf 113, 195-202.

[10] Subtelny, J. (1975). "Speech assessment of the deaf adult," J. Acad. Rehab. Audiol. 8, 110-116.

[11] Stone, M. (1981). "Evidence for a rhythm pattern in speech production: Observations on jaw movements," J. Phonetics 9, 109-120.

[12] Kiritani, S., and Hirose, H. (1979) "Effects of stress on jaw movements in American English," Annual Bulletin, Research Institute of Logopedics and Phoniatrics (University of Tokyo) 13, 53-59.

[13] Macchi, M. (1985). "Segmental and suprasegmental features and lip and jaw articulators," unpublished doctoral dissertation, New York University.

[14] Kelso, J. A. S., V.-Bateson, E., Saltzman, E., and Kay, B. (1985). "A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling," J. Acoust. Soc. Am. 77, 266-280.

[15] Harris, K. S., Rubin-Spitz, J., and McGarr, N. S. (1985). "The role of production variability in normal and deviant developing speech," in Proceedings of the Conference on the Planning and Production of Speech in Normal and Hearing-Impaired Individuals (ASHA Reports 15), edited by J. Lauter (ASHA, Rockville, MD).

[16] McGarr, N. S., and Löfqvist, A. (1982). "Obstruent production by hearing-impaired speakers: Interarticulator timing and acoustics," J. Acoust. Soc. Am. 72, 34-42.

SOME EFFECTS OF COCHLEAR IMPLANTATION ON SPEECH PRODUCTION

ANNE-MARIE ÖSTER

Dept. of Speech Communication & Music Acoustics
Royal Institute of Technology (KTH), Box 70014
S-100 44 Stockholm, Sweden

ABSTRACT

A new method for the treatment of acquired total deafness in adults is under probation in Sweden since 1983. The Vienna Cochlear Prosthesis is an extra-cochlear system comprising a single-channel implant with its active electrode placed in the round-window niche. The device functions on the basis of electrical stimulation of the cochlear nerve.

The present study reports on acoustical analyses of fundamental frequency of two patients' recorded readings of a familiar text consisting of 89 words and an unfamiliar text of 56 words. The recordings were made pre-implant and post-implant after 1, 3, 6, 12 and 24 months. We have also made recordings of the patients when they read the text without and with the implant.

The analyses made included the speech rate, phonation time as well as the mean and the standard deviation of the fundamental frequency. The results are shown in forms of F0-histograms. The main effect found is an improvement in F0-control which means a lowering in mean F0 and a more normal F0-distribution. A shift towards a more normal rate of articulation is also found.

INTRODUCTION

A cochlear implant is a new technical aid for the deaf based on direct electrical stimulation of the auditory nerve. An implant provides only limited auditory information but still most implanted patients report benefits by the implant. Hochmair-Desoyer /1/ gives details about a questionnaire to patients with one year's experience of a single-channel implant. The patients derived benefits in:

1) Provision of environmental (non-speech) sound.
2) Provision of speech sounds as an aid to lip reading.
3) Improved speech production.
4) Reduced awareness of tinnitus through distraction and/or suppression effects.

There are only a few reported studies which deal with aspects of speech production rather than speech perception of cochlear implantation with deafened adults. One reason to this research orientation is perhaps that.."many persons with an acquired hearing loss, as the result of an infection or through accident, continue to have great problems in receptive communication even after extensive training and the selection and fitting of prosthetic aids. In contrast the effects of acquired deafness on speech quality are far more subtle... Changes when they do occur tend to be acquired gradually." /2/. Nevertheless, those studies in the cochlear implant literature dealing with adventitiously deaf patients' speech production have reported some improvements in speech production after implantation.

CHANGES IN SPEECH WITH USE OF AN IMPLANT

In a study of four patients Iler-Kirk and Edgerton found that an improvement in voice parameters and fundamental frequency had taken place after implantation /3/. Waters studied the speech production of three cochlear implant wearers /4/. He assessed their speech pre-therapy and post-therapy and post-therapy while using the implant for six months. All three patients were judged to have improved production of speech after using their implants for six months. Not only the voice quality, which became less harsh and tense, improved but also overall timing and pitch control. Ball and Ison reported on a patient that showed a frequency range that approached normal with marked reduction in irregularity after electrical stimulation /5/. East and Cooper used a questionnaire one year following implantation for assessing the device /6/. The implant wearers and their families remarked that the improved modulation of speech volume when using the implant led to increased self-confidence. Plant and Öster (ref. /2/) found in a case study of a Swedish female speaker two years after implantation that ..."a number of changes had occurred after implantation. At the prosodic level these included a more normal range of fundamental frequency, improved F0-control in signalling emphatic stress contrasts and improvements in durational aspects."

An interesting and important question is, however, whether this improvement is attributable to speech training effects or to the information provided by the implant. Iler-Kirk and Edgerton (ref. /3/) analyzed speech samples of two men and two women reading a standard passage with and without their implants. The aided condition for the male patients resulted in a lower mean F0 and reduced variability in intensity level. The female patients showed a higher mean F0 and an increase in intensity variability. These results represent positive changes for all patients compared to normally hearing persons of the same sex.

METHOD

Subjects

A cochlear implant is of interest only to those patients who are deaf but still have an active cochlear nerve. No benefit from hearing aids is a criterion for operation as well as a strong motivation. The implanted patients must have auditory memories which exclude those who are born deaf. The two patients in this paper are a man born in 1930 and a woman born in 1955. They were both deafened post-lingually, the man at the age of 17 and the woman at the age of 26. Both had used their extra-cochlear implants for two years at the time of the last recording of their speech. Since the operation, they have obtained speech training one hour weekly including breathing and phonation exercises.

Materials

The recordings were made when the patients read a standard passage of 89 words. This text became over time very familiar to them since they read it at the time of recurrent testing that occurred prior to operation and 1, 3, 6, 12 and 24 months after operation. The two patients were recorded when they read the standard passage and an unfamiliar text of 56 words two years after implantation in order to study the direct feedback effect of the implant.

Recordings

The recordings were made in a sound-treated test room. A TEAC A-3340 four-channel tape recorder was used. A contact microphone attached to the patients' trachea recorded the synchronous larynx signal.

DATA ANALYSIS

Instrumental analysis

The fundamental frequency behavior was analyzed by using a computer program developed by S. Ternström of the Dept. of Communication and Music Acoustics. This fundamental frequency distribution analysis uses the tape-recorded signal from the contact microphone placed at the patient's larynx. The program gives graphic printouts of F0-contours as a function of time (pitch contour), mean F0, the most common frequency, standard deviation and F0-histograms with statistics.

RESULTS AND DISCUSSION

Fundamental frequency analysis

The results of the F0-analysis of the pre-implant reading and the three readings up to two years after implantation are presented in Table I for the female patient and Table II for the male patient. The measures show a considerable lowering in mean F0 for both patients. The pitch becomes more correct compared to normally-hearing persons of the same age and sex. There is also an noticeable decrease in standard deviations (SD) that, especially for the male patient, means that his intonation has improved. His intonation pre-implant was uncontrolled and extremely lively.

F0-histograms obtained for the standard passage pre- and post-implant of the female patient are presented in Fig. 1A. Fig. 2A shows the F0-histograms of the male patient. In Fig. 1A the pre-implant histogram is more positively skewed than the post-implant histogram two years after operation. A positive skewness is typical for normal distribution (see ref. /2/, p. 71). The post-implant histogram is more symmetrical due to a tendency towards instability and unperiodicity in the vocal cords that can be seen in the area around 100-150 Hz. Fig. 2A presents a post-implant histogram that is more peaky than the pre-implant histogram that is more diffuse indicating a wider range of commonly occurring frequencies.

	Mean F0	SD (Hz)
Pre-implant	265.39	53.40
6 months post-impl.	221.72	43.93
12 months post-impl.	238.65	45.81
24 months post-impl.	219.67	46.15

Table I. F0-measures for the female patient.

	Mean F0	SD (Hz)
Pre-implant	161.48	36.38
6 months post-impl.	141.71	23.95
12 months post-impl.	138.49	23.92
24 months post-impl.	144.40	27.72

Table II. F0-measures for the male patient.

Durational aspects

The F0-analysis also calculates the total duration of the speech sample, the per cent of pauses over 200 msec and the per cént of pauses under 200 msec. From these measures, the phonation time can be estimated which is done in Tables III and IV.

	Duration, sec	Phon.time, sec
Pre-implant	38.43	22.60
6 months post-impl	40.22	20.12
12 months post-impl	36.05	17.71
24 months post-impl	34.78	17.87

Table III. Duration time for the female patient's readings of the standard passage.

	Duration, sec	Phon.time, sec
Pre-implant	27.24	14.46
6 months post-impl.	37.69	20.45
12 months post-impl.	33.01	16.21
24 months post-impl.	36.64	17.31

Table IV. Duration time for the male patient's readings of the standard passage.

The results show that the two patients' phonation times 24 months after implantation have moved towards normal values. Their results can be compared with that obtained by a normally hearing 40-year old female, native speaker of Stockholm Swedish. In this case the phonation time was 16.77 sec. The results obtained two years after implantation for both patients indicate, therefore, a shift towards a more normal rate of articulation. The male patient's speech pre-implant was very fast and mechanical. Already after 6 months' post-implant, an extension in duration and phonation time can be observed. Over time he manages to control pausing and phrasing that improves his intonation considerably.

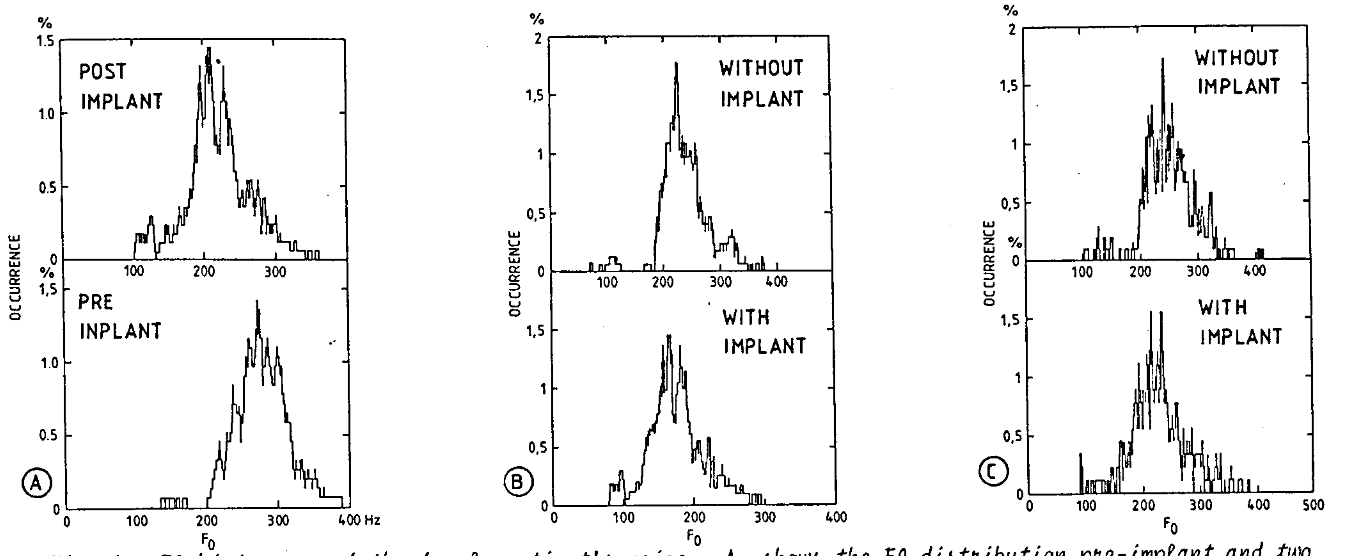


Fig. 1. F0-histograms of the female patient's voice. A: shows the F0-distribution pre-implant and two years after implantation. B: shows the patient reading a familiar text without and with the implant. C: shows the patient reading an unfamiliar text without and with the implant.

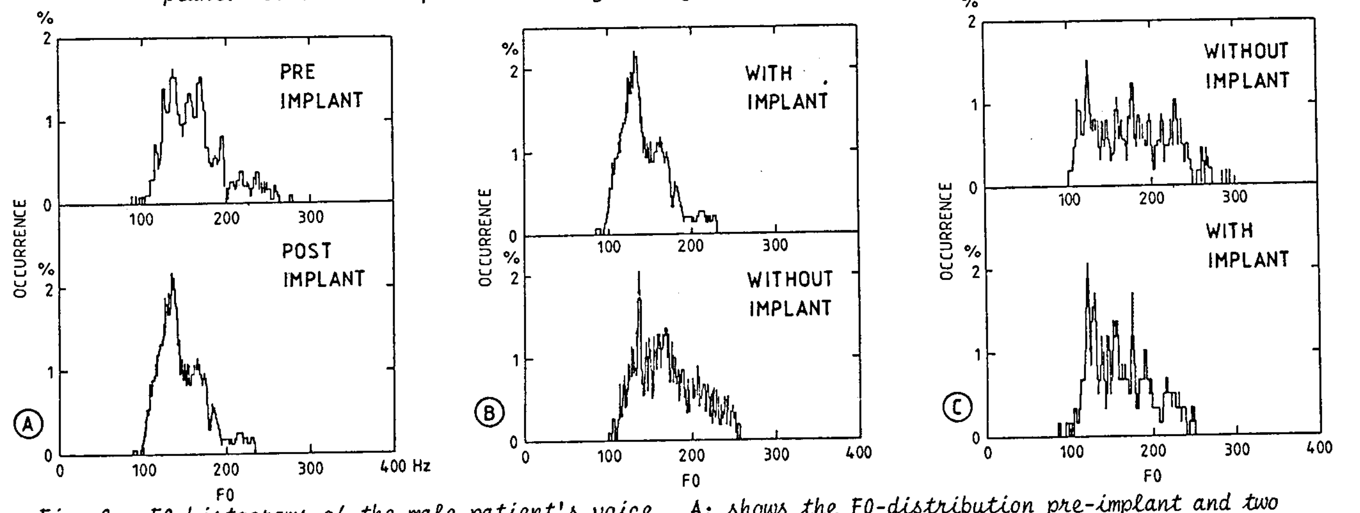


Fig. 2. F0-histograms of the male patient's voice. A: shows the F0-distribution pre-implant and two years after implantation. B: shows the patient reading a familiar text without and with the implant. C: shows the patient reading an unfamiliar text without and with the implant.

FAMILIAR TEXT	Mean F0 (Hz)	SD (Hz)
Without implant	239.58	41.95
With implant	219.67	46.15
UNFAMILIAR TEXT		
Without implant	252.55	46.07
With implant	229.61	50.01

Table V. Female patient's readings.

FAMILIAR TEXT	Duration, sec	Phon.time, sec
Without implant	37.90	18.76
With implant	34.78	17.87
UNFAMILIAR TEXT		
Without implant	25.61	11.23
With implant	22.87	9.53

Table VI. Female patient's readings.

FAMILIAR TEXT	Mean F0 (Hz)	SD (Hz)
Without implant	170.58	38.23
With implant	144.40	27.72
UNFAMILIAR TEXT		
Without implant	176.86	45.45
With implant	163.97	39.50

Table VII. Male patient's readings.

FAMILIAR TEXT	Duration, sec	Phon.time, sec
Without implant	36.09	11.96
With implant	36.64	17.31
UNFAMILIAR TEXT		
Without implant	20.95	7.07
With implant	21.00	6.27

Table VIII. Male patient's readings.

Direct feedback effect of the implant

The implant gives the subjects a direct feedback of their own voice production. To study this effect an analysis was made on the readings with and without implant, both on familiar text and unfamiliar text. The day of recording the patients came in the morning without their implants on. They read both texts without implants before they put them on and adjusted them to an appropriate level. After half an hour when the patients chatted with the experiment leader, the aided readings were recorded. In Tables V-VI the results for the female patient are shown and in Tables VII-VIII for the male patient. Fig. 1B and 1C shows the different histograms for the female subject's readings of the familiar and the unfamiliar texts. The histograms of the male subject's readings are presented in Fig. 2B and 2C.

The results from Tables V-VIII and Figs. 1B and 1C and 2B and 2C show that both patients' fundamental frequencies decrease when the implants are switched on. There are also changes in duration and phonation time for both patients in the aided readings that indicate more normal values. The histograms of the man's voice show in the two aided readings a more peaky distribution that indicates a more controlled behavior. The histograms of the woman's voice, Fig. 1B, however, shows distributions that in the aided readings become more symmetrical that indicates that her voice sometimes becomes unstable and creaky.

FINAL DISCUSSION

The results of the F0-analyses in this study show that both patients derived benefits in improved speech production thanks to their single-channel implant. The follow-up recordings show that the improvements are immediate and permanent. The most noticeable change seems to be a lowering in mean fundamental frequency to a more normal value considering age and sex. The implant provides some limited spectral information especially in the low frequencies, timing and intensity. The female patient s voice becomes sometimes with the implant very pressed and creaky. This is probably due to the fact that she strains her voice in order to get some low-frequency feed-back.

The most positive changes, however, occur in the male patient's speech. The benefits for this patient appear to derive from timing information provided by the implant. The pre-implant recording reveales a high tempo together with an uncontrolled intonation. In the part of the experiment when the implant is switched off, he returns to this way of speaking. After implantation it is obvious that the patient modifies and plans his speech production consciously. The possibility that the improvements in speech production is a result from the training provided can probably be excluded. All the measures and the histograms show that the implants have a direct feedback effect on both patients' speech production.

The implant is superior to ordinary speech-training devices, as for instance visual indicators, which are big, heavy and limited to clinical use. The patient will very often relapse into old habits as soon as he leaves the clinic. As the implant is wearable, it is always present and offers continuous training and monitoring to the patients.

Like previous studies dealing with the speech production of cochlear implant wearers, this study shows that improvements in durational aspects, a more normal range of fundamental frequency and an improved F0-control occur with an implant. The implant is an effective speech device as it offers feedback and voice control.

REFERENCES

/1/ I.J. Hochmair-Desoyer, "Fitting of an analogue cochlear prosthesis. Introduction of a new method and preliminar findings", British J. of Audiology 20, 45-53, 1986.

/2/ G. Plant, A-M. Öster, "The effects of cochlear implantation on speech production. A case study", STL-QPSR 1/1986 (KTH, Stockholm), 65-85.

/3/ K. Iler-Kirk, B.J. Edgerton, "The effects of cochlear implant use on voice parameters", Oto Laryng. Clinics of North America 16, 281-292, 1983.

/4/ T. Waters, "Speech therapy with cochlear implant wearers", British J. of Audiology 20, 35-43 , 1986.

/5/ V. Ball, K.T. Ison, "Speech production with electro-cochlear stimulation", British J. of Audiology 18, 18, 1984.

/6/ C.A. East, H.R. Cooper, "Extra-cochlear implants: the patient's viewpoint", British J. of Audiology 20, 55-59, 1986.

THE IDENTIFICATION OF SYNTHETIC VOWELS BY PATIENTS USING A SINGLE-CHANNEL COCHLEAR IMPLANT

EVA AGELFORS AND ARNE RISBERG

Department of Speech Communication and Music Acoustics
Royal Institute of Technology (KTH), Box 70014
S-100 44 STOCKHOLM, SWEDEN

ABSTRACT

Vowel perception in a /bV:b/-context of patients using a single-channel extra-cochlear implant developed in Vienna has been studied by means of synthetic speech. Three patients were asked to adjust the first and second formant of a synthetic vowel sound so that they perceived the sound as a given long Swedish vowel. To study the effect of training, the experiment was made at two sessions about a year apart. Normal hearing listeners identified the vowel sounds generated by the subjects. The results of the experiment were analyzed in confusion matrices and as F1-F2 plots.

A clear effect of training could be seen. Three years after implantation the best subject could adjust the frequency of F1 and F2 quite close to the correct values.

INTRODUCTION

It has long been known that stimulation of the auditory nerve with a weak electric current results in auditory sensation. During the late fifties the first experiments were made to use this effect in an aid for the deaf. Since the beginning of the seventies, House at the Ear Research Institute in Los Angelses has been implanting deaf subjects with a simple single channel cochlear implant /1/. At the same time research and development has been going on at several laboratories both on single-channel and multi-channel devices. In a single-channel cochlear implant, the electrode is either placed in the middle ear close to the round window, or inserted a few millimeters into the cochlea. In a multichannel device, the electrodes are placed in the cochlea at different positions along the basilar membrane.

In a single-channel implant, it seems that only time-intensity information in a speech signal can be transmitted. In a multiple-channel device, some frequency selectivity might be obtained by stimulating different nerve endings along the basilar membrane.

During the last years very good speech understanding without the support of lipreading has been reported from subjects using both single-channel and multi-channel devices. The good speech understanding reported from subjects using single-channel devices seems to indicate that they have some possibility to identify vowels. This has been studied by Doyle et al. /2/. They used natural vowels that had been equalized in duration and loudness. They concluded that "The vowel features of F0 and/or F1 (or information related to them) accounted primarily for the vowel confusions made by these single-channel cochlear implant subjects".

Dent /3/ studied vowel discrimination by subjects that used the same implant as in the above study (the House implant, ref. /1/). The results show that "monosyllables containing high back syllable nuclei and, to a lesser degree, those containing high front syllable nuclei, can be distinguished from monosyllables containing low syllable nuclei".

White /4/ studied vowel discrimination by one subject using another type of .intra-cochlear single-channel implant. He found a clear evidence that the subject used first formant information in discriminating between synthetic vowel pairs or identifying vowels in natural speech. However, he did not find any evidence that the subject could use information from the frequency of the second formant. That this to some extent is possible has been reported by the Vienna group /5/. The aim of the experiment described here is to shed more light on this problem.

METHOD

In the Swedish cochlear implant project, a single-channel implant developed in Vienna and manufactured by 3M in the USA is used /6/. The project is run at the Department of Audiology of the South Hospital (Södersjukhuset), Stockholm, in cooperation with our department. After implantation the subjects go through a longer, structured training and test program. Testing is made 1, 3, 6, 12, 24, and 36 months after surgery. In the test battery measurements of frequency and time discrimination and speech-perception ability with and without simultaneous lipreading are included. Here results on a vowel-identification experiment with three subjects is reported.

Subjects

Subject 1 was born in 1955. She had a progressive hearing loss that resulted in total deafness in 1981. She was implanted in 1984. She is an excellent lipreader and two years after implantation she had achieved some ability to understand speech without simultaneous lipreading.

Subject 2 was born in 1930. He become deaf in 1947 as a result of meningitis. He is a poor lipreader and has never used a hearing aid. He was implanted in 1984.

Subject 3 was born in 1932. He become deaf due to the effect of an ototoxic drug in 1977. He is a resonably good lipreader and uses a hearing aid in his left ear. His right ear was implanted in 1986.

Effect of training

Directly after implantation the subjects have great difficulties in using the information from the implant but this ability gradually develops. As an example, Fig. 1 shows frequency discrimination ability for a sinusoidal signal with the frequencies 125, 250, 500, 1000, and 2000. The figure shows the results for subject 2 at the test sessions 1, 3, 6, 12 and 24 months after implantation. From the beginning changes in frequency could only be detected for the lowest frequency, 125 Hz, but after 12 months, frequency discrimination ability was around 2-5 % for frequencies up to 1000 Hz. In the same way, the ability to identify speech sounds gradually improves.

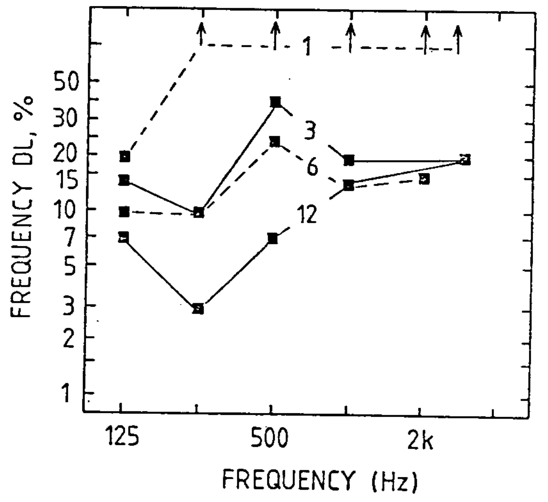


Fig. 1. Results from measurements of frequency discrimination ability for a sinusoidal signal by subject 2 at test sessions 1, 3, 6 and 12 months after implantation.

Test equipment and test procedure

In the experiment the text-to-speech equipment based on the OVE II synthesizer was used /7/. This is a cascade synthesizer with four-formant circuits. The frequency of the two lowest formants could be controlled by a joy-stick and at the same time the subjects could see the movements of the formants on an F1-F2-display. No frequency calibration was given in the plot, and none of the subjects had any knowledge in acoustic phonetics. Formant three and four were set by the computer program at the frequencies that was typical for the intended vowel, see Table I. The test vowels were the long Swedish vowels in context /bV:b/. The formant frequencies of the nine vowels are shown in Table I. The fundamental frequency variation was the same for all vowels.

The cochlear implant was coupled to the synthesis equipment over the line input of the implant. It was explained to the subjects that they by moving the joy-stick were to locate a specific vowel on the display. By pressing the space bar they could listen to the vowel in a /bv:b/ context. They were then allowed to play around with the joy-stick for some time. When we were sure that they had understood the task, one of the vowels was randomly selected and given in orthographic form in a syllable /bV:b/. The syllable was presented once with the formant frequencies of the vowel given in Table I. The subjects were then asked to adjust the joy-stick until they thought that the intended syllable was produced. Whenever they liked, they could listen to the synthetic syllable they had adjusted but they could not listen to the target syllable. They had to use their internal memory of the intended target.

Table I. Formant frequencies for the nine long Swedish vowels used in the experiment.

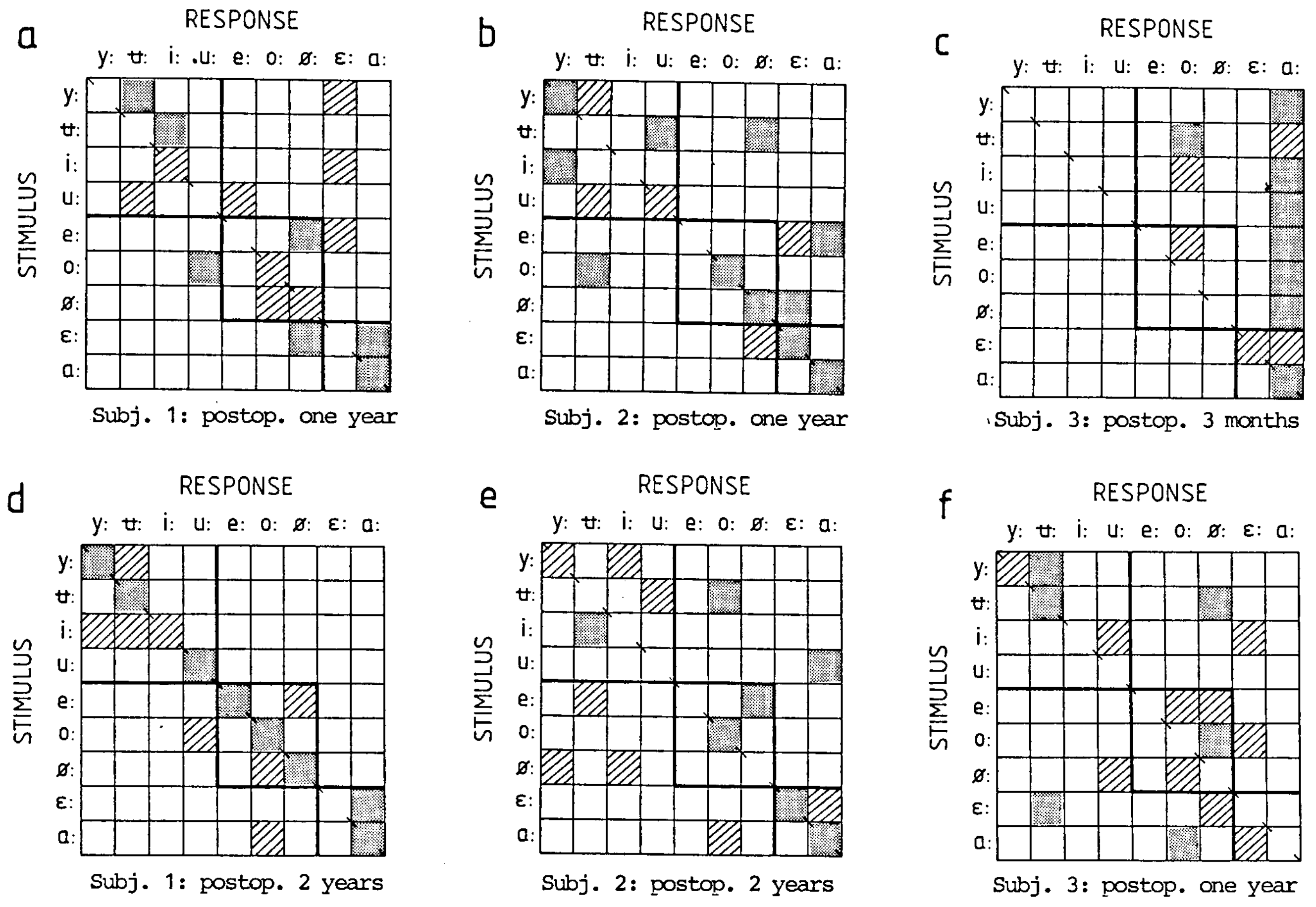
Vowel	Formant frequencies, Hz.			
	F1	F2	F3	F4
[ɑ:]	653	1000	2500	3200
[e:]	350	2200	2800	3450
[i:]	280	2200	3000	3700
[u:]	350	770	2800	3300
[ʉ:]	350	1750	2450	3150
[y:]	300	2000	2400	3400
[o:]	390	700	2400	3250
[ø:]	380	1750	2300	3350
[ɛ:]	450	1975	2550	3400

The matching for all nine long Swedish vowels presented in random order was made twice during each test session. The subjects were not told how well they had been able to adjust the formant frequencies of the intended vowel. For subjects 1 and 2 testing was made 12 and 24 months after surgery and for subject 3, three and 12 months after surgery. To get reference data five normal hearing subjects from our department were tested with the same program. They all had some knowledge in acoustics phonetics but as no frequency calibration was shown on the F1-F2 plot, it was difficult for them to use previous knowledge. The same five subjects were used in a listening test where the task was to identify the vowels in the syllables produced by the implanted subjects.

RESULTS AND DISCUSSION

The results from the experiment are shown as gross confusion matrices for the three subjects in Fig. 2a-2f. In the matrices, the vowels are arranged after increasing F1 based on the means of the results from the normal hearing subjects. The submatrices for vowels with about the same F1 are indicated. In Fig. 3a-3d the F1-F2-plots for subjects 1 and 3 are shown and the area where the normal hearing subjects placed their vowels is indicated.

A clear improvement over time can be seen especially in the results for subjects 1 and 3, see Fig. 3. This does not result in an increase in per cent correct identified vowels in the listening test with normal hearing persons. Three months after implantation subject 3 has almost no ability to identify the vowels but one year after implantation he makes clear differences in adjusting the formant frequencies but most of them are



Subj. 1: postop. one year

Subj. 2: postop. one year

Subj. 3: postop. 3 months

Subj. 1: postop. 2 years

Subj. 2: postop. 2 years

Subj. 3: postop. one year

Fig. 2a-f. Results from listening tests with five normal hearing persons. Gross vowel confusion matrices made by three cochlear implant subjects in the experiment at different times after implantations. The vowels are arranged after increasing frequency of the first formant.

▩ 50-100%   ▨ 25-49%   □ 0-24% correct

still far away from the correct values. Three months after implantation 17.4% of his vowels were correctly identified in the experiment and 12 months after implantation 9.6% was correctly identified. Fig. 2f and Fig. 3d, however, show that he adjusts the frequency of the first formant close to the correct value.

The results of subject 1 are good 12 months after implantation and are improving one year later. It is clear that she uses F2-information in finding the vowel on the display, see the vowels /i:/ and /y:/ in Fig 2a. All F2-values for these vowels are placed at high frequencies. For one /i:/- and one /y:/-vowel, however, F1 is placed too high. In the experiment three years after implantation almost all vowels are placed close to the correct targets.

Subject 2 has limited ability to perceive vowel information and he does not show any improvement in the experiment two years after implantation.

With subjects 1 and 2 a simple pitch-scaling experiment has been made. Subject 1 can scale pitch up to 2000 Hz but subject 2 only up to about 700 Hz.

It is apparent that a single-channel cochlear implant presents the central processor with an abnormal pattern of the acoustic signal. Timing information is reasonable well preserved but frequency information is very different from normal. From the beginning the subjects have great difficulties in interpreting the information but through learning they gradually improve their ability. Variations are, however, great between the subjects.

ACKNOWLEDGEMENTS

We would like to thank our subjects for their willingness to participate in an experiment that for some of them was very difficult. Special thanks to Lennart Nord for his help with the computer program The cochlear implant project at the South Hospital is supported by the Swedish Medical Research Council (project no 07500).

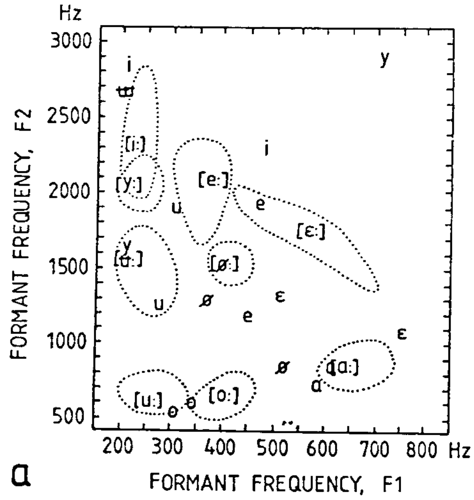


Fig. 3a. Subject 1: postop. one year

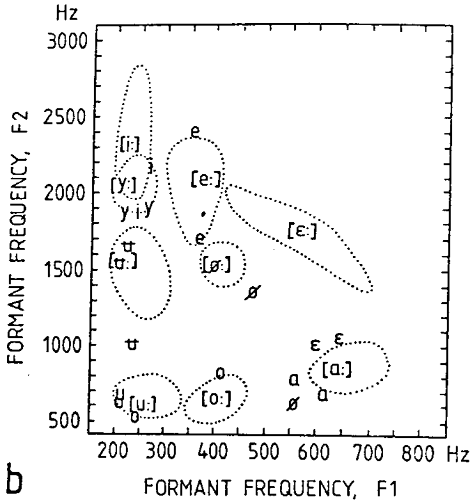


Fig. 3b. Subject 1: postop. two years

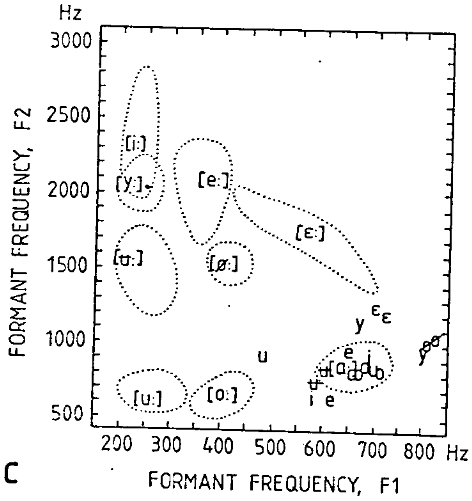


Fig. 3c. Subject 3: postop. three months

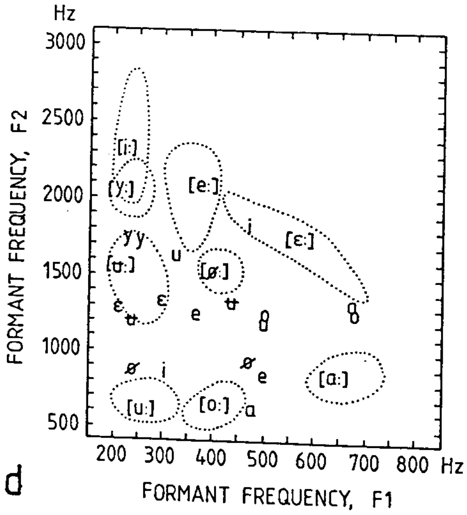


Fig. 3d. Subject 3: postop. one year

The normal hearing subjects' mean values are placed within square brackets.

REFERENCES

/1/ W.F. House, "Cochlear implants", Ann. of Otology, Rhinology and Laryngology, Suppl 27, 85, No 3:2, 1983.

/2/ K.J. Doyle, J.L. Danhauer, B.J. Edgerton, "Vowel perception: Experiments with a single-electrode cochlear implant", J Speech and Hearing Res., 29, 179-192, 1986.

/3/ L.J. Dent, "Vowel discrimination with the single-electrode cochlear implant: A pilot study", Ann. of Otology, Rhinology and Laryngology, Suppl 91, 91, No 2:3, 41-46, 1982.

/4/ M.W. White, "Formant frequency discrimination and recognition in subjects implanted with intracochlear stimulating electrodes". In C.W. Parkins and S.W. Anderson (Eds.), Cochlear protheses: An Internat. Symp. Annals of the New York Academy of Sciences, 405, 348-359, 1983.

/5/ E.L. v. Wallenberg, I.J. Hochmair-Desoyer, E.S. Hochmair, "Speech processing for cochlear implants", Proc. Seventh Annual Conference of the IEEE/Engineering in Medicine and Biology Society, Chicago, 1114-1118, 1985.

/6/ I.J. Hochmair-Desoyer, E.S. Hochmair, K. Burian, R.E. Fischer, "Four years of experience with cochlear prostheses". Medical Progress through Technology 8, 107-119, 1981.

/7/ R. Carlson, B. Granström, "Linguistic processing in the KTH multilingual text-to-speech system", Conf. Record, IEEE-ICASSP, Tokyo.

× × × × ×

PHONETICALLY BASED NEW METHOD FOR AUDIOMETRY: THE G-O-H MEASURING SYSTEM USING SYNTHETIC SPEECH

MÁRIA GÓSY GÁBOR OLASZY JENŐ HIRSCHBERG ZSOLT FARKAS

Dept. of Phonetics
Linguistics Institute
Budapest 1250 Pf. 19. Hungary

Heim Pál Children's
Hospital
Budapest 1089 Hungary

ABSTRACT

A special phonetically based method in generating artificial monosyllabic words was developed for audiometric measurements. Using this synthesized material, the hearing capacity and the speech perception and understanding level of children can be judged easily. It is very important to examine regularly the hearing capacity of children and the evaluation of speech understanding level from the point of view of learning writing and reading in school. The speciality of our artificial words is in their low-redundancy acoustic structure and in the special frequency structure.

INTRODUCTION

There is a close connection between the articulation and perception bases of the process of speech acquisition. The initial development of perception and understanding abilities precedes that of speech production, but this difference between them subsequently decreases: their further development is assumed to take place in a permanent interaction. The bases of speech perception and understanding is hearing; this does not mean, however, that good hearing automatically ensures the normal processing of speech perception/understanding. That is why regular examination of hearing and understanding is very important, particularly in the early years when acquisition of the mother tongue is in progress. The identification of speech production problems is easier than that of speech perception/understanding ones. The normal communication situations provide a better opportunity for adults to detect the speech errors of children, revealing articulatory or grammatical problems. However, perception and/or understanding/comprehension difficulties can remain hidden because of various supplementary and compensatory strategies of children. This fact leads to delayed diagnosis and to difficulties in carrying out the appropriate corrective procedures. There are a lot of well-known problems related to the hearing measurements and mass screening of children between the ages of 3 and 7. What are the criteria that a suitable method for auditory screening of these small children has to meet? First: the signal that is given to the child's ear should be natural and familiar for him. Second: the measuring task should be easy to understand, that is, we should make it easy for the child to understand what he has to do during examination. Third: the measuring method should yield the highest possible amount of information about the hearing mechanism operative between 200-8000 Hz. These expectations are all met by our new screening procedure: the G-O-H system.

METHOD AND MATERIAL

On the basis of the results of a perceptual examination of Hungarian speech sounds whereby the values of their invariant cues are determined [1], the process of speech understanding can be further studied: the hearing mechanism and the level of recognition of words can be examined. The examination of the two processes can be combined if we produce speech material which only involves acoustic values corresponding to invariant features (or hardly more than that). This condition is satisfied by computer-generated, artificial speech based on perceptual data.

Speech as an auditory stimulus is familiar for children, and repeating sound-sequences is a natural task when the child acquires his first language, and repeats the words of his mother. Human speech is suitable for judgement of understanding level, but the speech-audiometric results cannot give exact data about the hearing capacity or about the extent and type of impairement, because natural speech is very redundant as to its frequency structure. The redundancy of speech means that speech sounds contain far more building elements than would be necessary for understanding. That is why natural speech can be understood in the case of certain hearing impairements: the redundant elements give an opportunity to guess the meaning. Our specially synthesized words contain only the necessary frequency components of each sound. The difference between the natural and synthesized words is only in the redundancy of the frequency structure (cf. Fig. 1). In spite of this difference they sound very similar. The speech perception threshold proved to be the same as the normal one obtained by using natural words. How does the screening function of the artificially produced words work? Let us suppose that the system of speech understanding has to analyse data of quantity *x* to understand the word *bus*. But the acoustic structure of natural speech is highly redundant, i.e. it contains significantly more information (data) in the speech than is necessary for its safe recognition. But the word *bus* contains data of quantity *x+y*. The data surplus (*y*) becomes stored and can be immediately called out in case of any kind of 'disorder' (e.g. noise), to provide supplementary information. On the other hand the word *bus* we synthesize hardly contains more information than the necessary quantity *x*. Therefore, in case there is some 'disorder' at any point in the recognition process (hearing loss, central problem), *x+y* would make identification possible, but *x* itself does not, where comprehension will be mistaken (to some extent). *The comprehension of a signal sequence containing information x requires the processing of all information in a perfectly sound fashion*, e.g. by the help of normal hearing.

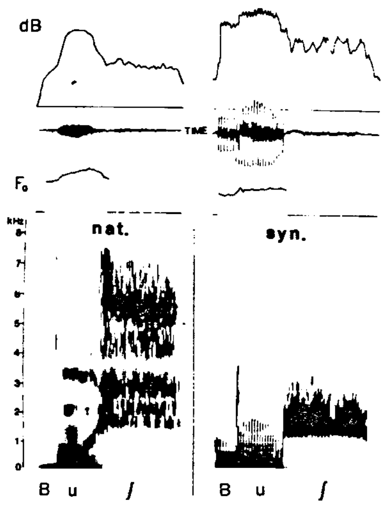


Fig. 1. Acoustic structures of natural and synthesized German word *Busch*

To provide a basis for the G-O-H method, a special test material was constructed which involved 44 meaningful monosyllabic Hungarian words. The criteria for choosing the words were as follows: (i) the monosyllabic words should have two or three speech sounds without consonant clusters, (ii) the words should contain speech sounds where the frequency parameters seem to serve as an acoustic cue for their identification, (iii) the test material includes three types of words containing only high-frequency sounds (like [sy:z]; words containing onl/ low-frequency sounds (like [bu:]); words containing both high and low frequency sounds (like [bus]); (iv) an effort was made to collect a material exhibiting almost all Hungarian speech sounds in different sound-contexts and phonetic positions, (v) most of the words should be familiar for children of ages 3-7, however, the sample should also include a few items that are meaningless sound-sequences for the children. Attention was also payed to the order of the words: low-frequency and high-frequency words alternate with one another. So all children have an experience of success, because they can understand and repeat correctly at least every second word. This is important for good co-operation between the child and the examiner.

The three frequencies (500, 1000, 4000 Hz) used in pure tone audiometry seem to be very insufficient for the evaluation of speech understanding level. In normal hearing the acoustic information received at these frequencies accounts for some 60% of understanding. This means that, in cases of hearing losses at other frequencies, the child - screened and judged to have normal hearing - cannot understand speech correctly. Our G-O-H method solves this problem as well. Experiments were carried out with the G-O-H test material in clinic and day-care-centers with the participation of 400 normal-hearing and 150 hearing impaired children. People with normal hearing understand both human speech and the special synthesized artificial words equally well. But the hearing impaired patients cannot correctly understand the synthesized words because of the lack of redundant building elements.

RESULTS

Speech synthesis gives us an opportunity to define the desired frequency bands in speech sounds. These facts lead to the perceptual/understanding differences between the normal hearing and impaired hearing listeners. For example, a high frequency hearing impaired child with hearing loss above 5000 Hz cannot understand the word [se:l] 'wind' if the noise component of the initial consonant of the word is above 5000 Hz. In this case, the child receives acoustic information that he identifies as consonant like [f,h] or [t] depending on the extent of the hearing loss of the child. In the case of a slight hearing loss above 5000 Hz, the child will identify the sound-sequence as [fe:l] 'he is afraid of sg' which is an existing Hungarian word. In the case of somewhat more severe loss of hearing above 5000 Hz, the child will identify the sound-sequence as [he:l] which has no meaning in Hungarian, and with more severe loss he will identify, instead of the spirant [s], a stop consonant like [t,p] or [k]. In the case mentioned, the child identifies the word as [te:l] 'winter', because it is a frequent item in children's vocabulary (Figure 2 shows a German example). From the answers of the listeners' judgements can be made about the extent of their hearing losses. The answers are regular consequences of hearing losses, they are not results of imagination. (Experiments with filtered words confirmed us about these regular changes in perception.) Mass-measurements were carried out together with a control pure-tone audiometry corresponded to the results gained by the G-O-H method. This method, however, gave information about the understanding level as well. (In some cases the child did not co-operate in pure-tone audiometric examination, but he repeated the artificial words of the G-O-H device, so his hearing could be measured.)

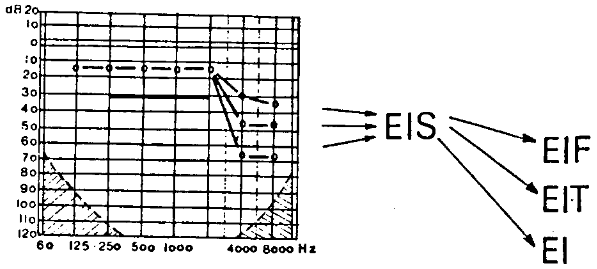


Fig. 2. Changes of possible responses in the case of different hearing losses

The possible answers of normal-hearing and impaired listeners were predicted theoretically on the basis of perceptual investigations concerning the acoustic cues of Hungarian speech sounds. Then, laboratory and clinical experiments were carried out, and the theoretically established types were slightly revised on the basis of the data obtained. Finally, the possible answers were arranged on an answer sheet according to the degree of hearing losses (Table 1).

Twenty monosyllabic German words have been developed so far in our laboratory. These synthesized words are suitable for application the G-O-H method in German too. Experiments were carried out with German-speaking children in kindergartens in Austria.

For the everyday use of this procedure, a measuring set has been developed. This portable case contains a playing system, a tape with the artificial words, an amplifier, a headphone and answer sheets. More than 150 of these sets with G-O-H system are being used in Hungary. To carry out the examinations there is no need for any expert, physician, phonetician or audiologist; it can be used by nurses, kindergarten teachers, speech therapists and so on.

The G-O-H method is a good tool for (i) finding out whether the child is mature enough to acquire writing and reading, (ii) detecting hearing problems, (iii) learning if the child with speech errors has perceptual problems too or not, (iv) detecting dyslexia, i.e. the disturbance in writing and reading.

Table 1. The Hungarian answer sheet (part) of the G-O-H system for measuring the hearing capacity from 200 Hz to 8000 Hz

Sor-szám	I. Jó hallás	II. Hallás- vagy beszédmegértési zavar állhat fenn, esetleg figyelmetlenség. A vizsgálat megismétlendő.	III. Hallászavar valószinű, orvosi vizsgálat javasolt.	IV. Hallászavar biztosra vehető, mielőbbi klinikai vizsgálat szükséges.
1.	meggy (megy)	begy legy vegy negy	egy	bó od e ó –
2.	síp (sík)	sít sűt sűp szíp szép	zúg suk sut su só víd fut hó	kút út tú ú –
3.	bú	dú bók but bot bó pók pú púk dú gú	tú tó pú pó út	ó ú –
4.	ász	ház pász áz	ás ágy áll áj	áf áh át áp á a ó ú –

The possible answer categories are the following: I. good hearing; II. slight hearing problem; III. hearing problem is probable; IV. hearing problem is certain, urgent clinical examination is required.

REFERENCE

[1] Gósy, Mária: Magyar beszédhangok felismerése, a kísérleti eredmények gyakorlati alkalmazása. Magyar Fonetikai Füzetek 15. 1986, 3-100.

ENCODING WITHOUT GRAMMAR: PHONIC ICONISM IN ENGLISH

ROGER W. WESCOTT

Linguistics Program
Drew University, Madison, NJ 07940, U.S.A.

ABSTRACT

In recent decades, interest in phonic iconism (or sound-symbolism) has revived. Phonologists have "re-discovered" direct mapping from sound to sense in the absence of conspicuous and arbitrary grammatical mediation.

Although phonic iconism is presumably detectable in all spoken languages, it is most easily demonstrated in languages that are widely used, well recorded, and intensively analyzed. Consequently, most examples of iconism in this presentation will be drawn from spoken English.

The presentation will conclude with citation of analogous examples of phonic iconism from other languages, some of them Indo-European and some non-Indo-European.

MICROLANGUAGE AND ALLOLANGUAGE

Microlanguage is the name given by George Trager to that core of spoken language which is subject to obvious and well known grammatical rules. /1/ In the view of Transformational linguists, this core is, in fact, the whole of language. But Trager also recognizes prelanguage, or "baby talk;" paralanguage, including exclamations; and metalanguage, or verbal art. Although he offers no cover-term for these three domains of speech, I refer to them as allolanguage and define them as speech that violates the rules of canonical utterance (while, in some cases, developing other rules peculiar to itself). /2/

ARCHAIC PHONOSEMY

The reason, I think, why so many languages can and do slight grammar is that grammar, unlike utterance or meaning, is only a means for multiplying the links between expression and content and does not in itself constitute either the expréssion or the content of spoken communication. Consequently, when language can minimize or dispense with grammatical mediation, it often does.

Direct mapping from sense to sound, without grammatical mediation, is sometimes referred to as phonosemy. Its ubiquity among the world's languages may be explained as a retention by those languages of an older and simpler manner of self-expression alongside one that is more recent and more complex. /3/

TYPES OF ICONICITY

Despite Ferdinand de Saussure's insistence on the arbitrariness of language, there is increasing evidence that all languages contain what Charles Peirce called icons, or utterances that mimic. /4/

In relation to any given utterance, however, the reality mimicked may be of any one of three types. The first such type of mimicry is *primary iconism*, exemplified by onomatopes like *buzz* or *hum*. The next type is *secondary iconism*, exemplified by syllabic phonesthemes like the -ash in *bash*, *dash*, and *gash*. And the last type is *tertiary iconism*, exemplified by infantile reduplicants like *booboo*, *doodoo*, and *googoo* or by palindromes like *pop*, *tot*, or *cock*. Words exemplifying primary iconism imitate non-linguistic reality; those exemplifying secondary iconism imitate other words; and those exemplifying tertiary iconism imitate (or, more precisely, repeat) segments of themselves.

One of the most striking areas of primary iconism in all spoken languages is that of bird vocabulary--words for birds themselves as well as for their vocalizations. Most birds that are small and produce high-pitched notes are represented, at least in part, by lexemes containing high front vowels. English examples are bird-names like *pewee* and *siskin*; verbs like *chirrup* and *twitter*; and echoics like *cheep!* and *tweet!*

Secondary iconism is well illustrated by a group of rhyming monosyllabic English nouns all of which denote something truncated: *bump*, *hump*, *lump*, *stump*, *clump*, etc.

Tertiary iconism of the reduplicative type is relatively straightforward in its structural derivation. But palindromy is more complex as regards the processes that give rise to it. It may be of any of four different subtypes, as follows:

A. progressive
    1. additive: pop < pa
    2. replacive: bub < bud
B. regressive
    3. additive: Nan < Ann
    4. replacive: Bob < Rob /5/

SEMANTIC PROCESSES

The semantic processes that produce iconic forms in English are of three types: monadic, dyadic, and triadic. Monadic processes produce forms that are not readily paired with other forms in an antithetical relation. Such a process is the use of the so-called "muffled" or "blurred" vowel /ə/ in the verbs *muffle* and *blur*.

Dyadic processes produce forms that are readily paired with corresponding forms in an antithetical relation. Such a process is the alternation of dorsal stops with dorsal fricatives in pairs like the following:

hack vs. hash
crack vs. crash
smack vs. smash
stack vs. stash

In each of the above cases, the form ending with a stop has a punctive force, expressing instantaneous action, while the form ending in a fricative has a durative force, expressing the result of the action. /6/

Tradic processes, like dyadic ones, generate antitheses. But, in addition to antithetical meanings, they also generate a neutral meaning, intermediate to the other two. An example is provided by the three nicknames *Hal*, *Hank*, and *Harry*, all hypocoristic variants of the forename *Henry*. In this case, the form containing the lateral is (or at least once was) diminutive, and that containing the vibrant is (or was) augmentative, while the form containing the nasal is, like its more formal source, neutral. /7/

PHONIC PROCESSES

Phonic processes yielding iconic effects are of two major types--phonosemic (or microlinguistic) and phonetic (or allolinguistic).

Phonemic processes, in turn, may be monadic, dyadic, or pluralic in subtype. An example of a monadic phonosemic process is the nasalization that adds sonority to the verb *clink* (from *click*). An example of dyadic process is the derogatory voicing in the verb *snivel* (as against *sniffle*). An example of a pluralic process is the labialization, apicalization, palatalization, and velarization encountered in the four provincial British nouns *craps*=*crits*=*crutchings*=*cracknels*, all denoting the cooked pig's intestines known in North America as *chitterlings* (and usually pronounced *chitlin's*). /8/

Three purely phonetic processes that normally go unrepresented in the standard orthography are:

gemination, as in [ɛnni] for *any*;

glottalization, as in [ʌˀow] for *oh-oh*; and

pharyngealization, as in [hʌˤri] for hurry /9/.

PSEUDO-MORPHOLOGICAL PROCESSES

Although, in most cases, allolanguage eliminates microlinguistic morphology completely, in others it substitutes a reduced and deviant morphology. This pseudo-morphology may be either replacive or additive.

Replacive pseudo-morphology is particularly evident in American slang, where it takes two highly specific forms. One is replacement of any consonant or consonant-cluster by /z/ (a process which I have nicknamed "zazzification") , as in *zillion* for *million*, *billion*, or *trillion*. /10/ The other is replacement of any syllabic nucleus--whether monophthong or diphthong--by /uw/, written oo (a process which I have nicknamed "ooglification"), as in *oogly* for *ugly*. /11/

Additive pseudo-morphology takes four forms, two of which are absent from microlanguage. These four are prefixation and suffixation (both present in microlanguage) plus infixation and interfixation (both absent). Examples follow.

1. prefixation: *smelt* from *melt*

2. suffixation: *kiddo* from *kid*

3. infixation: *purp* from *pup*

4. interfixation: *pit-a-pat* from *pat(ter)* /12/

The most distinctively allolinguistic of such affixes are syllabic prefixes consisting of a postalveolar consonant and a stressless blurred vowel. Examples follow:

derivative	source or cognate
kathob, "vague thing"	thob, "be credulous"
gazook, "tramp"	zook, "prostitute"
chewallop, "bang!"	wallop, "hit hard"
jamoke, "fellow"	moke, "dull person" /13/
yazunk, "plop!"	zonk, "to strike"

Such formative processes produce pseudo-morphological results. One of these results is the echo-compound, consisting typically of a pleremic, or obviously meaningful, base-word followed by a cenemic, or relatively meaningless, rime-tag. Examples follow:

roly-poly	(plump)
hurly-burly	(battle)
eenie-meenie	(one, two...)
palsy-walsy	(excessively friendly)

One thing that is noteworthy about such groupings is that the initial consonants of the rime-tags themselves form an apophonic series, ranging from surd through sonant and nasal to glide in the bilabial category. /14/

Another pseudo-morphological phenomenon is what I call a word-chain. Word-chains typically consist of three-word phrases, in which the first and third word lack phonic overlap but which are linked by phonic overlap with the second word. This overlap is either rime followed by alliteration or alliteration followed by rime, as below.

healthy, wealthy, and wise
grunt, groan, and moan /15/

PHONOSEMIC CORRESPONDENCE

One of the characteristics of allolanguage is a closer relation between sound and sense than obtains in microlanguage. Even in allolanguage, however, there is a disjunction of scale, in accordance with which a number of quite different phonic devices can produce a single semantic effect. Of no effect is this truer than of diminution, which can be achieved by syllabic elision, by cluster-reduction, by high-fronting of vowels, by lateralization of vibrants, by occlusivization of nasals, or by velarization of labials. Examples follow:

1.	Ed	<	Edward
2.	Kit	<	Christopher
3.	tip	<	top
4.	Sally	<	Sarah
5.	Peg	<	Meg
6.	dunk	<	dump

In a few cases, these devices may even be pitted against one another. An example is the name *Susan*, which, when hypocoristic, may take either of two forms. One of these, *Sue*, is more diminutive in terms of syllable-count, while the other, *Suzie*, is more diminutive in terms of vowel quality.

SEMANTIC WEIGHTING

A majority of the semantic categories in language generally are dyadic, involving such familiar pairings as nominal vs. verbal, subject vs. object, or active vs. passive. Allolanguage exhibits just as many such pairings as does microlanguage. But it weights them emotively in a more discriminatory direction, showing clear preference for diminutive over augmentative forms but for derogatory over plauditory forms.

SEMANTIC CROSS-CURRENTS

Some allolinguistic forms exhibit simultaneously positive and negative manifestations of the same semantic category. In the category of size (or seniority), the 17th century nickname *Poll(y)* is a striking example. In terms of its consonantism, it is doubly diminutive. In terms of its vocalism, however, it is augmentative. Reading from back to front, it means, literally, "(little) little big little Mary."

SEMANTIC FADING

Other allolinguistic forms, though distinctive as markers of non-canonical material, have so little intrinsic meaning as to be, in Glossematic terms, cenemic. Examples are the labial-onset prefixes in the following slang terms:

pizazz, "zest" (cf. zazzle, "sex appeal")
bazoo, "snout" (cf. kazoo, "mouth resonator")
fadoodle, "nonsense" (cf. doodle, "to scrawl")
vaboom! "a thunderous sound" (cf. boom)
magoo, "custard pie" (cf. goo, "slime") /16/

PHONIC ICONISM OUTSIDE ENGLISH

The polarity of high-front vowels versus low and/or back vowels seems universally to correlate with that between small and large. Indo-European examples are:

German Misch-masch: "heterogeneous mixture"

Italian bimbo, bambolo: "baby, child"

Russian pif da paf: "slam-bang"

Bihari din-un: "a day or so"

Non-Indo-European examples are:

Estonian vinderdi-vänderdi: "to and fro"

Basque bilin-balan: "ding-dong"

Mandarin ching, chung: "light, heavy"

Proto-Polynesian iˀi, oho: "small, large" /17/

In no language, however, is the diminutive/augmentative polarity more closely correlated with vocalic apophony than in English, where we encounter it equally often in derivative linkages like *sip* from *sup* and in echoic compounds like *zig-zag* and *see-saw*.

References

/1/ G.L. Trager, "Language," *The Encyclopedia Britannica*, 1955

/2/ R.W. Wescott, *Sound and Sense: Linguistic Essays on Phonosemic Subjects*, Jupiter Press, 1980, p. 19

/3/ ibid., pp. vii and 22

/4/ ibid., p. 3

/5/ R.W. Wescott, "The Iconicity of Consonant Alternation," The University of California Conference on Sound-Symbolism, Berkeley, January 1985, pp. 2 and 3

/6/ R.W. Wescott, *Sound and Sense* (as above, fn. 2), p. 340

/7/ R.W. Wescott, "Lateral/Vibrant Alternation in English Nicknames," *Comments on Etymology*, v. 9, n. 1, October 1979.

/8/ R.W. Wescott, *Sound and Sense* (as above), pp. 349 and 352.

/9/ R.W. Wescott, "The Iconicity of Consonant Alternation" (as above, fn. 5), p. 23

/10/ ibid., pp. 19 and 20

/11/ R.W. Wescott, *Sound and Sense* (as above), p. 389

/12/ R.W. Wescott, "Neglected Affixes in English," a lecture to the Linguistic Society at the State University College of New York at Buffalo, April 1978

/13/ R.W. Wescott, *Sound and Sense* (as above), p. 401

/14/ ibid., p. 402

/15/ ibid., p. 378

/16/ ibid., p. 401

/17/ ibid., p. 326

ESTONIAN ONOMATOPOEIA: A TYPOLOGICAL APPROACH

ENN VELDI

Dept. of English
Tartu State University
Tartu, Estonia, USSR 202400

ABSTRACT

The paper presents a classification of Estonian onomatopoeic words worked out by the author. The inventory of Estonian onomatopoeia (735 onomatopes) is examined within the framework of the universal classification of onomatopes worked out by S. Voronin.

INTRODUCTION

The past decade has seen growing interest in the study of various iconic aspects of language. Of particular interest is the emergence of typological studies in the field of phonosemantics. The groundwork for typological studies in the field of onomatopoeia was laid down by S. Voronin in his universal classification of onomatopes /1/.

Onomatopoeia as a phonosemantic subsystem can be viewed as modelling extralinguistic acoustic phenomena by means of the phonological structure of words in a given language. Accordingly, onomatopoeia is approached from the extralinguistic point of view, the classification of onomatopes being based on the typology of extralinguistic sounds. It should be emphasized that an onomatope as a model is a double approximation of an extralinguistic sound /2/. At first the extralinguistic sound is channelled by the human ear and then the model is built up by means of the phonological structure of a language. The linguistic sign being a linear entity, the individual properties of a complex sound can only be conveyed successively. The extralinguistic and typological approach in the study of onomatopoeia brings to the fore the deep-set universal isomorphic features that otherwise would not be self-evident.

The present paper is an attempt to apply the principles of the universal classification of onomatopes to the study of Estonian onomatopoeia. The subject is of considerable interest since Estonian, as well as other Balto-Finnic languages, is noted for its rich repertoire of onomatopoeic and descriptive resources /3/.

A CLASSIFICATION OF ESTONIAN ONOMATOPES

CLASS A. INSTANTS

Type I. Instants
Pattern 1

$$\begin{array}{l}\underline{(\mathrm{PLOS}+)\ (\mathrm{SON}^{\mathrm{LAT/NAS\ DENT}})}\\ \mathrm{FRIC}^{\hat{\mathrm{SIB}}}\\ \quad <\mathrm{AFFR}\\ \overline{\mathrm{FRIC}^{\check{\mathrm{LAB}}}}\end{array}\qquad +\breve{\mathrm{VOC}}+\mathrm{PLOS}^{(s)}$$

Examples: TIKK-TAKK - tick-tack (of a clock); KLOPS - a light clicking sound; KLUKSuma - to make a sound like water flowing unevenly out of an opening, esp. in the throat; NAPS - snap, the sound of closing the jaws quickly; SIDistama - (of a bird) to make a succession of short rapid sounds, twitter; VIDistama - (idem).

CLASS B. CONTINUANTS

Type II. Tonal Continuants
Pattern 2

$$\mathrm{CONS} + \overline{\mathrm{VOC}}\ (+\mathrm{CONS}^{(s)})$$

Examples: PIIP-PIIP - beep-beep, the high-pitched tone of a motor horn or telephone; TUUT-TUUT - toot-toot, the low-pitched sounds of a horn; SIITSuma - (of young birds) to cheep; VIIKSuma - (of small birds) to make a thin high-pitched sound, to squeak.

Pattern 3

$$(\mathrm{CONS}+)\ \breve{\mathrm{VOC}} + \mathrm{SON}^{\mathrm{LAT/NAS\ DENT}}$$

Examples: UNdama - to hoot (of sirens, communication lines); LLLLLLL - (occas.) the monotonous sound of communication lines.

Type III. Pure Noise Continuants
Pattern 4

$$\frac{\mathrm{FRIC}^{\wedge}}{\mathrm{PLOS}} + (\mathrm{SON}^{\mathrm{LAT}}+)\ \breve{\overline{\mathrm{VOC}}} + \frac{\mathrm{PLOS}}{\mathrm{FRIC}^{\times}}$$

Examples: SISisema - to give a high-pitched prolonged sibilant sound, hiss; SOSistama - to whisper; LÔHisema - tò make the sound of tempestuous flames; SIUH - the sound of cutting through the air with a high-pitched whistling noise, swish; KAHisema - to whisper gently through the leaves (of wind), to swish (of clothes); karSOSS - the sound of an object entering sand, swish; HABisema - (of wind) to whisper gently through the leaves.

Type IV. Tone-Noise Continuants
Pattern 5

$$\frac{\mathrm{FRIC}^{\vee}}{\mathrm{FRIC}^{\wedge}} + \overset{\smile}{\overline{\mathrm{VOC}}} + \frac{\mathrm{FRIC}^{\wedge}}{\mathrm{SON}^{\mathrm{NAS}}}$$

Examples: VIHisema - to make a noise by moving very fast, whizz, to whistle (of wind); VISisema - to make a continuous hissing sound (of wet logs burning), fizz; VIUH - (of an object) the sound of whizzing through the air; HUMisema - to make a low dull continuous buzz; SUMisema - to make a monotonous low hum (of bees in flight), buzz.

CLASS C. FREQUENTATIVES
Type V. Quasi-Instant Frequentatives
Pattern 6

$$\frac{(\mathrm{PLOS+})\ (\mathrm{SON}^{\mathrm{LAT}})}{\mathrm{FRIC}^{\mathrm{SIB}\wedge}} + \underleftrightarrow{\overset{\smile}{\mathrm{VOC}} + \mathrm{R}} + \mathrm{PLOS}^{(s)}$$

<AFFR

Examples: (P)RAKSuma - to make sudden repeated explosive sounds (of ice, burning logs), crack, crackle; KRAKSti - with a sudden sharp explosive sound (as of breaking), crack; PLARTSti - the sound of a body falling into water or sth. wet hitting a surface and being flattened; LÄRTSti - with a noise of sth. wet hitting a surface and being flattened; SIRTSuma - to make short sharp sound(s) (of small birds or insects), chirp; TÕRTS - a short sharp, usu. unpleasant low-pitched sound of a horn or a brass instrument.

Type VI. Pure Frequentatives
Pattern 7

$$(\mathrm{CONS+})\ (\mathrm{SON}^{\mathrm{LAT}}) + \overset{\smile}{\mathrm{VOC}} + \mathrm{R}$$

Examples: URisema - (of a dog) to make a deep rough throaty sound, growl; TIRisema - (of an alarm clock) to make a high-pitched continuous vibrating sound, (of a telephone) to ring; PÕRisema - (of an engine) to make a deep continuous vibrating sound by a succession of quick strokes, (of the drum) to roll; SIRistama (of small birds) to sing in a trilling manner, (of a grasshopper) to chirr; KÄRisema - to make a sound as of tearing (of cloth, paper); KÕRisema - (of dry peas moving in a bowl) to emit a hollow continuous sound made by a succession of quick strokes; PLÄRisema - to make a continuous harsh unpleasant dissonant blaring noise.

Type VII. Tonal Frequentatives
Pattern 8

$$(\mathrm{PLOS+})\ \mathrm{R} + \overline{\mathrm{VOC}}\ (+\mathrm{PLOS}^{(s)})$$

Examples: KRIIKSuma - to make a prolonged grating sound (of a badly-oiled door), to creak; KROOKSuma - to make a deep low noise such as a frog makes, to croak; PRÄÄKSuma - to make the characteristic cry of a duck, to quack; RÄÄKSuma - to make a harsh reedy cry (of a crake); RUIgama - (of a pig) to make deep throaty sounds, to grunt.

Type VIII. Frequentatives
Pure Noise Quasi-Continuants
Pattern 9

$$\underleftrightarrow{\mathrm{FRIC}^{\wedge} + \overset{\smile}{\mathrm{VOC}} + \mathrm{R}}$$

Examples: SURisema - to make a continuous dull vibrating sound, to whirr; RÖH RÖH, RÖH - (of a pig) short rough sounds in the throat; SORR, SORR, SORR - regular vibrating sounds made by the whirr of the spinning-wheel; HURisema - (of big insects) to make a continuous low-pitched vibrating sound.

Type IX. Frequentatives
Tone-Noise Quasi-Continuants
Pattern 10

$$\mathrm{FRIC}^{\mathrm{LAB}\vee} + \overset{\smile}{\mathrm{VOC}} + \mathrm{R}$$

Examples: VURR, VURR - regular vibrating whizzing sounds made by the rapid revolving of a spinning-wheel; VÕRRa-VÕRRa - the sound of the rapid motion of a threshing machine; VÜRama - to make a vibrating buzz.

HYPERCLASS AB. INSTANTS-CONTINUANTS
Type X. Tonal Postpulse Instants-Continuants
Pattern 11

$$(\mathrm{PLOS})\ (+\mathrm{SON}^{\mathrm{LAT}}) + \overset{\smile}{\mathrm{VOC}} + \mathrm{SON}^{\mathrm{NAS/LAT}} + (\mathrm{PLOS}^{(s)})$$

Sub-type A. Short

Examples: TÜMPSuma - to make repeated dull sounds when striking with a heavy blunt object; PONTSatama - (of a soft heavy object) to fall heavily with a resounding blow; KOLKSuma - to make repeated loud unpleasant sounds, as of metal striking against metal; PLÕNKSuma - (of playing a stringed instrument) to produce repeated dull resonant sounds, to plonk.

Sub-type B. Prolonged

Examples: PLÔNNima - to play a stringed instrument carelessly or informally, esp. without skill; PIMM-PAMM - slow and repeated deep resonant sounds of a large bell; KILL-KÕLL - clear resonant sounds of small bells; TILL - a prolonged high-pitched ringing sound (as of a small bell).

Pattern 12

$$\mathrm{PLOS}\ (+\mathrm{SON}^{\mathrm{LAT}}) + \overline{\mathrm{VOC}}$$

Examples: PIU-PAU - repeated resonant sound of bells, resounding discharges of guns; PLAU - (of the door) the sound of sudden shutting with a bang.

Type XI. Pure Noise Postpulse
Instants-Continuants
Pattern 13

$$\mathrm{PLOS}\ (+\mathrm{SON}^{\mathrm{LAT}}) + \overset{\smile}{\overline{\mathrm{VOC}}} + \mathrm{FRIC}^{\wedge}$$

Examples: KAUHti - the sound of a sudden fall into water, splash; PLAUHti - the loud sound of a sudden fall into water; TSUHH-TSUHH - a nursery name for a railway train (imitation of the sound of a steam-engine).

Type XII. Pure Noise Prepulse
Instants-Continuants
Pattern 14

$$\mathrm{FRIC}^{\wedge} + \overset{\smile}{\mathrm{VOC}} + \mathrm{PLOS}^{(s)}$$

Examples: SAPSima - to strike softly with a branch of twigs; SOPS - the sound of a soft blow with a birch-twig.

Type XIII. Tone-Noise Prepulse
Instants-Continuants
Pattern 15

$$\mathrm{FRIC}^{\mathrm{LAB}\vee} + \overset{\smile}{\mathrm{VOC}} + \mathrm{PLOS}^{(s)}$$

Examples: VOPSti - the sound of a swishing blow; VIPutama - (dial.) to make swishing blows with a branch of twigs; VAPP, VAPP - (of an owl) the sound of fluttering movements with wings.

Type XIV. Pure Noise-Tonal Prepulse-Postpulse Instants-Continuants
Pattern 16

$$\mathrm{FRIC}^{\mathrm{SIB}\wedge} + \overset{\smile}{\mathrm{VOC}} + \mathrm{SON}^{\mathrm{NAS}}\ (+\mathrm{PLOS})$$

Sub-type A. Short

Example: SUMPama - to make a muffled sound as in wading through deep snow or water.

Sub-type B. Prolonged

Example: SUMMima - to wade splashingly through water.

Type XV. Tone-Noise - Tonal Prepulse-Postpulse Instants-Continuants
Pattern 17

$$\mathrm{FRIC}^{\mathrm{LAB}\vee} + \overset{\smile}{\mathrm{VOC}} + \mathrm{SON}^{\mathrm{NAS}}$$

Sub-type A. Short (not found in Estonian; cf. Eng. zonk - (slang) the sound of a short resonant blow (esp. on the head).

Sub-type B. Prolonged

Examples: VINGuma - (of a bullet) to make the sound of moving quickly through the air; VONGuma - (of breaking ice) to make a booming sound.

HYPERCLASS CAB. FREQUENTATIVES
QUASI-INSTANTS-CONTINUANTS
Type XVI. Tonal Postpulse
Quasi-Instants-Continuants
Pattern 18

$$(\mathrm{PLOS+})\ \underleftrightarrow{\mathrm{R} + \overset{\smile}{\mathrm{VOC}} + \mathrm{SON}}^{\mathrm{NAS/LAT}}\ (+\mathrm{PLOS}^{(s)})$$

Sub-type A. Short

Examples: PRANTSti - the sound of a resounding vibrating blow caused by a heavy fall; RONTSatama - (of a heavy object) to fall with a muffled vibrating blow; TRAMPima - to step heavily with the feet; MÜRTSti - the sound of a thundering bang.

Sub-type B. Prolonged

Examples: PRÔMMima - to knock violently on a door with one's fist; TRÜMMima - to bang forcefully (on a door).

Type XVII. Pure Noise Postpulse
Quasi-Instants-Continuants
Pattern 19

$$\mathrm{PLOS} + \mathrm{R} + \overline{\mathrm{VOC}} + \mathrm{FRIC}^{\wedge}$$

Examples: PRAUH - a loud noise made by a violent blow, fall or break, crash; KRIUH-KRAUH - the sound of sudden rapid tearing of cloth.

Type XVIII. Pure Noise Prepulse
Quasi-Instants-Continuants
Pattern 20

$$\mathrm{FRIC}^{\mathrm{SIB}\wedge}\ (+\mathrm{SON}^{\mathrm{LAT}}) + \overset{\smile}{\mathrm{VOC}} + \mathrm{R} + \mathrm{PLOS}^{(s)}$$

Examples: karSORTS - the sudden noise of a pencil penetrating paper; SLORTS - the sound of a liquid falling on a surface.

DISCUSSION

The distribution of onomatopes within different types of onomatopoeia varies considerably. The most numerous types in Estonian are I and X (161 and 126 onomatopes respectively), which make up 39 per cent of the whole inventory. They are followed by types V (88), II (76), III (69) and VI (67) totalling 40.8 per cent. The remaining types cover about one fifth of the material. Estonian is poorly represented in Types XIV and XV. No Estonian examples for Sub-type A, Type XV, could be found in our material.

In modelling extralinguistic sounds the principle of homogeneity between the acoustic parameters of the extralinguistic sound and the corresponding phonemotypes is observed. This is one of the reasons why in onomatopoeia typologically isomorphic features dominate. As regards allomorphic features, they could be brought about by a variety of reasons. A possible source of allomorphism lies in the phonological redundancy of several types of onomatopes. For example, to render pure noise, in principle, only one voiceless fricative is needed; but since the majority of roots contain more than one consonant, the other can be considered to be redundant. Phonological redundancy may be manifested differently in various languages. Firstly, an extra (redündant) voiceless fricative may be supplied, e.g. SUSisema. Secondly, consonants can be found which do not fulfil any echoic function, e.g. the /k/ in KAHisema; cf. the Japanese designations of swishing, rustling movement kasa-kasa, goso-goso. Thirdly, the extra consonant may signal sound-symbolic values, e.g. the labial /p/: the protrusion of the lips in PUHuma - to blow.

Another possible source of allomorphism lies in the non-stable nature of affricates. In a group of Estonian onomatopes the initial affricate has undergone assibilation, and consequently the initial fricative sibilant in these onomatopes (Types I, V, and X) can no longer echo the plosive character of the sound.

A characteristic feature of Estonian Tonal Continuants and Tonal Frequentatives is the occurrence of doublets with long vowels and diphthongs. The onomatopes with diphthongs denote somewhat shorter sounds. e.g. KIIKSuma KIUKSuma; KRIIKSuma KRIUKSuma. The phoneme /s/ which occurs finally in the combinations -ks, -ps and-ts, e.g. KLUKSuma, NAPS, PRANTSti is of expressive character, and serves the purpose of intensifying the meaning of the word. In a few onomatopes the "natural" order of qualitative elements is reversed, e.g. MÜRTSuma, MÜRisema. A parallel to this phenomenon can be found in the Nenets onomatopes mirćedaś and mirneleś which have similar meanings (in Nenets /r/ does not occur in the initial position).

Finally, it should be pointed out that although in onomatopoeia auditory motivation prevails, the majority of onomatopes manifest the phenomena of sound-symbolic interference as well. As a rule, labial and guttural sounds in the designations of the sounds made by birds and animals denote the place of articulation. Sound-symbolic interference is especially strong in the designations of bubbling, e.g. PULisema, VULKSuma, MULKSuma.In such onomatopes the presence of an initial labial appears to be more important than that of an initial plosive. Thus, strong sound-symbolic interference is a force that makes the patterns of onomatopes more complicate. A characteristic feature of Estonian is the fact that the phoneme /v/ occurs initially in a number of onomatopes. In addition to its auditory value the initial /v/ in Estonian onomatopes is usually also sound-symbolically motivated.

By way of conclusion it should be said that the typological study of onomatopoeia is an emerging area of research and it remains to be hoped that new studies will bring to light interesting facts in various languages of the world.

References

/1/ S.V.Voronin, Fundamentals for a Universal Classification of Onomatopes.(Onomatopoeia and Phonosemantics). - In: Phonetics-83. - Papers presented for 11th International Congress of Phonetic Sciences. Moscow, 1983, p.44-55.(In Russian)

/2/ M.Grammont, Traité de phonétique, Paris, 1933, p. 378.

/3/ V.Põlma, Onomatopoeetilised verbid eesti kirjakeeles, Tallinn, 1967, 455 p. (Unpublished thesis)

Symbols

CONS	- consonant
PLOS	- plosive
AFFR	- affricate
FRIC	- fricative
SIB	- sibilant
R	- the phoneme /r/
SON	- sonorant
NAS	- nasal
DENT	- dental
LAB	- labial
LAT	- lateral: /l/
^	- voiceless (e.g. FRIC^- voiceless fricative)
ˇ	- voiced (e.g. FRIC - voiced fricative)
$\overset{\smile}{\text{VOC}}$	- short vowel
$\overline{\text{VOC}}$	- diphthong or long vowel
/	- "either...or"
()	- brackets for optional components

THE PHONEMOTYPE: A NEW LINGUISTIC NOTION (IMPLICATIONS FOR TYPOLOGICAL PHONOSEMANTICS)

STANISLAV V. VORONIN

Dept. of English Philology
Leningrad University
Leningrad, USSR, 199034

ABSTRACT

The problem of descriptive units is of paramount importance for any typology. Isomorphism (similarity), which prevails over allomorphism (dissimilarity) in the iconic (onomatopoeic and sound-symbolic) words of any two (unrelated) languages, cannot, as a rule, be revealed on the level of individual phonemes. The paper is a first report on the implications for typological phonosemantics of a notion introduced earlier by the author - the notion of the phonemotype (i.e., a "semantically loaded" acoustic or articulatory type of phonemes). The phonemo⁄type as a unit is shown to possess a number of unique features.

The emergence of the new linguistic science,phonosemantics (dealing with the iconic, i.e., onomatopoeic and sound-symbolic, system of language), necessitates the elaboration of typological phonosemantics, or a phonosemantic typology of the world's languages /1/.

Linguistic iconism is an absolute language universal, and the scope of the iconic system in language is, contrary to popular sentiment, extremely great /2/. This system does not include exclusively words that are felt to possess a phonetically motivated connection between sound and sense - it also embraces all those countless words where in the course of historical development, this connection has become obscured but where it can be uncovered with the aid of "deep down" etymological analysis buttressed by "external" typological data.

Invading the realm of iconicity, the researcher, like Alice in Wonderland, probes a world where many things are "so different" and "so unlike"; prepared to relinquish some of the hallowed age-long linguistic shibboleths and willing to work out a new set of values, the explorer presses on in his quest.

Phonetic (phonological) typology and semantic typology are venues for the study of sound and, disconnectedly, sense. The blazing gap is there - to be bridged by phonosemantic typology exploring the sound/sense connection in the lexis of different - primarily unrelated - languages.

The problem of descriptive units is of paramount importance for any typology. Isomorphism (similarity), which prevails over allomorphism (dissimilarity) in the iconic words of any two (unrelated) languages, cannot, as a rule, be revealed on the level of individual phonemes (instances like the English ting and Indonesian ting, both signifying the sound of a small bell, are very infrequent). This paper is a first report on the implications for typological phonosemantics of a notion introduced earlier by the author - the notion of the phonemotype (i.e., a "semantically loaded" acoustic or articulatory type of phonemes) /3/.

Taking by way of illustration a number of onomatopoeic groupings, I shall attempt to retrace the steps in arriving at the notion of the phonemotype in phonosemantic typology.

Illustration 1: Instants /4/. These onomatopes designate pulses (the pulse is an instant sound like a tap, tick, click or knock). Cf. examples from four languages (of diverse language families), viz. English (Eng.), Estonian (Est.), Bashkir (B.), Indonesian (Indon.) /5/.Eng. tap, tick, pat, pop, click, clop-clop, chop; Est. tikk-takk - tick-tack (of a clock), kɵp-kop - imitative of tapping on the door, klobisema --to go clop-clop (of wooden shoes), plōgisema - to click; chatter (of teeth); B. tap - instantaneous sound of hard object falling to the ground, dŏk - dull knock or tap (on the door), qup - sound of object striking wood, kelt-kelt - to tick; Indon. tuk - imitation of knocking, tak - sound of a stone striking wood, bap - imitative of an object falling on a soft surface, bak - a pat; sound of fruit falling on the ground, lepik - sound of matchbox falling on the floor. Listing the initial consonants in the roots of the onomatopes cited, we find them to be: /t,p,k,ḳ,d,b, tʃ/; the root-final consonants are: / p,k,t,g,b/. H/ere we see a great diversity of phonetic types: dentals, labials, velars; voiced and voiceless phonemes; stops, and even an affricate. The diversity within the initials list, as also in the finals list, is thus evident - but misleading. For in this diversity there is an underlying unity (less obvious but nevertheless tangible): what unifies these consonants (both initial and final) is the fact that they belong to one and the same type, viz. plosives (/tʃ/ is an affricate - but on this see below). Acoustically, plosives are essentially pulses, and it is only natural that they are used in onomatopoeic designations of a pulse. (As to affricates, the initial element here tends to be of a pulse-like nature; thus affricates, too, are a natural - if somewhat less accurate - rendering of a pulse.) Hence plosives (as well as affricates) in onomatopes designating pulses are not purely phonetical, asemantic groupings - they are "semantically loaded", and charged with the delicate task of conveying meaning; whenever an onomatope designates a pulse, it is primarily the plosives that do the semantic job for the entire onomatope. Plosives in Instants are an example of what I term the phonemotype. Summing up the essential components that go to make up the onomatopes cited above, we come to the general pattern followed (with remarkably few deviations) in the formation of onomatopoeic roots designating the pulse in the most diverse languages. In terms of phonemotypes, the general pattern for Instants is as follows (for symbols, see below):

$$\frac{\text{PLOS}}{\text{AFFR}} + \text{V}\breve{\text{O}}\text{C} + \text{PLOS}.$$

Illustration 2: Tonal Post-Pulse Instants-Continuants /6/. These onomatopes designate a complex sound - basically the combination of a pulse followed by a resonant tone (e.g. the ringing sound produced by string or bell). Cf. Eng. tang, ting, ping, bang, twang, clam, knell (O.E. cnyllan); Est. tinn - high-pitched ringing sound (of a string), plongutama - to ring (as of a string plucked), pumm - powerful resonating blow (as with a fist), till - prolonged high-pitched ringing sound (as of a small bell); B. tan - sound of metal struck, ton - imitative of resonant sound produced by heavy object striking smth hollow, ten - ringing sound (as of metal struck lightly), den-den - faint ringing sound (of a string); Indon. letang - sound of hammer on metal, ting - sound of a small bell, bong - imitation of sound produced by beating a large drum, lebam - loud sound of object falling on resonant surface, bum - sound of a gun or bomb. The root-initial and root-final consonants in these examples are, respectively, /t,p,b, k,d,tʃ/ and /ŋ,n,m,l/. The case for the initials is the same as in the above-mentioned Instants: they belong to the plosives phonemotype, and they render the initial pulse. The case for the finals is that they are all sonorants; acoustically, sonorants are predominantly tonal entities; it is therefore only natural that the sonorant phonemotype is used in onomatopoeic designations of tone. In terms of phonemotypes, the prevailing general pattern for Tonal Post-Pulse Instants-Continuants is this:

$$\frac{\text{PLOS}}{\text{AFFR}} + \text{V}\breve{\text{O}}\text{C} + \text{SON}^{\text{NAS/LAT}}.$$

Illustration 3: Pure Noise Continuants /7/. These onomatopes designate pure noise - that is, various hissing, swishing, whispering sounds. Cf. Eng. hiss, hush, huff, flush, slosh, swish, swash; Est. sahisema - to rustle, husisema - (dial.) to hiss, kahisema - to whisper gently through the leaves (of wind), to swish (of clothes), habisema - to whisper gently through the leaves (of wind); B.ysyldau - to hiss (of a goose or a snake), byšyldau - to hiss; to whisper, syj - swishing sound (caused by rapid movement), sajlau - (dial.) to whistle (of a bullet); Indon. desah - imitation of sound of polishing; the rustling of leaves in rain, sis - hissing, lesus - a whisper, kesik - rustling; whispering, kesu-kesi - leaves rustling in the wind. A cursory overview of root-initial and root-final phonemes gives a bizarre and discouraging picture. But a closer look yields two systematic subpatterns. Subpattern one is furnished by the entire English material and part of the Estonian (sahisema, husisema) and Indonesian (desah, sis) material: the initials /h,f,s/ and the finals /s, ʃ, f, h/ - different as they are, they all fall into the category of voiceless fricatives. Subpattern two does not have fricatives for both initials and finals, but it does consistently have one fricative - either initial or final - coupled practically with any other final consonant (as in Est. habisema, B. syj) or, respectively, any other initial consonant (as in Est. kabisema) or even with no final/initial consonant whatever (see Indon. kesu-kesi, B. ysyldau). The zigzag puzzle of the subpatterns resolves into the following comprehensive general pattern:

$$\frac{\text{FRIC}^{\wedge}}{\text{FRIC}^{\wedge}} + \text{V}\breve{\text{O}}\text{C} + \frac{\text{FRIC}^{\wedge}}{(\text{CONS})}\ .$$

The purport of this is that for the "portrayal" of pure noise at least one voiceless fricative (initial or final) is obligatory in the onomatopoeic roots of a given language (though some languages, like English, evince the redundant feature of employing even two voiceless fricatives,. initial and final). The voiceless fricative phonemotype, in itself acoustically pure noise, is the echoic correlate of pure noise designated by onomatopes of this kind.

It is hoped that illustrations 1,2 and 3 help to trace the logic in isolating the notion of the phonemotype.

One of the fundamental principles at work in the domain of onomatopoeia (and, mutatis mutandis, sound symbolism) is the principle of homogeneity: structural acoustic elements of the referent sound (i.e., the sound designated) are iconically rendered, in the corresponding onomatope, by structural phonetic elements belonging to the same acoustic type. Phonemes in the onomatopoeic root are thus correlated with the elements of the referent sound - but indirectly, via the phonemotype, the latter acting as a go-between or intermediary /8/.

Given the acoustic structure of the referent sound (together with the known phonetic peculiarities of the language in question) we can safely predict (in approx. 80-90 per cent of all cases) the phonemotype pattern of corresponding onomatopoeic roots (though not its concrete phonemic realization). The crucial unit in an onomatope's structure is, then, the phonemotype - and not the phoneme.

* *
*

The articulatory phonemotype in sound-symbolic words, though differing somewhat from the acoustic phonemotype of onomatopes, is fundamentally the same entity as the one outlined above (a detailed analysis calls for discussion in a separate paper).

* *
*

The phonemotype in the iconic vocabulary of languages possesses a number of highly specific features. To name just a few:

- The phonemotype is a semanticized entity.
- It is a two-faceted entity, both phonetical and semantic. (Here one might even be tempted to introduce the sesquipedalian term "phonemosemotype", or rather "phonosemotype").
- The phonemotype is able to dissect phonological space in a manner impossible for phonemes, a manner peculiar only to itself; cf. the phonemotype of labials in designations of rounded shape: the fundamental phonetic dichotomy of consonant/vowel is here irrelevant. /9/
- The phonemotype is a psycholinguistic reality.
- It is, further, inter-disciplinary in essence.
- The phonemotype is a cross-linguistic phenomenon.
- Being basically an ontological entity, it may be, and is, employed as a methodological instrument.

Further evolvement of the notion entails discussion of such problems as fuzzy sets and language as choice and chánce.

* *
*

The notion of the "semantically loaded" phonemotype (coupled with that of onomatopoeic patterns) leads us to realize the intrinsic limitations of the longstanding belief that root morphemes, though divisible phonetically or semantically, are allegedly indivisible phoneto-semantically. Root morphemes can to a large extent be structured in terms of phonemotypes.

As demonstrated by recent research, units like the phonemotype are proving themselves adequate instruments not only in language-specific phonosemantics, but also in typological phonosemantics /10/ as well as in typological paleolinguistics. For the latter, cf. Prof.R.Wescott's view: "... sound correlations in ... language families of great internal time depth must be formulated either subphonemically, in terms of articulatory or acoustic features, or transphonemically, in terms of morphophonemes" /11/. This transphonemic reference is, as has been shown, the very essence of the phonemotype, instrumental in tapping the largely untapped iconic (onomatopoeic and sound-symbolic) resources of the world's languages.

Symbols

CONS	-	(any) consonant
PLOS	-	plosive
AFFR	-	affricate
SON	-	sonorant
NAS	-	nasal
LAT	-	lateral : /l/
^	-	voiceless : FRIC^- voiceless fricative
VŎC	-	short vowel
( )	-	brackets for optional components

References

/1/ Vide: S.V.Voronin. Fundamentals of Phonosemantics. Leningrad, 1982. (In Russian).

/2/ O.Jespersen. Symbolic Value of the Vowel i. - O.Jespersen. Linguistica. Copenhagen, 1933; G.Ramstedt. Ueber onomatopoetische Wörter in den altaischen Sprachen. - J.Soc.Finno-Ougrienne, No 55, 1951; A.M.Gazov-Ginzberg. Is Language Imitative by Origin? (Evidence from Common Semitic Stock

of Roots). Moscow, 1965. (In Russian) B.A.Serebrennikov. Miscellanea. I. - Sovietskoye finno-ougrovedeniye, 1976, No 4. (In Russian).

/3/ Vide: S.V.Voronin. English Onomatopes. (Types and Structure). Unpublished. M.Phil.Diss. Leningrad, 1969. (In Russian).

/4/ Ibid.; S.V.Voronin. Fundamentals for a Universal Classification of Onomatopoeias. - Phonetica-83. Papers Presented for 11th International Congress of Phonetic Sciences. Moscow, 1983. (In Russian).

/5/ For material, see largely: S.V.Voronin. English Onomatopes; I.B.Bratoes. Acoustic Onomatopes in Indonesian. Unpublished M.Phil.Diss. Leningrad, 1976. (In Russian); L.Z.Lapkina. English and Bashkir Acoustic Onomatopes. Unpublished M.Phil.Diss. Leningrad, 1979. (In Russian); E. Veldi. Estonian Onomatopes: A Classification. - Linguistica. Tartu, 1986. (In Russian)

/6/ Vide: S.V.Voronin. English Onomatopes.

/7/ Ibid.

/8/ Ibid.

/9/ Cf.: Ye.I.Kuznetsova, S.V.Voronin. Symbolism in Designations of Roundedness. - Systemic Description of Germanic Vocabulary, vol.4. Leningrad, 1981. (In Russian).

/10/ Vide inter alia: O.A.Kazakevich. Single-Syllable Sound-Imitative Ideophones in Zulu. - Papers in Structural and Applied Linguistics. Moscow, 1975. (In Russian); A.Ju. Afanasyev, Problems in Semantic Evolution of Vocabulary. M.Phil. Diss. Abstract. Leningrad, 1984. (In Russian); O.D.Kuleshova. The Text and Its Phonosemantic Structure. M.Phil.Diss. Abstract. Moscow, 1985. (In Russian); I.A.Mazanayev. Chief Groupings in Sound Symbolic Words. M.Phil.Diss.Abstract. Leningrad, 1985. (In Russian); L.A.Komarnitskaya. Subjective and Objective Sound Symbolism in English. M.Phil. Diss.Abstract. Odessa, 1985. (In Russian); L.F.Likhomanova. Semantic Filiation of English Iconic Verba Movendi. M.Phil.Diss. Abstract. Leningrad, 1986. (In Russian); S.V. Klimova. Verbs of Obscure Origin in Shorter Oxford English Dictionary. (Elements of Etymological Phonosemantics). M.Phil.Diss.Abstract. Leningrad, 1986. (In Russian); T.Koibayeva. Sound-Symbolic Words in English and Ossetian. M.Phil. Diss. Abstract. Leningrad, 1987. (In Russian).

/11/ R.Wescott. Protolinguistics: The Study of Protolanguages as an Aid to Glossogonic Research. - In: Origins and Evolution of Language and Speech. Annals of the New York Academy of Sciences, vol.280. New York, 1976.

PRINCIPLES OF INTONATIONAL STRUCTURING OF THE SPONTANEOUS MONOLOGUE

NATALIA BARDINA

Russian Language Department
Odessa State University
Odessa, Ukraine, USSR 270015

ABSTRACT

The spontaneous speech intonation has generally been viewed as a modification, or a version of the intonational structure of an audible written text. In actual fact, being of different nature, these systems possess intrinsic properties enabling them to regulate their elements' structuring and functioning capability.

INTRODUCTION

The 20th century linguistics has focused upon language as a system of historically evolved means of communication. Modelling language stratificationally has proved, however, to be imperfect and perfunctory in a great number of aspects, with the worked-out model failing to incorporate all facts of authentic human communication.

Attempts to make linguistics' subject matter more expanded and comprehensive, which brought about the emergence of speechology and psycholinguistics in particular, ought to be given credit to as constructive and generative.

On the other hand, it should be noted that an approach to subsystems dissimilar both qualitatively and functionally will call for diversity of the research method applied. An element of a system cannot be forcibly transposed into a different system wherein it is likely to acquire new qualities. Such ecclectics could lead to an inadequate analysis of the object under study as its integrity, structure and dynamism are supposed to be reflected by each element analysed.

A series of experiments carried out by the Experimental Phonetics Laboratory, University of Odessa, made it possible to establish specific features of spontaneous speech which manifest themselves through a direct conjugation of mental continuum and discrete language means.

The monologue seems to be the most spontaneous in this respect, being distinguished on the precept of independence of motivation and stability of concept, as well as its tendency toward informational adequacy alongside with the programme's simultaneous composition and implementation.

PSYCHOLINGUISTIC PRINCIPLES OF INTONATION DIVISION IN SPONTANEOUS MONOLOGUE

The basic divergence of the spontaneous monologue from other types of speech subsystems lies in its intonational structuring. In an extemporaneous speech its intonation functions as a means of conveying expression along with other language units, thereby reflecting the mode in which semantic categories within a described extralinguistic situation are being grouped in a speech/thought stream.

It has been established experimentally that, on being represented graphically, i.e. deprived og its intonation, a spontaneously generated utterance is often perceived as inconsistent and meaningless even by the speaker himself.

In reading, intonation patterns "wander over the grammatical surface of language" /1/, thus effectuating a mediated segmentation of sense pointed out and included in the text by the author. Figuratively speaking, the difference between the spontaneous speech intonation and reading intonation is similar to that between a living bear's skin which is essential for keeping the animal alive and enables one to make judgement concerning the condition of the whole of the animal's body - and a fur-coat made from this skin just to be sold. In fact, intonational segmentation of a spontaneous monologue points to the "apportioned" character of text composition.

In all existing speech generating models the function of intonation stands in need of a clear definition, being wrongly treated as identical to functioning of purely articulatory means. At the same time, an analysis of authentic monologues has suggested that communicative and expressive components of intonation emerge at different levels of speech generation and are subject to diverse psycholinguistic phenomena.

A generalized communicative and intonational model of a phrase is issued as far back as the communication level where a generally subjective sense is modelled. For instance, an individual wishing to obtain information conceives a notion of a question which is normally followed by the emergence of an intonational question marker preceding the utterance of an undetermined "a-a-a..." type with a rising tone, whereupon the question is modelled into an articulated language form.

In the process of semantic and grammatical programming based on the primary semantic pattern including separate words and their potential correlations /2/, intonation acts as a part of an operational unit belonging to this level, i.e. a syntagma. Effecting the same expressive function, the intonational and semantic aspects of the syntagma become interdependent and interfacial within a system. It is the syntagmas that provide the transposition of a semantically implicit pattern into an explicitly continuous text.

The psycholinguistic significance of the syntagma has been displayed by L.A.Chistovich and A.A.Leontiev, whereas this unit's intonational and syntactic dualism has been neglected only to lead to identification of the syntagma with C.Osgood's functional class or a UC of a sentence.

As both grammatical and semantic programming, as well as motor programming occur within a syntagma /3/, with intonation being far more variable than syntactic and lexical units, it is only through a complex analysis of a speech unit or "block" that some thought-formation functions of language could be observed "at the output"

FUNCTIONAL TYPES OF INTONATIONAL-SEMANTIC UNITS

Intonational and semantic variability of these "blocks" having been analysed, their functional heterogeneity has been established. The spontaneous speech creative components acting as both linear and motor programming units, i.e. syntagmas per se, are opposed by certain conjunctions, linking words and cliches whose form diversity is provided by the diversity of psycholinguistic values of intonation-sense complexes being created and reproduced, as well as the diversity of their correlations with the utterance generation levels.

Emerging at the articulatory level only, linking words adhere to syntagmas, thus providing their rhythmic completion, filling in hesitation pauses and relieving the speaker's operative memory. In this case a certain disagreement between mental and speaking activities of man is observed -- used in authentic spontaneous speech, linking words ("of course" "certainly", "first of all", "generally", etc.) possess considerable automatism and often convey no modality into the utterance; while they are being uttered, the speaker is busy planning the following utterance which results in a noticeable change of intonation characteristics, such as tempo abatement, a shorter fundamental frequency range and intensity range, as well as undivided tone contour and dynamic structure. As a rule, linking words overlay syntagmatic borders making them diffusive, as the operational unit's junction line lies, in fact, "under" the linking words which make no part of either syntagma: для меня это было, в общем, неожиданно; я пробыл, скажем, двадцать лет.

The information obtained from listeners' reports upon hearing these diffusive syntagmatic border audio strips seems to be rather contradictory: disjunction areas were detected either before or after linking words, or both. Some listeners deemed it necessary to call attention to the ambiguous status of these words. In the process of reading the same excerpts, the informants qualified linking words as parenthetical and expressing the speaker's personal estimations of the utterance. This is what determined the relevant intonational expression.

The use of linking words is supported by the speaker-listener antinomy: within a syntagma, the essential sense is marked by discrete language means, whereas the introduction of words conveying new information obscures the message being perceived. Hence, the spontaneous speech is characterized by rhythmization of senses, i.e. an alternation of more and less significant units within a syntagma or phrase framework.

The second type of the spontaneous monologue automatized units includes conjunctions used in complex and compound sentences of disjunctive class /4/: и, или, но, если, когда (and, or, but, if, when) etc. This automatization is not caused, however, by the same provisions as motivate the use of linking words. The conjunctions listed above seem to emerge in an individual's conscience as far back as at the primary record level. Their emergence marks junctions between the principal syntagmas, i.e. microthematic units. As for the linking words, they function as articulatory units only and thus belong to a more "superficial" level than conjunctions and syntagmas. In a spontaneous monologue, conjunctions effectuate "direct" connections of inner and motor programming levels, which is similar in effect to the function of linking words, i.e. relieving the linear expansion level, overlaying syntagmatic border lines and filling in hesitation pauses. Intonationwise, conjunctions are characterized by a wider melody range as well as a noticeable rise in the acoustic force of the utterance. Thus, an increased loudness along with an abated tempo of the conjunction "u" ("and") in a narrative monologue act as a marker for spontaneous speech.

The third type of operational units opposed to syntagmas is the cliche, i.e. an intonational/lexical/grammatical phraseological unit belonging to the motor level. These may be both general and idiolectal units created by frequent repetitions or habitual word combinations used by the speaker. Occasionally, cliches may enfold syntagmas positionally, covering up to 80 per cent of units in some microthematic unities; yet they do not equal syntagmas functionally since they are not fashioned in an actual speech act.

Cliches are determined by distinct intonation characteristics, i.e. a short fundamental frequency range, a tone contour with distinguishable termination tone modifications, a short intensity range, a faster tempo, etc. Since equal characteristics are possessed by syntagmas containing familiar, or thematic information at the whole text level,one can assert that the familiar/novel opposition in the spontaneous monologue is also associated with a partial discharge of the linear programming process as well as an accent shift to the articulatory level.

In the event of multisyllable cliche units, these may parcel out melodically into 3 to 4 phonetic-word segments, the latter fact being caused by the isochronism of intonational segments. This parcellation is of superficial, or motor nature, the cliche's dynamic integrity pointing to its intactness as a separate operational unit. Consequently, the spontaneous monologue's intonational variability is caused, primarily, by the functional heterogeneity of its formative units along with the value disparity of the motor programming and linear prognostication units.

SEMANTIC AND SYNTACTIC FACTORS OF INTONATIONAL STRUCTURING

Along with the abovesaid features common for all spontaneous monologues, their intonational structuring is determined by the peculiar component grouping for the described situation in the speaker's conscience. The syntagma's capacity and content as well as minimum and maximum melodical and dynamic distribution within it are determined by the semantic and syntactic aspects of the utterance; the syntagma's general tone model depends on how independent semantically and syntactically it is, e.g. the tone used for autosemantic units will be falling and falling-rising while the rising and rising-level tone will convey synsemantic units.

The way a correlative semantic category is formed into a separate syntagma depends, above all, on whether the speaker's conscience is focused on it or not and whether other semantic categories are relevant or irrelevant for the utterance.

An experimental investigation of the spontaneous monologue's syntagmatic structuring has brought about conclusions concerning two essential issues, i.e. intonational and semantic structures' symmetry within a syntagma and the syntagma's intonational variability determined by its semantic and syntactic purport. It was established that a disjunctive intonational presentation of the described situation's components is determined by these categories' semantic types as well as their informativeness. Thus, a neutral, or standard intonation similar to the like units in reading is typical for syntagma situations including all the necessary components of the semantic structure arranged neutrally. The purport of this type of syntagmas lies in the nomination of a complete situation or event which results, in the majority of cases, in the fact that these syntagmas are previously contrived and initiate a microthematic unity acting as a kind of antecedent, or follow a protracted pause. All syntagma situations operate on a falling or level melody scale with a rise-fall termination alternated by a falling termination (full stop) in reading. Separate words carry no accentuated stress. The whole dynamic structure attenuating with the maximum emphasis laid on the initial syllables, the final stressed vowel stands longest.

The subject-predicate syntagmas may include a varied number of principal,or elementary and auxiliary, or non-elementary categories. A distinctive feature allowing to oppose this type of syntagmas to other types is the imperative accentuation of the subject and the predicate which obviously points to the speaker's conscience focusing on these principal situational components. As a rule, the subject in this type of syntagmas is expressed by a semantically autonomous word, while the predicate's content is presented by qualificative relations: /жизнь ведь настолько сложна/, /программа очень интересная/.

All syntagmas of this type possess a double-peak intonational structure, with the melodical and dynamic curves being isomorphous. In reading, the segments reproducing qualificative syntagmas do not hold a specific intonation structure.

Words like очень, настолько ("very", "so") etc. expressing a higher degree of attributed quality, are accentuated on a regular basis. This points out the fact that the lexical meaning of a separate word is realized more distinctly in reading, while this word, along with the syntactic arrangement of the utterance, determines the choice' of an intonational structure, with

with the semantic and syntactic factor retreating to the secondary plane.

The intonational prominence of action verbs in short stories disguises the absence of norm-premeditated dependent wordforms. The rising-falling termination tone used to shape up a vacant-valence verb syntagma provides an opportunity to allege that this type of predicate could operate as an autosemantic unit in the speaker's conscience.

The logical and intonational division of the predicate may stand out as a specific feature of the spontaneous speech, the two-component predicate including a modal element like можно, нужно, необходимо ("possible", "necessary") plus an infinitive.

An expanded subject in the spontaneous speech prevails as an independent sense-intonation unity. Apparently this may be determined by the fact that the subject or object attribute acts, in effeet, as a disguised predicate while a syntagma is capable of including a single vectorial (predicative) element only, with the programme unfolding extemporaneously.

In a number of cases, an insufficiently informative subject "attracts" the various predicate elements like auxiliary verbs, linking verbs with modal, phase and emotional meaning, adverbial modifiers, demonstrative pronouns, etc. In the vector syntagmas thus formed, it is the predicative elements that carry the maximum sense, with the subject's position being optional and determined, for the most part, by the tendency toward phonetic and sense rhythmisation.

The data obtained in the process of our investigation revealed the syntagma being center-oriented in extemporaneous conceptual speech. This is caused by the fact that sense accentuation is expressed by intonational accentuation in a progressively unfolding speech. Consequently, an informatively prominent object is generally represented by a separate syntagma.

It should be noted that the predicate-object division and subject-predicate syntagmization are ensued by the same phenomena, i.e., provided the emergence of an object expressed by an explanatory clause be rather indeterminate, the predicate syntagma is adjoined by the "object index" -- the conjunction ("that"): /представляла, что/ считаю, что/ думаю, что/. Based on the regularity of this peculiarity of intonational segmentation, one can assume that explanatory verbs are recognized by the speaker not only as lexemes but as syntaxemes as well; in other words, the verbs' semantics includes their valences in the mind of the speaker. Therefore, the dialectic unity of the speech stream's discreteness and continuity brings about the emergence of assymmetric syntagmas wherein the intonation value is shifted to the right of syntactic structures.

The tendency of the spontaneous monologue toward continuity is materialized in the conduct of the acoustic constituents. In extemporaneous speech, the opposition of logical and hesitation pauses appears to be practically obliterated, as it is these types of pauses that mark the junctions of psycholinguistic units. The purely syntactic pauses observed in reading appear to be non-existent in spontaneous speech. If the speaker relates some facts that he is well aware of and convinced in ("communication of events"), neglectful of whether these facts are to be correctly apprehended, it is only physiological pauses that he is likely to make in the speaking process. The use of linking verbs, conjunctions and "superfluous" pronouns reduces the number of hesitation pauses considerably.

Fundamental frequency variation depends on the syntagmas' autosemantics/synsemantics and their modality as well as indicates the significance of the semantic categories in utterance generation and compensates for grammatically inadequate structures.

The type of dynamic curve is called forth by the reproducibility of the intonation-sense unities in the text.

The spontaneous syntagma's tempo arrangement depends on the probability of the subsequent unit's emergence, whereas in reading it is the isochronal factor that is more operative.

CONCLUSION

Generally speaking, the spontaneous monologue's intonational structure is modelled by the direct thought-formation and extralinguistic apprehension process whereas a usage-bound interface of lexical, grammatical and intonational means can be observed in reading.

REFERENCES

/1/ A.M.Peshkovsky."Intonatsia i grammatica". In: "Izbrannye trudy", Moscow, 1959, p. 191.

/2/ A.A.Leontiev. "Psikholingvisticheskiye yedinitsy i porozhdeniye rechevogo vyskazyvaniya", Moscow, 1969.

/3/ V.A.Kozhevnikov and L.A.Chistovich (ed.) "Rech. Artikuliatsiya i vospriyatiye", Moscow - Leningrad, 1965.

/4/ V.N.Beloshapkova (ed.)"Sovremennyi russkii yazyk", Moscow, 1981, p. 537.

АКУСТИЧЕСКОЕ ВЫРАЖЕНИЕ "БЛОЧНОГО" ХАРАКТЕРА СПОНТАННОЙ РЕЧИ

АЛЛА БАГМУТ

Институт языковедения
Академия наук УССР
Киев, СССР, 252001

РЕЗЮМЕ

Спонтанная речь рассматривается как состоящая из блоков, имеющих определенную просодическую структуру и характеризующихся семантической неполнотой. Блочный характер спонтанной речи позволяет говорящему строить речевой поток как непрерывную семантическую цепь. Блок спонтанного текста не полностью коррелирует с синтагмой. Паузальное вычленение блока создает его четкие акустические границы.

ВВЕДЕНИЕ

Процесс речеобразования и формирование смысловой структуры высказывания осуществляется как создание внутренней программы /А.А.Леонтьев; В.А.Патрушев/. Одновременность создания внутренней программы речи, выбор лексических средств, коррекция речи - все это проявляется в ряде таких черт, как лексическая и семантическая повторяемость элементов речи, ее структурная и смысловая незавершенность, прерванность, самоперебивание, уточнение, разъяснение и дополнение сказанного ранее, наличие непроективных синтаксических конструкций, хезитационная паузация, нарушение связности элементов текста. При таком смысловом и синтаксическом построении текста нахождение определенной единицы /предложения, resp. фразы/ очень затруднительно. Исследователи русской разговорной речи указывали на сложность определения границ предложения в спонтанном тексте /А.Б.Шапиро; О.А.Лаптева/.

Слитный речевой текст членится паузами различной длительности на большие или меньшие отрезки - блоки спонтанной речи. Эти объективно представленные в речи структурные единицы не совпадают ни с фразой /предложением/, ни с синтагмой, ни со словосочетанием или словом. Вместе с тем блоки речи обладают большей или меньшей смысловой самостоятельностью, являясь средством развертывания речевой цепи.

МАТЕРИАЛ ИССЛЕДОВАНИЯ

Украинская спонтанная /неподготовленная/ речь была записана от 8 информантов в различных ситуациях: при обсуждении агротехнических вопросов на собрании агрономов колхоза /Киевская область/; воспоминания матери о войне /запись проведена ее сыном дома /Черкасская обл./; беседа со служащей в ее учреждении /Черниговская обл./; разговор на улице четырех колхозниц /Черкасская обл./. Во всех случая.. украинский язык был родным языком информантов. Представлено 5 женских и 3 мужских голоса.

Пример спонтанного текста: I/300 мс/ наш Володя та візьме /540 мс/ гм /260 мс/ таке зробив /70 мс/ жорна /70 мс/ Отак візьму та на... /I00 мс/ та надеру ж і на... /320 мс/ на те та й ізварю /I000 мс/ та такий шестилітровий /230 мс/ чавунець /220 мс/ і це /280 мс/ і це /I60 мс/ на три рази /230 мс/ треба /380 мс/ То вони сидять та кажуть /I80 мс/ а /I60 мс/ скоро /I70 мс/ мамо будемо /I40 мс/ заколоту їсти /70 мс/ Ні зажарене ні зашкварене /230 мс/.

Text: Our Volodya made a millstone. So used to grind the grain with it. Then would put the flower into a six litre pot and boi it to make soup. And the children were sitting and waiting. When shall have the soup? and in it there was neither meat nor any fat.

По данным перцептивного анализа определены мелодические изменения и паузация текста; по данным инструментального анализа указана длительность пауз /мс/, темп речи информанта /=средняя величина длительности слога /мс/, тональный диапазон /Гц/, максимальные величины ч.о.т. /Гц/, интенсивность звучания /мм/.

РЕЗУЛЬТАТЫ ИНТОНОГРАФИЧЕСКОГО АНАЛИЗА

В состав блока может входить от I до 9 фонетических слов /большее количество слов встречается редко/. Лексическая длина блока находится в прямой зависимости от темпа речи информанта: при более быстром темпе длина блока обычно больше. Следует иметь в виду обусловленность темпа речи психологическими особенностями говорящего, ситуацией и темой общения. Например,в рассказе о себе, о своей работе и семье темп речи говорящего /женщина/ был равен 197,3 мс /блок содержит в среднем I,98 слова/; при взволнованном рассказе о родных, погибших во время войны /речь иногда перебивается плачем/, темп речи замедленный - 223,I мс /блок равен 2,42 слова/; спонтанная речь на собрании, требующая от говорящего /мужчина/ убеждения слушателей, имела более быстрый темп - II3,6 мс /блок равен 3,3I слова/. Хотя во всех текстах немало блоков, имеющих по одному слову, в среднем блок содержит 2-3 фонетических слова. В интонации славянских литературных языков /Николаева/ отмечено паузальное вычленение последнего слова фразы. Такое же явление наблюдается и в спонтанной речи. При этом нередко нарушаются семантические и грамматические связи: пауза появляется между определяющим и определяемым словом /під цукровий /I50 мс/ буряк; ... під першу-ліпшу /230 мс/ культуру/; между предикатом и субъектом речи, /а ви шо/ самі/ говорите?//, а кругом/ ліс//; перед обстоятельством /...можна покататися /390 мс/ на річці//,.. а я працюю /326 мс/ в горсобезі//.

Минимальный блок, состоящий из одного слова, чаще всего представлен междометием, союзом, местоимением/ну, так, ось, і, але, а, я/. Время звучания односложного блока обычно увеличено - 3I0 мс при среднеслоговой длительности речи 2I2 мс; 296 мс при I87 мс и т.д./. При этом месторасположение блока в речи - в начале, середине или конце фоноабзаца - не влияет на его длительность. В однословном блоке, имеющем 3-5 слогов, можно проследить тенденцию к уменьшению среднеслоговой длительности в начальной позиции. В речевом тексте можно наблюдать последовательность однословных блоков, либо же чередование более длинного речевого блока с несколькими краткими. Отмечено, что говорящий изменяет темп звучания кратких блоков таким образом, что длительность звучания каждого из последовательно расположенных блоков почти уравнивается. При семантически равной нагрузке блоков существует их темпоральное уподобление: розкачуються в Драбові /I22,4 мс/, розкачуються в Золотоноші /I02,2 мс/, розкачуються в Смілі, пон'мате /I20,8 мс/. Значительная вариативность в акустической структуре блока имеет как функциональное, так и собственно конструктивное значение, важное для речевого потока. В завершающей части блока может быть представлен восходящий или ровный тон на среднем тональном уровне, определяемый как неполное завершение /ожидание продолжения речи/. На конечном ударном гласном может быть представлено восходяще-нисходящее движение тона, где нисходящий тон охватывает не больше I/3 звука и не влияет на перцептивную характеристику гласного как восходящего. Довольно последовательной является зависимость, при которой начало следующего блока тонально ниже завершения предшествующего блока. При этом пауза, разделяющая блоки, может быть длительной, очень длительной и краткой. В спонтанной речи возможны также различные виды соотношения конца предшествующего и начала последующего блоков.

Тональные изменения в блоке, состоящем из нескольких слов, чаще выступают в виде волнообразного контура с высоким завершением. Довольно четко проявляется компенсирующий характер временных, динамических и тональных характеристик.

Анализ частотного диапазона блоков спонтанной речи, проведенный в связи с семантикой текста, позволил установить большую частотность блоков с диапазоном I5-20 пт; блоки с узким диапазоном /I0 пт и меньше/ содержали либо вводную конструкцию, либо союзы и частицы, то есть семантически и интонационно отличались от основного текста. Блоки с более широким тональным диапазоном оказываются маркерами начала фразы, установленной аудитивным путем.

Интенсивность спонтанной речи превышает, по нашим наблюдениям, интенсивность литературной речи. В каждом блоке спонтанной речи увеличение амплитуды интенсивности подчеркивает информативный /семантический/ центр, выделяемый информантом. Амплитуда интенсивности очень вариативна, максимальная величина и ее местоположение в блоке трудно предсказуемо. Вместе с тем можно отметить некоторое чередование интенсивных и слабоинтенсивных участков речи в одном блоке. Уровень интенсивности в завершении блока достаточно высок и обычно соответствует высокому тональному завершению. Можно также отметить конструктивное тяготение пика интенсивности к начало блока. Интенсивность возникает в каждом блоке изначально, величина ее соотносима с семантико-коммуникативной нагрузкой блока.

ЛИТЕРАТУРА

Багмут А.И., Борисюк И.В., Олейник Г.П. Інтонація спонтанного мовлення. – Киев: Наукова думка, 1985. – 215 с.

Лаптев О.А. Некоторые эквиваленты общелитературных подчинительных конструкций в разговорной речи. – В кн.: Развитие синтаксиса современного русского языка. М.: Наука, 1966, с. 53-60.

Леонтьев А.А. Язык, речь, речевая деятельность. М., 1969.

Николаева Т.М. Фразовая интонация славянских языков. – М.: Наука, 1977.

Патрушев В.А. Структурно-семантические различия письменной и устной речи. Автореферат дис. ... канд. филол. наук. М., 1978. – 24 с.

Скребнев Ю.М. Введение в коллоквиалистику. – Саратов, изд-во Саратовского ун-та, 1985. – 208 с.

Черемисина Н.В.Русская интонация: поэзия, проза, разговорная речь. – М.: Рус. язык, 1982. – 206 с.

Шапиро А.Б. Очерки по синтаксису русских народных говоров. – М.: изд-во АН СССР, 1953. – 317 с.

ATTRIBUTION OF RUSSIAN LITERARY PARAPHONETICS

Zh.V.GANIYEV

MOSCOW STATE PEDAGOGIC INSTITUTE
OF FOREIGN LANGUAGES,USSR, 119034

ABSTRACT

What remains in the significant after the language-specific information is "taken away" - viz types and number of gap fillers, prolongation of vowels and consonants, unusual vowel variations at the end of sense-groups, non-linguistic pauses etc. - functions in conformity with the general norms of human behaviour as part of language etiquette.

INTRODUCTION

A full description of text phonetics includes, apart from the positional realization of phonemes and prosodic means, a specific "remainder" which means types and number of non-linguistic speech gap fillers, prolongation and distinctive segment variations at the end of speech segments etc. At the level of phonetics a linguistic message is accompanied by parentheses and non-linguistic segment variations (paraphonetic means). An investigation of reading and unprepared speech (an interview and text rendering, 10 hours of taped materials in total) has shown that the Russian literary language speakers, examined in the experiment used in their unprepared speech from 15 to 20 different paraphonetisms per minute. In reading this figure is 8-10 times less. Paraphonetisms are as if attached to the end of speech segments and are, therefore, of supersegmental nature, even though they do not accompany phonemes as tone or intensity do. Like social and situational variations of explicative means, paraphonetic phenomena are never rigidly determined by stratification and/or situational factors. Yet, one may speak of a certain regular correspondence and subsequently of a sociolinguistic value of probability correlations, as applied to paraphonetic phenomena.

PARENTHESES

a) We observe introduction of er sounds (refered as er'ling later on in the paper) in all types of unprepared speech as a sign of reflection (hesitation) in a certain emotional state which can range from emotional strain to neglect of the situation. In the overwhelming majority of cases these sounds are of incomplete formation and are considerably less intense than linguistic sounds.(This is additionally indicated with round brackets). Not any er'ing strikes one's ears with one and the same effect. There are moderate types and there are "monstrous" types which can be heard right away and are unpleasant (Cf. [(ə), (m:) and[ə:üə:mn'i],[ə:ufə:m]. Naturally enough first types are more frequent). Er'ing also depends on the functional state of the examinees. They had different attitudes towards their interviewers and the experiment itself since their social experience made them assess differently one and the same situation even though the experiment took place in surroundings quite usual for them, i.e. they were interviewed at the plant, in the theatre or in the college where the informants work or study. Moreover no-one of them had not met the interviewer before. One part of them who were young workers from Moscow enterprises were somewhat embarrassed in their speech behaviour, taking the interview as something unusual. By intuition they tried to improve the aestheticism of their speech and er'ed in answering questions and rendering me text far less frequently than those who were more at ease, i.e. students and actors.

Students proved to be more at ease of them all: their noticeable unpleasant er'ing accounts for 40 per cent of all parentheses. In actors and young workers moderate and hardly noticeable er'ing accounts for two thirds of all parentheses. It is probable that part of the non-too-pleasant-for-ear er'ing was controled by the speaker. He as if shifted off his responsability for his unpleasant mumbling onto the interviewer, as if saying: You ask me a question and I ponder and search. You can see that, so just have patience. Maybe in different, less nervous surroundings, when they had no strict necessity to follow the thread of the interview their er'ing would be more moderate in terms of quality and quantity.

b) Prolongation of vowels (final vowels mostly) is a phenomenon similar to er'ing. Prolongation of consonants is somewhat further from er'ing as vocal phenomenon since also voiceless consonants can be prolonged. Prolongation of vowels is observed most often in the conjunctions что,и, parasite words ну,вот, in certain prepositions and the particle это. Sometimes it is also observed in personal pronouns. Prolongation of speech-sounds is observed in the unprepared speech roughly as often as er'ing (100 times an hour). Vowels are prolonged 4 times more often than consonants. At any rate, correlation between the numer of cases of er'ing and the prolongation of speech-sounds in individual informants proved to be inversely proportional: those who abused of er'ing used prolongation less frequently whereas those who er'ed less often and unnoticeably prolonged speech-sounds more frequently.

c,d) Physiologically the phenomena of the breath and voiceless explosion of vocal cords are similar to prolongation of consonants. The breath is held subconsiously when speech sounds are on the verge to appear and the stage of control selection of speech units is not yet finished. Cord explosions are due to similar reasons. Six hours of unprepared speech yielded 70 cases of breath holding phenomenon. This phenomenon is practically non-existant in the speech of those who are used to answer questions and to solve speech problems. Naturally enough, best of all these requirements are met by students. As to the voiceless explosion of cords, most cases of this phenomenon are observed in those who have a habit to hold breath.

e) Noisy sighs when inhaling or exhaling, similar to noticeable er'ing, are out of place and indestrable in the context of official surroundings. The experiment has shown that er'ing and noisy sighs have a statistical correlation: when choosing the next words in unprepared speech young workers produced noisy sighs twice less frequently as students and actors did.

f) When the front of the tongue comes unstuck from the hard palate and the sides of the tongue come unstuck of the cheeks they produce smacking which fills the pause necessary to ponder over a phrase. Besides this, smacking typical of a tired reader when he starts a new paragraph. Smacking, if audible, is just as unpleasant as mumbling. Smacking can mean both neglect towards interlocutor or disappointment with the subject of the discussion. Actors use smacking to create an image of uncultured Philistine. At the last stage of the experiment after 40-45 minutes of work the informants were asked to read aloud two texts (one of them jocular, as if produced by this very Philistine and the other, a serious one where the informant immitates a new announcer). The informants were getting tired as they were reading the second text after the first one. This was evidenced by the number of smacking in the students and workers, but not the actors. Those actors who use smacking to make the speech more colourful smacked practically in the jocular text every time they took the breath. Reading the announcer text none of them smacked at all. And yet, unprepared speech of actors produced different results. Here all of them smack. The experiment has shown that students smacked more than others (twice as frequently as actors and 8-9 times as frequently as young workers).

g) Non-linguistic pauses are also known as hesitation pauses, pauses of pondering or word selection etc. Retelling a six-hundred-word text the workers had more such difficulties than young intellectuals: the workers had 50 and almost 30 per cent more of non-linguistic pauses if compared with the actors and students respectively. If it is reasonable that in conformity with official etiquette a silent pause (er'ing for example), then this etiquette was more strictly observed by the young workers. The number of their silent pauses is 8 times as great as their er'ing. The same indices for actors and students are 2.2 and 5 times respectively.

VARIATIONS

Unlike parentheses, variations are not sighs of reflection, selection or hesitation. The-э paraphonetic phenomena are widely spread in reading and are socially determined as sighs of speech carelessness.

a) Nasalization is probably the most harmful variation for the etiquette and it is considered to be even vulgar. Nasalization does not depend on the vicinity of a nasal consonant. It is produced by air escaping simultaneously through nasal and mouth cavities (...традиционный праздник кни[г'ъ̃]). This is caused by a careless pre-emptive lowering of the soft pulate (and the tongue) before the speech segment is ended. The actors did not have any nasalization at all when reading, even though in general this phenomenon was registered 600 times in the course of the experiment. Both reading and unprepared speech of the workers account for 80 per cent of all nasalization cases. In each individual group the females nasalized pre-pause sounds 5-6 times more frequently than the males.

b) The speech organs wanting to take their neutral position pre-emptively "demobilize" post-stressed high vowels at the end of sense-groups: Ларис,а где твои нарцисс[ъ]? Красивые они бы[л'ъ^{е}], Дорогие москвичи и гости столиц[ъ̃]!etc. In reading such phenomena are much less frequent than in unprepared speech (1:5). Professional young actors read without any "demobilization" at all. The students account for 10 per cent of the cases, the remaining 90 per cent are accounted by the workers. The actors account for 15 per cent of all cases of "demobilization" in unprepared speech. Students and young workers account for the remaining

30 and 65 per cent of cases respectively. Furthermore it should be mentioned that high vowels deformation is predominantly typical of females.

c) Hemming at the end of sense-groups is also another result of speech carelessness. It is produced when vocal cord continue to vibrate whereas the lips are already closed in the neutral position. 180 cases of strong and the same number of weak cases of hemming were registered in the experiment: в городе(м), девушки(м), о женитьбе(м),про [н'иво́(м)] etc. This phenomenon prevails in the females of all social groups. Also it is less marked in students and most important in workers.

CONCLUSIONS

The experiment was conducted with informants part of whom are actors speaking professionally stage Russian, others are Russian language students who had studied Russian orphoepy and expressive reading and still others are workers who have no special knowledge in Russian orphoepy. All the informants can be considered young people by the standards of Soviet psycology. All of them had spent their early years in Moscow (such selection allowed to prevent undesirable distorsions of results). Each of the three groups consists of an equal number of informants having general secondary education at least. The groups were equally devided into males and females. The experiment has shown that paraphonetics in speech is governed by the general norms of human behaviour as part of the speech etiquette. Wherever the linguistic norm is based, among other things, on aesthetic factors, viz those of taste, cultural tradition, pretige of established standards, the non-linguistic norms are somewhat similar to linguistic norms or rather to the most elementary norm of usage.

Apparently innate, since they manifest their existence in non-linguistic functioning of the speech apparatus, paraphonetic means are, in fact, dependent on the current conventions. Reactions to them are conditioned by historically formed attitudes of the nation to specific verbal behaviour patterns.

PERCEPTION OF ENGLISH WORD ACCENTUAL PATTERNS IN THE SPEECH OF L_2 LEARNERS OF ENGLISH

L.Y.KUKOLSHCHIKOVA I.V.PANKOVA

Dept. of Phonetics
Leningrad State Univ.
Leningrad, USSR, 199034

ABSTRACT

Perception by American subjects of English iambic and trochaic words spoken by L_2 learners of English whose native tongues were Russian,Chinese and Kirghiz has shed some light both on universal and specific properties of English word stress. The results seem to be helpful in practical pedagogical work.

INTRODUCTION

In English word stress,as is well known, one can find the whole array of phonological problems: first and foremost, it plays an important role in organizing, recognition and discrimination of words. Another problem concerns the existence of degrees of stress. A specific feature of English stress is the occurrence of full grade quality vowels in pretonic and post-tonic syllables. In the latter case Gimson /2/ speaks about prominence of unstressed syllables. Then a question arises as to whether these two linguistic notions have anything in common with linguistic experience of English speakers. An experimental study conducted by Nadibaidze /5/ provided no evidence in support of the existence of four distinctive degrees of stress in AE and seemed to confirm the view expressed earlier by Zinder /7/ that discrimination of the degrees of stress in phonetic terms is hardly possible. It may only have phonemic value.

Recent findings in the domain of word stress have been mostly connected with its perceptual aspect, human linguistic experience being in its centre, as Bondarko /1/ puts it. Within the frame of such an approach one can place studies on perception of distorted word accentual patterns. It is a widely known fact that target-language word accentual patterns suffer great changes in the speech of foreign language learners. Experimental perceptual studies of such patterns have become popular and contribute both to understanding stress as a language universal and to providing guide-lines in practical pedagogical work.

METHODS AND MATERIAL

The present study was aimed at investigating the perception by American subjects of accentually distorted words in non-native pronunciation. The non-natives chosen for the experiments were Russian, Chinese and Kirghiz speakers. The languages they represented either distantly resembled English or crucially differed from it in terms of stress function in the phonetic systems of the languages under study as well as in the stress pattern manifestation.

Russian word stress is known to be an efficient means of combining syllables of a word into a close-knit unit. The existence of stress in Chinese has been proved by experimental investigations /6/. However, the viewpoint depriving Chinese of both the word in its classical form and word stress is not uncommon /4/. In a vowel harmony bound word of the Kirghiz language the main stress is traditionally assigned to the last syllable. Coexistence of two prosodic layers/stress and vowel harmony/ seems to be doubtful.

Cross-language investigation in this study was not an end in itself. Cross-language interference has been used here as a natural means of modifying the accentual structure of a word in a predictable direction.

The inventory of linguistic material consisted of 135 English iambic and trochaic words, two syllable in size, with all the monophthongs possible both in stressed and unstressed syllables. The words were embedded in a contextually neutral carrier-phrase. The material was recorded by two Americans, two Russians, two Chinese and a Kirghiz speaker. The tokens produced by Americans were used as reference material. While recording the material,care was taken to ensure that the target-words were in a phrase-nuclear position. The tone was falling.
The team of trained phoneticians has chosen a set of 500 stimuli both close to the standard stress patterns and deviant from them.
The first listening test /Experiment I/ was designed to test American subjects on their ability to define accentual patterns in the presented material. The subjects were American students enrolled in a Russian language programme at The Leningrad State University. The procedure was the following: the subjects were asked to listen to experimental phrases and define the accentual patterns of the target words, i.e. the number of syllables and stress placement, using the symbols ´–, –´, ´´. /To keep the experiment within the limits of feasibility, several subsets were prepared/.

RESULTS AND DISCUSSION

The results were evaluated by a signtest. The stimuli were divided into two major groups: the first, where our subjects displayed complete consistency in their judgements disregarding whether they identified the right or the wrong stress pattern, and the second group where their responces were inconsistent. /In TABLES I-IV: S1,S2 - American speakers, S3, S4 - Russian speakers, S5, S6 - Chinese speakers, S7 - Kirghiz speaker/.

TABLE I. Data on the perception of all the stimuli.

Speakers	Patterns	Consistent judgement,%		Inconsistent judgement,%
		right	wrong	
S1	´–	78	0	22
	–´	74	0	26
S2	´–	91	0	9
	–´	88	0	12
S3	´–	87	0	13
	–´	97	0	3
S4	´–	90	0	10
	–´	88	0	12
S5	´–	70	2	28
	–´	96	0	4
S6	´–	68	0	32
	–´	56	0	44
S7	´–	39	47	4
	–´	59	29	12

The following observations can be made from TABLE I. There were no wrong judgements of accentual patterns in the stimuli pronounced by Russian and Chinese speakers. The opposite results were obtained for the stimuli spoken by the Kirghiz speaker. He was unable to realize the required accentual pattern, stress phenomenon being apparently of no linguistic importance for him.
The group of stimuli inconsistently judged by the subjects was not uncommon even within the group of tokens produced by American speakers. It is in that group that the subjects either placed two stress marks in one word, or shifted the stress mark from the syllable prescribed by the norm.

TABLE II. Data on perception of consistently judged stimuli.

Speakers	Patterns	PERCEPTION,%		
		Standard	Non-standard	
			one stress	double-stress
S1	´–	61	3	36
	–´	59	3	38
S2	´–	59	3	38
	–´	66	8	26
S3	´–	61	8	31
	–´	71	8	21
S4	´–	62	12	26
	–´	63	10	27
S5	´–	59	9	32
	–´	49	12	39
S6	´–	59	4	37
	–´	59	5	36
S7	´–	52	6	42
	–´	60	4	36

When a t test was applied to the data of TABLE II, the values obtained never reached significance /t_{05} = 1.96/, but for all the speakers the values were rather close to the critical value, especially so for the Kirghiz speaker. It seems that the subjects, when evaluating the accentual patterns of the group under study, found themselves in an ambiguous situation.This ambiguity, or entropy /H/, was subject to testing /3/. In our experiment H_{max}=1.58, H_{min}=0.

TABLE III. Entropy data on the consistently/A/ and inconsistently/B/ judged stimuli.

Speakers	H, BITS OF INFORMATION			
	A:		B:	
	´–	–´	´–	–´
S1	0.67	0.79	1.11	1.16
S2	0.59	0.60	1.13	1.21
S3	0.47	0.63	1.25	1.11
S4	0.65	0.56	1.30	1.26
S5	0.58	0.64	1.29	1.40
S6	0.54	0.52	1.17	1.20
S7	0.69	0.48	1.25	1.16

From comparison of the data it is obvious that the subjects experienced greater difficulties when defining the patterns of the stimuli in group B. To put it figuratively, those stimuli possessed a kind of "eroded", or "loose" structure.
It remained to be seen whether there were any phonetic grounds for assigning the stimuli to that group. At the present stage of the study we were content with evaluating the quality of vowels in unstressed syllables by ear.
The stimuli obtained from American speakers were those with full vowels in unstressed syllables. It is of interest that the entropy values calculated for all the stimuli produced by S1 were different for tokens containing /ɪ/, /ə/, on the one hand, and for full vowels in unstressed syllables, on the other hand, the values being 0.66, 0.78, 0.97 for iambic words and 0.65, 0.65, 0.95 for trochaic ones, respectively. The occurrence of full vowels in unstressed syllables seems to bring about ambiguity in defining the accentual pattern.
Unstressed vowels produced by Russian speakers were considerably reduced both in quality and quantity.Nevertheless there was a group of stimuli where unstressed vowels seemed to be less obscured, where our subjects displayed a tendency to the increase of inconsistency.
As TABLE I indicates, inconsistency was most pronounced for Chinese speakers. The explanation of this tendency might be the preservation of tonal features in both stressed and unstressed syllables, tone I and tone IV being the most common associations with characteristics of both syllables of the word, as well as the relative pitch difference between them, as judged by ear.
In the experimental words pronounced by the Kirghiz speaker the unstressed vowels stood out both for quality and quantity which led the subjects to perceive them as stressed.
In planning Experiment I we started from the idea that a thick foreign accent would obscure the segmental structure and hence the meaning of the words under study. Our subjects were, therefore,expected to evaluate the accentual structure proper.
Experiment II was designed to test the validity of the results of Experiment I. The experimental material composed of randomly chosen stimuli of both A and B groups was presented to a group of 6 subjects in white noise / s/n ratio = -4dB/. Subjects were asked to tick the word in the carrier-phrase and mark stress /stresses/. Comparison of data of both experiments is given in TABLE IV.
As can be seen, the subjects' responces in Experiment I and Experiment II collapsed into the same groups with respect to their consistency and inconsistency.Thus, the procedure in Experiment I seems to receive a certain support. /It is worth noting that the data on perception of both iambic and trochaic words, pronounced by speakers 3-7 as well as of the whole experimental set of stimuli of reference speakers were combined/.

TABLE IV. Data on perception of stimuli in Experiment I and Experiment II.

Speakers	PERCEPTION, %					
	Experiment I			Experiment II		
	Standard	Non-standard		Standard	Non-standard	
		one/double-stress			one/double-stress	
S1, S2	88	4	13	72	10	18
S3	88	3	9	94	1	5
S4	86	4	10	86	8	6
S5	78	5	17	69	5	26
S6	77	3	20	70	5	25
S7	81	2	17	68	6	26

In summary, the results of both experiments seem to shed some light on universal and specific properties of English word stress. From the point of view of native subjects' perception the linguistic material was divided into two main groups. In patterns with /ɪ/ and /ə/ in unstressed syllables the subjects found no difficulty in stress placement, though one might expect phonetic manifestation of stress to vary in the production of non-native speakers. These findings can be treated as manifestations of the universal character of the stress phenomenon. In structures with unstressed full vowels, stress failed to perform its organizing function. It is this group of stimuli that gives us a hint of the specific nature of English word stress.The findings reported here do not contradict what is known about intimate links of word and phrase prosody in English. In a once-observed production of a sentence "We've got a 'canteen,'too", the nuclear tone occurred on a pretonic /according to the norm/ syllable /kæn-/. May not this shift of stress be attributed to the fact that for the English language speaker it is the rhythmic group and not the word that is of great importance both in language production and perception, as has been put by Kassevich? /personal communication/.
The reported data seem to have a certain instructional value for those involved in foreign language teaching. The knowledge of the universal and specific nature of word stress in a target and in native languages may contribute to devising a successful strategy in word accentual pattern training. Especially reassuring seems to be the finding that perception of stress

may not be hindered by its varying phonetic manifestation in the speech of L_2 learners of English.

REFERENCES

/1/ L. Bondarko, Phonetic Description of Language and Phonemic Description of Speech, Leningradsky Universitet, Leningrad, 1981 (in Russian).

/2/ A.C.Gimson, An Introduction to the Pronunciation of English, Second Edition, Edward Arnold, 1970.

/3/ A. Yaglom, I. Yaglom, Probability and Information, Moskva, 1957 (in Russian)

/4/ V. Kassevich, Phonological Problems of General and Oriental Linguistics, Moskva, Nauka, 1983 (in Russian).

/5/ Y. Nadibaidze, Degrees of Stress: their Functions and Manifestations in AE, Leningrad, 1982, unpublished dissertation (in Russian).

/6/ N. Speshnev, Chinese Phonetics, Leningradsky Universitet, Leningrad, 1980 (in Russian).

/7/ L. Zinder, General Phonetics, Moskva, Vysshaja Shkola, 1979 (in Russian).

AN EXPERIMENT ON PERCEPTION OF LITHUANIAN VOWELS BY ENGLISH SUBJECTS

BRONIUS SVECEVIČIUS

Experimental Phonetics Laboratory
Vilnius State University
Vilnius, Lithuania, USSR, 232000

ABSTRACT

The English base of perception has been checked by an experiment on the correspondence between Standard Lithuanian vowels and Standard English ones. It was found that the Lithuanian vowels /i̇̀/ and /ù/ were the most difficult for the English listeners to respond to. It is suggested that the above vowels are characterized by some specific features which are not common in the vowel system of English.

INTRODUCTION

The present experiment was stimulated by a lack of objective data on vowel perception in the methods of teaching English pronunciation to Lithuanians. Such experiments also provide a useful check in elucidating the process by which vowel sounds of a foreign (quite unfamiliar) language are perceived and compared to the native ones. According to the theses of perceptive phonetics /1/, the perception of segmental and suprasegmental speech units is based on the comparison of these with the "standards". The rules of comparison stand for a program of operations by which a given signal is being compared with the standards.

The linguistic perceptive basis is formed in man in the process of mastering a given language (dialect). Different language systems correspond to different perceptive linguistic bases.

With this in view, we find it rather difficult to select a uniform group of naive listeners with a uniform perceptive basis of English though all the subjects might understand RP and use it in practice.

EXPERIMENTAL TECHNIQUE

A tape recording of Standard Lithuanian monophthongs and two "diphthongoids" (isolated and in /t/+V position) spoken by an experienced phonetician was presented to 25 phonetically naive listeners (first year students from Great Britain staying in Minsk State Pedagogical Institute of Foreign Languages). Each vowel with a normal duration was repeated three times and followed by a two-second pause. An instruction was played by tape recorder using RP (the text spoken by an Englishman, a teacher of English).

The first session of the experiment ended in failure. The listeners were not adapted to listen to the signals presented. Then the second session was arranged in which the subjects were asked to listen to the whole recording twice for audial adaptation before being given an answer sheet.

THE RESULTS

The results of the second session revealed a group of 10 listeners who were rather constant in their responses. Basing on the data obtained we find that all Standard Lithuanian vowels (including the "diphthongoids" /ie/ and /uo/) as perceived by the English subjects fall into 3 groups:

1. Vowels that readily find their correspondences in RP. Such Standard Lithuanian vowels are /a:/, /u:/, /o:/, /i:/, /ò/, /è/.

2. Vowels that are perceived as similar correspondences in English. Such are Lithuanian long diphthongized vowels /e̤:/, /æ :/ (which were accepted as the English diphthongs /ei/ and /Eə/, respectively) and the "diphthongoids" /ie/, /uo/ recognized as /ıɜ/ and /uɜ/.

3. Vowels that find no correspondences in English. No regularities were observed in perceiving the short Lithuanian vowels /ì/ and /ù/. This may be explained by different spectral as well as prosodic features for these vowels as compared to the English /I/ and /U/.

The alternative perception of /a/ (marked by "æ/a" in most answers) confirms the tendency of /æ/ being retracted to /a/.

The data obtained suggested some new clues to the methods of teaching English pronunciation to Lithuanians. These data were also applied in compiling a simplified phonetic transcription for a Lithuanian-English phrase-book /2/.

REFERENCES

/1/ Z. Japaridze, Perceptivnaya fonetika, Tbilisi, 1985. 96-104.

/2/ B. Svecevičius, Lithuanian - English phrase-book, Vilnius, "Mokslas", 1980. 3-233.

ON SOME PEDAGOGICAL ASPECTS OF THE VOWEL SYSTEM IN SPANISH, REGARDING CZECH LANGUAGE AS THE MOTHER TONGUE

JANA KULLOVÁ

Faculty of Philosophy
Charles University
Prague, CSSR, 110 00

ABSTRACT

Comparison of the acoustic structure of Spanish and Czech vowels su‸gest that even in textbooks designed for a wider public appropriate attention should be devoted to pronountiation.

Correct pronountiation of Spanish vowels has been given only little attention in Czech instruction books on Spanish. The authors have mainly relied on the facts that both in Spanish and Czech there are five vowel phonemes /i/, /e/, /a/, /o/, /u/ and the vowel position does not affect its auditive characteristics [1], [2], [3]. Major distinctions were seen in the Czech phonological quantity and numerous tautosyllabic vowel groups /di- and triphtongs/ in Spanish. In textbooks designed for a wider public there have been only scarce mentions about Spanish having vowel phonemes /e/ and /o/ with both open [ɛ], [ɔ] and closed [e], [o] combinatoric variants. Thus the authors usually stated that in general the Spanish and Czech vowel pronuntiations correspond.

On the other hand, auditive analysis by native speakers of Spanish discourses produced by Czech speakers, as well as experiences in teaching Czech to Spanish-speaking population have revealed a more complicated relationship between the two vowel systems.

Comparing the F1 and F2 values of Spanish and Czech vowels we obtain the following:

Czech vowels [7]:

[i]	- F1	300 - 500 Hz
	F2	2100 - 2700 Hz
[e]	- F1	500 - 700 Hz
	F2	1600 - 2100 Hz
[a]	- F1	800 - 1000 Hz
	F2	1200 - 1400 Hz
[o]	- F1	500 - 700 Hz
	F2	900 - 1200 Hz
[u]	- F1	300 - 500 Hz
	F2	600 - 1000 Hz

Spanish vowels [4], [5]:

[i]	- F1	202 - 243 Hz
	- F2	2308 - 2422 Hz
[e]	- F1	283 - 405 Hz
	F2	1822 - 2349 Hz
[a]	- F1	607 - 729 Hz
	F2	1012 - 1417 Hz
[o]	- F1	283 - 505 Hz

- F2 850 - 1012 Hz

[u] - F1 203 - 243 Hz

F2 576 - 850 Hz

It is evident that F1 has considerably lower values in Spanish that in Czech. The Czech vowels F2 values indicate, generally, mayor dispersion than that of Spanish vowels.

Thus the acoustic structure of Czech i-vowels falls into the dispersion area of Spanish e-vowels; the dispersion area of Czech e- and o-vowels partly cuts across that of Spanish a-vowels; the dispersion area of Czech u-vowel covers partly the dispersion area of Spanish o-vowels.

Czech a-vowels thus reveal incorparably higher F1 than their Spanish counterparts.

Bearing on mind that the resulting F1 and F2 values are in any case result of an overall configuration of the oral cavity [6] generalizing the statement about direct proportions between F1 and oral cavity opening, and between F2 and front articulation, we may suggest, especially for teaching purposes, that compared to Spanish the Czech vowel articulation is more open; thus thus Czech i-vowels reach the acoustic values of Spanish e- vowels, etc.

Under certain circumstances, inaccurate pronuntiation may hamper communication. This fact is a sufficient reason for maximal extension of both instructions and practical drilling of Spanish vowel pronuntiation in textbooks in preparation.

REFERENCES:

[1] J. Dubský, J. Carrasco, K. Hoyer, Španělština pro jazykové školy I., Prague, 1963.

[2] J. Dubský et al., Moderní učebnice španělštiny, Prague, 1974.

[3] L. Prokopová, Španělština pro samouky, Prague, 1983.

[4] A. Quilis, Fonética acústica de la lengua española, Madrid, 1981.

[5] M. Esgueva, M. Cantarero /Eds./, Estudios de fonética I., Madrid, 1983.

[6] J. Llisteri, D. Poch, Análisis acústico del timbre vocálico en las realizaciones normativas del plural en andaluz oriental; paper red at Simposio de la Sociedad Española de Lingüística.

[7] M. Romportl, Základy fonetiky, Prague 1981.

FAKTOREN DER PHONETISCHEN (WORT-)VERSTÄNDLICHKEIT IN DER FREMDSPRACHE (DEUTSCH)

URSULA HIRSCHFELD

Herder-Institut der Karl-Marx-
Universität Leipzig 7022, DDR

ZUSAMMENFASSUNG

Eine experimentelle Untersuchung des Zusammenhangs von Art und Grad der Ausspracheabweichungen in der Fremdsprache und der Verständlichkeit der fehlerhaften Äußerungen beim Muttersprachler ist für eine Neubestimmung von Ziel und Inhalt des Phonetikunterrichts von großer Bedeutung.

1. DAS PROBLEM DER PHONETISCHEN VERSTÄNDLICHKEIT IM FREMDSPRACHENUNTERRICHT

Fortschritte in der fremdsprachigen Artikulation und Intonation lassen sich meist nur mit erheblichem Zeit- und Konzentrationsaufwand erreichen. Es ist deshalb wichtig, ein gleichermaßen realistisches wie akzeptables Unterrichtsziel festzulegen. Lehrprogramme orientieren immer häufiger auf eine "verständliche Aussprache". Was gehört dazu? Ist phonetische Verständlichkeit - wenn schon realistisch, erreichbar - auch akzeptabel? Sollte man diesem Ziel nicht kritisch gegenüberstehen, wenn man bedenkt, daß man sich auch mit großen phonetischen Unzulänglichkeiten verständigen kann? Es scheint notwendig zu sein, das Phänomen "phonetische Verständlichkeit" zunächst zu untersuchen und zu erfassen, sein Verhältnis zur Korrektheit zu bestimmen, Faktoren und Bedingungen herauszuarbeiten.

Mit Hilfe zweier von mir vorgenommener Untersuchungen möchte ich die Problematik verdeutlichen. In der ersten Untersuchung ging es um die Beurteilung der Aussprache Deutschlernender durch Muttersprachler. Die phonetischen Leistungen von elf ausländischen Studenten der Mittelstufe wurden von fünf Phonetikern nach einer Skala von 1 (sehr gut) bis 5 (ungenügend) bewertet. 150 deutsche Studenten nahmen ihre Einschätzung mit Hilfe eines Polaritätsprofils vor, das 15 Merkmalspaare enthielt; eins davon betraf die Verständlichkeit, die Skala reichte von "sehr gut verständlich" (5 Punkte) bis "sehr schlecht verständlich" (1 Punkt). Die Untersuchung brachte folgende Ergebnisse:

1. Zwischen der phonetischen Leistung der Sprecher und deren Einschätzung durch deutsche Hörer läßt sich eine Korrelation nachweisen. Bei den Merkmalen "gut verständlich"/"schlecht verständlich" beträgt der Korrelationskoeffizient 0,48.
2. Die aus den Polaritätsprofilen errechneten Mittelwerte belegen, daß Sprecher mit einer schlechteren phonetischen Leistung von der jeweiligen Hörergruppe teilweise als besser verständlich eingeschätzt wurden als solche mit einem höheren Leistungsniveau. Das ist ein Hinweis auf das komplizierte Bedingungsgefüge, es gibt kein direktes, gradliniges Abhängigkeitsverhältnis zwischen phonetischer Leistung und Verständlichkeit.
3. Ein und derselbe Sprecher wurde von seinen Hörern ganz unterschiedlich beurteilt, in der Regel wurden mehrere Skalenwerte markiert, z.T. (bei fünf von elf Sprechern) wurden beide Extreme angegeben, also "sehr gut verständlich" und "sehr schlecht verständlich". Der subjektive Faktor spielt eine große Rolle.

Wie man sieht, ist es sehr schwierig und durchaus nicht unproblematisch, Äußerungen in der Fremdsprache im Hinblick auf ihre Verständlichkeit zu beurteilen.

Das zweite Experiment beschäftigte sich mit den Auswirkungen phonetisch fehlerhafter Äußerungen. Durch die erschwerten, vom Gewohnten abweichenden Perzeptionsbedingungen kommt es beim Hörer zu einem erhöhten zentralen Aufwand, zur Verlagerung der Aufmerksamkeit vom Inhalt auf die Form - und somit nicht selten zu Informationsverlusten /1/. Im Experiment sollte untersucht werden, wie hoch die Informationsverluste sind, wenn die Aussprache eines Deutschlernenden nicht sehr gut und nicht sehr schlecht, sondern "verständlich" ist. Zwei laotische Studenten, deren phonetische Leistungen bei 2,4 und 3,6 lagen (bewertet von fünf Phonetikern nach der Skala von 1 bis 5), lasen einen Sachtext, der von je etwa 20 deutschen Studenten einmal gehört wurde. Im Anschluß daran mußten sie Fragen zu den im Text enthaltenen Informationen beantworten. In einem Kontrolltest mit einem deutschen Sprecher wurden max. 24 der 26 Fragen richtig beantwortet (24 Punkte = 92,3 %). Der leistungsstärkere Sprecher verhalf seinen Hörern zu durchschnittlich 17,1 Punkten (65,6 %), der leistungsschwächere zu 12,8 Punkten (42,7 %). Die mittlere Standardabweichung betrug beim ersten s= 2,1; die Spanne der richtigen Antworten reichte von 13 bis 20. Beim zweiten lag die Streuung bei 3,5; es wurden zwischen 6 und 17 richtige Antworten gegeben. Der Informationsverlust kann eindeutig auf die schlechten phonetischen Leistungen zurückgeführt werden. Bei den Hörern beider Sprecher zeigte sich außerdem ein deutlicher Abfall bei den Fragen zum zweiten Teil des Textes; das zeugt davon, daß Aufmerksamkeit und Konzentrationsfähigkeit nachließen.

Beide Experimente bestätigen, daß man "phonetische Verständlichkeit" nicht von vornherein und ohne nähere Betrachtung zum Ziel des Fremdsprachenunterrichts erklären darf. Es gilt zunächst, Begriff und Bedingungsgefüge zu bestimmen und schließlich Faktoren und Kriterien zu ermitteln, die die Verständlichkeit befördern oder beeinträchtigen, diejenigen Merkmale der Vokale und Konsonanten festzustellen, die für die Perzeption besonders wichtig sind und nicht bzw. nur bis zu einem bestimmten Grad in ihrer korrekten Realisierung beeinträchtigt sein dürfen. Davon ausgehend, unter Berücksichtigung verschiedener anderer Anforderungen an eine akzeptable Aussprache, könnte das Ziel des Phonetikunterrichts neu durchdacht werden.

2. BEGRIFF UND BEDINGUNGSGEFÜGE

Beim Verstehen von Sprache wirken verschiedene sprachliche und außersprachliche Komponenten, zu denen auch die phonetische gehört. Sie wirken miteinander, ergänzen sich, und Abweichungen in einem Bereich können von den anderen Bereichen kompensiert werden. Auch phonetische Fehler können durch den Kontext ausgeglichen werden; phonetische Verständlichkeit kann mit artikulatorischen und intonatorischen Abweichungen erreicht werden. Diese Abweichungen bewegen sich in einem Toleranzbereich, der umso kleiner wird, je weniger Entschlüsselungshilfen anderer Art vorhanden sind, d.h. je mehr der phonetische Faktor isoliert ist. Die phonetische Verständlichkeit kann eigentlich nur an Minimalpaaren, Nonsens- und Quasinonsenswörtern (Familiennamen) überprüft werden. Man kann eine solche Äußerung dann als phonetisch verständlich bezeichnen, wenn ein Hörer Laute/Lautverbindungen/Lautfolgen und suprasegmentale Merkmale akustisch so deutlich wahrnehmen kann, daß er sie als Realisationen eines bestimmten Phonems/einer Phonemverbindung/Phonemfolge oder eines Intonems identifizieren kann. Dabei muß ein bestimmtes Verhältnis zwischen dem Grad der sprachlichen Korrektheit, bezogen auf das Phonemsystem bzw. die standardsprachliche orthoepische Norm, und der Verständlichkeit bestehen. In der normalen Kommunikation ist dieses Verhältnis von Sprecher zu Sprecher, von Hörer zu Hörer, von Äußerung zu Äußerung und von Situation zu Situation verschieden. Damit ist zugleich das Bedingungsgefüge angedeutet, durch das der Grad der Verständlichkeit bestimmt wird. Im folgenden möchte ich nur auf einen dieser Faktoren eingehen, auf die Äußerung, die vom Lernenden spezifisch gestört wird. Einige Überlegungen und Beobachtungen, auch offene Probleme seien zusammengefaßt:

a) Suprasegmentalia

Akzentuierung, Melodisierung, Gliederung und Rhythmisierung sind erwiesenermaßen von hervorragender Bedeutung für die Sprachauffassung. Abweichungen im suprasegmentalen Bereich erschweren das Sprachverstehen, sie können - z.B. im Falle der Akzentuierung - zur Bedeutungsveränderung führen, sie wirken irritierend und können negative Emotionen provozieren. Ungeklärt ist Frage, ob segmentale oder suprasegmentale Fehlleistungen schwerer wiegen, - wenn auch vieles darauf hindeutet, daß das Suprasegmentale das Segmentale überlagert und bspw. in Wörtern mit falschem Akzent die lautliche Ebene der neuen Akzentstruktur angepaßt/verändert wird.

b) Segmentalia

Es wird angenommen, daß phonologische Fehler (Systemverstöße) die Verständlichkeit mehr beeinträchtigen als phonetische Fehler (Normverstöße), obwohl das Problem noch nicht eingehend untersucht wurde. Desgleichen wird in der Literatur die Frage aufgeworfen, ob vokalische oder konsonantische Verstöße schwerer wiegen. Im vokalischen Bereich interessiert neben der Artikulationspräzision (Zungenhebung, Lippenrundung) vor allem das Verhältnis von Qualität und Quantität, bei den Konsonanten die Fortis-Lenis-Korrelation und die Artikulationspräzision in bezug auf Artikulationsart und -stelle.

c) Struktur

Bestimmte strukturelle Eigenschaften der Äußerung wirken begünstigend bzw. benachteiligend auf ihre Verständlichkeit. Dazu gehören distributionelle Gegebenheiten, aber auch die Silbenstruktur, die Silbenzahl, die Stellung der Akzentsilbe, die Lautzahl der Silbe und des Wortes, das Verhältnis Vokal - Konsonant. Es spielt sicher auch eine Rolle, ob sich der gestörte Laut in einer akzentuierten oder einer nichtakzentuierten Silbe befindet, ob er in der ersten, zweiten oder dritten Silbe des Wortes vorkommt u. ä. Von Bedeutung ist weiter die Gebrauchshäufigkeit eines Wortes sowie seine Zugehörigkeit zu semantisch und kommunikativ bestimmten Kategorien (wie Namen, Zahlen, Wörter, Sätze).

3. STAND DER EXPERIMENTELLEN UNTERSUCHUNGEN

Die phonetische Verständlichkeit ist Gegenstand zahlreicher Versuche in der Logopädie und in der Technik, denen man wichtige Einzelerkenntnisse entnehmen kann. Sie lassen sich jedoch nur in geringem Umfang auf die Fremdsprachenproblematik übertragen. Die Ursache dafür liegt in der grundsätzlich anderen Art der Störung; es handelt sich meist um die Veränderung nur einer oder nur weniger Komponenten (einzelner Laute/Lautgruppen bzw. physikalisch-akustischer Parameter). Zudem ist der grammatische und semantische Kontext meist fehlerfrei und eine zuverlässige Hilfe. Diese Bedingungen sind bei der Auffassung der phonetisch gestörten Fremdsprache nicht gegeben. Die Fehler treten für einen nativen und naiven Hörer unerwartet und unsystematisch auf. Neben der phonetischen sind auch alle anderen Sprachebenen betroffen, so daß der Kontext seine normale Funktion nicht erfüllen kann. Es gibt ein kompliziertes Bedingungsgefüge, ein Zusammen- und Nebeneinanderwirken verschiedener Faktoren. Bisher wurde kein Meßverfahren entwickelt, gibt es kein linguistisches oder kommunikationstheoretisches Modell, das im ganzen diejenigen Faktoren berücksichtigt, die für die sprachliche Verständigung in Realsituationen entscheidend sind /2/. So ist es nicht verwunderlich, daß der Untersuchungsstand noch völlig unbefriedigend ist, daß nicht nur gegensätzliche Auffassungen, unterschiedliches Vorgehen und widersprüchliche Ergebnisse zu konstatieren sind, sondern daß bisherige Untersuchungen fast ausschließlich auf Einzelerscheinungen gerichtet waren und oft eine sehr schmale experimentelle Basis zugrundegelegt wurde. Tiefer, komplexer in die Zusammenhänge vorzudringen, versuchen u.a. Marouškova (1982) und Bannert (1984), die bereits Zwischenergebnisse veröffentlichten /3/. Von der am Herder-Institut in Angriff genommenen relativ breiten Untersuchung seien im folgenden erste Teilergebnisse dargestellt /4/.

4. UNTERSUCHUNG DER PHONETISCHEN WORTVERSTÄNDLICHKEIT

Um die Wirkungsweise der verschiedenen phonetischen Faktoren zu erfassen, muß man zunächst andere, situative und kontextuelle Faktoren, semantische Hilfen, auszuschalten versuchen. Ein erster Schritt zur Untersuchung der phonetischen Verständlichkeit ist die Untersuchung der phonetischen Wortverständlichkeit. Deshalb bilden vorwiegend Einzelwörter, von Deutschlernenden der Mittelstufe gelesen/gesprochen, den Gegenstand einer umfangreichen experimentellen Untersuchung, an der etwa 50 ausländische Studenten und 700 deutsche Hörer beteiligt sind. Da die Datenauswertung noch nicht abgeschlossen ist, können noch keine Endergebnisse vorgestellt werden. In dieser Untersuchung kann es auch noch nicht darum gehen, Zusammenhänge nachzuweisen, dafür ist die experimentelle Basis nicht breit genug, sondern sie aufzudecken, Abhängigkeiten sichtbar zu machen; Faktoren und Bedingungen für die phonetische Verständlichkeit sollen gefunden, Toleranzschwellen für abweichende lautliche und intonatorische Realisierungen bestimmt, die phonetischen Leistungen (einschließlich der Akzeptabilitätsgrenze) durch Muttersprachler eingeschätzt und bewertet werden. Ziel soll sein, weiterführende Untersuchungen vorzubereiten, erste Konsequenzen für die phonetische Arbeit im Fremdsprachenunterricht zu ermöglichen. Auf zwei der Experimente möchte ich näher eingehen, sie dienen der Bestimmung einer Reihe von Bedingungen und Faktoren der phonetischen Verständlichkeit.

4.1. UNTERSUCHUNGSBEDINGUNGEN

a) Testmaterial

Es wurden je 52 zweisilbige Einzelwörter und Familiennamen (alles Teile von Minimalpaaren) überprüft, die die im Deutschen wichtigsten phonologischen Oppositionen enthielten. Die vokalischen Distinktionen wurden außerdem in 32 Logatomen erfaßt.

b) Versuchspersonen

Im Experiment I waren drei nikaraguanische und drei laotische Studenten der Mittelstufe die Sprecher und 127 deutsche Studenten (jeweils etwa 20 pro Sprecher) die Hörer. Im Experiment II wurde die Aufnahme eines der im Experiment I mitwirkenden Nikaraguaner von 40 deutschen Hörern abgehört, die acht Gruppen á fünf Personen bildeten: 8-jährige und 12-jährige (Schüler), 22-jährige (Studenten), 30-jährige und 40-jährige (Berufstätige), 65-jährige (Rentner), sowie Phonetiker und Sprachlehrer für Deutsch als Fremdsprache.

c) Versuchsablauf

Im Experiment I lagen die Minimalpaare, von denen jeweils ein Wort/Name von den Ausländern auf Tonband gesprochen worden war, den Hörern vor. Sie sollten nach einmaligem Hören markieren, welches Wort gesprochen wurde. Im Experiment II konnte die Aufnahme beliebig oft gehört werden, es war (ohne Vorlage) zu notieren, welches Wort/welcher Name verstanden worden war.

4.2. ERGEBNISSE

Für das Experiment I lassen sich beim gegenwärtigen Untersuchungsstand folgende Teilergebnisse nennen:

1. Es wurde angenommen, daß die Studenten mit der Muttersprache Spanisch aufgrund der größeren phonetischen Ähnlichkeit von Ausgangs- und Zielsprache besser verständlich sind als die Laoten. Das bestätigte sich nicht, wie die folgende Tabelle 1 zeigt: (Fehlerrate in %)

	insges.	Vokale			Konsonanten	
		Wörter	Namen	Logat.	Wörter	Namen
laot.	21,5	18,9	26,1	22,9	17,6	19,9
span.	23,1	26,5	28,7	19,3	22,4	17,6

Tabelle 1

2. Ein Vergleich der Fehlerraten bei sinnvollem (Wörter) und sinnleerem Material (Namen) ergab nur einen geringen durchschnittlichen Unterschied von 0,2 %. Aus Tabelle 2 geht hervor, daß bei den Spanischsprachigen die Namen sogar besser verständlich waren als die Wörter. Auch im Bereich der konsonantischen Distinktionen waren die Namen insgesamt gesehen besser verständlich als die Wörter, während es bei den Vokalen umgekehrt war: (Fehlerrate in %)

	insges.	laot.	span.	Vokale	Konsonant.
Wörter	21,0	18,2	24,1	22,5	19,9
Namen	21,2	19,2	22,3	26,5	16,3

Tabelle 2

Der sinnvolle Kontext erwies sich in der Regel nicht als Perzeptionshilfe.

3. Der Zusammenhang zwischen der Verletzung bestimmter distinktiver Merkmale und der Verständlichkeit wird noch überprüft, er soll exemplarisch für /ɑ:/ dargestellt werden:

realisierte Variante	Hörfehler		Hörfehler	
	in %	absolut	in %	absolut
ɑ:	22,2	4/18	11,1	2/18
a·	41,7	35/84	37,7	23/61
a	66,7	16/24	93,2	44/47
	"staatlich"		"Schahler"	

Tabelle 3

Auch wenn das distinktive Merkmal, hier die Quantität, eindeutig (kurz oder lang) realisiert wurde, treten zu einem nicht geringen Anteil Hörfehler auf; in dem in Tabelle 3 dargestellten Beispiel bei dem Wort "staatlich/stattlich" mehr als beim Namen "Schahler/Schaller". Daß auch korrekte Realisationen falsch interpretiert wurden, ließ sich in diesem Versuch generell beobachten, es hängt zweifellos auch mit der durch eine Reihe weiterer Fehler verursachten ungewohnt klingenden Aussprache des gesamten Wortes bzw. Familiennamens zusammen.

Bei der bisherigen Auswertung des Experiments II ergaben sich folgende Teilergebnisse:
1. Zunächst wurden die Fehlerraten mit denen des Experiments I verglichen, und zwar für den nikaraguanischen Sprecher, dessen Aufnahme in beiden Experimenten verwendet worden war. Es zeigten sich für das Hören mit und für das Hören ohne Vorlage ganz unterschiedliche Resultate:
(Fehlerrate in %)

	insges.	Vokale			Konsonanten	
		Wörter	Namen	Logat.	Wörter	Namen
I	20,5	17,7	29,3	10,4	21,3	16,3
II	52,3	39,9	75,3	46,6	48,3	55,3

Tabelle 4

Die aus Tabelle 4 ersichtlichen Unterschiede machen deutlich, welchen Einfluß der gestörte Kontert hat, der im Experiment I in schriftlicher Form korrekt vorlag, so daß diese Störungen teilweise neutralisiert wurden. Hier werden auch die Unterschiede in der Perzeption sinvollen und sinnleeren sprachlichen Materials deutlich, bei den Namen war die Fehlerrate sehr viel höher als bei den Wörtern.

2. Bei der Überprüfung der Hörergruppen zeigten sich gruppenspezifische Unterschiede im Grad der Verständlichkeit und in der Zahl und Art der Substitute. Für Wörter und Namen ergibt sich hinsichtlich des Verständlichkeitsgrades folgende Rangfolge der Hörergruppen: (Fehlerrate in %)

1. Deutschlehrer	47,9
2. 40-jährige	50,2
3. Phonetiker	50,8
4. Studenten	51,0
5. 30-jährige	55,0
6. Rentner	55,4
7. 12-jährige	59,0
8. 8-jährige	65,8

Darin läßt sich ein Zusammenhang mit der muttersprachlichen Kompetenz erkennen. Die Annahme, daß Erfahrungen in der Kommunikation mit Ausländern sich sehr positiv auswirken würden, hat sich nicht bestätigt - die die Testperson unterrichtenden Lehrer und Phonetiker haben sie nicht besser verstanden als andere. Diese Erfahrungen wirken sich im phonetischen Bereich offensichtlich nicht so stark aus. Und allgemeine Kommunikationserfahrungen, über die Fremdsprachenlehrer verfügen und die ihnen sehr helfen, den Lernenden zu verstehen, kamen in diesem Versuch mit Einzelwörtern nicht zum Tragen.

3. Alle Substitute wurden auf die erhalten gebliebenen Laute und Lautmerkmale hin überprüft, weil man davon ausgehen kann, daß die für die Perzeption wichtigen lautlichen Strukturen sich im Substitut wiederfinden. Besonders deutlich wird das bei den Namen, weil die semantische Klammer, der Zwang, ein sinnvolles Wort zu "verstehen", wegfällt. Bei den Vokalen erwies sich die Quantität, die in 96,9 % aller verstandenen Namen erhalten geblieben ist, als das stabilere Merkmal im Vergleich zur Qualität, die nur in 46,5 % aller Fälle beibehalten wurde. Ein langer/offener Vokal wurde also fast generell als langer/geschlossener aufgefaßt, ein kurzer/geschlossener als kurzer/offener Vokal. Hebungsrichtung und Lippenrundung erwiesen sich ebenfalls als stabil mit 96,7 bzw. 97,0 %, wenn substituiert wurde, dann durch Vokale mit den gleichen Merkmalen. Bei den (prävokalischen) Konsonanten steht die Artikulationsart an erster Stelle (95,7 %), dann folgt die Artikulationsstelle (93,4 %); die Fortis-Lenis-Korrelation blieb nur in 78,4 % der verstandenen Namen erhalten.

Soweit zu einigen ersten Zwischenergebnissen. In der weiteren Auswertung sollen die angedeuteten Zusammenhänge weiter geprüft, sollen Faktoren und Bedingungen einbezogen werden, die hier noch nicht erwähnt wurden, es sollen Hierarchisierungen vorgenommen und Signifikanzen festgestellt werden.

Anmerkungen

/1/ siehe z.B. G. Lindner, "Hören und Verstehen", Berlin 1977, Kap. 5.3.; G. Meinhold, "Die Meßbarkeit von Teilleistungen bei der Verarbeitung gesprochener Sprache", in: Sprechwirkung, Univ. Halle 1976, 60-63.

/2/ P. W. Kahl, " Die Erfassung der mündlichen Kommunikationsfähigkeit im Fremdsprachenunterricht", in: Kommunikation in Europa, Frankfurt u. a. 1981, 156-173.

/3/ R. Bannert, "Intelligibility of Foreign Accent", in: Abstracts of 10th ICPS 1984, 600.

M. Maroušková, "Zur Erforschung der Einstellungen gebürtiger Deutschsprachiger zu verschiedenen Varianten der deutschen Aussprache", in: Beiträge zur Sprachwissenschaft, Potsdam 1982, 131-147.

/4/ U. Hirschfeld, "Zur phonetischen Verständlichkeit deutschsprechender Ausländer", in: Wissenschaftliche Beiträge der Univ. Halle (im Druck).

THE DANISH "STØD"
THE PROBLEM OF PHONETIC REALIZATION AND PERCEPTION

Y.V. KRASNOVA

Department of Scandinavian Languages Leningrad State
University, USSR, Leningrad, 199164

ABSTRACT

The present investigation of the Danish stød is based on recordings by 13 Danish speakers. Duration, Fo, intensity, formant frequences of the stød-vowels has been compared with those of the stødless vowels. The stød-vowels are distinguished from the stødless vowels first of all by Fo pattern, than by quality and intensity changes. It has been found out how the stød-vowels are perceived by the groups of subjects with different level of Danish knowlegde.

A specific feature of the Danish prosodic system is that unlike closely related Swedish and Norwegian, there exists stød in Danish, which is a dynamic syllabic accent resulting in a brief vokal folds compression or a complete closure with an additional distinctive value.

The stød may occur in a long vowel, in a diphtong, almost in all sonorant consonants, in / ð / and only in a stressed position. By many phonologists the stød in Danish is considered to be phonemically distinguishing element. O.Jespersen represented 400 minimal pairs distinguished only by presence or absense of the stød. The stød is commonly found in monosyllables, but the majority of oppositions stød-word/stødless-word are disyllabic words: (jeg) læser [lɛʔsər]- (I) read, læser [lɛ:sər] - reader. The stød, however, here is not an independent distinctive element, it plays only a secondary part in the system of expressive means of the language, because the differences in these minimal pairs are obvious from word order in the sentence. Danish linguist A.Hansen pointed out that the stød is not absolutely necessary for understanding Danish /1/.

Historically, the stød is correlated with pitch accents of other Scandinavian languages. While comparing Norwegian, Swedish and some Southern Danish dialects, one finds some similarity in opposition of stød presence and absence in Danish words and that of presence and absence of pitch accents in the words of Scandinavian origin. For instance, a simple pitch accent in modern Swedish is manifested in a falling tone and a complex one in a falling-rising tone. These accents were replaced in the 13-14 centures by the opposition stød-word/stødless-word, the stød corresponding to the accent 1, the absence of stød respectively to the accent 2.

The Danish stød has always been in the centre of word prosody studies. The problems of its origin and realization were interpreted in the works of R.Rask, K.Verner, L.Hjelmslev, A.Martinet,O.Jespersen and other linguists, but still the Danish stød evokes a lively discussion.

Formerly, the stød was considered to be a complete closure, hence it was called "a glottal stop". S.Smith found that a complete closure is as a rule rare, a brief and intense innervation of the expiratory muscles takes place. It abruptly initiates and briefly terminates /2/. Two or three stød phases may be observed: the 1st phase - regular oscillations, the 2nd phase - oscillations weaken or disappear (in a case of complete closure) and the 3rd phase is possible when new oscillations appear.

Recent phonetic investigations have shown, that there is great variability in the phonetic manifestation of the stød, produced by different speakers or the same speaker in different words /3/. The difference may be manifested in duration, pitch, intensity.There are several ways of stød realization and the question arises, whether there are any common characteristics of stød and what helps us to perceive these different acoustic signals as stød.

This paper is concerned with the study of stød, occuring in vowels. 800 Danish words and 96 sentences were chosen for the study. Most words presented pairs (with a long vowel and a stød-vowel) or three words with a long, short and stød-vowel, so that the consonants, surrounding vowel under study were the same. Some words were put into phrase constructions of different length. Others were single words, containing stød-vowels. 13 speakers, students and staff members from the universities of Denmark were recorded. The instrumental and audio analysis were taken up.

To study the perception of the Danish stød the combinations, consisting of a consonant and a long vowel, a stød vowel or a short vowel, were cut out of 45 words, performed by two speakers. These combinations were then presented to 3 groups of subjects with different level of Danish knowledge.

The curves of single words and phrases with stød vowel make it possible to distinguish different types of stød pronunciation: from the complete closure to the curve with no changes.

Out of many stød types 3 main types seem to be pointed out according to 3 phases. Two-phased stød vowels prevails in this material (69% for the isolated words and 96% for the statements), while there are 27% of three-phased stød-vowels in the isolated words and only 2% in the statements. The 2nd phase is likely to be the stød itself, so it is produced in the end of a vowel and seldom in the middle (in a case of 3 phases). One-phased stød-vowels, i.e. vowels with no change in the oscillatory pattern (4% for the isolated words, 2% for the sentences), will be regarded further.

In our investigation the length differences between long and stød-vowels are not significant in almost all cases (All the exeptions are single words). The duration of the stød-vowel comprises 93% of the long vowel in statements, 84% in single words.

The phase duration in the stød-vowel follows next sequences: phase 1 takes 2/3 of the whole stød-vowel (in a case of 2-phased stød-vowel); this length stretches in from the 1st to the 3rd phase (in a case of 3-phased stød-vowel), the first one being the most long (50% or so). With this in mind, the stød takes 1/3 of the whole vowel length.

Average Fo in the stød-vowel is likely to be higher than that of the stødless vowel, the range of Fo changes being wider in the stød-vowel.

Stød-vowel can also change the fundamental frequency of the statement due to its frequency characteristics. At the end of a statement amplified Fo may be observed instead of subdued Fo because of the stød in the last word. This kind of amplification occurs, when a speaker reenforces his vocal folds while producing stød.

The intensity in the stød-vowel undergoes various changes because such vowel has 2 or 3 phases. Besides these changes, the intensity may be marked in comparison with stødless vowels. The higher intensity in the 1st phase of a stød-vowel can influence the dynamic changes in the statement.

F1 and F2 comparison in stød and stødless vowels revealed some quality differences, though not very significant ones. The stød-vowel may be described as more open, front stød-vowels are a little more retracted, back stød-vowels are a little more advanced.

The prime importance in the research is attached to the stød-vowels with no visible changes in the curve. Though the group is small, its investigation is very important, because a weak stød could be heard during the listening-test in spite of stød absence in the curve. These types of stød-vowels have a higher Fo level in comparison with the stødless vowels, whereas the average intensity, duration, F1 and F2 do not differ.

Therefore, the main characteristic feature of the stød-vowel is the Fo change, as the stød-vowel in all cases is accompanied with higher Fo and a wider range of changes than the stødless vowel has. Nevertheless there are some difficulties in outlining the peculiarities of Fo changes in the stød-vowel. Fo can be falling and rising, more complicated changes can be observed. Evidently, the determining factor is the change itself and the average higher Fo of the stød-vowel, than that of the stødless one. The quality and intensity changes are of minor importance as they don't always take place.

The research of the Danish stød perception by those informants whose native language has different segmental and suprasegmental features is believed to be of use in the definition of specific and universal perception abilities and in picking the most effective methods of training.

Russian students come across great difficulties in learning Danish vowel system both on articulation and perception levels. A student may hear no variations in native and foreign sound patterns, here the sensory abilities of a speaker are misleading. When the articulatory basis of the native language dominates, the motor mistakes may take place. A distinct boundary of these mistakes is the way for their improvement.

The perception in its turn depends either on the universal abilities of a person (for example, an ability to organize a word, to oppose vowel/consonant, coarticulation) or an individual abilities characteristic of the system of the native language and formed under the influence of this language phonological system (for example, oppositions of long and short vowels, stød).

To study the mechanisms of the stød perception in Russian class, to define the perception connection with the level of language knowlegde, to outline the acoustic features, stimuli, containing combinations of consonant and long, short or stød-vowel were suggested for 3 groups of subjects: 1) the 3rd and the 5th year students of Danish department of Leningrad State University; 2) students unfamiliar with Danish; 3) staff members of the Department of Phonetics, unfamiliar with Danish.

The listening-test in the 1st group has shown that stød-vowels are identified only in half of all cases by the 3rd year and 5-th year students (54% and 53% respectively), "pure" long vowels are identified better (71% and 84%). Thus, the stød is identified rather badly, though the students can pronounce it. At the same time they have difficulties, connected with the normal realization of the stød-vowels in a coherent text.

The most "favourable" for identification are vowels: [u?] - 6%, [œ?] - 6%, [a?] - 12%, [ø?] - 12% of errors.

Stød-vowels are often identified as short ones, evidently, because of the fact, that stød-vowels in the coherent text seem to be shorter than corresponding long vowels. The duration of the first vowel phase (2/3 of the total stød-vowel length) is that part which is heard by the Russian subjects, when they identify these vowels as short ones.

Broadening of the phonetic context has improved the audibility of the stød (only 22% of errors). The subjects in this case were asked to identify words, containing investigated stimuli.

Thus the opportunity to appraise the type of the vowel contact with the following consonant influenced stød audibility. A pause, an interval in the vowel, additional noise or a kind of creaky voice - all this can be the indication of the stød for the Russian subjects. Particular characteristics of the stød-vowels (Fo, intensity, quality) evidently don't play the leading role for these subjects.

It was noticed at the same time that in some stimuli, containing the short vowel, but cut out of the words, where this short vowel is followed by the stød-consonant, the short vowel was identified with a stød-vowel in 76% of all cases. These errors can be explained in such a way: the short vowel, followed by the stød consonant in the words of the type hyld [ˈhyl?] - elder, mild [mil?] - soft has usually the same average pitch level and in many cases also intensity level, than the stød vowel in words of the type hyl [hy?l] howl , mil [mi?l] mile. Thus also Fo and intensity are of certain importance for subjects, who knows Danish.

The results of the perception tests'in the 2nd and 3rd groups shows that subjects, who don't know Danish, but were only given some information about stød, identify stød-vowels badly (64% of errors in the 2nd group and 62% - in the 3rd).

The stød audibility is not always high in the case when Danish subjects are asked to identify the stød. In the investigation of P.Riber Petersen the listening-test gave from 8 to 50% of errors /4/.

The received facts should be taken into consideration, when Danish is taught. Though the most essential stød characteristic is Fo, important are also intensity changes, phase distribution, quality differences. As an additional cue for the stød in the vowel the type of contact with the following consonant must be also taken into account.

References

/1/ Hansen A. Stødet i Dansk. Det Kgl. Danske Videnskabernes Selskab hist.-filolog. Meddelelser, XXIX, 5, København, 1943.

/2/ Smith S. Bidrag til Løsning af Problemer vedrørende Stødet i dansk Rigssprog. København, 1944.

/3/ Riber Petersen P. An instrumental investigation of the Danish "stød". Annual Report of the Institute of Phonetics University of Copenhagen 7, 1973, p.195-234.

/4/ Ibid., p.209-210.

THE ROLE OF ARTICULATORY EMPATHY IN THE TEACHING OF PRONUNCIATION

JANINA OZGA

Institute of English
Jagiellonian University
Kraków, Poland

ABSTRACT

The concept of articulatory empathy is discussed in the context of teaching FL pronunciation to learners with poor phonetic ability. Successful learners are able to empathise with a variety of models (even ones with voice sets radically different from their own) or with a "generalized" model. The underachievers are not so flexible: a randomly chosen model - even one that they find attractive - will not lead to a permanent empathic response which can only be evoked by a suitably matched voice set. Practical implications of this fact (or postulate) are considered.

INTRODUCTION

In this paper I am concerned with the pronunciation training of FL learners whose phonetic ability is rated low because of marked foreign accent which they are (apparently) unable to drop. This category of underachievers does not seem to attract either researchers or FL methodologists. SLA research connected with pronunciation capability has concentrated (justly or unjustly) on the successful learner (e.g. Guiora et al./1/, Sutor /2/, Purcell /3/). FL methodology considers the problem marginal, which is not surprising in view of the general insistence on cost-effectiveness: prgrammes devised to improve the accent of underachievers imply long hours of extra work with uncertain results, particularly with learners past the critical age of puberty. Moreover, while foreign accent weakens FL performance, it does not, by itself, preclude successful communication (cf.Brown: "We all know people who have less than perfect pronunciation but who also have a magnificent and fluent control of a second language" /4/).

I hope to demonstrate here that (1) the study of the poor learner may contribute to SLA research, by analysing the concept of articulatory empathy, and (2) the approximation of TL accent is not beyond the learner written off as unteachable. What he needs, however, is not a multiplication of drills and exercises administered to his better-endowed colleagues, but a qualitatively different instruction.

My interest in the problem arose in the context of my teaching a remedial course of phonetics to Polish students of English philology. My students are future teachers: their pronunciation should be decent if they are to serve as models to their pupils. But the examination boards of my Institute are surely not alone in tolerating imperfect pronunciation in candidates with high proficiency in other areas. Since "pronunciation capability and overall proficiency in a given language are independent capacities" (Guiora /5/), there are many good, ambitious students whose accent remains their weak point despite their strong "concern for pronunciation accuracy" (Suter, op.cit.) and wholehearted participation in the remedial programme (individual sounds, intonation, stress, rhythm, weak forms, assimilations, etc.) and in the parallel course in general phonetics and English phonology. It is this - admittedly small - group of students that commands my immediate interest. I hope, however, that the issues I intend to raise apply to other groups of learners as well.

ARTICULATORY BASIS IN FL PRONUNCIATION TEACHING

Foreign accent results from the interaction of many factors. Knowing that the majority of these was attended to in our remedial course, I concentrated in my earlier work on one factor that was ignored in our teaching: articulatory basis. According to Honikman /6/, "where two languages are disparate in articulatory setting, it is not possible completely to master the pronunciation of one whilst maintaining the articulatory setting of the other". Assuming this to be true, I described and compared the articulatory bases of English and Polish, using Honikman's parameters (plus state of glottis, Ozga /7/). That study was followed by a report (Ozga /8/), in which I checked the adequacy of my descriptions by testing the success of the articulatory training based on them. There were a number of procedural errors and contaminating variables in my "experiment", but one thing was clear: students in the experimental group, who had the additional drills and exercises connected with the acquisition of the English articulatory basis achieved greater phonetic accuracy and naturalness than the control group taught only by standard auditory and postural methods.

In a remedial course the articulatory "programming" is, of necessity, an exercise in re-orientation. Actually, as was demonstrated by Kolosov /9/, the optimal time to introduce it is the very beginning of a FL course (artikulac'ionaya priorientirovka). I have tried, over the years, to convince teachers of the usefulness of such pre-orientation exercises, -especially for children, who respond more readily than adults to this kind of treatment. However, few teachers are prepared to start a course by "making faces" and my programme has elicited practically no response in the teaching profession.

For my part, I have continued to include the articulatory training in the remedial course, though I have been careful not to administer it wholesale to student groups since the time of the experiment. The relative mean achievement of the experimental group concealed individual differences. That was inevitable as "in experiments we are limited to the statistical averages of a group and not individual factors" (Ochsner /10/), but he teacher in me refused to acknowledge this. I was worried by those cases which spoilt the neat picture of overall success. Some students obviously did not profit from the articulatory training which I devised and in a - mercifully small - number of cases their pronunciation actually deteriorated. Their attempts to reduce lip, cheek, and jaw mobility (which is less pronounced in English than in Polish) produced a peculiar "frozen" lockjaw effect, with open vowels flattened and distorted. This overkill took a lot of individualised instruction to undo. Since then I have always tried to deal with the "hard cases" on the on the individual instruction basis, concentrating on the physical and psychological conditioning of particular learners. That called for inquiry into personality characteristics, notably into the affective learner variable referred to by the term empathy.

EMPATHY IN SECOND LANGUAGE ACQUISITION STUDIES

The notion of empathy appeared in SLA studies in the context of interaction between personality and language behaviour. Empathy is a transactional factor in the affective domain which has been defined as "the projection of one's own personality into the personality of another in order to understand him better" (Brown, op.cit.). Numerous studies have attempted to show that there is a direct relation between this ability "to put oneself into another's head" and language learning success: empathy is said to be a characteristic of the good language learner (see Reves /11/ and works quoted therein) and a valuable predictor of LL success (Guiora and Acton /12/, but see Brown, op.cit. p.109). Of particular relevance to this paper is the early research of Guiora et al./13/, which investigated the relation between empathic capacity and ability to pronounce a SL accurately. The study demonstrated that high degree of empathy, as measured by the Micro-Momentary Expression test (MME), is a predictor of authenticity of SL pronunciation. Empathy is described as related to the flexibility (permeability) of language ego boundaries, which accounts for the ease with which SL pronunciation is assimilated before the age of puberty. Since to speak a SL authentically is "to take on a new identity", around puberty, when the ego boundaries are firmer, this flexibility is said to be drastically reduced and it is more difficult to "move back and forth between languages and the presonalities that seem to come with them" (Guiora and Acton, op.cit.). Empathic ability appears to vary not only globally with age but also between individual speakers and under experimentally induced conditions (alcohol, hypnosis, see Guiora and Acton, op.cit.).

IMPLICATIONS OF RESEARCH ON EMPATHY FOR THE TEACHING OF FL PRONUNCIATION

Research on affective learner variables like empathy (I should also add self-esteem, anxiety, aggression, etc., which surely interact with empathy), which influence FL pronunciation accuracy, helps us to understand how it happens that the ability to acquire native-like pronunciation varies when such factors as cognitive styles, motivation, exposure to training are held constant or are comparable across learners. The problem is is that "if indeed a high degree of empathy is predictive of success in language learning, it would be invaluable to discover how one could capitalize on that possiblity in language teaching... One would need to determine if empathy is something one can "learn' in the adult years, especially cross-culturally" (Brown, op.cit.p.109). I am interested in these questions in so far as they are related to the development of empathic capacity in the underachievers, but obviously there are numerous other areas where empathy studies are relevant to FL teaching. Let me refer to just a few of these. The phase-specific empathic ability of young children manifests itself in acquiring native-like pronunciation, when the FL is learnt in its native environment (actually SL). I doubt whether this ability manifests itself so strongly as a gruup variable when pronunciation is acquired in a foreign country in the context of formal instruction. Individual differences are sharper then and the influence of the teacher as a pronunciation model is of crucial importance (in my experience young children usually get the worst teachers!). There are other interesting questions connected with pre-puberty pronunciation acquisition (e.g.durability of early model-based habits, later modifications, fossilization of infantile habits in children who acquired a SL in native environments and were later taught it as a Fl).

Guiora et al.(op.cit.) stress the drastically reduced ability to assimilate native-like FL pronunciation after the age of 12. This bodes ill for Polish learners who generally start to learn English at the age of 15. And yet, if puberty contributes in an important way to the completion of the articulatory profile of a person in first-language acquisition (see Birnbaum /14/), should we exclude the influence of this phase on the formation of the correct FL pronunciation profile, even in a formal teaching programme? According to Birnbaum, "the modification of the articulatory manners and preferences affecting these young people are more radical, since they are deliberate, than the difficulties in imitation and pronunciation adjustment encountered in early childhood" (op.cit.). As teenagers are a model-seeking generation, responsive to fads and fashions, there are excellent possibilities to capitalize on this in teaching pronunciation (but also great dangers, if unattractive, unimaginatively selected models are offered).

Let me now consider the problem from the point of view of individual learners. Paradoxically, it is the higly empathic learners that are likely to suffer failure in acquiring a native-like pronuciation of a FL, if - as is too often the case - they have an influential but inadequate single early model to empathise with. Thus, many underachievers among my students are "hidden empathics". Fortunately, they respond satisfactorily to the remedial course and to articulatory basis training; but why should they have to unlearn bad habits and arrive at the native-like approximation so late, when their empathic potential qualifies them for much earlier success? They had the bad luck of having fashioned their pronunciation habits on teachers (very often good teachers) with poor accent. Those teachers were recruited from the ranks of the underachieving students: the vicious circle is closed.

That the situation of the higly empathic learner is generally not so dramatic as in the above account is due to the fact that successful language learners are able to empathise with a variety of models. Their empathic response - demonstrated in terms of articulatory adjustments - is evoked by models with voice sets radically different from their own. Their FL pronunciation profiles are acquired through two strategies: (1) the choice - usually deliberate - of a single, attractive model (persona adoption), or (2) the elaboration of a generalised model, resulting from the combined influence of several models. Empathy in such learners appears to generate aptitude for oral mimicry and also to be connected with tonal memory, musical abilities and certain perceptual qualities which enable them to empathise with disembodied voices on tapes and records, without the reinforcing presence of visual cues.

The underachievers are much less flexible: a randomly selected model - even one that they find attractive - will not lead to a permanent empathic response. Forcing models on such learners ends in a sad caricature. The only way to ensure successful teaching is to find suitable models for the learners to rely upon. I look for such models in a principled way, acting on the assumption that underachievers must have at least a modicum of empathic capacity, i.e. ability to empathise with models whose voice sets are similar to their own. I refer to this type of empathic ability as "articulatory empathy".

ARTICULATORY EMPATHY

Although I know a female Polish student whose English pronunciation training was based solely on the Laurence Olivier films and she indeed sounds like the famous actor when she speaks English, I would insist that imitating very remote models (also age- and sex-wise) is to be avoided even with good students. That is why the middle-aged, precise, dignified voices regularly heard on records and tapes of phonetic material are so exasperating to the students in general, and fad-sensitive teenagers in particular. While this is the question of teaching materials rather than methods, it certainly does have a bearing on the success of the teaching process and cannot be disregarded.

The importance of well-matched models for pronunciation struck me with full force in an anecdotal context. Over the period of some fifteen years over twenty students have formerly been the pupils of the same teacher (from one of the Kraków secondary schools). In assessing their pronunciation on admission to the Institute I observed a certain regularity: those whose pronunciation was poor did not possess certain vocal characteristics of the teacher (clear, high, precise voice, with a slight glottalization). on the other hand, most of those whose pronunciation was good shared these characteristics and all of them clearly "inherited" certain personal pronunciation mannerisms of the teacher. In a few cases it was, in fact, possible to guess which school they had attended because the pronunciation profile of the teacher came through very clearly.

These observations indicate, however, that in talking about articulatory empathy it is not sufficient to refer to the "voiceset" or "voice-quality" understood as "permanently present background person-identifying vocal characteristic" which is "biologically controlled" (Crystal /15/ but rather to a person's "habitual mode of phonation" (Laver's description), including the pausal profile, speech-rate, and articulatory /permanent lip-rounding/ as well as pitch-related (drawling, clipping) mannerisms.

In my remedial work I have to use a fairly small inventory of terms, which are, of necessity, often impressionistic. I also make use of the set of 24 descriptive parameters proposed by Kelz /16/ for the description of articulatory bases of languages. The teaching relies on improvisation to a large extent and resembles psychotherapy more than anything else. But it is not time-consuming and, more importantly, it works. I hope to report on the details of the training and on the framework that underlies it after I have managed to give it a more stabilised and efficient shape.

CONCLUSION

Ideally, a course intended to eliminate pronunciation inaccuracies and foreign accent in underachievers, should rely on a well-stocked library of video-cassettes and recordings of phonetic teaching materials made not by the usual one-male-one-female team, but representing different voice-types. But the collection would be just as useful to them as to all other learners. In fact, it is not altogether utopian to expect that at some point in the future the teaching of FL pronunciation will be based precisely on such model-oriented materials.

But today's underachievers cannot wait. It is for them that I have undertaken a "small scale intervention" (Brown, op.cit.), without waiting for the corroboration of large experimental designs which I am unable to undertake myself. In any case, if I relied on experimental evidence in my teaching, I would have to believe Purcell (op.cit.), who says that "classroom learning just does not seem to have much to do with pronunciation accuracy" and leave things as they are. Instead, being sympathetic to hermeneutic rather than nomothetic mode of inquiry (see Ochsner, op.cit.) I have tried to use intuition, common sense and experience to develop a teaching framework that works.

REFERENCES

/1/ A.Z.Guiora, M.Paluszny, B.Beit-Hallahmi, C.Y. Dull, J.C.Catford, R.E.Cooley, "Language and person, studies in language behaviour", Language Learning 25.1.1975

/2/ R.W.Suter, "Predictors of pronunciation accuracy in second language learning",Language Learning 26.1976

/3/ E.T.Purcell, "Models of pronunciation accuracy", Issues in language testing research, Newbury House, 1983

/4/ H.D.Brown, "Principles of language learning and teaching", Prentice Hall, 1980

/5/ A.Z.Guiora, "Is there a general capability to pronounce a foreign language: an experimental inquiry", Proceedings of the 6th AILA Congress, 1981

/6/ B.Honikman, "Articulatory settings", En honour of Daniel Jones, Longmans, 1964

/7/ J.Ozga, "The relevance of the notion 'basis of articulation' to contrastive phonetics", PSiCL IV, 1976

/8/ J.Ozga, "Teaching English articulatory settings to Polish learners", PSiCL VI, 1977

/9/ K.M.Kolosov, "O roli artikulac'ionnoy bazy v obucenii proiznosyeniu", Inostrannye jazyki v skole, 5/71, 1971

/10/R.Ochsner, "A poetics of second language acquisition", Language Learning, 29, 1979

/11/T.Reves, "A new attempt of testing empathy - an assumed characteristic of the good language learner", Proceedings of the 6th AILA Congress, 1981

/12/A.Z.Guiora, W.R.Acton, "Personality and language: a restatement", Language Learning 29, 19879

/13/A.Z.Guiora, R.Brannon, C.Y.Dull, "Empathy and second language learning", Language Learning 22, 1972

/14/H.Birnbaum, "Ongoing sound change and the abductive model: some social constraints and implications", Prpceedings of the 9th ICPhS, 1979

/15/D.Crystal, "The English tone of voice", Edward Arnold, 1975

/16/H.P.Kelz, "Binary features for the description of the basis of articulation", Study of Sounds, XVIII, 1978

TEACHING PHONETICS FOR LINGUISTIC FIELD WORK

KATHRYN C. KELLER

Summer Institute of Linguistics
Apartado 22067
14000 Mexico, D.F., Mexico

ABSTRACT

This paper reports on a successful training program for linguistic field workers, for learning to transcribe and speak unwritten languages. The emphasis is not only on transcription, analysis and theory, but also on individual production of the various sounds that the human vocal tract is capable of producing, which may be encountered in languages. Emphasis is placed also on controlling longer utterances with proper pitch patterns, stress, and rhythm.

INTRODUCTION

Beginning phonetics courses are offered by the Summer Institute of Linguistics (SIL) at the Universities of Oklahoma, North Dakota, Oregon, and Texas at Arlington, as well as in England, France, Germany, Australia, Canada, New Zealand and in other parts of the world. Although the courses in the different places vary in detail, the overall structuring and methods are similar. I am reporting from my own particular viewpoint and experience.

Through the years many hundreds of students have taken these courses and have done successful field work, including language learning, in exotic languages around the world. The pedagogical approach has benefitted from feedback from these field workers, adding to the basic course design as originally developed by Evelyn G. Pike. We have also made use of the many excellent suggestions from skilled British phoneticians, going back to Daniel Jones and Westermann and Ward, as well as more recent ones.

Some aspects of the course have been reported on in *Language Learning*: K.L. Pike, "Problems in the Teaching of Practical Phonemics" [14]; George M. Cowan, "An Experiment with a Wire Recorder in Teaching General Phonetics" [3]; Euniee V. Pike, "A Test for Predicting Phonetic Ability" [9]; Frank E. Robbins, "A Ten Day Program of Preparation for Language Learning" [17].

PURPOSE AND SCOPE OF THE COURSE

The SIL approach to teaching articulatory phonetics has its theoretical base in K.L. Pike *Phonetics* [11], and in K.L. Pike *Phonemics* [12], with additional input from later materials of Abercrombie [1], Ladefoged [7], and others. It is designed to prepare students for field work in unwritten minority languages and for being able to speak those languages with as little mother-tongue interference as possible. The methods can be applied to learning any foreign language.

Emphasis is placed on the recognition and production of all types of sounds used in languages, and on giving the necessary flexibility to arrive at the specific qualities of sounds in individual languages.

In addition to extensive drill on individual sounds, emphasis is placed on awareness and control of the rhythm and pitch patterns of longer segments of natural speech. Conversations in a variety of languages are learned by mimicry from tapes. Theoretical advances in understanding the larger units in the phonological hierarchy come from Pike *Language in Relation to a Unified Theory of the Structure of Human Behavior*, chapters 8 and 9 [15], and from Pike *The Intonation of American English* [13].

During the last segment of the course, students work with a native speaker of a non-European language, to determine the sound system from their own transcribed data and to learn to carry on a conversation with the speaker in his language. They also prepare and practice drills to sharpen their hearing and to help in sorting out the raw phonetic data.

MECHANICS OF THE COURSE

The 35 teaching periods are given over primarily to drill sessions by small groups of a maximum of ten students, to give the individual attention needed for mastering the material. Teachers are rotated so that students hear the voices of several different instructors.

Instructors are mostly workers of the SIL with extensive phonetic experience in at least one language. Those who have not yet worked in the field, or who are teaching for the first time, are assigned to an experienced staff member for checking dictation in preparation for the class session.

Uniformity in the drill sessions for an enrollment of 80 or 100 students is facilitated by detailed lesson plans and by daily briefing sessions for the instructors. At the briefing sessions we have an opportunity not only to discuss the new lesson and to check each other for a mastery of the material, but to share with each other the things that worked or didn't work in our classes. The briefings constitute in-service training for the instructors, where we all learn, experienced and inexperienced alike.

Daily assignments, weekly testing, and material on tape are planned to help the students master the material.

Most of the students are also enrolled in courses in phonology and grammar for a well-rounded introductory course in linguistics. Thus the role of phonetics is seen in its larger setting.

Lectures are relatively few in comparison with the drill sessions. They present some basic theory, and make use of the specialties of other SIL staff members. For instance, Kenneth Gregerson lectured on tongue root function in languages of southeast Asia. The students have as a textbook *Dictation Exercises in Phonetics* by Eunice V. Pike [10]. This is supplemented by assigned readings in phonetic theory. The SIL courses in the different locations differ as to what book is used as a text.

METHODS AND TECHNIQUES

In each drill session a specific sound type is presented. The presentation has three elements: theory, production of the sounds, and ear training and transcription. Theory and practice go hand in hand. Review, especially of the sounds studied at the last class hour, is included to help fix the sounds in the minds of the students.

Theory

During the nine-week time span of the course we cover quite thoroughly the range of basic points of reference in sounds. In addition to the theory presented in the lectures, background theory including the physiology of speech production is presented in the drill sessions as each new type of sound is taught. The features which identify a sound and make it contrast with another are emphasized. Students are required to be able to illustrate these features by making face diagrams of the sounds and by knowing the technical name for each sound. Dynamic diagrams of strings of sounds are presented.

Some films are shown. One of the most useful for beginners is "Velopharyngeal Function in Normal Speakers", produced by Kenneth Moll (distributed by Bureau of Audio-Visual Instruction, State University of Iowa), which presents the action of the velopharyngeal mechanism by animated drawings and by cinefluorographic film sequences. This film is an excellent way to show the students the action of the velic and the coordinated movements of the organs of speech.

For other films which we have used from time to time see Keller *Instrumental Articulatory Phonetics* [6]. I am indebted to the University of Edinburgh for the stimulus for the use of films in teaching phonetics, as well as for a broader perspective in phonetics and for a background in experimental articulatory phonetics.

The videotape of K.L. Pike on phonetics [16] is invaluable in giving an idea of the range of human speech sounds, and techniques for flexibility. Also illustrated in the videotape are languages with dynamic phonetic features very different from those of English.

Production of Speech Sounds

There are two basic problems faced by the student in learning to produce new sounds or to speak a second language. One is to overcome the psychological barriers to making new and different sounds which may sound very "queer" to him, or which may turn out very different from the model thus making him feel inferior or inadequate. The second is to overcome interference from the mother tongue, ingrained habits which are difficult to break and of which the student is not even aware.

In the very beginning of the course we try to deal with the psychological blocks. The small classes and an atmosphere of informality help put the students at their ease and make possible student-teacher interaction. In teaching the production of specific new sounds, we have the class repeat in unison, mimicking the teacher, before having each student in turn try to produce the sound. This gives students a chance to try making the sound before being called on individually. Learning to laugh at oneself helps in the learning process. During the group mimicry, the teacher listens critically, commending good performance or making suggestions for improvement. By the end of the class the student knows whether he can make the sounds satisfactorily, or how to go about working on their production.

To overcome interference from the mother tongue the student needs to understand the differences between the new sounds and those of his mother tongue, and be exposed to extensive drill. In the context of SIL we have learned that we need to give hands-on time to students to help them overcome ingrained mother-tongue influences and help them master sounds maximally different from sounds in their first language. Students often have problems with ejectives and implosives, for example, and in controlling vowel glides. They also may have problems controlling sounds of their mother tongue when they occur in positions different from where they occur in their original language. English speakers have to learn to produce a velar nasal initially in an utterance even though they can do it easily when it occurs syllable final.

For the mastery of individual sounds, we give in-class help individually to the students as well as collectively in the group mimicry. We emphasize learning by mimicry. If that fails, we give suggestions on how the student could produce the sound, going from the known to the unknown. Simple nontechnical hints often work. To help the student produce a bilabial fricative, for instance, in contrast to the labiodental fricative he is accustomed to making, we get results by saying, "Grin and blow the hair off your forehead." Many phonetic exercises for pronouncing sound types are included in chaper 2 of Pike *Phonemics* [12]. We have gotten hints on how to produce sounds from Daniel Jones [5], Westermann and Ward [19], Ladefoged [8], Smalley [18] and others. Feedback from staff members is an additional source of ways to help students master sounds.

In addition to in-class help, tapes are made available to the students. Students are encouraged to work in pairs, so that they can help each other. Their progress is monitored through weekly check-ins with a staff member, in which the sounds studied that week are gone over with the student. Any sound not yet mastered is recorded for further work, along with suggestions on how to proceed to master the sound.

For practicing sounds in context, frame drills of nonsense combinations are assigned from time to time. Also a list of words from actual languages, covering the sounds currently being studied, is given to the students each week. The students are subsequently given a production test on these words.

In teaching students to hear and control intonation, rhythm and tone, we begin with exercises in English intonation. The students are given a selection with the intonation, stresses and pauses marked, and are asked to practice reading it until they can read it as marked. Meanings conveyed by different contours are pointed out. If one is aware of what he does in his own language, he can more readily recognize patterns in another language. One lecture is devoted to what we term *speech styles*, in which students are made aware of overall phonetic features which may color a language and make it sound different from another or which may be contrastive within a single language. Examples collected from field experience are presented. Required reading is the article "Articulatory Settings" by Beatrice Honikman [4], which emphasizes the importance of "shifting gears" in going from the mother tongue to a new language.

For the mastery of longer utterances we have found buildup drills helpful. Another valuable technique is tracking, referred to as "shadowing" by L.A. Chistovich in "Relation between Speech Production and Speech Perception" [2] at the Tenth International Congress of Phonetic Sciences. This rapid imitation of speech, following along with a speaker a half syllable behind, can be done silently or out loud and helps a student get into the rhythm, pitch patterns, and speed of the target language.

To give the students practice in the mastery of longer utterances, conversations in a variety of languages have been recorded on tape. The students are assigned a different language each week. They listen over and over to the phrases, mimicking until they can control the intonation, speed, and rhythm. To facilitate the learning, each phrase of the conversation is repeated on the tape about five times, with a pause for mimicry between each utterance; then the conversation as a whole is given. At the end of the week the students are tested on their mastery of the conversation. We have used such diverse languages as Chinese, Isneg (Philippines), Turkish, Basque, Totonac (Mexico), Igbo (Africa).

Ear Training and Transcription

Drills are those from *Dictation Exercises in Phonetics* by Eunice Pike [10], with additional language materials which we have collected from field members down through the years.

For each sound type, we begin with oral ear training, contrasting the new sound with a similar known sound with which the student would be most likely to confuse it. Transcription practice begins with single-syllable *differential drills*, contrasting the sounds just worked on orally. When the students can distinguish the sounds in minimal contrast, harder drills are given. *Recognition drills* are made up of nonsense words containing the new sounds along with sounds studied to date. Each lesson also contains drills of actual language words. The drills are dictated by the teacher for transcription, then students are asked to read back the words. Dictation tests are given about once a week.

Tapes of controlled material by native speakers illustrating specific sound contrasts are available for extra ear training. These include Amharic, Finnish, Gugarati, Zulu.

APPLICATION OF THE TRAINING

The last segment of the course consists of an eight-day linguistic field problem. Instead of classes, the student is assigned a non-European language to work on.

The students are divided into groups of five or six to work with the native speaker of the language. Staff members are assigned to each group to advise and to encourage. Each student is with the native speaker for three periods during the day.

During the first period the entire group is together for an assimilation session, in which material is elicited in context, learned and practiced as a conversation. Tangible objects to work with help make real the situations around which the conversations are built. Writing is discouraged, but the phrases are taped for later mimicry and practice.

Each student has a period in which he can elicit material for analysis and another period in which he is an observer when another student is eliciting. The students are to learn whatever they can about the sound system of the language from their own elicited data. They write this up for their phonology course and look for grammatical patterns for their grammar course.

The phonetics requirements are twofold. The students are to construct drills which will help them in pronunciation or which will help to sharpen their hearing. They are also to learn enough of the language to be able to carry on a five-minute conversation with the native speaker from materials learned during the assimilation periods and practiced in their eliciting sessions. On the last day, each student is asked to converse with the language speaker, using props but nothing written.

SUMMARY

By small classes, individual help, extensive drill, and carefully monitored student progress we try to give students an understanding of the whole gamut of possible speech sounds, help them to make and recognize these sounds, and help them to have flexibility to reach shades of sounds not specifically taught in class that they might come across in a language.

REFERENCES

[1] Abercrombie, David (1967) *Elements of General Phonetics*. Chicago: Aldine.

[2] Chistovich, L.A. (1983) Relation between Speech Production and Speech Perception. In *Proceedings of the Tenth International Congress of Phonetic Sciences*. Holland: Foris Publications.

[3] Cowan, George M. (1949) An Experiment with a Wire Recorder in Teaching General Phonetics. *Language Learning* 2 (3), 76-82.

[4] Honikman, Beatrice (1964) Articulatory Settings. In Abercrombie, D. et al. *In Honour of Daniel Jones*. London: Longmans.

[5] Jones, Daniel (1956) *An Outline of English Phonetics*. 8th edition. Cambridge: Heffer.

[6] Keller, Kathryn C. (1971) *Instrumental Articulatory Phonetics*. Norman, Ok: Summer Institute of Linguistics.

[7] Ladefoged, Peter (1964) *A Phonetic Study of West African Languages*. Cambridge: Cambridge University Press.

[8] Ladefoged, Peter (1975) *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

[9] Pike, Eunice V. (1959) A Test for Predicting Phonetic Ability. *Language Learning* 9, 35-41.

[10] Pike, Eunice V. (1963) *Dictation Exercises in Phonetics*. Santa Ana, CA: Summer Institute of Linguistics.

[11] Pike, Kenneth L. (1943) *Phonetics*. Ann Arbor: University of Michigan Press.

[12] Pike, Kenneth L. (1947) *Phonemics*. Ann Arbor: University of Michigan Press.

[13] Pike, Kenneth L. (1947) *The Intonation of American English*. Ann Arbor: University of Michigan Press.

[14] Pike, Kenneth L. (1948) Problems in the Teaching of Practical Phonemics. *Language Learning* 1 (2), 3-8.

[15] Pike, Kenneth L. (1967) *Languages in Relation to a Unified Theory of the Structure of Human Behavior*. The Hague: Mouton.

[16] Pike, Kenneth L. (1977) *Pike on Language*, series. Videocassette Program No. 1: Voices at Work (Phonetics). Ann Arbor: University of Michigan Television Center.

[17] Robbins, Frank E. (1960) A Ten Day Program of Preparation for Language Learning. *Language Learning* 10, 157-163.

[18] Smalley, William A. (1963) *Manual of Articulatory Phonetics*. revised edition. Pasadena, CA: William Carey Library.

[19] Westermann, D. and Ward, I.C. (1933) *Practical Phonetics for Students of African Languages*. London: Oxford University Press.

МОДЕЛИРОВАНИЕ ЗВУЧАЩЕЙ СТОРОНЫ ИНОСТРАННОГО ЯЗЫКА В ЦЕЛЯХ ОБУЧЕНИЯ

ДЕНКА СПИРОВА ЯНКОВА

Катедра "Методика", ФСФ, СУ "Климент Охридски"
София, 1000, бул. "Руски" № 15, България

РЕЗЮМЕ

В данной работе приводятся результаты теоретического и экспериментального исследования систем обучения звучащей стороне иностранного /русского/ языка.

ИЗЛОЖЕНИЕ

1. Принцип последовательного моделирования в методике иноязычного обучения. "Глобальная" речь является объектом исследования ряда наук, которые, в зависимости от своего предмета, строят различные исследовательские модели, не входящие между собой в субординативные отношения.Поэтому полное описание понятий методики иноязычного обучения связано с учетом ряда моделей: педагогического, лингвистического, психолингвистического, социолингвистического, социологического, психологического и т.д., которые при этом касаются двух разных систем общения, контактирующих в процессе обучения. Подобный учет ряда моделей при построении модели обучения иноязычной речи в методике мы называем принципом последовательного моделирования.Поскольку при построении модели обучения возможен учет разных моделей в разных комбинациях, можно постулировать тезис о неединственности методической модели обучения звучащей стороне иноязычной речи. Таким образом, в методике реализуется одно из основных положений гносеологии - о моделирующем характере познания.

2. Обучение фонетике в методике иноязычного обучения.В соответствии с принципом последовательного моделирования в теории и практике обучения следует по-новому оценить роль и место лингвистических моделей звучащей стороны иностранного языка.Обучение фонетике нужно рассматривать как процесс преподавания и изучения определенной модели звучащей стороны иностранного языка в сопоставлении или вне сопоставления с соответствующей моделью звучащей стороны родного языка. Т.е. нужно связывать цели и содержание обучения фонетике с формированием только языковой /лингвистической/ компетенции учащихся. В рамках утвердившегося взаимосвязанного обучения видам речевой деятельности - говорению, аудированию, чтению, письменной речи ,- т.е. коммуникативного подхода в иноязычном обучении, проблематика обучения фонетике не в состоянии охватить все стороны проявления звучащей стороны речи.

3. Гипотеза фоноречевого характера обучения звучащей стороне иноязычной речи в системе коммуникативного подхода. Видя цель обучения языку в подготовке к общению через посредство иностранного языка в разных видах интеллектуальной и практической деятельности, базируясь на разных стратегиях использования языка как средства в деятельности, нужно обратиться к интегрированной презентации материальной структуры языка и ее общественного функционирования,к овладению языком,речевым поведением и культурой носителей языка.

На основе данной целевой установки, принципа последовательного моделирования и концепции обучения "выразительной" стороне речи Е.И.Пассова, которую можно "собрать" на базе его работ /1/,мы в процессе работы над учебным комплексом по русскому языку для III класса болгарской средней школы подошли к рабочей гипотезе фоноречевого характера обучения звучащей стороне иноязычной речи в системе коммуникативного подхода. Основные положения гипотезы следующие: 1/ Речь, являющаяся способом формирования и формулирования мысли посредством языка, представляет собой орудие, средство выполнения всех видов речевой деятельности. В значимых единицах речевой деятельности можно выявить "означаемое", т.е. коммуникативное намерение, и "означающее", представляющее собой физическую/акустическую или графическую/ его реализацию./2/ 2/ Звучащая сторона речи является подсистемой в системе речи и через нее - в системе речевой деятельности. 3/ Внутреннее проговаривание является общим психофизиологическим фактором механизмов всех видов речевой деятельности. Оно касается не только звукового, но и интонационного оформления речи. Механизм реакции на слово представляет собой единую функциональную систему, образующуюся благодаря запечатлению в мозгу непосредственного раздражителя и его устного и письменного символа. Таким образом, в целях и в процессе коммуникативно ориентированного обучения звучащая сторона речи выступает в единстве с представлением о звучании /при адекватном восприятии речи/, с графическим изображением этого звучания в соответствии с орфографическими правилами, а также с представлением об этом графическом изображении/на письме и при чтении вслух и про себя/. Считается, что конкретные параметры этого единства и степень его выраженности обусловлены ситуацией общения и социальной ролью коммуникантов. Это единство мы называем фоноречевым комплексом. Фоноречевой комплекс объединяет "означающие" четырех возможных способов выражения определенного речевого действия в процессе речевой деятельности. 4/ Фоноречевой комплекс - основная единица обучения форме иноязычной речи и через нее - системе форм проявления видов речевой деятельности. Это свидетельствует об атрибутивном характере фоноречевого комплекса. Выступая в статусе операции он является компонентом любого речевого действия, имеет специфический состав в каждом виде речевой деятельности. Совокупность характеристик этого состава, в зависимости от национальной аудитории, образует в процессе обучения своеобразную подсистему в системе обучения речевой деятельности на иностранном языке. Эту систему мы называем фоноречевой подсистемой. 5/ Фоноречевой комплекс реализуется в исполнительском звене/ части/ структуры деятельностного акта. 6/ Фоноречевая подсистема является предметом обучения форме проявления видов речевой деятельности/т.е. тем, чему мы обучаем/. Через процедуру отбора она превращается в содержание обучения в этой области/ т.е. в минимум, который есть то, на основе чего мы обучаем/. Все, что касается этого предмета обучения, составляет проблематику фоноречевого обучения. 7/ Целью фоноречевого обучения,выявленной на фоне цели обучения иностранному языку вообще, считаем максимальную степень развития индивидуальных умений и навыков учащихся устанавливать адекватную связь между формой и содержанием речевой деятельности в разных ее видах,т.е. между их "означаемым" и "означающим".

Описание фоноречевого комплекса и фоноречевого минимума на языке методики иноязычного обучения возможно только в системе более общей теории обучения иностранному языку в рамках коммуникативного подхода, частным случаем которой будет теория фоноречевого обучения. На настоящем этапе рабочая гипотеза фоноречевого характера обучения звучащей стороне иностранного языка подчеркивает направление методических поисков в области обучения форме проявления видов речевой деятельности. Это направление требует переосмысления понятия аспекта обучения. Аспект обучения должен быть связан с некоторыми характеристиками или подсистемами именно процесса обучения речевой деятельности, т.е. быть методической, а не какой-либо другой, в т.ч. лингвистической реальностью. С этой точки зрения неправомерным является выделение в качестве аспекта обучения "обучение фонетике"/ а также "обучение грамматике", "обучение лексике" и т.д./. В соответствии с различением лингвистической и коммуникативной компетенции в иноязычном обучении следует считать "обучение фонетике" аспектом изучения определенной лингвистической модели/звучащей стороны/иностранного языка в сопоставлении с соответствующей моделью родного языка, а фоноречевое обучение - аспектом обучения видам речевой деятельности в рамках системе коммуникативного подхода. В связи с этим является необходимым уточнение терминологического порядка: фонетика преподается и изучается, усваивается, как система знаний, фоноречевая система овладевается, как система навыков и умений, ей обучают. Соотношение представленности фонетической и фоноречевой систем в определенной модели обучения иностранному языку для каждой национальной аудитории зависит от этапа и цели обучения, от контингента учащихся, от состояния фонетических и фоноречевых исследований. Что касается обучения русскому языку как иностранному на настоящем этапе развития методики , в рамках научного обоснования коммуникативного подхода, очевидно,имеется острая необходимость в пересмотре лингвистических моделей, используемых для презентации русского языкового материала, с одной стороны, и в том, чтобы точно определить функции работы над языковым материалом в общей системе обучения и ее соотношение с целями обучения.

4. Фоноречевое обучение и преподавание фонетики в III классе действующей системы обучения в болгарской средней школе.

В болгарской средней школе русский язык изучается с третьего класса. Третий и четвертый классы считаются начальным этапом обучения и на русский язык отводятся три часа в неделю. В настоящее время в связи с реформой образовательной системы идет работа по новой серии учебников. В этом 1987 году ученики и учителя будут работать по новому учебнику для VII класса. Седьмой класс является последним классом среднего этапа обучения, когда на язык отводятся два часа в неделю.

Впервые в новой серии учебников русского языка для средней школы представлена определенная система обучения звучащей стороне русской речи. Ввиду недостаточной разработки гипотезы фоноречевого обучения - отсутствия сопоставительных описаний фоноречевых комплексов компонентов коммуникативного минимума и самого минимума -, а также отсутствия сопоставительных лингвистических описаний определенных фонетических систем русского и болгарского языков с точки зрения единых принципов, существующую систему можно охарактеризовать как компилятивную и, в некоторой степени, эклектическую. Мы считаем, однако, что подобное явление можно назвать противоречием роста, поскольку над этими проблемами уже работается. Кроме того, в учебном комплексе для III класса/3/ работа в области овладения навыками адекватного оформления русской речи в большой мере учитывала основные положения гипотезы фоноречевого характера обучения звучащей стороне иностранного языка. Материал, который был отобран, охватывает звуки, ударение, звукосочетания, звуко-буквенные отношения, интонацию, синтагматическое членение фразы, темп, слитность, стиль произнесения. Кроме существующих лингвистических описаний фонетических систем русского и болгарского языков и описаний ошибок болгарских учащихся, впервые учитывались статистические данные частотности слогов, звуков и букв русского и болгарского языков/4/, а также результаты анализа лексического минимума для болгарской средней школы и для III класса. Были учтены также текстовой материал учебника и уровень лингвистических знаний, полученных при обучении родному языку. Таким образом, был определен фоноречевой материал/минимум/, который указывал не только список звуков, как в известных до этого времени фонетических минимумах, но учитывал звуко-буквенные отношения, структуру и интонацию фразы, а также включал определенный орфографический материал. Т.е. при составлении фоноречевого минимума и при его разработке учитывалась структура фоноречевого комплекса.

Весь материал разработан в устном вводном курсе/ 9 уроков/, в цикле"Мы читаем по-русски"/ 6 уроков/, в комплексном курсе "Говорим.Читаем.Пишем.Играем."/36 уроков/. Вводный курс распространяется на формирование элементарных навыков и умений решать определенную речевую задачу в рамках говорения и аудирования. Далее во всех уроках обучения чтению, разработанных на знакомом лексическом и потенциально знакомом грамматическом материале, сочетается работа по формированию и овладению звучащим и графическим образами слов, путем произнесения слов, чтения их вслух, распознавания среди сходных болгарских, списывания и снова чтения и произнесения.

В работе по овладению звуко-буквенными отношениями используются специальная учебная тетрадь, магнитофонные записи, сопровождающие некоторые задания в ней, набор букв и слов для частичной учебной транскрипции, набор предложений с указанием движения тона при чтении и произнесении. Примеры в таблицах к урокам повторения также сопровождаются интонационным контуром. Заботой авторов учебного комплекса была коммуникативная и личностная мотивированность заданий по овладению навыками адекватного оформления русской речи. Осознание элементов системы лингвистической компетенции в этой области является следствием фоноречевого характера обучения, а не целенаправленной работы. Из-за ограниченного объема доклада трудно привести подробные примеры и их анализ.

В заключение необходимо отметить целесообразность развития фонетических и фоноречевых исследований в следующих направлениях: 1/ Описание контактирующих в обучении фонетических систем иностранных языков с точки зрения единых принципов, в тождественных системах понятий и терминов. 2/ Конкретизация гипотезы фоноречевого характера обучения звучащей стороне иноязычной речи на основе законченной теории коммуникативного обучения с учетом принципа последовательного моделирования. 3/ Создание фонетических и фоноречевых минимумов для конкретной национальной аудитории, этапа и цели обучения, и контингента учащихся.

ЛИТЕРАТУРА

/1/ Пассов Е.И. Основы методики обучения иностранным языкам, М., 1977 ; Теоретические основы обучения иноязычному говорению, Воронеж, 1983.

/2/ Зимняя И.А. Психологическая характеристика слушания и говорения как видов речевой деятельности, "Иностранные языки в школе", 1973, № 4

/3/Цанкова М, Гочева Э., Янкова Д., Дукадинов Л., Русский язык для III класса ЕСПУ, "Народна просвета", С., 1983 .

/4/ Янкова Д. Возможности применения критерия частотности звуков в целях обучения / на материале болгарского и русского языков/; Linguistique balkanique , (1985) ,3

FUNCTIONAL LOAD AND THE TEACHING OF PRONUNCIATION

ADAM BROWN

Language Studies Unit,
Aston University,
Birmingham, U.K.

ABSTRACT

The concept of functional load has been used by various writers in various linguistic fields. It has consequently received differing definitions and methods of calculation. It has not, however, been applied to the teaching of pronunciation. In this paper are discussed several aspects of functional load which may be relevant for the assessment of the relative importance of segmental features of learners' speech.

Introduction

'Suppose you are teaching English to foreign students, on a tight schedule, with no special time for pronunciation teaching,' writes Gillian Brown [1] p.53. 'Which of the following problems would you tackle first? Discrimination of /θ/ and /ð/, [etc.].'

Her answer: 'When time is short it is probably not worthwhile spending time on teaching /θ/ and /ð/ if the students find them difficult, but be sure that the sounds substituted by the students are /f/ and /v/ sounds which are acoustically similar to /θ/ and /ð/ and bear a low functional load in English (i.e. don't distinguish many words), and not /s/ and /z/, which are acoustically very different from /θ/ and /ð/ and bear a much higher functional load.'

Many writers have made appeal to the notion of functional load (FL), and for various purposes. However, the precise definition given to the concept has varied from writer to writer [2]. King [3] p.831 writes that 'in its simplest expression, functional load is a measure of the number of minimal pairs which can be found for a given opposition. More generally, in phonology, it is a measure of the work which two phonemes (or a distinctive feature) do in keeping utterances apart - in other words, a gauge of the frequency with which two phonemes contrast in all possible environments.'

It is not clear how much thought has been given to the problem of definition by writers making appeal to the notion. For instance, we could disagree with Brown above, in that phonemes such as /f/ and /v/ do not have FLs in isolation; it is only the contrasts between pairs of phonemes which can carry FLs.

King [4] p.7 proposes a formula for the calculation of FL which 'is the product of two factors: the first measures the global text frequencies of the two phonemes in the opposition; the second measures the degree to which the two phonemes contrast in all possible environments, where environment means, roughly speaking, one phoneme to the left and right'. As Vachek [5] p.65 points out, although environment is of obvious importance, King's definition of this as one phoneme to the left and right should have been stated in finer terms.

The main difference between King's formulation and those of other writers is that it is based on conditional probabilities instead of being an information theory approach. Wang [6] (see also [7]) compares four information theory measures of FL, concluding 'more important than the development of a measure that is internally consistent and which conforms to certain linguistic requirements is the task of providing empirical justification for the measure' (p.50).

The value of the concept of FL has been recognised in other linguistic fields, including general descriptive linguistics [8], diachronic phonology [3], automatic speech synthesis and recognition [9, 10] and spelling reform [11]. It has not, however, been applied to the question of language teaching. In this paper, I therefore wish to explore certain aspects of FL which are of use in the teaching of pronunciation. This discussion owes much to the ideas of Avram [12]. For illustration, I shall deal in particular with the following pairs of (RP) phonemes, which are often conflated by learners: /iː, ɪ; ɪə, eə; e, æ; ɔː, ɔɪ; uː, ʊ; p, b; ð, d; n, ŋ; tʃ, dʒ/.

Cumulative text frequency

In the table below, I give the cumulative frequencies for these pairs of RP phonemes based on the figures for connected speech given by Denes [13]. Thus, for example, the cumulative frequency for the pair /e, æ/ (11.05%) is calculated by adding the individual text frequencies of 7.16% for /e/ and 3.89% for /æ/. On the basis of these calculations, we may then propose that a pair with a high cumulative frequency (e.g. /e, æ/, 11.05%) is of greater importance than one with a low (e.g. /ɪə, eə/, 1.83%). That is, over one in every ten vowels is either /e/ or /æ/, whereas under one in every fifty vowels is either /ɪə/ or /eə/. The risks, as far as loss of intelligibility is concerned, of conflating /e, æ/ may thus be considered greater than those of conflating /ɪə, eə/.

Probability of occurrence

These cumulative frequencies disguise the fact that one member of a conflated pair may occur much more frequently than the other. For example, /iː, ɪ/ have a high cumulative frequency (25.57%); one in four of all vowels in connected speech is either /iː/ or /ɪ/. Given that a learner has produced a vowel of the [i] type, it is, however, four times more likely that this corresponds to /ɪ/ than to /iː/. The basic text frequencies are 21.02% for /ɪ/ and 4.55% for /iː/.

The closer to 0.50, the more equal are the individual frequencies, and the greater is the potential confusion to be caused by the conflation of the pair. (The probability of the more frequent member is one minus the probability of the less frequent). In this way, we may distinguish four extremes:

(i) pairs with a high cumulative frequency and relatively equal probability, e.g. /ð, d/,

(ii) pairs with a high cumulative frequency but unequal probability, e.g. /iː, ɪ/, /n, ŋ/,

(iii) pairs with a low cumulative frequency but relatively equal probability, e.g. /ɪə, eə/, /tʃ, dʒ/, and

(iv) pairs with a low cumulative frequency and unequal probability, e.g. /ɔː, ɔɪ/.

It would seem reasonable to rank them as above in decreasing order of importance for learners and teachers.

Occurrence and stigmatisation in native accents

Whilst RP has been used as the reference accent in this paper, certain of the learners' conflations are to be found in other native accents. /uː, ʊ/ conflation is widespread in Scotland; /ɪə, eə/ conflation is an increasingly common phenomenon in New Zealand, the West Indies and East Anglia; and /ð, d/ conflation is found, if only sporadically, in the Republic of Ireland, although it is heavily stigmatised. We may conclude that listeners are accustomed to making the perceptual adjustment necessary for intelligibility of these conflations, but not for the others.

Acoustic similarity

As Brown quoted above notes, acoustic similarity between sounds is a relevant factor. That is, /θ, f/ and /ð, v/ are more acoustically similar than /θ, s/ and /ð, z/. For example, /θ, f/ may be difficult to distinguish in bad transmission conditions, as on a telephone line; listeners are therefore already familiar with recognising the intended sound from context. On the other hand, /θ, s/ are more distinct, even on noisy telephone lines; listeners are therefore unaccustomed to realising that a misinterpretation or conflation may have taken place. Comparable acoustic similarity is found between the nasal consonants /m, n, ŋ/.

The structural distribution of phonemes

It is a phenomenon of English syllable structure that /ŋ/ only occurs in syllables containing short vowel phonemes (/ɪ, æ, ʌ, ɒ/). /n/, on the other hand, occurs in syllables with either long or short vowel phonemes. Thus, a learner who conflates /n, ŋ/ will not be open to misunderstanding all the time; his conflation may only lead to confusion where it occurs after a short vowel phoneme, since any occurrence after a long vowel must be /n/ not /ŋ/.

In similar vein, it is a feature of English that stressed word-final syllables do not contain short vowel phonemes unless they also contain a final consonant. Thus, /bɪt/ is permissible (bit), but not */bɪ/. Long vowel phonemes are not subject to this constraint, e.g. /biː/, bee. Thus, any vowel in a stressed word-final syllable without a final consonant cannot be a short vowel phoneme.

Syllable structure constraints therefore limit the potential confusion of conflated pairs (/n, ŋ/, /iː, ɪ/) in particular environments.

Lexical sets

We must not lose sight of the fact that phonemes combine to create the actual words of the English lexicon. There are some phonemes which are not contained in many words. For instance, Wells [14] p.133 notes that the lexical set for the phoneme /ʊ/ is relatively small - around 40 words. The frequency of this phoneme is a mere 1.95%, and would be even lower were it not for the fact that this lexical set includes a number of words of very frequent occurrence, such as put, good, look, would.

The number of minimal pairs

The simplest expression of the FL of a phonemic contrast is the number of minimal pairs which this contrast serves to distinguish. For some English phonemic contrasts, there are plenty of minimal pairs; for others, there are relatively few. For /uː, ʊ/, the only minimal pairs involving common modern words are pool, pull; fool, full; who'd, hood; suit (if pronounced /suːt/), soot. Minimal pairs are similarly scarce for /ʃ, ʒ/ and /θ, ð/. Misunderstanding is therefore very unlikely to occur for these contrasts and on this basis, we may consider them to be relatively unimportant. The following table shows the relative importance of all the vowel and consonant contrasts introduced earlier, in terms of the number of minimal pairs exemplifying the contrasts. The criterion has been set, somewhat arbitrarily, at 20 minimal pairs. Fewer than 20 pairs can be found for those contrasts marked -, while over 20 pairs can be found for those marked +. Minimal pairs for consonants in word-initial position and in word-final position have been calculated separately.

The number of minimal pairs belonging to the same part of speech

Following on from the previous section, we may note that although there are certain contrasts for which there are several minimal pairs, sometimes these minimal pairs involve few words from the same part of speech. These pairs are therefore unlikely to cause confusion in the context of a sentence. For example, there are several minimal pairs for initial /ð, d/. However, it is a phenomenon of English that words beginning with /ð/ are grammatical words, such as the, those, they, then, though. They are thus unlikely to be confused in context with the corresponding /d/ words, which are virtually all lexical words, such as doze, day,

	1	2	3	4
/iː, ɪ/	25.57%	0.18	+	-
/ɪə, eə/	1.83%	0.40	-	+
/e, æ/	11.05%	0.35	+	-
/ɔː, ɔɪ/	3.28%	0.07	-	-
/uː, ʊ/	5.57%	0.35	-	+
/p, b/	6.34%	0.46	+ +	-
/ð, d/	11.81%	0.42	- -	+
/n, ŋ/	13.72%	0.15	* -	-
/tʃ, dʒ/	1.46%	0.42	- -	-

Column 1 = cumulative text frequency, expressed as a percentage of the occurrence of all vowels, or of all consonants.
Column 2 = probability of the less frequent member of the pair.
Column 3 = whether 20 minimal pairs can be found. For consonants, this is given for word-initial and word-final positions. * indicates that /ŋ/ does not occur initially in English words.
Column 4 = occurrence in native accents.

----

den, dough).

Consideration ought also to be given to the fact that the frequency of occurrence of members of the closed set of grammatical words is higher than for lexical words.

The number of inflections of minimal pairs

One problem in counting the number of minimal pairs relying on particular phonemic contrasts is the use which English makes of inflections such as the suffixes for plural, past tense, -ing forms. Thus, for example, for the /ɪə, eə/ contrast, several pairs take /z/, /d/ and /ɪŋ/ endings, e.g. fear, fare; spear, spare; steer, stare. Whether these should be counted as separate minimal pairs or not in the calculation of FL is a somewhat arbitrary methodological consideration.

The frequency of members of minimal pairs

Minimal pairs for the English contrast /uː, ʊ/ are scarce. A few examples exist, further to those quoted above, but in which one member is of such infrequent occurrence that the minimal pair can hardly be said to have any importance. Thus, while the /ʊ/ words would, could, should, look may be considered frequent, the corresponding /uː/ words wooed, cooed, shoed/shooed, Luke are so infrequent as to be almost contrived.

The number of common contexts in which the members of minimal pairs occur

It is also worthwhile to consider whether the members of minimal pairs belong to the same semantic field or not, i.e. whether contexts can be easily supplied in which both members of a minimal pair are plausible alternatives, both grammatically and semantically. Such contexts are easily supplied for English pairs such as fate, faith; trek, track; sherry, cherry; shin, chin; cheer, jeer, but this is not possible for the majority of minimal pairs in English.

Conclusion

In summary, it should be clear that more advanced analysis than a counting of the number of minimal pairs is involved in the calculation of FL. Avram [12] summarises this point succinctly: 'if we suppose that one opposition is illustrated by ten minimal pairs and another by twenty, it does not necessarily mean that the second opposition is twice as important as the first. Starting from minimal pairs, the successive application of certain correctives is essential if we wish to establish the actual value of an opposition more clearly' (p.42).

On the basis of the above observations on FL, we may propose that the relative importance of the phonemic RP contrasts discussed in this paper can be ranked as follows, most important first: /p, b; e, æ; iː, ɪ; ð, d; n, ŋ; tʃ, dʒ; uː, ʊ; ɪə, eə; ɔː, ɔɪ/.

References

[1] Brown, G. (1974) 'Practical phonetics and phonology' in J.P.B. Allen & S. Pit Corder (eds.) The Edinburgh Course in Applied Linguistics (vol.3: Techniques in Applied Linguistics). Oxford University Press, pp.24-58.
[2] Meyerstein, R.S. (1970) Functional Load. Janua Linguarum Series Minor no.99, Mouton, The Hague.
[3] King, R.D. (1967) 'Functional load and sound change' Language 43:831-852.
[4] King, R.D. (1967) 'A measure for functional load' Studia Linguistica 21:1-14.
[5] Vachek, J. (1969) 'On the explanatory power of the functional load of phonemes' Slavica Pragensia 11:63-71.
[6] Wang, W.S.-Y. (1967) 'The measurement of functional load' Phonetica 16:36-54.
[7] Wang, W.S.-Y. & Thatcher, J.W. (1962) 'The measurement of functional load' Report no.8, Communication Sciences Laboratory, University of Michigan, Ann Arbor.
[8] Hockett, C.F. (1955) A Manual of Phonology. Memoir no.11, International Journal of American Linguistics, Baltimore.
[9] Fry, D.B. & Denes, P.B. (1957) 'On presenting the output of a mechanical speech recogniser' Journal of the Acoustical Society of America 29:364-367.
[10] Fry, D.B. & Denes, P.B. (1958) 'The solution of some fundamental problems in mechanical speech recognition' Language & Speech 1:35-58.
[11] Wells, J.C. (1986) 'English accents and their implications for spelling reform' Simplified Spelling Society Newsletter no.2:5-13.
[12] Avram, A. (1964) 'Some thoughts on the functional yield of phonemic oppositions' Linguistics 5:40-47.
[13] Denes, P.B. (1963) 'On the statistics of spoken English' Journal of the Acoustical Society of America 35:892-904.
[14] Wells, J.C. (1982) Accents of English (3 vols.) Cambridge University Press.

VOWEL SHIFT AND LONG-TERM AVERAGE SPECTRA IN THE SURVEY OF VANCOUVER ENGLISH

JOHN H. ESLING

Department of Linguistics
University of Victoria
Victoria, B.C. V8W 2Y2 Canada

ABSTRACT

To investigate the relationship between long-term (voice setting) and short-term (segmental) components of accent in social varieties of Vancouver English, formant analysis of digitally sampled vowels and long-term average spectral (LTAS) analysis from context-controlled readings are compared. Four contrasting patterns of vowel formant frequency shift result for the four survey groups divided by socio-economic index. LTAS peaks for UWC and UMC subjects are significantly differentiated, paralleling consistent vowel system differences between these groups. Comparisons with articulatorily performed models permit tentative identification of supralaryngeal settings corresponding to each acoustic pattern. An explanation is offered of the potential effect of long-term configuration on the measurement of individual vowel formants.

SAMPLING AND SPEECH ANALYSIS

The objective of this research is to determine whether socio-economic divisions of an urban linguistic community can be distinguished on the basis of voice setting shifts as well as in terms of differences in individual vowels. Sociolinguistic data for acoustic analysis are drawn from the Survey of Vancouver English carried out by Gregg et al. at the University of British Columbia [1] and archived at the University of Victoria, which includes tape-recorded interviews with 240 native speakers of Canadian English. Subjects chosen for investigation are 32 female and 32 male natives of Greater Vancouver, from the youngest of the three age divisions (16-35) in the survey. Female and male subjects are divided into four socio-economic groups of 8 subjects each on the basis of social index scores established in the original survey using the Blishen & McRoberts [2] occupation scale and other social indicators. Group 1 represents low social index scores (Lower Working Class), and group 4 represents high social index scores (Upper Middle Class).

To compare vowel clusters across the four groups, vocalic nuclei are computed for two tokens of each of ten vowel phonemes for each speaker, from identical environments of the same text in reading style. Using ILS speech processing algorithms to determine formant frequencies, speech samples digitized at 10K samples per second are analyzed using 12-pole autoregressive linear predictive coding [3]. The analysis results in 12 reflection coefficients (K's) per frame (200 points/frame; 50 frames/sec). The K's are converted to filter coefficients (A's) to represent the vocal tract's filtering effects, and the filter response of the A's in each frame is calculated and displayed in a spectral array showing up to five resonant peaks (formants) in the 0-5000Hz range. The peaks' centre frequencies are calculated based on a -3dB shoulder and listed. Target vowels are isolated from remaining speech data auditorily, and mean F1,F2 frequencies are calculated and filed by group for statistical processing and plotting. Follow-up vowel measurements and data collection are now performed more expediently on the Micro Speech Lab package developed in the Centre for Speech Technology Research at the University of Victoria on the IBM-PC microcomputer.

For LTAS analysis, a 45sec sample of continuous speech for each speaker, from the same text used for vowel measurements, is digitized with a PDP-11 time-series data-capturing program. One long-term spectrum is computed for each voice, using a main-frame program accepting only voiced frames while excluding voiceless and low-energy frames. Power spectra of non-overlapping 20msec windows at 50Hz resolution and pre-emphasis factor 1 are integrated to obtain final LTAS.

STATISTICAL ANALYSIS

Statistics are performed on log-mean normalized F1,F2 data for approximately 600 female and 600 male vowels, respectively [4]. To compute distance between group vowel clusters, principal component analysis and canonical discriminant analysis are applied to the four female and four male groups, with the Mahalanobis distance calculated between each group. This yields a probability relating collections of vowels to each other, first as complete vocalic inventories by social group, then as individual vowel phoneme clusters by group.

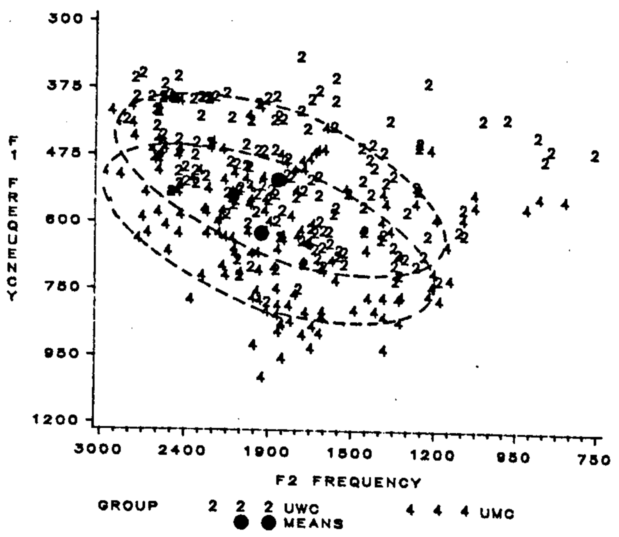
A generalized squared distance measure is used to classify F1,F2 coordinates, as unknown test values, into one of the four social groups as known reference cells. Vocalic inventories of the four male groups are also compared with equivalent vowels from texts performed by the author as models representing contrasting articulatory settings. In this case, test values are assigned to known reference models to yield numbers of vowels from each group that associate most closely with each model [5].

In LTAS evaluation, the same procedure is used to compute probabilities and distance relating spectra in the four female and four male groups, although statistics operate on unnormalized data. Male LTAS are compared with LTAS of the articulatory models using generalized squared distance to identify clustering patterns and to relate LTAS shift to vowel formant shift.

VOWEL FORMANT ANALYSIS

For female subjects, the complete vocalic inventories of all four social groups are significantly differentiated (p<0.001), and a majority of individually compared vowel phoneme clusters are also separated across socio-economic group. The acoustic characteristics of each group's vowels match the four corners of the two-dimensional vowel space: Group 1 (high F1,low F2); Group 2 (low F1,low F2); Group 3 (low F1,high F2); Group 4 (high F1,high F2). The most coherent and best differentiated groups are groups 2 (Upper Working Class) and 4 (Upper Middle Class), illustrated in figure 1. Linguistic contexts are identical; only speakers vary by group affiliation.

FIGURE 1.
SURVEY OF VANCOUVER ENGLISH, FEMALE UWC AND UMC. IDENTICAL VOWELS OF SOCIO-ECONOMIC GROUPS 2 AND 4.

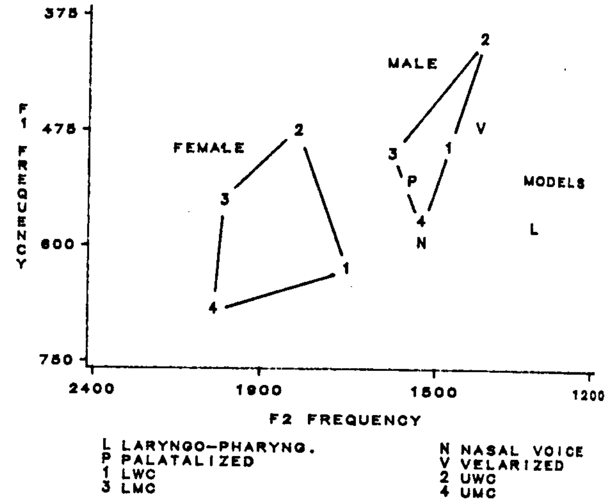


Male vowel cluster values follow the pattern of female vowels except that differentiation between groups 1 and 3 is marginal for speaker-normalized vowels, and not significant using unnormalized data. All other pairings show significant separation (p<0.001). As with female groups, male UWC is furthest separated from other male groups particularly UMC. Figure 2 illustrates normalized means of the four socio-economic groups by sex, and also vocalic means of four comparable model settings.

In the analysis of individual vowel phoneme clusters by group, 77% of all possible pairings for the ten vowels are significantly differentiated for female speakers across the four survey groups (p<0.05), and 43% of all pairings remain separated at the p<0.001 level. Social groups 2 and 4 are successfully differentiated for all ten vowels individually (p<0.01). For groups 1 and 3, which are most difficult to differentiate, only four of the ten vowels show no separation. This supports the distinctions reported for the complete vowel systems of these groups. The rank order of most significantly separated vowels across·groups for female speakers, /ʊ/ /e/ /ɛ/ /ʌ/ /i/ /o/ /æ/ /u/ /ɪ/ /ɒ/, suggests no obvious principles, except that mid, front to central vowels tend to be better differentiated than peripheral, especially open vowels.

Individual vowels for male speakers demonstrate less separation than female speakers' vowels across the four groups. At the p<0.05 level of significance, 62% of all possible pairings for male vowels are differentiated, while only 27% separate at the p<0.001 level. The analysis of individual vowels positively separates male groups 2 and 4, where all vowels differentiate significantly (p<0.001) except /i/, but is not successful in separating the individual vowels of groups 1 and 3. The rank order of socially best differentiated vowels for male speakers is: /ɛ/ /ɪ/ /ʌ/ /ɒ/ /æ/ /u/ /ʊ/ /e/ /o/ /i/. The Spearman rank order correlation coefficient relating male and female rank orders (rho=-.24) indicates that the two lists do not correlate, suggesting that those vowels which function as salient social markers for female speakers are not the same vowels that function as principal social markers for male speakers in the same social classes.

FIGURE 2.
FEMALE AND MALE NORMALIZED GROUP VOWEL MEANS, VOCALIC MEANS OF FOUR VOICE SETTING MODELS.



One possible interpretation of the male order is that /i/ functions as a pivotal vowel, virtually identical in all groups, and that peripheral tense vowels /e/ and /o/ remain more or less the same across groups, while the majority of shifting occurs on open or mid-open vowels. Greatest differentiation appears in the area of /ɪ/ /ɛ/ /ʌ/ /æ/ /ɒ/, where a decrease in F1,F2 accompanies raising and backing for group 2, and an increase in F1,F2 accompanies fronting with nasalization for group 4.

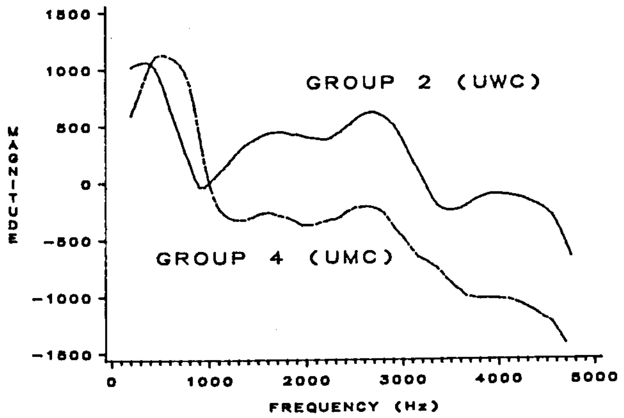
LONG-TERM AVERAGE SPECTRAL ANALYSIS

For LTAS analysis, 45-60sec of the 64 subjects' voices are low-pass filtered at 5KHz and digitized at 10K samples/sec with high-shaping to accentuate frequency information and remove DC. Digitized data are processed in 200 sample point frames through a Hamming window and FFT routine to obtain 20msec power spectral arrays. After unvoiced and silent frames are removed, a swept filter adjusted according to expected harmonic spacing produces smoothed spectra accumulated in a single array to represent the average vocal tract response of the utterance.

For articulatory identification, LTAS of three 40sec phonetic texts performed by the author using controlled voice settings described by Laver [6] and Esling [7] are analyzed: close rounding (CLR), close jaw (CLJ), dentalization (DEN), retroflexion (RET), palatalization (PAL), uvularization (UVU), velarization (VEL), laryngo-pharyngalization (LAR), nasalization (NAS), faucal constriction (FAU), raised larynx (RLX) and lowered larynx (LLX). Root-mean-squared distance measures indicate that each text resembles more closely other texts with the same voice setting than it does identical texts with different settings. Speaker recognition research corroborates that samples of this length are relatively text-independent [8].

The first two formants of four settings, LAR, VEL, PAL, NAS, parallel F1,F2 plots of survey data (see figure 2). The first two dominant LTAS peaks (P1,P2) of these models also correspond to F1,F2 in their relative acoustic orientation, but with P1,P2 systematically lower in frequency than F1,F2. Superimposing laryngo-pharyngalization on a given text increases P1 and decreases P2, which conforms with acoustic predictions for extreme tongue retraction [6]; velarization produces an approximation of P1 and P2 as for an [u]-quality vowel; palatalization results in a systematic shift in mean spectral peaks as for an [i]-quality vowel; and a nasal setting results in higher-frequency P1, with attenuation in the magnitude of P1 relative to P2. In evaluating LTAS data for Vancouver survey groups, it is expected that group 1 will demonstrate high P1,low P2; that group 2 will demonstrate low P1,P2; that group 3 will demonstrate low P1,high P2; and that group 4 will demonstrate high P1,P2. The relative influence of each of the first four LTAS peaks in distinguishing the social divisions of the survey will also be determined.

FIGURE 3.
LTAS OF FEMALE SOCIAL GROUPS 2(UWC) AND 4(UMC).



For female groups, LTAS data significantly differentiate social group 1 from group 2 and group 2 from group 4 (p<0.01) as in figure 3, while other relationships show no significant separation. Spectra are set to zero magnitude at 1000Hz for comparability and to minimize the effect of amplitude variation. Female LTAS data corroborate socio-economic distributions of vowel formant data in that groups 2 and 4 are separated by both measures. Due to the presence of voiced obstruents in LTAS, frequencies are predictably lower than for vowel nuclei. Relative F1,F2 orientations are preserved primarily in P1 values and not in P2, as much of the difference between groups is therefore present in third and fourth LTAS peaks.

Table 1. Female vowel formant and LTAS means.

	F1 , F2 (Hz)	P1 , P2 (Hz)
Group 1(LWC):	631 , 1702	450 , 1600
Group 2(UWC):	477 , 1813	350 , 1725
Group 3(LMC):	552 , 2006	400 , 1600
Group 4(UMC):	683 , 2039	550 , 1600

Male LTAS results are also successful in significantly differentiating group 2 from group 4 and group 3 from group 4 (p<0.05). Other relationships again are not significant. The relationship between F1,F2 values and LTAS P1,P2 values is clearer for male groups than for female groups. Both F1,F2 and P1,P2 for male group 2 are low, resembling the predicted pattern of velarization, while F1,F2 and P1,P2 for group 4 increase, coinciding with the shift predicted for nasalization. P1,P2 are systematically lower than F1,F2, confirming that LTAS data include voiced speech information which has the effect of lowering average frequencies.

INTERPRETATION OF RESULTS

An articulatory interpretation of the acoustic differentiation of vowels across the social scale of Vancouver English is proposed which associates LWC vowel clusters with tongue backing and lowering (laryngo-pharyngalization); UWC with tongue backing and raising (velarization); LMC with tongue fronting and raising (palatalization); and UMC with tongue fronting and nasal voice setting. To quantify these associations, male survey data are compared with equivalent vowel systems of four articulatorily modelled settings which are included in the male normalization routine. The generalized squared distance algorithm takes the four models as reference cells and forces tokens from survey data into one of the four cells. Internally, there is considerable misclassification of vowel tokens among the four settings, and the majority of survey values cluster with the velarized model. However, classification of survey data differentiates significantly in the case of groups 2 (UWC) and 4 (UMC) and the VEL and NAS models as tabulated below.

Table 2. Assignments of male vowels by group to model setting vowel sets (rounded %).

	LAR	VEL	PAL	NAS	n
Group 1(LWC):	13%	68%	14%	5%	139
Group 2(UWC):	3%	97%	0%	0%	145
Group 3(LMC):	14%	67%	12%	8%	153
Group 4(UMC):	19%	56%	10%	15%	145
Totals:	12%	72%	9%	7%	582

These distributions reflect the same articulatory pattern as female vowel clusters. Individual vowel phonemes classify primarily into VEL from group 2, and into NAS from group 4. Chi-squared tests indicate that there is significant evidence for an association between groups 2 and 4 and the four reference models LAR, VEL, PAL, NAS (3 d.f., p<0.001) and, furthermore, that the two groups are significantly differentiated on the basis of assignment into VEL, NAS (1 d.f., p<0.001). Broader interpretations of these results depend on variables such as performance conditions of the models and limitations of using only two formants. Nevertheless, they permit identification of the relative susceptibility of vowels to the shift from UWC to UMC quality, reflected in the acoustic shift from low to high F1,F2 values.

LTAS data support conclusions reached on vowel formant evidence. Tukey's test for variable effect is applied to the four models, LAR, VEL, PAL, NAS, to assess the relative influence of each LTAS peak. The result indicates that P1 is a better predictor of VEL or NAS than is P2. P3 is also a successful variable in separating VEL and NAS settings, and in separating fronting from backing. P4 does not distinguish PAL from VEL or NAS, but does separate it from LAR, as does P2. This suggests that P3 adds information to P1, and that P4 adds to P2, when LTAS data are used in addition to F1,F2 to distinguish voices.

Statistical comparisons of male LTAS data with the 12 models indicate that the models as a set are significantly differentiated from the four survey groups ($p<0.05$). The generalized squared distance function indicates high internal coherence for each survey group, and yields similar associations to those previously discovered by vowel formant analysis, namely the association of tongue-retracted settings UVU, VEL with groups 1 and 2 (LWC/UWC) and of NAS, PAL with group 4 (UMC), shown in table 3.

Table 3. Distance between voice setting models and male Vancouver social groups in %.

	1(LWC)	2(UWC)	3(LMC)	4(UMC)
UVU	0.50	0.38	0.02	0.09
VEL	0.51	0.23	0.00	0.25
LAR	0.05	0.06	0.83	0.07
LLX	0.09	0.02	0.85	0.05
FAU	0.03	0.02	0.95	0.00
DEN	0.22	0.17	0.32	0.29
CLR	0.31	0.16	0.02	0.51
CLJ	0.32	0.09	0.01	0.58
LLX	0.09	0.19	0.12	0.59
RET	0.20	0.10	0.02	0.68
PAL	0.17	0.06	0.01	0.77
NAS	0.02	0.01	0.00	0.97

Bearing in mind the significant separation of groups 2 and 4, that groups 1 and 3 are not distinguished except for certain vowels, and that group 3 LTAS are more coherent than group 1 LTAS, assignments to group 3 (e.g., LAR) must be treated circumspectly. Assignment of VEL and UVU to both groups 1 and 2, on the other hand, provides supporting evidence to the vowel formant procedure that these groups occupy a different acoustic space from group 4 (if not from group 3) with its closer association to NAS and PAL. Despite the single-speaker limitations of the performed model approach, the associations suggested here are a positive indication that sociolinguistically obtained dialect survey groups can be analyzed, differentiated and tentatively classified using both vowel formant analysis and LTAS analysis techniques.

EFFECTS ON FORMANT MEASUREMENT

There is evidence in this study that long-term settings may influence formant frequency measurement, contributing to why vocalic data values are often difficult to measure. Monsen & Engebretson [9], comparing spectrographic with linear prediction techniques of formant analysis, find that "for fundamental frequencies between 100 and 300Hz, both methods are accurate to within approximately ±60Hz for both first and second formants." They also observe that formant frequencies can be obscured by masking from the fundamental or by broadening of bandwidths.

> It may be easier or harder to accurately recover the resonances of the vocal tract in the vowel sound wave depending on objective factors such as the fundamental frequency, the degree of nasalization of the vowel, or the position of the articulators.

The ILS peak-picking routine used here is observed to encounter masking problems of just this sort. Group 1 vowels produce greatest loss of second formant, resulting in a smaller number of tokens that are acceptable for inclusion, and (perhaps not incidentally) in wider deviation of the tokens that remain. Group 2 is the easiest group to measure, with all formant peaks and bandwidths clearly distinguishable, and has correspondingly the most coherent set of formant values. Group 3 is also not difficult to measure, but group 4 begins to demonstrate the appearance of an intermediate peak and widening bandwidths in all vowels for the largest number of speakers both male and female. This secondary, usually higher amplitude peak overlaps in bandwidth with peak 1, and has therefore been averaged into the computation of F1 since it is distinctly not associated with F2. This phenomenon occurs only rarely in other groups and when it does the voice demonstrates pronounced nasality. It seems likely, therefore, that a generalized low back position of the articulators in group 1, evident in the F1,F2 values of retained vowels, causes a decreasing F2 peak to merge with an increasing F1 peak for many tokens. The fronted and nasalized setting of group 4, implied by the damped but increased values of F1 due to the combined calculation, and the slightly higher values of F2, would not be apparent if these somewhat spectrally confusing tokens had to be eliminated. In this way, the results of this study help to isolate those contributions of vocal tract resonance that are of longer-term duration than individual vowels, and also help to identify how contrasting articulatory configurations affect otherwise identical vowels.

REFERENCES

[1] R. Gregg et al., An urban dialect survey of the English spoken in Vancouver, in H. Warkentyne (Ed.), Papers from the Fourth International Conference on Methods in Dialectology (pp. 41-65), University of Victoria, 1981.

[2] B. Blishen, H. McRoberts, A revised socioeconomic index for occupations in Canada, Canadian Review of Sociology and Anthropology, 13, 71-79, 1976.

[3] J. Markel, A. Gray, Linear prediction of speech, Springer-Verlag, 1976.

[4] D. Hindle, Approaches to vowel normalization in the study of natural speech, in D. Sankoff (Ed.), Linguistic variation: Models and methods (pp. 161-171), Academic Press, 1978.

[5] J. Esling, C. Dickson, Acoustical procedures for articulatory setting analysis in accent, in H. Warkentyne (Ed.), Papers from the Fifth International Conference on Methods in Dialectology (pp. 155-170), University of Victoria, 1985.

[6] J. Laver, The phonetic description of voice quality, Cambridge University Press, 1980.

[7] J. Esling, The identification of features of voice quality in social groups, JIPA, 8, 18-23, 1978.

[8] C. Dickson, An investigation of theories and parameters pertaining to speaker recognition, University of Victoria, 1980.

[9] R. Monsen, M. Engebretson, The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction, J. Speech and Hearing Research, 26, 89-97, 1983.

Acknowledgements. Thanks are due to Prof. R.J. Gregg, Dr. S.K. Wong, and J. Clayards. The cooperation of B. Craig Dickson and the Centre for Speech Technology Research, and grants from the President's Committee on Research, University of Victoria, are gratefully acknowledged. The present investigation was supported by grant #410-85-0431 from the Social Sciences and Humanities Research Council of Canada.

PHONETIC INVENTORY AND TABOO

ROBERT K. HERBERT

Program in Linguistics, State University of New York
Binghamton, New York 13901 U.S.A.

The practice of linguistic taboo, i.e. the avoidance of specific words because of associations with forbidden or sacred things, is well-documented. However, the implications of taboo have not received serious attention in historical studies, especially historical phonetics. While it is recognized that taboo may effect lexical replacement and shift, the role of taboo in explaining irregular sound correspondences (in single words or sets) is an equally important, though neglected, aspect of taboo [5]. These two effects correspond to the strategies open to a speaker when a word is taboo: (1) replace the word with an alternate (synonym, archaicism, borrowing), and (2) modify the pronunciation. There is yet another possible consequence of widescale taboo on the phonetic system of a language, namely an increase in the size and complexity of the sound system. Such an effect is limited to situations of extensive language contact where one language provides the resources for avoiding taboo words -- either a stock of alternate vocabulary or new phonetic units to be exploited in phonetic modification. Given sufficient time and institutionalization of such practice, these foreign sounds may be incorporated into the host sound system. This paper explores this role of taboo in the historical expansion of a phonetic inventory, using the dramatic example of click incorporation in Southern Bantu languages.

One of the most striking and well-known examples of phonetic contamination due to language contact involves the Bantu languages of southern Africa. This group of languages is typologically distinct from the Khoisan languages that surround it in most major features with the exception of the regular exploitation of velaric ingressive consonants, i.e. click sounds, within their phonetic and phonological inventories. This feature is so pervasive in these groups and so rare elsewhere that these languages are sometimes known as "the click languages", although this time is sometimes reserved for Khoisan languages. Clicks have been reported in various languages outside Africa, but they do not function within normal phonology and the number of oppositions never approaches that found in southern Africa.

It is well-established that clicks are not inherited elements in Bantu. They were borrowed from Khoisan, probably Khoikhoi (Hottentot). The Bantu languages most affected by this contact include the Nguni group and S. Sotho. The Nguni group is subdivided into a number of language units, including Zulu, Xhosa, Swazi, Ndebele. This paper provides a new explanation for the incorporation of such highly marked units as clicks into the phonetic inventories of S. Bantu, an explanation that goes beyond reference to language contact. Most documented cases of phonetic/phonological influence due to borrowing are confined to instances of filling inventory gaps, restructuring of constraints, etc. Apart from such instances, non-native phonetic elements are often subjected to loan phonology, i.e., elements from the native system substitute for them. What sets the S. Bantu case apart from others is the enormity and the peculiar nature of the contact effects [4]. For example, it is estimated that one-sixth of Xhosa words contain clicks. The vast majority of these words are of demonstrable or *presumed* Khoisan origin, but there are examples of clicks substituting for inherited Bantu consonants. Almost half of the 55 consonants of Xhosa are almost exclusively confined to the borrowed vocabulary [4]; these are the non-inherited consonants, indicated within parentheses in the chart below.

ɓ									
p	t	(t͡s	t͡y)	c	k	(k͡x)	(ʇ	ʗ	ʖ)
b	d	(d͡z	d͡y)	j	g		(ʇ͡g	ʗ͡g	ʖ͡g)
pʰ	tʰ	(t͡sʰ	t͡yʰ)	cʰ	kʰ		(ʇʰ	ʗʰ	ʖʰ)
f	s	ɬ		s		(x)			h
v	z	ɮ				(γ)			ɦ
w		l	y						
m	n		ɲ		ŋ		(ʇ͡ŋ	ʗ͡ŋ	ʖ͡ŋ)
mʰ			(ɲʰ)				(ʇ͡ŋʰ	ʗ͡ŋʰ	ʖ͡ŋʰ)

An initial question concerns the reasons for this widescale phonetic influence of one language upon another. The usual explanation has to do with the taking of Khoisan wives by Bantu-speaking males [2, 4]. According to this theory, S. Bantu males were polygamous, and the father was only an occasional visitor to his families. The dominant linguistic influence was therefore that of the mother. Such intermarriage had a high incidence and existed over a period of centuries. The details of this explanation, in particular the polygamy of Bantu males, are not universally accepted, but all agree that widescale and enduring contact must be reconstructed. Oral history among several Bantu groups relates the incorporation of Khoisan-speaking clans.

A number of features have enshrined the Khoisan-Bantu contact in the linguistic literature. First, the majority of borrowed sounds are clicks, which are incorporated at three places of articulation with a number of distinct qualities, e.g., Xhosa exhibits 15 distinct click sounds. The mere receptivity of a language to such unusual sound types requires explanation, especially in view of their high markedness value. Second, borrowed consonants appear in inherited Bantu lexical items. Cf. internal correspondences such as Zulu kh:xh as in xhopha 'to hurt the eye' vs. ukhophe 'eyelash'; c:th as in -consa 'fall, drip, leak' vs. ilithonsi 'a drop of liquid'. Lexical reconstructions occasionally show the same bizarre correspondences, e.g. *-tima > -cima 'to extinguish'. Commonly, both an inherited Bantu form and a modified form co-exist with differentiated meanings, e.g. chela:thela 'to sprinkle (ceremonially?):to pour, pour out'. Third, the phonological influence of Khoisan is confined to consonant borrowing. The nasalized vowels of Khoisan are not borrowed, and there is no influence on canonical Bantu phonotactics; vowel sequences and word-final consonants, both pervasive Khoisan traits, are absent in Bantu. Finally, there is no significant Khoisan influence on the very distinctive and highly resilient Bantu morphological and morphosyntactic systems. Thus, if one assumes some intense brand of bilingualism to explain the borrowing of such a large number of exotic consonants, one is hard pressed to explain the absence of other significant influence.

A contributing factor in the incorporation of Khoisan sounds into Bantu phonetic systems must have been the very distinct acoustic quality of clicks. Clicks are perceptually sharp and distinct as a class. The nature of the bilingualism present in the contact situation (whatever its details), the sharp quality of the clicks, and the absence of any inherited Bantu sounds with which they might be easily matched are all factors contributing to their incorporation.

Little can be said with certainty about the linguistic prehistory of southern Africa, including the identity and nature of the Khoisan contact languages. Five click types are found within Khoisan: bilabial, dental, alveolar, palatal, and lateral. No Bantu language displays more than three, and only Xhosa and Zulu exhibit a three-way opposition: dental,(pre-)palatal, and lateral. In other Nguni languages, the inventory is reduced or eliminated. The only non-Nguni language to have acquired clicks is S. Sotho, which displays voiceless, aspirated, and nasal forms of the palatal click. It is generally assumed that S. Sotho acquired only this one type, and there is no good reason to argue otherwise; other demonstrable effects of Khoisan contact are slight when compared with Nguni, e.g. in borrowed vocabulary.

There is very good reason to believe, based on studies of "gene flow", that there is no relationship between Khoisan admixture (as a measure of population absorption) and linguistic borrowing. Studies of gene flow are relevant only if one assumes an prehistoric state of affairs in which San physical types spoke San languages. (Obviously, genes do not speak languages.) Bantu languages spoken by populations with little biological admixture exhibit clicks, and populations with extensive admixture speak click-less languages, e.g. Kgalagadi, Tswana. Thus, the absorption of Khoisan populations cannot in itself explain click incorporation. There must be more to the sociohistory of clicks than Beach's view that "clicking is to some extent contagious" [1].

The most plausible explanation for the peculiar results of this contact situation refers to hlonipha (also hlonipa, hlonepha), customs observed by married women with regard to their male relatives-in-law (and sometimes the mother-in-law), especially the father-in-law. In addition to rules having to do with dress, access to areas of the homestead, etc., hlonipha involves the avoidance of the names of a husband's father and other senior male members of the male line. The custom appears strongest and most extensive among the Zulu and Xhosa, where it is not only the individual's name that must be avoided but also any of its composite syllables (except for suffixal elements). For example, Finlayson [3] discusses the case of a Xhosa woman who must avoid, inter alia, the names Dike, Ntlokwana, Nina, and Saki: she must not utter the syllables di, ke, ntlo, kwa, sa, ki, ni, na. A number of distinct strategies are employed to this end:

(1) deformation by consonant substitution

Xhosa	*hl'onipha*	
idikazi	ishikazi	'unmarried woman'
unina	utsitsa	'your mother'
sam	tyam	'my' (cl. 7)

(2) morphophonetic deformation by usana

intsana		'baby' (cl. 11>9)
usapho	intsapho	'family'

In the above examples (2), the syllable is avoided by a morphophonological change due to noun class transfer.

(3) use of a semantically related word:

intsasa	'brushwood'	>	iinkuni	'firewood'
iswekile	'sugar'	>	intlabathi	'sand'

(4) neologism

ukusaba	'to flee'	>	ukulimelela
ukusa	'to take to'	>	ukunawukisa

(5) use of an archaic or borrowed word.

Phonetic substitutions are relevant to present concerns. This strategy is most common when the initial syllable of a stem requires avoidance. Only consonant substitution is involved; there are no cases where vowel substitution alone deforms a syllable sufficiently.

The suggestion is advanced here that the process of hlonipha is the essential part of any explanation for click incorporation. There is no way to understand the extensive (yet restricted) Khoisan influence without recourse to some peculiar linguistic feature of the sociohistorical context. Specifically, the claim here is that native (i.e. Khoisan) phonologies provided Khoikhoi and/or San women with a ready-made and "natural" source for consonant substitution required by hlonipha. That is, a women enjoying a prohibition against uttering particular syllables would look to her own phonetic inventory in order to find alternates. On the one hand, the substitution of a foreign element such as a click is perceptually salient and deforms the syllable acceptably. On the other hand, the use of non-Bantu consonants for this purpose precludes the possibility of homophony with existing words. The existence of an extraordinary phonetic inventory therefore served an important sociolinguistic function.

Several advantages derive from this explanation. First, the presence of clicks in inherited Bantu words is explained. The seemingly random substitution of a click for an inherited consonant represents the historical "fixing" of a hlonipha form. As mentioned above, co-existence of an inherited form and a hlonipha alternate with semantic differentiation is more common.

One striking fact not mentioned in the literature is that there is a direct correlation between the existence of hlonipha in a language and the extensiveness of consonant incorporation. Hlonipha is most pervasive in the same Nguni languages that exhibit the greatest number of click types, i.e. Xhosa and Zulu. It is surely not accidental that the languages in which syllable avoidance is most widely practiced are the same languages that have incorporated three click types and other Khoisan consonants. Apart from Nguni, hlonipha is practiced only by the S. Sotho, but it is less extensive both in terms of the range of individuals whose names must be avoided and the rules of practice. Note that a single click type occurs in S. Sotho. The languages most closely related to S. Sotho, viz. Tswana and N. Sotho, exhibit neither click incorporation nor hlonipha.

The proposed connection between hlonipha and consonant incorporation is further supported by the nonclick consonants that act as favored substitutes in hlonipha. Although no firm patterns appear [2,3], two of the most common Xhosa substitutes are ty [c'] and dy [j]; these consonants are not reflexes of Proto-Bantu consonants. The preferred status of these sounds in hlonipha is like the status of clicks, i.e., they became established as preferred substitutes precisely because they did not occur in native Bantu words. Also, in earlier times, these Khoisan consonants did not themselves require avoidance since they did not occur in Bantu personal names.

A fundamental problem in any attempt to gauge the climate and mechanisms of earlier hlonipha as practiced by Khoisan women is the lack of written records. The linguistic and cultural prehistory of southern Africa is an enormously complex web of migrations, conquests, assimilations, and diversifications. One can say more about the current status of hlonipha, and it is clear that its strength is waning through the area. The literature is full of anecdotal reports of situations in which individuals are forced to violate the taboo. "The custom, once broken, steadily loses its peculiar power over the person breaking it." [2] Urbanization and the consequent weakening of tribal traditions also contribute to the decline of hlonipha.

The conclusion here is that clicks (and other Khoisan consonants) may originally have been restricted to a supplementary vocabulary, i.e. a vocabulary recognized as being outside of "normal" language. However, over the course time, this special status disappeared or was blurred, and the consonants were absorbed into the native inventory, leading the way for lexical borrowings without the expected patterns of loan phonology. As it is impossible to assert anything about the contact situation with certainty, it may be instructive to compare briefly this proposed analysis with other cases of linguistic taboo that have left lasting imprints.

Simons [5] provides a comprehensive survey of taboo in Austronesian languages. The details on what is taboo and the strategies of word avoidance vary from language to language, but there are certain similarities with the S. Bantu data. First, in many instances, it is not only the individual's name that is taboo but also common words from which that name is formed or words that "sound like" the taboo name. For example, all Owa words sharing the initial syllable with a taboo name must be avoided. This parallels the use of phonetic deformation in Bantu when the offending syllable is the first root syllable. Second, there are cases in which specific avoidance forms become conventional. Such cases demonstrate that inherited words can be replaced over time by avoidance forms even when the replaced word is not universally taboo in the community. In one instance, that of Muyuw, 19% of the **basic** lexicon was replaced over a span of 50 years. Third, the effects of naming taboos may be widespread indeed; by some estimates, nearly two-thirds of the basic vocabulary may be potentially taboo (for various individuals) in a community. It must be noted that although the Austronesian examples testify to the potentially considerable influence of word taboo, they are unlike the S. Bantu case in that contact languages had broadly similar sound inventories, and one does not observe the restructuring seen in Bantu. The necessary conditions for the S. Bantu type of contact influence seem to include: (1) intense language contact or bilingualism, (2) radically different phonetic systems in the contact languages, (3) the long-term practice of taboo. Situations in which two of these three conditions obtain are not uncommon, and they may result in significant externally-induced change, including an enrichment of the phonetic inventory. However, it is claimed that the three conditions must be jointly invoked to explain the very peculiar nature of the S. Bantu case, i.e. the incorporation of very heavily marked phonetic items and relatively little influence elsewhere in linguistic structure.

Not all Khoisan words in Bantu are <u>hlonipha</u> forms. The claim is rather that the practice of <u>hlonipha</u> "primed" the languages to be receptive to click incorporation, especially if, as has been traditionally maintained, children's main linguistic influence was that of <u>hlonipha</u>-practicing mothers. The sociolinguistic history of southern Africa is considerably more complex than traditional accounts (oral history and early ethnography) present, but the extensive practice of linguistic taboo has been underappreciated in explaining the outcome of language contact. The reconstruction of sociolinguistic prehistory in the area poses a continuing challenge to linguists and anthropologists alike.

[1] D.M. Beach. *The Phonetics of the Hottentot Language*. Heffers, 1938.

[2] C.U. Faye. The influence of "hlonipa" on the Zulu clicks. *BSOAS* 3:757-782. 1923-25.

[3] R. Finlayson. Hlonipha - the women's language of avoidance among the Xhosa. *South African Journal of African Languages*, Supplement 1982, 35-60.

[4] L.W. Lanham. The proliferation and extension of Bantu phonemic systems influenced by Bushman and Hottentot. *Proc. Ninth Int'l. Congress of Linguists*, 382-391. Mouton, 1964.

[5] G.F. Simons. Word taboo and comparative Austronesian linguistics. *Pacific Linguistics* C-76, 157-226. 1982.

CONNECTED SPEECH PROCESSES: A SOCIOPHONETIC APPROACH

SUSAN WRIGHT

Department of Linguistics
University of Cambridge
Cambridge CB3 9DA. UK

PAUL KERSWILL

Dept. of Linguistic Science
University of Reading
Reading RG6 2AH. UK

ABSTRACT

This paper reports results from an electropalatographic study of coarticulatory phenomena in the speech of 2 speakers of Cambridge English. These are alveolar place assimilation and /l/vocalisation, which occur in connected speech (connected speech processes or CSPs). The principal aim of the study was to investigate the articulatory gradualness of these CSPs and to determine the effects of speaking rate and care of articulation on their application. Assimilation is shown to function as a fast speech process, strongly influenced by speech rate, whereas /l/vocalisation is sociolinguistically salient -- its application being more affected by care than by rate.

1. INTRODUCTION

1.1 Phonetic and Phonological Description

Both phoneticians and phonologists have been concerned with the description of coarticulatory phenomena -- as phonetically motivated processes and in terms of their description within a phonological theory /1,2/. Connected speech processes (CSPs) have been classified in terms of phonemic and allophonic variation /3/. The former subsumes processes such as assimilation (a segment changes phonemic identity under the influence of an adjacent segment), coalescence (segments combine to form one segment, yet retaining articulatory and auditory features of both) and deletion. Allophonic variation includes feature-spreading, lenition and reduction processes (where segments fail to reach articulatory or auditory targets in production).

This dichotomy however, implies that so-called 'phonemic' CSPs may be discrete or categorical, applying in an on-off fashion. In addition, the fact that some CSPs may result in a segment's coming to resemble phonetically a different phoneme could be viewed as a matter of chance, depending on the phonemic inventory the language happens to have. A further difficulty is related to the problem of discreteness. Some CSPs become phonologised or 'fossilised' in the course of linguistic change, leading to morphophonemic alternations which are always discrete. 'Fossilisation' then is the residual effect of coarticulation after the factors conditioning its application in connected speech have disappeared. The transition from processes motivated by 'phonetic' factors such as speaking rate, care of articulation and environment, to morphophonological rules does not seem to have been studied in any great detail from a phonetic point of view (although results from experimental work on phonetic motivation have been applied to the explanation of sound change /4/).

1.2 CSPs as Sociolinguistic Variables

Sociolinguistic studies /5/ provide strong evidence that sound change does not occur uniformly and imperceptibly in a language or speech community. Instead, two forms (older and newer variants) of a sound may co-exist within the community, not randomly, but showing systematic patterning. This patterning is manifested as either i) linguistic differences between groups of speakers (distinguished by sociological criteria like sex, age, class) or between individuals; or ii) style or register-bound variation in an individual's speech or in the speech of a group.

Kerswill /6/ found that the working-class vernacular English of Durham contains a number of CSPs differentiating it from RP. Regressive voicing assimilation ([laɪg meɪ] for like me) and the deletion of the final vowel of into in the phrase into the car are processes not found in RP. Alveolar place assimilation (giving [bæggai] for bad guy) on the other hand, does not occur in Durham vernacular though it is widespread in RP. The fact that RP is used by some speakers in Durham while others use a broad 'vernacular' or intermediate variety, suggests the presence of a socially-stratified system as a possible model of variation. Within such a model, the CSPs mentioned function as linguistic variables. In Durham vernacular, some processes seem to be deliberately avoided in formal speech styles as well as being less widespread in middle-class speech, while alveolar place assimilation is a prerequisite of the RP spoken in Durham.

One framework for investigating the sociolinguistic salience of CSPs is the extent to which their articulation is discrete or gradual: If a variable with clear social differentiation thought to involve discrete alternation (on the basis of careful auditory analysis) is actually articulatorily

gradual (that is, its variants show intermediate articulations), it would be reasonable to conclude that the process was both a linguistic variable and a phonetically-motivated CSP.
However, for a CSP in the process of fossilisation to be sociolinguistically salient, one would expect the auditory distinction between its application and its non-application to be more marked than if the CSP was a completely 'natural' coarticulatory process. Consequently, we would expect a tendency towards auditory discreteness while the CSP is still influenced by speaking rate and care of articulation. But the actual articulation may contain elements basic to both segments to differing degrees. The hybrid nature of the gradual articulations contrasts with the discrete auditory percept.
We focus on the instrumental investigation of two CSPs in Cambridge English: alveolar place assimilation and /l/vocalisation /7/. We attempt to identify the effects of speaking rate and care, which provide some insight into the nature of the interaction of phonetic and social factors influencing CSPs. Alveolar place assimilation has been identified as a phonetically-conditioned coarticulatory process in RP, and as a socially-stratified variable in the Durham speech community. /l/vocalisation converts postvocalic [ɫ] in prepausal and preconsonantal environments into a nonsyllabic back vocoid [ɤ] or (rounded) [o]. It has been identified as a quite recent development in local southern varieties of English /8/ -- treated as an optional process found in rapid casual speech, which may be influenced by non-linguistic factors like style, age and class in Norwich English /9/.

2. METHOD

Two speakers of Cambridge city English, aged 18 and 22 years, read a set of sentences designed to elicit these CSPs. Since one possible index of sociolinguistic salience would be the extent to which a CSP is applied (both completely and partly) across a range of speaking conditions or 'phonological styles' /10/, the subjects were instructed to read the sentences in four different modes, differentiated on the basis of speech rate and care: 'slowly and carefully', 'at a normal, comfortable speed', 'as fast as possible, but carefully at the same time' and 'as fast as possible'.
The sentences contained 17 positions where an alveolar assimilation could occur, and non-assimilating environments as a basis for comparison. The alveolar consonants were /d, n/. There were 10 opportunities for /l/vocalisation, in preconsonantal and prepausal environments:

Assimilating environments:	Control environments:
DB lad passed; bad place	lab passed
Fleetwood Park	
DB red banner	ebb back
DG bad guy; shed got;	Rag guide
DK bad car; good clothes;	Craig couldn't
maid couldn't	
DM orchard management;	Saab motor
retard motion	
NB phone box	handsome boy
screen back	cream back
NM manmade	ham mayonnaise
NK Ron comfort	wrong company
NK Van Causeway	Hang Corner
Jason caught	

/l/vocalisation environments: calls from; calls upset; sold; told; well retard; well#; table#; Castle#; Wilson.

A 3-way classification for the analysis of the articulation types on the electropalatographic (EPG) record was adopted (see Appendix for illustration):

absence of assimilation/vocalisation: the EPG record shows a complete alveolar closure at some point during the articulation (score = 0);

partial assimilation/vocalisation: the record shows more lateral and/or alveolar contact than the non-assimilating environment, but no complete alveolar closure at any point during the articulation (score = 1);

complete assimilation/vocalisation: assimilation: the record is either identical with the non-assimilating environment; or shows less lateral and/or alveolar contact; /l/vocalisation: the record shows either total absence of lingual contact for vocalisation, or contact characteristic of back vocoids /11, 12/ (score = 2).

Partial assimilation/vocalisation will not be reflected as a uniform pattern of lingual-palatal contact on the EPG record. An individual's record may be marked by idiosyncratic articulatory patterns and asymmetries, and at times obscured by articulatory environment /12/. Thus EPG analysis involves interpretation (abstraction and normalisation) of the lingual-palatal 'plan' of each subject. Partial assimilations include instances of articulatory 'gradualness' as exemplified in the Appendix, but exclude double articulations.

3. RESULTS

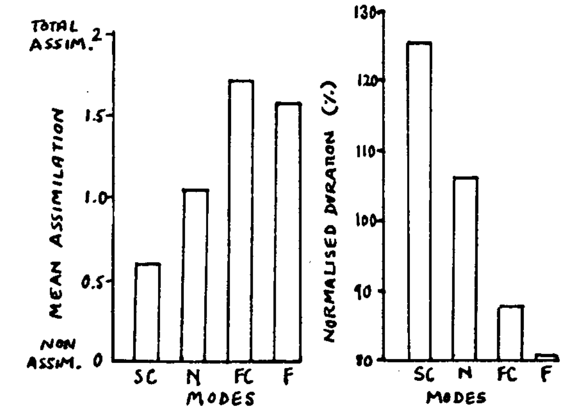
3.1 Alveolar Place Assimilation

Table 1: Number of occurrences of articulation types across speaking modes (2 speakers):

Reading modes:	slow careful	normal	fast careful	fast
complete assim.	8	16	27	24
partial assim.	6	6	5	7
non-assim.	20	12	2	3

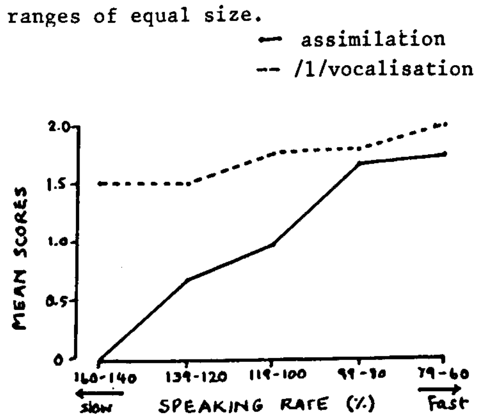
Table 1 shows the number of articulation types for both speakers across the four reading conditions. It shows a large increase in the number of assimilated tokens as the speaking mode becomes faster. This increase occurs at the expense of non-assimilations, the proportion of partial assimilations remaining quite constant across all modes. Consequently, although there is a marked shift overall, from frequent non-assimilation to complete assimilation in faster rates, the application of the process is by no means discrete. This is indicated by the gradual articulations in each mode (which never fall below 14% of the total). Figure 1 shows a comparison of the speakers' mean assimilation scores with their mean speaking rate (normalised duration for each speaking condition). Speaking rate is expressed as a mean percentage of the normalized duration for each mode.

Figure 1: Comparison of normalised durations and mean assimilation scores across speaking modes:



A sharp increase in complete assimilations occurs with a shift to fast speaking modes. According to this profile, the possible effect of care is not separable from that of speaking rate. Indeed, complete assimilations seem to be applied without reference to 'care' of articulation, which might be expected to reduce lenition of the sort involved in this CSP. A possible explanation is that speakers actually paid less attention to articulation in 'fast, careful' mode. Alternatively, (since this condition was tested after 'fast' mode) it is possible that speakers' habituation /10/ to both the test material and the task reduced their attention to their actual speech.
To test whether effects of rate can be distinguished from care of articulation, the duration of each reading of each sentence was measured against assimilation scores. (Because the sentences varied markedly in syllable number, each was measured in ms, then its duration was calculated as a percentage of the mean duration of 4 readings, arriving at a normalised index of duration.) The speaking rate varied considerably, by no means being isomorphic with the speaking conditions in all cases (so some sentences in 'normal' mode are shorter than they are in 'fast' mode). Figure 2 shows mean scores plotted against normalised duration (speaking rate). The solid line indicates the distribution of assimilations across five ranges of equal size.

Figure 2:



Around the mean normalised duration (100.3%) the type of articulation changes dramatically -- from partial assimilations (scores below 1) to close to complete assimilations. This pattern suggests that speaking rate has a direct effect on the degree of assimilation above a certain range. More importantly, actual rate contrasts markedly with the rate speakers intend or identify with the flags 'slow', 'normal' and 'fast'. This suggests that assimilation occurs sporadically at lower rates, but once a certain rate is reached, complete assimilations seem to be applied almost without exception.

3.2 /l/vocalisations

Table 2: Number of occurrences of articulation types across speaking modes (2 speakers):

Reading modes:	slow careful	normal	fast careful	fast
complete voc.	14	15	15	18
partial voc.	5	5	5	1
non-voc.	1	0	0	1

In contrast with the distribution of assimilations across speaking conditions, which appears to be influenced by a speaker's actual rate of utterance, the comparatively low variation in the application of complete vocalisations suggests that speaking rate does not have the same influence on /l/vocalisation. In particular, the frequency of complete vocalisations is consistently high in all 4 modes. This uniformity suggests that vocalisation is a CSP which is becoming phonologised, and is characteristic of a range of speaking conditions including 'phonetic' criteria like rate and care of articulation.
There are some effects of rate and care however. The incidence of partial /l/vocalisations is quite consistent across all modes except the fastest. In this condition, complete vocalisations increase at the expense of partial /l/vocalisations, indicating a possible tendency for speakers to apply vocalisation discretely above a certain speaking rate. This might be interpreted as evidence that rate has a direct influence on the transition from gradual to discrete application of this CSP.
However, as indicated in Figure 2 (/l/vocalisation represented by the dotted line), actual speaking rate is not isomorphic with the rate speakers believe they are adhering to as complete vocalisations are by no means correlated with increasing rate. Consequently, some sentences produced in 'fast, careful' mode are actually of shorter duration than those in 'fast as possible' mode. If speakers can be assumed to be paying more 'attention' to speech in terms of care of articulation in 'fast,careful' mode than in 'fast' mode, then it is reasonable to conclude that, with fast speaking rates, care of articulation may reduce the incidence of partial /l/vocalisations.

4. CONCLUSION

The comparison of the scores for /l/vocalisation and alveolar place assimilation provides a partial profile of CSPs with different functions within the same speech community (see /13/ for a detailed analysis and supplementary evidence from an auditory study). These processes contrast in the degree to which they are influenced by phonetic factors like speaking rate and care of articulation.
Assimilations apply gradually and not uniformly across a range of speaking conditions. Speaking rate appears to influence the application of complete assimilations, whereas their occurrence is not markedly reduced by shifts to 'careful' modes of speech.
Vocalisation, on the other hand, shows a tendency to occur with a consistently high frequency across all modes. The increase of complete vocalisations in fast modes -- not restricted by the criterion of 'care of articulation' or 'attention to speech' -- indicates that /l/vocalisation is more affected by care than by speaking rate.
These results indicate that alveolar place assimilation functions as a 'fast speech process' directly influenced by phonetic factors, whereas /l/vocalisation seems to be a sociolinguistically salient CSP in the process of fossilisation.

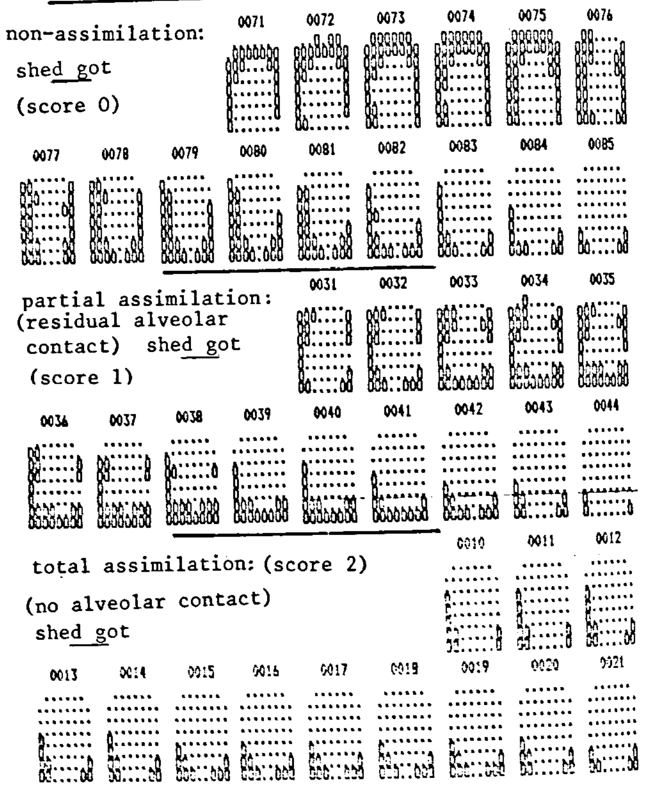
ACKNOWLEDGEMENTS
This paper is based on research funded by the Economic and Social Research Council (ESRC) ref. no. C00232227. The project is being carried out jointly by Dr S. Wright, Dr P. Kerswill, Dr S. Levinson and Dr F. Nolan. We are very grateful to Dr S. Hawkins for her comments and advice on earlier versions of this paper.

REFERENCES

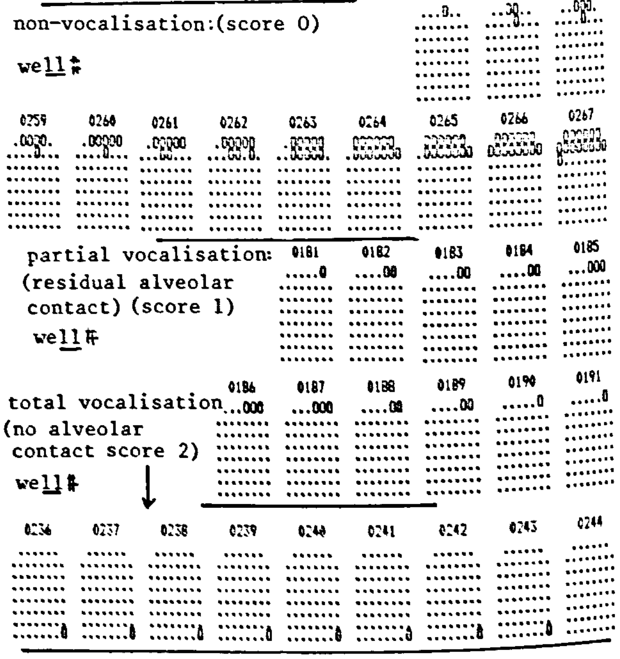
/1/ P. Linell, Psychological Reality in Phonology: A Theoretical Study, Cambridge University Press 1979.
/2/ R. Lass, Phonology. An Introduction to Basic Concepts. Cambridge University Press, 1984.
/3/ A.C. Gimson, An Introduction to the Pronunciation of English (3rd ed.), Edward Arnold, 1980.
/4/ J. Ohala, Phonetic explanation in phonology. In Papers from the Parasession on Phonology. CLS, 1974.
/5/ W. Labov, The social motivation of a sound change. Word 19, 1963.
/6/ P. Kerswill, Levels of linguistic variation in Durham. Cambridge Papers in Phonetics and Experimental Linguistics 3, 1984.
/7/ S. Wright, An experimental profile of CSPs in Cambridge English. Cambridge Papers in Phonetics and Experimental Linguistics 5, 1987.
/8/ J. Wells, Accents of English. 3 vols. Cambridge University Press, 1982.
/9/ P. Trudgill, The Social Differentiation of English in Norwich, Cambridge University Press, 1974.
/10/ W. Dressler & R. Wodak, Sociophonological methods in the study of sociolinguistic variation in Viennese German, Language in Society, 1982.
/11/ W. Hardcastle & P. Roach, An instrumental investigation of coarticulation in stop consonant sequences, Work in Progress 1, University of Reading Phonetics Laboratory, 1977.
/12/ W. Hardcastle & W. Barry, Articulatory and perceptual factors in /l/vocalisation in English Work in Progress 5, University of Reading Phonetics Laboratory, 1985.

APPENDIX: Palatograms showing articulation types

A: degrees of assimilation:



B: degrees of /l/vocalisation:



PERCEPTUAL DIMENSIONS OF LAUGHTER AND THEIR ACOUSTIC CORRELATES

SHIRO KORI

Osaka University of Foreign Studies
Minoo, Osaka, JAPAN 567

ABSTRACT

This study described the acoustic correlates of two perceptual factors that were found to determine the recognition of laughter. 16 tokens of laughter and a non-laughter control token simulated by a Japanese male performer were presented to 10 Japanese who rated the appropriateness of the tokens to each of 12 labels of laughter.
A factor analysis of the appropriateness scores yielded two factors that have been labeled pleasant-unpleasant and superior-inferior, respectively. Correlation analysis of the factor scores and the acoustic data showed that pleasantness vs unpleasantness has significant correlation with the long vs short duration of the strong expiratory noise that may occur at the beginning of laughter, and with the large vs small rate of overall diminishment of the vowel amplitude. Superiority vs inferiority was highly correlated with the long vs short interval between vowels, the high vs low F0 max. or mean value, and the small vs large rate of overall vowel amplitude diminishment.

INTRODUCTION

People laugh for various reasons. Funny or 'incongruous' situations have been considered to be the most characteristic stimulator of laughter. From a communicative point of view, however, laughter is most frequently a signal of the well-being and friendliness of the person who is laughing. It may also used either as a sign of superiority, a device to communicate contempt, or as a submissive vocal gesture.

In most cases the situational cues help us to interpret why a man is laughing. But we are also capable of more or less inferring the reason for the laughter by only listening to the sound. However, very little is known about the acoustic correlates of laughter.

In this paper I will first check the transmissibility of laughter content through the auditory channel, then describe the acoustic correlates of the factors which account for the results of an auditory recognition test on several tokens of laughter.

MATERIAL AND METHOD OF THE RECOGNITION TEST

A thirty year old Japanese male simulated a set of laughters in an anechoic chamber imagining various kinds of laughter in his mind. 16 tokens which seemed to cover a wide range of laughter types were selected from the recordings. These tokens were used in an auditory recognition test along with a non-laughter [hahahahaha]utterance as a control stimulus.

The auditory recognition test was designed to obtain a perceptual characterization of the tokens and to examine the extent to which the performer's intent in laughter transmits through the auditory channel. 17 cassette tapes, each containing more than 50 repetitions of one of the 17 tokens, were prepared and presented to 10 Japanese subjects. The subjects were 20-40 years and 4 of them were male. They were instructed to listen to each of the cassette tapes through headphones as many times as they needed, and to judge, in the first place, whether the token was laughter or not. When the subjects judged the token to be laughter, they were then asked to describe freely the type of laughter and judge whether it was spontaneous laughter or forced laughter. After the judgment of spontaneousness, the subjects were invited to rate the degree of appropriateness of the sound stimulus to each of the 12 Japanese labels of laughter on a 3-point scale. A typical question given in the questionnaire was as follows: "Do you think that this is happy laughter? Please check one. / No. Somewhat. Very much." When the subjects did not judge the token as laughter, no more judgments or ratings were made on that token and the next token was presented to them. The order of the presentation of tokens was random and it was different from one subject to another.

The 12 laughter lablels used in the test were tentative ones. They were chosen by the present author from a list of some forty idiomatic or nearly idiomatic descriptions of laughter collected mainly in a questionnaire of several university students. An attempt was made to include labels which could cover as wide a range of laughter as possible. Most of the labels used here consisted of an adjective and a noun warai (laughter), or of a compound noun whose last element was warai.

They can be translated as follows: /1/ happy laughter (ab. happy), /2/ laughter for funny situation (funny), /3/ mocking laughter (mocking), /4/ ingratiating laughter or friendly laughter (ingratiating), /5/ triumphant laughter (triumphant), /6/ boisterous laughter or heroic laughter (boisterous), /7/ bawdy laughter (bawdy), /8/ laughter to cover one's awkwardness or shy laughter (awkward-covering), /9/ self-deprecating laughter (self-deprecating), /10/ cold-hearted laughter (cold-hearted), /11/ embarrassed laughter or uncomfortable laughter (embarrassed), and /12/ defiant laughter or challenging laughter (defiant). Since many of the labels did not seem to have equivalent expressions in everyday English, the translations given here are not necessarily 10C% accurate. In the following discussion, however, I will use these English approximation for convenience's sake.

RESULTS AND DISCUSSION

Appropriateness scores of the tokens for each of the 12 labels were calculated in the following manner. Responses of maximum and medium favor were given 1 and 0.5 points, respectively. Negative responses were given 0 points. Sums of these scores across the subjects were then calculated and divided by the total number of the subjects and transformed into percentages. Tab.1 shows the obtained scores together with the spontaneousness scores computed in a similar way. The performer's intent is indicated by underlining. Note that a token which was judged to be laughter by X% of the subjects could at most obtain the score 'X' for any of the 12 labels, even though all these subjects judged it with maximum favor, because the subjects who did not judge it to be laughter did not make any ratings about the content of the laughter in that token. The results of the free description and the spontaneousness scores will not be discussed in this paper.

Transmissibility of laughter content through the auditory channel

12 tokens out of 17 obtained favorable judgment as laughter ()=8C%) and 2 tokens registered very unfavorable scores ((=20%). One of the two tokens which were not heard as laughter was the non-laughter control stimulus (No.7 in Tab.1), and another one was intended and judged to be laughter by the performer. These two tokens has two unique acoustic characteristics: the respiratory noise between vowel-like component was all voiced, and the range of vowel amplitude change was very small.

Of the 12 tokens which were heard as laughter, 3 showed a strong agreement between the performer's intent and the judged content by the subjects ()= 75%, No.6,8,16). They were two tokens of happy and one token of mocking laughter. The happy tokens were also judged to be funny. The mocking token was also heard as cold-hearted. There was one token (No.9) in which the judgment of the subjects (mocking) opposed to the performer's intent and perception (funny).

These results show that it is possible, though not always, to identify the content of laughter through the auditory channel at least for happy and mocking laughter. However, the results suggest at the same time that there are individual differences in the way of laughing that might prevent the hearer from detecting its real intent: funny laughter for one person might be mocking one for others. In other words, while there are individually independent expressions with respect to different types of laughter, there is also room for the individuals to express their feeling and attitude in their own way.

Other types of laughter failed to be recognized correctly. Some of them were given completely different interpretations by the listeners (e.g. No. 10, funny or happy and cold-hearted or mocking). This may either be explained by the bad quality of performance, or this may suggest that laughter itself does not necessarily have the function of communicating such intents. This may highlight the importance of contextual cues. Moreover, the intent of laughter may also be hidden intentionally. For example, the intention of the person who laughs to please someone on purpose is presumably to express apparent happiness or friendliness regardless of his actual disposition. In such a case it would be only the artificiality of the laughter or the situational cues that could signal the real intent. Anyway, we can conclude that the content of laughter is not always unambiguously encoded in its vocal output. Facial expression and body movement may provide cues to detect the laughter content, although laughter is most characteristically manifested in the voice. At this point, however, we cannot make any further assumptions about the relative importance of different channels in decoding laughter content.

Table 1

Recognition scores of laughter
Percent appropriateness of the labels to the tokens

token \ label	L	S	1	2	3	4	5	6	7	8	9	10	11	12
(1)	60	50	35	35	20	25	10	5	20	10	5	10	10	20
(2)	100	70	30	45	35	20	5	0	40	30	20	20	25	20
(3)	50	20	10	15	30	5	40	30	0	0	5	20	5	45
(4)	100	40	25	30	30	50	10	10	15	15	5	25	10	20
(5)	20	0	0	0	10	15	0	0	5	5	5	10	5	0
(6)	100	100	75	60	20	5	10	20	5	10	10	15	15	15
(7)	10	0	0	0	0	10	0	0	0	0	0	10	0	0
(8)	100	30	20	25	75	10	25	5	15	0	15	60	0	55
(9)	80	20	0	5	70	10	10	0	5	15	30	65	25	35
(10)	90	70	35	45	50	15	35	10	10	15	20	55	10	50
(11)	80	30	15	35	50	0	50	25	20	5	15	35	5	50
(12)	80	30	10	5	45	35	20	5	35	25	40	40	45	25
(13)	90	50	25	40	55	5	20	10	10	5	25	20	5	30
(14)	80	20	15	20	20	40	10	0	5	25	20	25	35	5
(15)	100	50	40	50	35	35	5	10	5	20	15	25	25	20
(16)	100	100	80	85	15	10	20	40	15	20	0	5	10	25
(17)	50	10	15	10	15	0	10	5	0	0	0	30	5	20

L : whether or not the token expressed laughter
S : spontaneousness

Perceptual dimensions of laughter

In judging the content of laughter, such pairs of labels as happy and funny, mocking and cold-hearted and triumphant and defiant were used similarly. This suggests that the two laughter-types of each pair are realized similarly. Evidently, the two labels of each pair share a common psychological meaning or 'factor'. Most likely, the judgment of the laughter content of the sound tokens was made according to such underlying psychological factors. In order to extract such factors that determined the listeners' strategy in judging the tokens in terms of appropriateness to the laughter labels, the appropriateness scores were analyzed in a factor analysis.

Of the 17 tokens used in the recognition test, two tokens, including the control token, were excluded from the data set for the factor analysis, because they obtained very unfavorable judgment as laughter ((20%).

Principal factor analysis using squared multiple correlations for the prior communality estimates yielded two major factors. These two factors together accounted for 70% of the total variance. They were then rotated using the varimax procedure. Fig. 1 presents the rotated factor loading pattern. Both factors turned out to be bipolar. The first factor has been labeled pleasant-unpleasant, because it loaded happy, funny, and boisterous on one side and cold-hearted, mocking and self-deprecating on the other side. The second factor has been named superior-inferior, because triumphant and defiant were opposed to embarrassed, awkward-covering and ingratiating. These two factors correspond to the two fundamental psychological dimensions of interpersonal behavior (love-hostility and dominance-submission). Pleasant-unpleasant is also the most important dimension in the recognition of emotionss. It is labeled also as evaluative or friendly-hostile.

As a final step of factor analysis, the factor scores of the tokens were calculated for each of the two factors.

Correlation analysis of the perceptual factors and the acoustic variables

15 acoustic measures were defined and calculated for each of the tokens and then correlated with the factor scores. These variables consisted of 5 durational, 4 FO and 3 amplitude characteristics as well as the first three formant frequencies.

Durational variables are: /1/ duration of the initial strong expiratory noise, /2/ duration of the initial high FO (500-1000Hz), /3/ number of alternations of respiratory noise ([h]) and vowel, /4/ mean duration of vowels and /5/ mean interval from the end of a vowel to the beginning of the next one (ab. mean vowel interval). See spectrograms for the variables /1/, /2/ and /3/.

FO variables are /6/ the highest FO value of the vowels in the token (ab. FO max.), /7/ the mean of the vowel maximum FO values on a logarithmic scale (ab. FO mean), /8/ the range of FO movement on a logarithmic scale, defined as the difference between the highest and the lowest vowel maximum values, /9/ the normalized FO range defined as the measure /8/ divided by the total duration of the noise-vowel alternations in the token. Since the vowel maximum FO values monotonously declined from the beginning to the end of the noise-vowel cycles, this measure roughly correspond to the rate of overall FO declination. The pattern of FO movement during the noise-vowel alternations was not included in the variables to be correlated with the factor scores, because in these tokens it declined almost monotonously from the beginning to the end.

Amplitude variables are: /10/ the range of amplitude change in the token, defined as the difference between the highest and the lowest vowel maximum velue, /11/ the normalized amplitude range of vowel defined as the measure /10/ divided by the total duration of the noise-vowel alternations, and /12/ the mean amplitude difference between the maximum value of the fricative noise and that of the following vowel (ab. noise-vowel amp. difference). Since in most cases the vowel maximum amplitude diminished monotonously from the beginning to the end of laughter, the measure /11/ roughly correspond to the rate of overall vowel amplitude diminishment.

Formant frequencies are the mean values of steady state portions (/13/ F1, /14/ F2, /15/ F3).

Figure 1

Rotated factor pattern

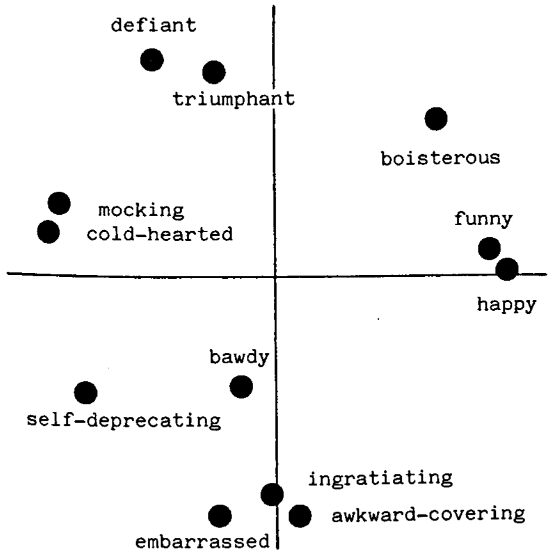


Table 2

Correlation coefficients between 15 acoustic variables and factor scores

accustic variable	Factor 1		Factor 2	
/ 1/	.646	**	-.166	
/ 2/	.148		.270	
/ 3/	.185		.419	
/ 4/	-.351		.469	
/ 5/	-.157		.761	**
/ 6/	-.091		.570	*
/ 7/	-.249		.558	*
/ 8/	.267		.081	
/ 9/	.425		-.290	
/10/	.322		.090	
/11/	.570	*	-.555	*
/12/	-.084		.511	
/13/	-.102		.445	
/14/	-.348		.210	
/15/	-.234		.507	

* p(.05 ** p(.01 HO : r=0

In tab.2 the correlations between the two sets of factor scores and the 15 acoustic variables is presented together with the statistical significance (HO: r=0). Since one of the tokens consisted of only one vowel-noise cycle (No.9), it was excluded from the calculation of correlation coefficients for the variables /5/, /8/, /9/, /10/ and /11/. Acoustic variables which showed significant correlation with the first perceptual factor (pleasant-unpleasant) were /1/ initial expiratory noise duration (r=.646) and /11/ normalized vowel amplitude range (r=.570). These two variables were significantly correlated with each other (r=.750, p(.01). The next highest correlation was found in /8/ FO range (r=.425), but it did not result statistically significant.

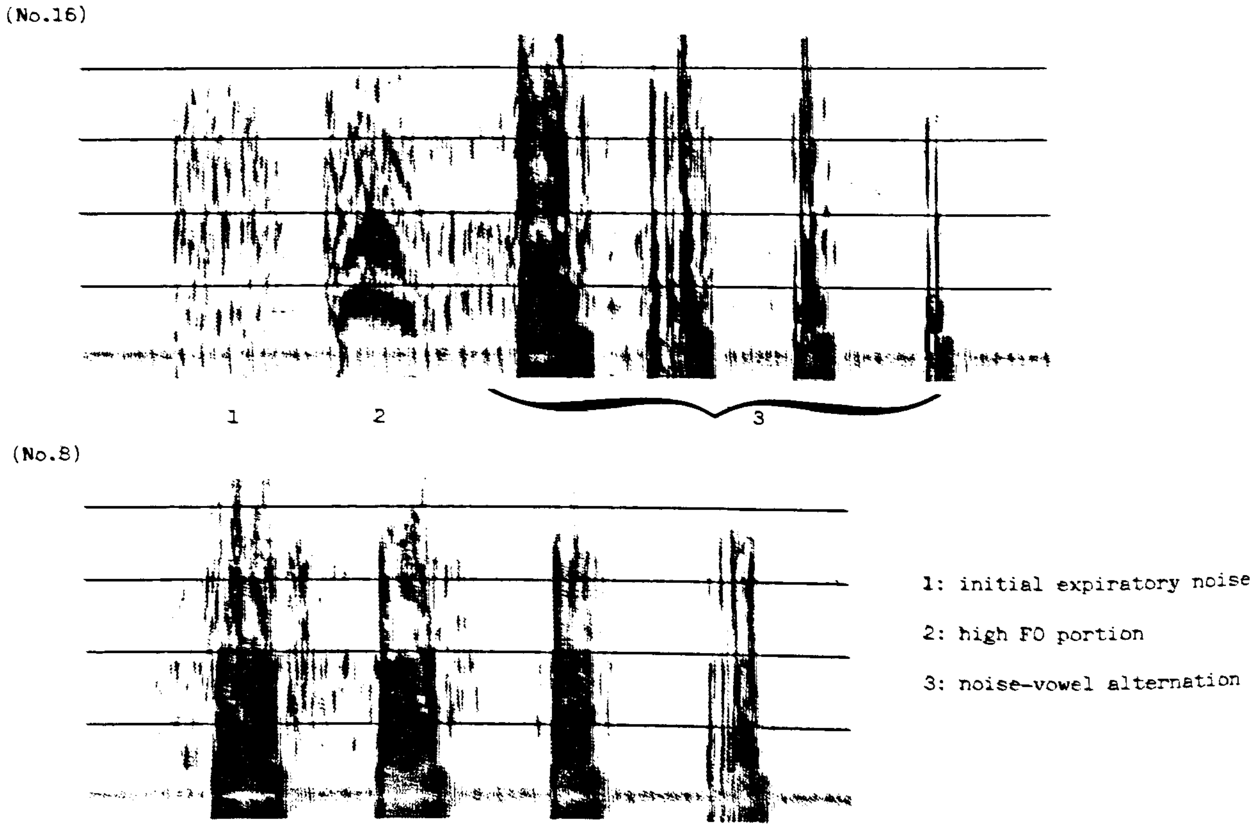
The second perceptual factor (superior-inferior) was significantly correlated with /5/ mean vowel interval (r=.761), /6/ FO max. (r=.570), /7/ FO mean (r=.558) and /11/ normalized vowel amplitude range (r=-.555). /12/ noise-vowel amplitude difference (r=.511) and /15/ F3 (r=.507) had relatively high correlation though they were slightly lower than the significance level. Of these variables, /6/ FOmax and /7/ FO mean (r=.951, p(.001), /5/ mean vowel interval and /11/ normalized vowel amplitude range (r=-.671, p(.01), /5/ mean vowel interval and /15/ F3 (r=.554, p(.05), /7/ FO mean and /15/ F3 (r=.560, p 05), /11/ normalized vowel amplitude range and /15/ F3 (r=-.659, p(.05) showed significant inter-correlation.

These results suggest that the pleasantness vs. unpleasantness of laughter was acoustically characterized in part by the long vs short duration of the initial expiratory noise (42%) and by the large vs small normalized amplitude range, i.e. the rate of amplitude diminishment from the beginning to the end of the noise-vowel cycles (32%). Since these two variables were highly correlated with each other, they accounted only for 44% of the total variance. Even adding the normalized FO range (/9/) which showed the next highest correlation, did not improve the R^2. On the other hand, the superiority vs inferiority of laughter was determined well by the long vs short interval from vowel to vowel (58%), the high vs low maximum or mean FO value of the noise-vowel reiteration (32, 31% respectively) and by the small vs large rate of amplitude diminishment (31%). These three variables accounted for 85% of the total variance.

The present results, however, do not ensure that those acoustic variables found to be correlated with the hypothetical perceptual factors are the 'real' perceptual correlates. A research using synthesized stimuli will be necessary in order to evaluate the perceptual effect of those acoustic variables.

Figure 2

Wide-band spectrograms of a typical happy or funny laughter (No.16) and a mocking laughter (No.8)



DES MECANISMES DE CONSTITUTION DU CONTOUR MELODIQUE DANS LES QUESTIONS "OUI-NON" EN FRANÇAIS

MICHEL NIKOV

Faculté des lettres classiques et modernes,
Université de Sofia, 1000 Sofia, Bulgarie

RESUME

La communication étudie les rapports entre la visée communicative, la perspective fonctionnelle et la structure mélodique des questions "oui-non"en français moderne. La visée communicative s'exprimant par le concours de moyens intonatifs et lexico-grammaticaux,on se penche sur les rapports entre ces deux types de moyens expressifs. Outre les rapports bien connus,décrits par Pechkovskij, certains types de questions françaises fournissent un exemple d'interaction que l'on pourrait appeler "le principe de solidarité".

INTRODUCTION

Des études expérimentales récentes (1) témoignent qu'il convient (du moins pour certaines langues) de rechercher des traits intonatifs de la visée communicative (2,pp.22-25)(ou modalité d'énonciation (3,p.145))dans trois dimensions,notamment, dans l'organisation mélodique, temporelle et dynamique de l'énoncé. Ces recherches mettent plus particilièrement en valeur le rôle de la structure temporelle - résultat du fonctionnement en temps réel du mécanisme de production de la parole.Même sous cette optique, le paramètre de fréquence fondamentale continue à apparaître comme le principal fournisseur de traits intonatifs pour la distinction des intentions communicatives du locuteur. Des recherches bien connues ont, d'un autre coté, prouvé qu'un modèle théorique de l'intonation (destiné, dans une optique plus moderne, à servir de point de départ à une procédure expérimentale) peut être établi à partir d'observations auditives, portant exclusivement sur la mélodie.Dans notre cas, ces dernières ont été complétées par l'observation visuelle d'environ mille intonogrammes. Plusieurs de nos hypothèses ont été confirmées en outre par les témoignages d'auditeurs d'origine française.

Nous nous sommes posé la tâche de rechercher les mécanismes par lesquels la visée communicative contribue à la constitution du contour mélodique dans les phrases interrogatives françaises. La présente étude porte sur la large catégorie des questions "oui-non". Nous les avons choisies pour la diversité qu'elles présentent aussi bien sur le plan du contenu que sur celui de l'expression. Nous avons apprécié, en particulier, la possibilité d'étudier les manifestations de l'intonation dans des phrases soit dotées, soit dépourvues de marqueurs lexico-syntaxiques de l'interrogation.

LA VISEE COMMUNICATIVE DES ENONCES

Nous entendons sous visée communicative cette composante du contenu de l'énoncé qui lui est conférée par l'intention communicative du locuteur. Elle est conforme au but que celui-ci a poursuivi en projetant d'exercer sur l'interlocuteur une forme particulière d'influence (4,pp.156-157). L'idée que la visée communicative est une structure complexe du contenu n'est pas neuve (5). Une analyse en contexte peut révéler ses composantes. Nous présentons ci-dessous les résultats d'une telle analyse; nous l'avons réalisée dans le but de subdiviser la large classe des questions "oui-non" en catégories communicatives plus étroites. Ayant consulté plusieurs ouvrages,[1] nous nous sommes penché sur l'intention communicative de locuteurs qui ont posé des questions "oui-non" dans environ mille contextes (extraits d'oeuvres de la littérature française ou "piqués sur le vif" au cours de contacts de l'auteur avec des Français). Le tableau 1(voir à la page suivante) reflète notre démarche. L'analyse nous a permis de définir 13 critères de classification (leurs numéros figurent à l'horizontale, en haut, au milieu du tableau). Voici ces critères:

1) le locuteur pose la question dans l'intention d'obtenir une réponse (ce critère s'applique à toutes les "vraies" questions);

[1] Nous ne mentionnerons que les noms de quelques-uns de leurs auteurs: Bally,Ch.; Bliznitchenko,L.;Chéviakova,V.;Jespersen, O.;Kingdon,R.;Pankratz,G.;Pilipenko,O.;Pinaiéva,V.;Poukelis,V.;Reingand,P.;Restan, P.;Talandene,M.;Teriochkina,L.;Travkina,A.

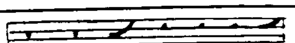
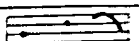
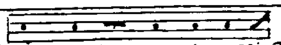
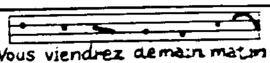
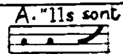
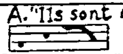
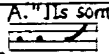
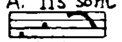
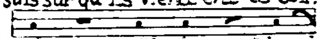
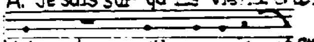
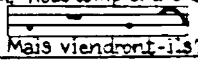
	catégories		traits	1	2	3	4	5	6	7	8	9	10	11	12	13	
I	Véritables questions "oui-non"	totales	neutres	+	+	-	-	+	-	**+**	+	-	-	-	-	-	exemple:  A."Vous viendrez demain matin?" - B."Oui.(Non)."
II			dem.de confirm.	+	+	-	+	+	-	**+**	+	-	-	-	-	-	exemple:  A."Vous viendrez (n'est-ce pas)?" - B."Oui".
III		parti-elles	neutres	+	+	+	-	-	+	**+**	+	-	-	-	-	-	exemple:  A."Vous viendrez demain matin?" - B."Oui.(Non)."
IV			dem.de confirm.	+	+	-	+	-	+	**+**	+	-	-	-	-	-	exemple:  A."Vous viendrez demain matin(n'est-ce pas)?" - B."Oui".
V	Demandes de précision		neutres	+	+	+	-	-	+	-	-	**+**	-	-	-	-	exemple: A."Ils sont rentrés dans leur pays." -  B." Par le train?" - A."Oui (Non)."
VI			dem.de confirm.	+	+	-	+	-	+	-	-	**+**	-	-	-	-	exemple: A."Ils sont rentrés dans leur pays." -  B." Par le train?" - A."Oui."
VII	Questions "écho"		neutres	+	+	+	-	+	-	-	-	-	**+**	-	-	-	exemple: A."Ils sont partis par le train." -  B."Par le train?" - A."Oui.(Non)."
VIII			dem.de confirm.	+	+	-	+	+	-	-	-	-	**+**	-	-	-	exemple: A."Ils sont partis par le train." -  B."Par le train?" - A."Oui".
IX	Q."oui-non" de deuxième instance avec mise en relief d'un élément autre que le verbe ou que le marqueur de la modalité interrogative			+	+	+	-	-	+	-	-	-	-	**+**	-	-	exemple: A."Je suis sûr qu'ils viendront ce soir." -  B."Bon, d'accord! Mais est-ce qu'ils viendront lundi?"
X	Q."oui-non" de deuxième instance avec mise en relief de l'élément notionnel de la forme verbale			+	+	+	-	-	+	-	+	-	-	-	**+**	-	exemple: A."Je suis sûr qu'ils viendront." -  B."Bon! Mais est-ce qu'ils repartiront?"
XI	Questions"oui-non" insistantes			+	+	+	-	+	-	-	-	-	-	-	-	**+**	exemple: A."Nous comprendrons quand ils seront venus." -  B."Bon! Mais viendront-ils?"

tableau 1

2) le locuteur est informé à un degré qui lui permet de formuler une supposition; l'interlocuteur devra la confirmer ou la nier;

3) le locuteur est informé à un degré qui lui permet de considérer la réponse "oui" et la réponse "non" comme également probables;

4) le locuteur est informé à un degré qui lui permet de considérer l'une des deux réponses opposées comme nettement plus probable que l'autre;

5) ne disposant d'aucune information certaine, le locuteur fait porter l'interrogation sur l'ensemble de l'énoncé interrogatif;

6) le locuteur fait porter l'interrogation seulement sur certains éléments de l'énoncé, disposant pour les autres éléments d'une information qu'il juge suffisante;

7) le locuteur interroge sans se référer obligatoirement et d'une façon spécifique au contexte précédent;

8) le locuteur introduit dans le dialogue un nouveau sujet de conversation;

9) le locuteur demande des précisions en se référant à un fait énoncé précédemment;

10) reprenant partiellement ou dans leur ensemble les paroles de l'interlocuteur, l'auteur de la question se fait confirmer qu'il a correctement compris ces paroles;

11) le locuteur signifie avec insistance qu'il s'intéresse à un détail précis concernant l'action qui fait l'objet de la conversation - il oppose l'objet de son attention à tout le reste, qu'il signale comme dénué d'intérêt;

12) le locuteur signifie avec insistance qu'il s'intéresse à une action autre que celle qui fait l'objet de la conversation;

13) le locuteur signifie avec insistance son désir d'obtenir une information digne de foi (soit qu'il la considère comme particulièrement importante, soit qu'il éprouve le doute).

Le tableau 1 présente de haut en bas les 11 catégories de questions "oui-non" dont notre corpus de contextes atteste l'existence. Les résultats de l'analyse de leur visée communicative sont présentés sous la forme d'une matrice de traits distinctifs. Les(+)indiquent, de gauche à droite, les critères de classification, valables pour chaque catégorie. En considérant chaque critère comme une composante simple de la visée communicative (comme une signification communicative), on obtient une description analytique de cette visée pour chaque classe. Une signification est considérée chaque fois comme la plus caractéristique pour une classe donnée; elle est signalée par un caractère gras (**+**).

LE ROLE DE LA PERSPECTIVE FONCTIONNELLE DE LA PHRASE

Selon notre point de vue, le contour mélodique d'un énoncé est constitué d'un certain nombre de proéminences plus ou moins importantes; dans celles-ci se réalisent des tons à configuration différente qui remplissent différentes fonctions.Les tons des syllabes inaccentuées relient les tons proéminents; suivant la langue et le fragment de la chaîne parlée, ils fonctionnent ou non comme des éléments linguistiquement pertinents. Il en découle que les rapports entre la visée communicative de l'énoncé et la forme de son contour mélodique ne sont pas directs. Nos observations nous ont mené à la conclusion que les mécanismes de constitution de ce contour doivent se décrire en deux temps: il faut rendre compte, dans un premier temps, des mécanismes grâce auxquels un type particulier de visée communicative conditionne une distribution caractéristique des proéminences le long de la chaîne de l'énoncé; dans un deuxième temps, on décrira les mécanismes qui permettent à la visée communicative de conditionner l'apparition dans des proéminences déterminées de tons à configuration caractéristique.

La théorie de la perspective fonctionnelle de la phrase, exposée dans les ouvrages de J. Firbas(6),(7),(8) nous proposait un modèle élaboré de la distribution des proéminences prosodiques dans l'énoncé. Voici les principales raisons qui nous ont décidé à l'appliquer en tant que méthode à la première étape de notre analyse intonative: a) J. Firbas décrit un modèle de distribution du dynamisme communicatif à plusieurs degrés - modèle perfectionné qui, d'après nous, promettait de rendre plus fidèlement compte de la nature des phénomènes; b) il décrit les rapports qui relient dans l'énoncé la distribution du dynamisme communicatif à celle du poids prosodique,nous fournissant de la sorte le "pourquoi" de la distribution des proéminences;c)il fait des observations sur la perspective fonctionnelle des phrases interrogatives qui nous ont été extrêmement profitables.

La première étape de notre analyse intonative se fonde sur l'hypothèse suivante: la visée communicative d'une question "oui-non" dépend étroitement de la situation (extra-linguistique) et du contexte (linguistique,dialogique) précis dans lesquels se concrétise l'intention communicative du locuteur; d'autre part, la même situation et le même contexte exercent des contraintes sur la perspective fonctionnelle de l'énoncé interrogatif et de là, sur la distribution des proéminences prosodiques. Ce qui permettait d'espérer que l'existence probable de rapports étroits entre le type de visée communicative d'un énoncé et la distribution caractéristique des proéminences prosodiques dans celui-ci peut s'expliquer, compte tenu de la structure accentuelle de la langue donnée, par la théorie de J. Firbas.

CONCLUSIONS

Faute de place, nous n'exposerons que nos conclusions générales.

I. La première étape de notre analyse a prouvé que le modèle de Firbas est applicable au français. Les proéminences prosodiques ont été dans la plupart des cas produites par nos informateurs français et perçues par nos auditeurs au cours des tests en ces points de la chaîne parlée où l'analyse théorique préalable permettait de les prévoir. (Cette analyse partait chaque fois de la visée communicative de l'énoncé, étudié dans son contexte, pour déterminer la distribution du dynamisme communicatif parmi ses éléments et prévoir la distribution appropriée des proéminences prosodiques).

L'importance relative des proéminences indiquée par les auditeurs n'était pas toujours celle que la théorie prévoyait (nous examinerons ce problème ailleurs). Firbas montre pour l'anglais que la structure prosodique ne signale pas dans tous ses détails la distribution des degrés de dynamisme communicatif parmi les éléments de l'énoncé. Ceci s'avère encore plus vrai pour le français, langue dépourvue d'accent de mot, qui opère au niveau prosodique avec des fragments relativement plus étendus de la chaîne parlée.

Notons en conclusion que la première partie de notre analyse semble porter appui à l'affirmation de Firbas que les mécanismes qu'il a découverts sont valables au moins pour l'ensemble des langues indo-européennes.

II. Nous accordons une attention plus grande à la deuxième partie de notre analyse. Alors que la distribution des proéminences prosodiques manifeste des tendances interlinguistiques bien prononcées, la configuration des tons qui se réalisent dans ces proéminences semble varier considérablement d'une langue à l'autre. Il ne nous semble pas possible de pouvoir affirmer avant vérification que ces tons apparaissent dans la chaîne parlée des différentes langues sous l'action de mécanismes semblables. Les mécanismes dont nous parlons ci-dessous sont valables pour les questions "oui-non" en français.

D'après nous, la configuration du ton interrogatif pertinent, utilisé dans chaque catégorie communicative de ces questions dépend étroitement du degré maximal de dynamisme communicatif que peuvent atteindre les énoncés de cette catégorie.

L'analyse de la perspective fonctionnelle nous a mené à la conclusion que selon le degré de dynamisme communicatif (DC) qui leur est propre, les onze catégories étudiées ici peuvent se répartir en trois groupes:

- le premier englobe les questions "oui-non" à degré de DC relativement bas,dites "neutres" - celles-ci ne dépassent en degré de DC les énoncés assertifs que parce qu'elles expriment une incitation, adressée à l'interlocuteur, à confirmer ou à nier une supposition (il s'agit des catégories I,III,V,VII de notre classification); (nous laissons de côté, comme peu importante à cette étape de l'étude, la question de savoir si les classes mentionnées ci-dessus se distinguent les unes des autres par le degré maximal de DC que les énoncés qui les composent atteignent).
- le deuxième groupe comprend les questions dites "demandes de confirmation" dont le degré maximal de DC peut être qualifié de "moyen" dans le système étudié - dans ce cas, le sujet parlant ne fait pas qu'inciter l'interlocuteur à fournir une réponse - il lui suggère aussi, avec une insistance plus ou moins grande, la réponse qu'il s'attend à recevoir (catégories II,IV,VI et VIII de notre classification);
- le plus haut degré de DC est atteint, pour des raisons bien connues, dans les questions "oui-non" de deuxième instance (catégories IX,X et XI de notre classification).

Les tests auditifs et l'examen des intonogrammes nous ont permis de conclure, d'autre part, que dans les questions "oui-non" en français se réalisent deux mélodèmes "interrogatifs": 1)le ton ascendant [/](doublé dans certains types d'énoncés par le ton [¯](l'"écho" de P.Delattre)) et le ton ascendant-descendant [⌒].

Le ton ascendant est un trait distinctif des questions neutres à bas degré de DC (catégories I,III,V et VII);la coïncidence de ce ton soit avec la syllabe accentuée de la forme verbale (catégorie 1), soit avec la syllabe finale (catégorie 2) est un autre trait, qui assure la distinction des véritables questions neutres totales de leurs homologues partiels.

Le ton ascendant-descendant est caractéristique pour les questions "oui-non" à haut degré de DC (les demandes de confirmation et les questions de deuxième instance). L'analyse intonative portant uniquement sur la mélodie ne nous révèle donc pas de traits intonatifs, permettant de distinguer ces deux catégories. On remarque cependant que les questions de deuxième instance (qui figurent parmi les questions neutres - tableau 1) contiennent obligatoirement un marqueur lexico-syntaxique de la question neutre-"est-ce que" ou l'inversion du pronom-sujet,alors que dans les questions neutres à bas degré de DC (c.I,III,V,VII) ces marqueurs sont facultatifs et le second se fait rare en français moderne. La combinaison obligatoire du ton [⌒] avec "est-ce que" ou l'inversion du pronom-sujet est donc un procédé qui permet d'éviter l'interprétation des questions "oui-non" de deuxième instance comme des demandes de confirmation (dans ces dernières, "est-ce que" et l'inversion du sujet, porteurs de la signification de q. neutre, n'apparaissent jamais).

Si l'on se penche maintenant sur l'interaction des traits intonatifs et lexico-syntaxiques dans les questions "oui-non" en français, on remarquera dans les différentes classes le fonctionnement de mécanismes différents:

- dans les questions neutres à bas degré de DC se manifeste le "principe de substitution" de A.M.Pechkovskij (9); ainsi, la présence dans la question de "est-ce que" rend le ton [/], porteur de la même signification, redondant - raison pour laquelle ce ton n'est dans ce cas utilisé que facultativement dans la dernière syllabe;
- le propre des questions des catégories IX,X et XI - le contraste de deuxième instance - est signalé toujours par le fonctionnement conjoint de deux traits formels de différente nature ("est-ce que" + le ton [⌒] ou l'inversion du pronom-sujet + ce même ton); ces deux traits, pris à part, expriment des significations différentes, si bien qu'utilisés dans un même énoncé, ils n'entrent pas en rapport de substitution (l'un ne s'efface jamais au profit de l'autre); ils fonctionnent en solidarité,comme les deux composantes obligatoires d'un signifiant complexe - c'est ce que nous avons appelé "le principe de solidarité".

BIBLIOGRAPHIE

(1) А.Мишева, "Интон. система на съвр.бълг. книж. език",докт. дисертация, София, 1987.
(2) A.Andrievskaïa, "Syntaxe du français moderne,Киев,1973.

(3) В.Гак, "Теоретическая грамматика французского языка. Синтаксис".Москва, 1981.
(4) И.Зимняя, "Психология слушания и говорения", докт. диссертация, Москва, 1973.
(5) В.Артемов, "Психология обучения иностранным языкам, Москва, 1969.
(6) J.Firbas, "On the prosodic features of the modern English finite verb as means of FSP", Brno studies in English,7,Brno,1968.
(7) J.Firbas, "On the interplay of prosod.and n.-pros. means of FSP", The Prague Sch. of Ling.and lang.teach.,Fried,London,1972.
(8) J.Firbas,"A study in the func.persp.of the Engl.and the Slav.interr.sent.", Brno studies in English,12,Brno,1976.
(9) А.Пешковский, "Интонация и грамматика", Избранные труды, Москва, 1959.

THE ROLE OF SENTENCE INTONATION IN SEMANTIC INTERPRETATION OF NORWEGIAN NEGATIVE DECLARATIVES

Thorstein Fretheim

University of Trondheim

ABSTRACT

It is demonstrated how a particular model of intonation-syntax interaction will account for the ways in which Norwegian sentence intonation affects one's understanding of the relative semantic scopes of negator and quantifier/adverb in negative sentences with straight and inverted word order. The central prosodic unit referred to in this study is the Intonational Phrase, which is an immediate constituent of the largest intonational unit.

STATING THE PROBLEM

The data with which I am concerned are spoken utterances of Norwegian sentences containing two semantic operators, one of which is the negation marker ikke (or enclitic 'ke) and the other one a quantifier or a time or frequency adverb. Cf. English.

(1) It didn't happen often.
(= It happened seldom)

(2) We didn't find many.
(= We found few)

If often and many are preposed, the scope relations are reversed. The operator to the left takes priority over the one to the right. While often and many are NEG-INTERNAL (i.e. inside the scope of n't) in (1) and (2), they are NEG-EXTERNAL in (1')and (2').

(1') Often it didn't happen.
(2') Many we didn't find.

Substituting a time adverb like yesterday for the frequency adverb of (1)-(1'), it is possible to get a NEG-external interpretation of the adverb even in (3) where the negator precedes it. (This is hardly possible with often in (1) or many in (2).)

(3) He didn't come yesterday.
(3') Yesterday he didn't come.

A falling nuclear tone on come followed by a low rise on yesterday favours an interpretation of (3) that makes it synonymous with (3'). All other intonation patterns communicate that the adverb is supposed to be in the scope of not.
The corresponding types of scope assignment in spoken Norwegian are mental tasks that rely rather more on the employment, and recognition of intonational devices. An adverb/quantifier may be NEG-internal even if it is placed to the left of the negator in the linear syntactic string of words. Conversely, an adverbial operator may be NEG-external even if it is located to the right of the negator, provided a specific intonation structure is assigned to the sentence. The rules according to which a Norwegian adverb/quantifier is understood to be NEG-internal or NEG-external will have to refer to properties of intonational as well as syntactic form.
In our discussion of relative semantic scope determined by the interaction of word order and the intonation structure of utterances, we shall refer to the following three pairs of sentences.

(4) Han kom ikke i går.
(He didn't come yesterday)

(4') I går kom han ikke.
(Yesterday he didn't come)

(5) Det skjer ikke ofte.
(It doesn't happen often)

(5') Ofte skjer det ikke.
(Often it doesn't happen)

(6) Vi fant ikke mange.
(We didn't find many)

(6') Mange fant vi ikke.
(Many we didn't find)

The intonation structures that we are going to impose on these syntactic structures all share certain important features. They all contain one very prominent rising pitch accent movement at a fairly early point in the utterance.

THE INTONATION MODEL

Cruttenden, in his textbook on intonation [1], distinguishes between intonation languages, pitch accent languages, and tone languages. He classifies Norwegian and Swedish as 'predominantly intonational languages' in which 'a limited number of words are distinguished by tone alone'.
His remarks on Norwegian and Swedish prosody

are unfortunately marred by his expressed belief that one of the two word accents, the so-called Accent 1, is somehow 'the common accentual pattern', and that Accent 2 has a much more limited range of occurrence than Accent 1. Cruttenden shows a lack of appreciation of the fact that any assignment of pitch accent to a word form in a Norwegian or Swedish utterance entails the use of one or the other of the two opposing word accents. The pitch contours determined by the word accents are always present in spoken signals, and they are phonologically distinctive quite independently of their differentiating morphological function.

The actual number of minimal pairs whose members are distinguished solely by word accent - say a limited number like five hundred, or a larger number like five or ten thousand - is quite irrelevant if the issue is whether Norwegian and Swedish are tone, pitch accent, or intonation languages. It has no bearing on the structural relations between pitch profiles determined by word accent and the pitch profiles that make up the global intonation patterns of utterances.

In East Norwegian, on which the present study is based, there is no neutralisation of the word accent distinction in any environment. The two paradigmatically opposed pitch profiles can actually be said to shape the various sentence intonation patterns of East Norwegian to a large extent. The word accent dichotomy is an invariant phonological feature of accented words in actual utterances. Not even tonal 'perturbations' caused by the global intonation structure can ever modify the fixed pitch accent contours of Accent 1 and Accent 2 for any specific linguistic purpose.

In the West and the North of Norway, where accented syllables are associated with high tone (Accent 1) or a rise to high tone (Accent 2), there is definitely a distinction between a rising and a falling NUCLEAR tone, but in the East (including the capital Oslo), where accented syllables are low-pitched (Accent 1) or gliding down to a F0 minimum point (Accent 2), there is no nuclear tone in the proper sense of the term, and rising vs. falling intonation only plays a subsidiary role, in a small subsection of the intonation system. It is fair to describe the essence of the East Norwegian intonation system as being encapsuled in a specific structural property of the prosodic FOOT unit (i.e. the stretch of syllables from one accented syllable up to, but not including, the next accented syllable of the segmental chain). Any foot is assigned either the plus or the minus value of the binary tonal feature of [±raised peak] (the term 'raised peak' being due to Ladd [2]). Morpholexically a foot encompasses a whole word, just part of a word, or a sequence of words. Phonetically it is mono- or polysyllabic. No matter how extensive or how short a foot is, it consists of two parts, one in which the intonational distinction between [-raised peak] and [+raised peak] is realised and one in which word accents occur.

Let us refer to a [+raised peak] foot as a FOCAL foot, and a [-raised peak] foot with only a moderate, or even nonexistent end-peak as a NONFOCAL foot. The foot contours of Figure 1a and b display Accent 1 melodies before the dotted vertical line, and focal and nonfocal accent, respectively, after the dotted line. Figure 2a and b show the corresponding Accent 2 patterns.

Figure 1a Accent 1 Figure 1b

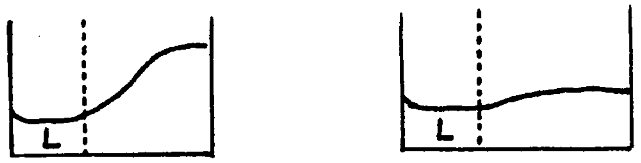
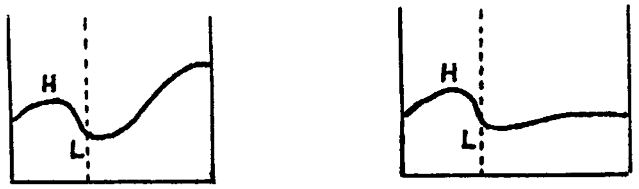


Figure 2a Accent 2 Figure 2b



That part of the F0 curve that appears to the left of the dotted line is intonationally irrelevant, and the part that appears to the right of the dotted line is word-prosodically irrelevant.

Observe that scholars like Selkirk [3] and Nespor & Vogel [4] use the term 'foot' with a meaning that differs from the meaning attributed to Norwegian feet in the present study. 'Clitic groups' necessarily contain more than one syllable. My foot contains n syllables (with the monosyllable as the minimum foot), and, for that matter, an indefinite number of unaccented words after the accented word, which is an obligatory element of the foot. For me, a given clitic group is either equal to, or smaller than the foot in which it appears. For the above-mentioned authors, however, a foot is a unit below, and the clitic group a unit above the word level.

In Gårding's model of intonation developed principally for Swedish [5], overall sentence intonation patterns are generated independently of the local highs and lows of the Swedish word accent melodies.

Matsunaga [6] considers it necessary to separate accent from intonation in Japanese, a pitch accent language. Accent and intonation are treated as independent cooccurring prosodic systems. I am arguing that East Norwegian (word) accent and intonation should in fact not be viewed as mutually independent. The intonational distinction between [+raised peak] and [-raised peak] in East Norwegian involves the presence vs. absence of a LH pitch movement in the latter part of a foot. These are features of the sound wave which are clearly dictated by the fact that you go down in pitch when you move from a foot F_i to F_{i+1}, that is, when you produce the East Norwegian word accent in the initial part of F_{i+1}. In West and North Norwegian where the word accents are associated with high pitch, [±raised peak] coincides with the word accent realisation in the initial part of the foot. Hence, where East Norwegian has intonationally significant pitch movements up to a raised peak, West and North Norwegian have intonationally significant falls from a raised peak.

The exact length of a Norwegian foot can be ascertained quite easily due to the two contrasting word accent melodies appearing early in the foot. Because the word accents are so easy to perceive for the (native) listener, the word accent melodies in foot contours, which are associated primarily with the accented syllable heading the foot, have an important function apart from the lexical one. They also function as juncture markers for prosodic feet, and the tonal structure of the foot is of paramount importance for Norwegian sentence intonation. A raised foot-final peak is an important juncture marker, too. The turning-point where the pitch starts to drop from a focal peak marks the boundary between two INTONATIONAL PHRASES (IPs). In East Norwegian intonation there is a phonetic difference between PRE-focal and POST-focal in the category of nonfocal. Pre-focal feet permit a mild tonal upglide after the F0 minimum point in the foot. Post-focal feet are generally quite even and low in pitch after the F0 minima, and successive feet after a focal IP boundary usually exhibit a marked F0 declination, both through minima and maxima. The systematic phonetic difference between pre-focality and post-focality provides evidence that nonfocal feet before and after a particular raised, focal peak belong to the same larger intonation pattern. Since pre-focal and post-focal foot profiles are in complementary distribution, they are both phonologically nonfocal, but the auditory difference between those two types of nonfocal foot may ease the listener's identification of the locus of a focal peak representing the end of an IP.

I shall refer to the intonational category above the IP as the INTONATIONAL UTTERANCE (IU). An IU can contain from one to three IPs but only two focal feet. The final IP in an IU made up of three IPs may not include a focal foot at the end, and when the IU consists of two IPs, the later one may or may not terminate in a focal foot.

INTONATIONAL PHRASES AS INFORMATION UNITS

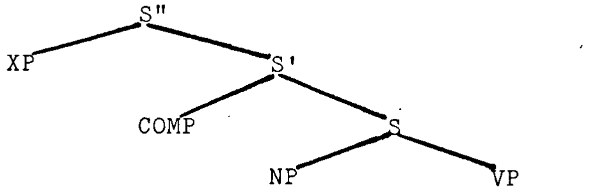
Differences in intonational phrasing bear directly on the information structure of utterances. For any IP ending in a foot specified as [+raised peak] there is a corresponding FOCUS DOMAIN in the surface-syntactic representation of the sentence. Terminal symbols of surface phrase-markers are enriched with the feature [+raised peak] if they head a focal foot in intonation structure. (Cf. Selkirk's concept of 'intonated surface structure'.)

I shall propose the following operational definition of the concept of 'focus domain':

> A focus domain is the highest syntactic node up in the syntactic tree from a given instance of [+raised peak] assigned to a terminal symbol, which
> i) dominates no other instance of [+raised peak],
> and
> ii) dominates no symbol to the right of (i.e. temporally succeeding) the symbol specified as [+raised peak].

There are certain weaknesses pertaining to this definition but it will work in the context of the present study.

I assume that Scandinavian main clauses should be represented syntactically by a phrase-marker in which there is an XP node to the left and an S' node to the right:



In declaratives, the XP position is filled by a subject (coming from NP under S) or a nonsubject (coming from somewhere within VP under S). COMP is filled by the finite verb of the sentence.

I also assume that there are certain FOCUS INTERPRETATION rules applying to focus domains of the surface-syntactic structure. One rule says, if there are two focus domains both of which are part of S', then the first one is a RHEME and the last one a THEME. On the other hand, if XP is one focus domain and the other one is S' or part of S', then the former is a theme and the latter a rheme.

Thus the syntactic focus domains (henceforth FDs) associated with IPs of IUs are considered to be the smallest information units in a discourse. FDs comprising the whole S" exemplify BROAD FOCUS as described by Ladd [7] and others. In Norwegian the contrary phenomenon of NARROW FOCUS is a result of splitting the IU into two or three IPs by assigning [+raised peak] to prosodic feet that are not utterance-final.

NEG-INTERNAL AND NEG-EXTERNAL THEMES

Let us return to the main topic of this paper, and to the sentence pairs of (4)-(4'), (5)-(5') and (6)-(6').

If you assign focal accent to the pronominal subject of (4) - Han kom ikke i går -

you have thematised the subject, intonationally as well as syntactically, and a later focal accent will coincide with the rhematic syntactic element sometimes referred to as the information focus of the sentence or the item with the 'highest degree of communicative dynamism (CD)' in Firbas' sense. The following three distinct IUs contain the same number of IPs and display the same basic type of theme-rheme structure. (I am using a self-explanatory labelled bracketing notation where focally accented words are written in capitals.)

```
(7)  ( ( (HÁN-kom) ) ( (ÌKKE-i) ) ( (gǻr)))
     IU IP F           IP F          IP  F

(8)  ( ( (HÁN) ) ( (KÓM-ikke-i-gǻr) ) )
     IU IP F       IP F

(9)  ( ( (HÁN) ) ( (kóm-ikke-i) (GǺR) ) )
     IU IP F       IP F            F
```

Though all three versions are negations of the proposition ⌜he came yesterday⌝, they do not answer the same questions. The underlined words constitute syntactic phrases which are separate focus domains according to the definition of focus domain offered above. (7) is an example of 'phrasal negation', which means that there is a positive CONVERSATIONAL IMPLICATURE, namely There were others who did come yesterday, attached to the negative statement, an implicature that is lacking in (8)/(9). In (7) the scope of the negator ikke includes the thematic FD in the XP position. Similar things can be achieved in English by means of one tonal nucleus placed on he and another on not: HE did NOT come yesterday. Ikke is focally accented in (7) in order to underscore the negative polarity of the statement. The denial of the proposition ⌜he came yesterday⌝ is the only directly conveyed new information in (7). In (8) the finite verb, kom, carries focal accent for a similar reason. Here the rhematic FD dominated by S' is the COMP node. That FD includes the negator, because any occurrence of unaccented ikke is verb-enclitic, even if it retains its segmentally full form. The verb gets focal accent in (8) for the same reason that the negator got it in (7). It is made accentually prominent for modal reasons, and there may be no 'contrastive stress' involved here. There is, however, one interesting functional difference between assigning POLARITY FOCUS [8] to ikke and assigning it to the verb. Focal accent placed on ikke links the negator with some other FD in the sentence in such a way that we readily interpret the syntactic material of that FD as a phrase which is inside the scope of ikke - a NEG-INTERNAL phrase, and in our example (7), a NEG-internal theme, the subject han. When it is the verb that carries focal accent in order to highlight the (negative) polarity of the sentence, we tend to understand the semantic scope relations differently. Now the scope of ikke is generally taken to cover no syntactic items to the left of the negative operator, and the subject han of (8) is therefore a NEG-EXTERNAL theme.

Suppose we retain the polarity focus on the verb/negator in sentence (4) but assign focal accent to the AdvP i går ('yesterday') instead of the subject. The actual denial will again be the new information conveyed in the speech act, and focal accent on ikke may still be felt to connect the negator more closely to the succeeding thematic FD i går than in the alternative version where kom gets the focal accent. And indeed, it is possible to interpret the FD i går as either a NEG-internal or a NEG-external theme if ikke is the focussed polarity item, but only the NEG-internal interpretation is possible if the focal accent is on kom. Preposing the AdvP, as in (4'), we find a potential meaning difference between letting polarity focus be carried by the verb or by the negator, but this time (10) is the ambiguous structure and (11) the unambiguous form which only admits a NEG-external interpretation of the phrase i går.

```
(10)  ( (i (GǺR-kom-han) ) ( (ÌKKE) ) )
      IU IP  F              IP  F

(11)  ( (i (GǺR) ) ( (KÓM-han-ikke) ) )
      IU IP F        IP F
```

The pair (5)-(5') contains a frequency Adv - ofte - where (4)-(4') had a time Adv. The only difference between (5) and (4) is that with a frequency Adv it is impossible to get a NEG-external interpretation of a sentence-final theme even if there is focal accent on the verb. (5') differs markedly from the English sentence Often it does not happen, where often is NEG-external regardless of the intonation employed. The theme ofte in (12) is ambiguously NEG-external or NEG-internal, depending on the context.

```
(12)  ( ( (ÒFTE-skjer-det) ) ( (ÌKKE) ) )
      IU IP F                  IP  F
```

If the accent is shifted from ikke to skjer, ofte is outside the scope of negation, as in the English translation.

Our decision to distinguish NEG-external from NEG-internal themes is strongly supported by data like (6)-(6') where a quantifier - mange ('many') - interacts with the negator. When mange is in the XP position and receives focal accent there is a clear-cut semantic difference between the negator-in-focus version, which means either Many we didn't find or We found few (=not many), and the verb-in-focus version, in which mange only has the former reading with a NEG-external theme.

REFERENCES

[1] A.Cruttenden,"Intonation", CUP, 1986.
[2] D.R.Ladd,"Phonological features of int. peaks",1985.
[3] E.Selkirk,"Phonology and Syntax",MIT,1984.
[4] M.Nespor&I.Vogel,"Prosodic Phonology",1986.
[5] E.Gårding,"Contrastive Prosody", SL,1981.
[6] T.Matsunaga,"Prosodic Systems of Jap."1984.
[7] D.R.Ladd,"The Structure of Int. Meaning" IUP, 1980.
[8] C.Gussenhoven,"On the Grammar and Semantics of Sentence Accents",Foris, 1984.

TYPES OF SEMANTIC RELATIONS
BETWEEN INTONATION AND LEXICO-GRAMMATICAL MEANS (LGM) OF LANGUAGE

V.I. PETRYANKINA

MOSCOW PEOPLES' FRIENDSHIP UNIVERSITY
DEPARTMENT OF GENERAL LINGUISTICS

ABSTRACT

The report generalizes from the results of a long-term investigation carried on by the author. This research is based on the material of languages belonging to different morphologo-syntactical types. The general problem is outlined as "The interaction of intonational and lexico-grammatical means of language", the particular one being "The establishment of the types of semantic relations between intonation and lexico-grammatical means" with regard to the character of distribution, opposition and the amount of semantic meaning. The relations holding between intonation and lexico-grammatical means are typified oriented on "deep" interpretation of intonational facts their presentation at the abstract level irrespectivs of details. These types of semantic relations are regarded as language universals.
The analysis of the types of semantic relations such as semantic harmony, sameness, inclusion, overlapping, exclusion is designed to disclose purely intonational semantics.
Theoretical assumptions are based on the experimental data obtained by means of auditory and electro-acoustic analyses and processed with the computer.

Different types of semantic relationships holding between intonation and LGM can be established taking into consideration the character and amount of semantic meaning, the character of opposition and distribution, their interrelation and interdependence.
Intohational and LGM in the flow of speech may stand out as semantically one-directional possessing in their meanings some common semantic elements and multidirectional, those not having common semantic elements. The former are classified as being comparable presented in all their variaties depending on whether they coincide or do not coincide in meanings, the latter - as being incomparable.
This opposition comprises intonation and LGM as two objects knit together in such a way as their general meaning can not represent one object without representing the other, that is to say, the members of the opposition form a single unity. The analysis of the semantic features of objects under comparison discloses their different oppositions, being realized in the flow of speech.
The following types of semantic relation between intonation and LGM seem to be most essential: semantic "harmony", the relations of sameness, inclusion, overlapping, exclusion.

The relation of "semantic harmony" between intonation and LGM. I. "Semantic harmony" is relevant for semantically one-directional relations between intonation and LGM when all the means merge entering into an integral unity and without diminishing functional value of each other make up a single semantic whole. Intionational features being constant, and to a lesser extent dependent (or independent) on the context.
This type of relationship between intonation and LGM is clearly traced in the utterances, formed on the basis of sentences of different typical meaning, and, hence, different semantico-syntactic structure (1).
Observing intonation with reference to its "harmony" with LGM resulted in what has proved to be essential - a syntactic hierarchy of classes of sentences being different as to the oorrelation with different types of the process of thinking.
As it is known, the whole domain of what can be manifested in sentences is divided into two different classes: the class of thoughts and the class of facts and events. "Thoughts are the products of rational, analysing and generalising activity" (2; 320). Facts are all that is becoming the property of our consciousness through immediate observation and perception, "sensual perception of reality at the speech moment" (3; 144). Accordingly distinction is made between two kinds of sentence (4; 139 и др.): perceptive (Идет дождь. Меня знобит.) and logically mediat-

ed (Наташа - доктор. Волга - река).
Assuming the heterogeneous level system of language means used to express cognitive power of consciousness the hypothesis is adduced that the forms of thinking, on the one hand, at the emperical level of cognition, and, on the other hand, at the level of logical analysis and theoretical generalisations, having different linguistic forms are differentiated intonationally which is conditioned, in first turn, by the nature of accentuation. The sentences containing the verbal result of mental operations on the signifier denoting in particular notions, laws, categories, and also such mental operations as comparison,. contrasting, logical classification etc., are shaped by means of two-accented intonation, each of the predicative components - subject and predicate are said to be accentually marked.

Accentually-articulated intonation of these utterances is opposed to unarticulated one-accented intonation represented by sentences of the perceptive type verbally reflecting phenomena, facts, features, links of denotations being observed in real world.

One-accented/two-accented intonation in addition to LGM is a formal indicator of such interrelated language features setting off sentences to different categories as temporal realization of predicativeness, namely, localization/non-localization in time, attachment/non-attachment to a speech moment and also abstractness/concreteness of an action, state or event. Cf.: Мне больно - a concrete single state of mind of a person, experienced at the moment of speaking; Чайка - птица - a constant feature, abstracted from a definite place or time.

Thus intonation being one-accented or two-accented is one of the intonational features which is in full accord with LGM of the classes of sentences differing as to character of reflecting phenomena in the real world and its correlation with different types of the process of thinking.

2. In case of intonation formed by means of accentuation there seem essential two closely connected features of intonation, i.e. the degree of prominence regarding accents and the presence/absence of a pause between accents which correlates with the semantico-syntactic structure of the utterance concerned (in its predicative minimum and out of context).

In sentences of the perceptive type the absence of a carrier of the predicative feature - (Холодно), its diminished semantic significance (it is not an agent) - (Меня знобит), inability to stand out as an expander of the word-predicative at the intonational level correlate with the absence of accentual prominence and a greater semantic significance of the predicative component correlates with the perception of it as a semantic centre and a syllable bearing the sentence stress (SS) but with a greater degree of expresiveness than the neutral SS. The types of accentuation: S P[1] (Меня бьет дрожь); ( ) P (Люблю тебя. Холодно). These utterances are likely to be treated as contexually independent, global, bearing unarticulated notion which is reflected in accentual-intonational unarticulateness and in the absense of a pause between the components.

In logically mediated syntactically two-member sentences the degree of expressiveness of accents and connected with it their articulateness/anarticulateness on syntagms depends on semantic significance of each of the components and language means designed to render them.

The striking contrast can be found with two-member noun patterns, having the meaning of the qualifier of the subject (in the type: Волга - река), where the presence of both the componente being related to each other as parts and the whole is obligatory. Intonational articulateness on syntagms correlates with the absence of morphological and syntactical links between the components, independent status of the subject, its inability at the syntactic level to be a subjugated enlargement on the predicative, the pause (at a zero juncture) is intensifying contrast between the components. The type of accentuation is as follows: S ‖ P.

In articulated syntactically two-member utterances the components being closely connected there exists lexical, morphological and syntactic concordance. The subject is likely to be the expander of the predicative and there is no pause between the components. The type of accentuation is: S P. It holds true, for instance, for the utterances based on sentences with isosemic patterns having the typical meaning of an action characterizing the subject (Люди работают), the property of the subject (Вода замерзает) and the like.

Thus, different degree of accentual prominence of the components and the presence/absence of a pause between them seem to be the manifestation of semantic harmony between intonation and lexico-grammatical structure of the sentence the reflection of close links between the components depending upon the semantic value of words expressing them.

3. We can also see differences in accentual-intonational structure of utterances formed on the basis of sentences in their isosemic/non-isosemic patterns in which nouns in conformity with their prepositional-case forms and categorical-semantic content stand out as a typical nomination means regarding the features of an object (in a set with adjectives, numerale and adverbs): Джинсы - в заплатках.

Two-accented intonation of these utterances with the syntagmatic division of the type: S // P, is distinguished from two-accented intonation without the syntagmatic division - S P solely by the presence of a pause and from two-accented, two-member nominal sentences having the meaning of the qualifier of the subject S // P by a smaller degree of accentual expressiveness of the predicative components.

Such are the cases with a division on syntagms in predicatively connected patterns consisting of two noun-forms excluding a notional verb from the structure (Наташа - из дома моделей), in non-verbal patterns admitting formal verbs which do not render any information but some stylistic colouring and "omitted" due to redundancy (Театр - на площади. Театр находится на площади).

Due to the fact that the combination of word-forms in these cases is sufficient to form predicative minimum intonation in itself does not render any relevant information and is in semantic harmony with LGM forming the general semantic essence of the utterance.

The relation of sameness between intonation and LGM. This type of relation is relevant for the cases when intonational and LGM are semantically equivalent to each other the two entities having the same meanings and being equivalent in their distribution.

Thus, a rising intonation used to shape predicatives in interrogative sentences and LGM, namely, the expressing of the predicative by an interrogative pronoun or an adverb are semantically equivalent, i.e. each of them is likely to convey the meaning of a question.

Full substitution of intonation and LGM in a context without any detriment to its sense takes place, for instance, in those languages in which the meaning of the communicative design of the utterance is rendered either by means of intonation or LGM, i.e. there formed a zero opposition the members of which are equivalent in distribution (for instance, in the language of Bamana the particle "wa" may replace a rising intonation in any position of the text). In the Russian context which comprises information constituting a great amount of the speakers' knowledge the meaning of a question has been minimized to the that of a word represented by an interrogative pronoun or an adverb, intonation of a question does not work and the "compensatory law" is operative. Nevertheless, due to a specific character of categorical-semantic meaning of words-predicates interacting with intonation complete semantic subsistution does not occur (the semantics of pronominal or non-pronominal questions is different). In case of the least dependence on the context and the absence of contexually formed knowledge and, accordingly, a great amount of information required pronominal and non-pronominal questions are characterized by identical intonational forms of semantically meaningful parts of a text. Therefore, intonation and LGM as the opposites are not equivalent as far as the amount of meaning and distribution are concerned.

The relation of inclusion. This type of relation is commonly found in cases when intonation and LGM are semantically one-directional and the meaning of one of the components represented as a carrier of an additional semantic feature is contained in a wider scope of meaning created by the other component. Those may be, for instance, the relations between the intonation of a question with the meaning of problematic or categoric reliability of epistemic modality and formal means for expressing modal meanings (particles, paranthetic and modal words, grading a degree of certainty on the part of the speaker in the truth of the utterance: hardly possible assumption, hesitant assumption, assumption with doubt in plausibility of the fact required etc.), thus, introducing a degree of assumption into the suppositional meaning.

There formed the so-called preventive opposition based on the presence/absence of an additional semantic feature where the LSM are presented as a marked member of the opposition which is richer in semantic features (it includes the meaning of supposition plus an additional meaning - the degree of supposition (or assumption). But because of a narrower amount of meaning in the given marked member of the preventive opposition it is more restricted in terms of distribution and on this ground any guestion formally indicating the degree of reliability is likely to be replaced by the question the modal meaning of which is rendered solely by means of intonation (to add that intonation does not depend upon the presence/absence of LGM), that is to say, the distribution of the unmarked member of the opposition includes in it the distribution of the marked one. Opposition with included distribution reflects the relation of compatibility of the amounts of meanings (the notion of suppositional modality includes a notion of any of its varieties).

So there exists an inversely proportional dependence between the amount of semantic meaning and the amount of distribution. The relation of inclusion also manifests itself in the interrelation of

[1] Here and further: (S) - subject, (P) - predicate, ( ) - zero subject, (`) - neutral SS; (\\) SS with a greater degree of expressiveness; (_) - element bearing a primary stress, (¯) - absence of an accent; // - pause.

intonation correlated with the expression of subjective evaluative modality and LGM for expressing it.

The relation of overlapping. This type of relation holds provided that intonational and LGM are semantically multi-directional being related to each other as incompatible opposites and there appears a new "average" meaning between the amounts of meanings of the components.

This phenomenon can be observed, for instance, in utterances whose semantic structure combines both the meaning of exhortation expressed by either a grammatical or a lexical form and the meaning of uncertainty that the exhortation would be performed (realized) and, hence, the stumulus to a verbal reaction (whether the speaker is able to perform this action), expressed by a rising intonation. The main representative of the group of stimuli evoking a response is a question (syncretically combining the elements of intellect and volition) which enables as to speak about concurrence in such utterance of the meanings of exhortation and a question: requests - Разрешите посмотреть журнал? Помогите мне; offers - Хотите посмотреть? Купите книгу; invitation - Не хотите потанцевать. Приходите в гости; advice - Не читайте по вечерам and the like.

As the opposites - intonation (as a general semantic feature) a stimulus to a speech reaction and LGM differentiating between semantic meanings are different but equal in rank features they are related to each other as equippolent. Their distribution does not fully coincide and is correlated with the degree of intensity of exhortation: the greater the degree of intensity the more definite is the speech context in which exhortation is sent to a definite performer (Cf.: Не хотите поговорить? - Не разговаривайте. - Не разговаривать!).

The relation of exclusion. This relation is relevant for the cases when intonational and LGM are semantically multi-directional, when the meanings they represent are remote from each other having no common elements and being in disjunctive opposition.

With their interaction in producing the general semantic effect the dominant role is played by intonation: Как можно молчать! (Нельзя молчать). Хорош друг (in the sense of "Плохой друг"), Вот Пушкин (in the sense of "Хотите купить Пушкина?" - in a bookshop).

Intonation (as well as true sense) are actualized in speech situation. Since the members of this opposition do not have common contexts they are in complementary distribution to each other.

The types of semantic relationships holding between intonation and LGM are regarded as language universals. The disclosure of them makes it possible to establish regular connections between the character of realization of intonation in its interraction with LGM in a context, the character of opposition of interrelated means in different positions in the text, their semiotic relevancy/irrelevancy.

Semiotically relevant is considered to be the orientation on the functional actualization of a speech signal for the distinctive function of intonation which is opposed to semiotically irrelevant purely intonational function of identifying the utterance.

Literature

1 For a detailed discussion see: Золотова Г.А. Коммуникативные аспекты русского синтаксиса. М., 1982.

2 Виноградов В.В. Из истории изучения русского синтаксиса. М., 1957.

3 Щерба Л.В. Восточно-лужицкое наречие. Пг., 1915.

4 Арутюнова Н.Д. Предложение и его смысл. М., 1976.

ПОИСК ВАРИАТИВНЫХ И ИНВАРИАНТНЫХ ПРОСОДИЧЕСКИХ СРЕДСТВ ДИФФЕРЕНЦИАЦИИ ПОБУДИТЕЛЬНОЙ МОДАЛЬНОСТИ В АНГЛИЙСКОЙ ДИАЛОГИЧЕСКОЙ РЕЧИ

МАРИНА СОКОЛОВА

Кафедра фонетики
Педагогический институт
им.В.И.Ленина
Москва,СССР,II9435

КАРИНЭ МАХМУРЯН

Кафедра иностранных языков
Педагогический институт
им.В.И.Ленина
Москва,СССР,II9435

РЕЗЮМЕ

Выделены фонологически релевантные просодические характеристики, способствующие дифференциации вариантов побуждения, установлены зоны их действия.

ВСТУПЛЕНИЕ

Изучение интонологической функции просодических параметров предполагает разрешение целого ряда вопросов, связанных с проблемой выделения инвариантных и вариативных просодических средств и структурными особенностями их реализации, соответствующими передаче того или иного значения. Попытка изучения этого вопроса на материале побудительных высказываний позволила решить некоторые задачи, связанные с изучением функциональной нагрузки просодических характеристик в создании трех вариантов побуждений: приказания, совета, просьбы. Выбор такой универсальной и социально-значимой категории как побуждение связан с тем, что быстрое и правильное декодирование побуждений невозможно без изучения их вариативных и инвариантных просодических характеристик. Для выявления наборов просодических дифференторов каждого варианта и их функциональной нагрузки, а также для определения зоны допустимой вариативности просодических характеристик каждого варианта необходимо было провести экспериментально-фонетическое исследование с целью выделения ведущих и неведущих, постоянных и переменных специфических и неспецифических просодических признаков. В данной работе под ведущими параметрами понимаются просодические характеристики, релевантные для различения того или иного варианта побуждения. Неведущими являются параметры, принимающие участие в создании вариантов и их конкретных реализаций, но нерелевантные для их дифференциации. Ведущие и неведущие параметры могут быть постоянными и переменными, при этом один и тот же просодический параметр в одних вариантах может выступать как переменный, в других - как постоянный. Под постоянными параметрами понимаются параметры, которые присутствуют в вариантах без изменений, а вариативные те, которые могут изменяться в зависимости от лексического наполнения, грамматической структуры и влияния экстралингвистических факторов. Специфические параметры - это характеристики, присущие только этому варианту, в то время как неспецифические или инвариантные признаки являются общими для всех вариантов побуждения.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНОГО ИССЛЕДОВАНИЯ

Анализ I20 диалогов с побудительными стимулами показал, что для исследуемых трех вариантов побуждений характерны следующие наборы просодических параметров. Для приказания - четыре ведущих постоянных параметра (направление движения основного тона в предъядерной части, направление движения и объем основного тона в ядре, скорость произнесения фразы, локализация ядерного тона), три ведущих переменных параметра (максимальный тональный уровень, тональный диапазон фразы и максимальная громкость), два неведущих постоянных параметра (локализация максимального и минимального тонального уровня, локализация максимальной и минимальной громкости). Для совета характерно наличие трех ведущих постоянных параметров (направление движения основного тона в предъядерной части, объем основного тона в ядре, максимальный тональный уровень), пяти ведущих переменных просодических характеристик (направление движения основного тона в ядре, тональный диапазон фразы, максимальная громкость, скорость произнесения фразы, локализация ядерного тона), двух неведущих постоянных параметров (локализация максимального и минимального тональных уровней, локализация максимальной и минимальной громкости). Просьба имеет один ведущий постоянный параметр (направление движения основного тона в предъядерной части), шесть ведущих переменных просодических характеристик (направление движения и объем основного тона в ядре, максимальный тональный уровень фразы, тональный диапазон фразы, максимальная громкость, скорость произнесения фразы, локализация ядерного тона), два неведущих постоянных параметра (локализация максимального и минимального тонального уровня, локализация максимальной и минимальной громкости во фразе). К неспецифическим или инвариантным признакам можно отнести один ведущий постоянный параметр – нисходящее движение основного тона в предъядерной части побуждений, а также все неведущие постоянные просодические характеристики. К специфическим признакам относятся направление движения и объем основного тона в ядре, тональный уровень и тональный диапазон фразы, максимальная громкость, скорость произнесения фразы, локализация ядерного тона во фразе. Экспериментальный анализ показал, что эти характеристики могут выступать в различных конкретных реализациях то как абсолютно специфические, то как относительно специфические, вследствие влияния экстралингвистических факторов и лексико-грамматической структуры. Однако набор данных параметров всегда будет абсолютно специфическим для каждого варианта. Следовательно, можно говорить о взаимодополнении и взаимокомпенсации одного параметра другим.

Различия в качественном составе набора просодических характеристик каждого варианта и различия активности каждого из параметров предопределяют зоны действия вариантов в поле побудительности и зону их допустимой вариативности. Экспериментальный материал дал возможность выделить следующие зоны действия релевантных просодических характеристик трех вариантов побудительных стимулов с различной лексико-грамматической структурой. Приказание имеет высокий или средний максимальный тональный уровень, широкий диапазон фразы, большой или средний объем нисходящего тона в ядре, высокую или среднюю максимальную громкость, большую скорость произнесения фразы. Совет имеет средний максимальный тональный уровень, средний тональный диапазон фразы, малый или средний объем основного тона в ядре, низкую или среднюю максимальную громкость, большую или среднюю скорость произнесения фразы. Просьба имеет высокий или средний максимальный тональный уровень, узкий, средний, широкий тональный диапазон фразы, малый, средний, большой объем основного тона в ядре, низкую, среднюю, высокую максимальную громкость фразы, малую, среднюю, большую скорость произнесения фразы. Сопоставление вариантов показывает, что самая большая зона пересечения просодических параметров у просьб, самая малая – у приказаний. Значительная вариативность в первом случае объясняется преобладанием эмоционального над логическим и волевым воздействием. Зоны действия каждого из вариантов расплывчаты, зыбки и частично пересекаются. В центре поля побудительности действуют инвариантные просодические характеристики, выполняющие интегрирующую функцию и объединяющие варианты в поле побудительности. На периферии поля находятся абсолютно и относительно специфические просодические характеристики, выполняющие смыслоразличительную функцию и дифференцирующие варианты. Поиск специфических просодических признаков реализуется через систему оппозиций, чаще всего градуальных. Наиболее регулярно прослеживаются оппозиции таких просодических параметров как тональный уровень и тональный диапазон фразы, объем основного тона в ядре, максимальная громкость. Стимулы, содержащие императивные структуры, могут быть членами большего количества градуальных оппозиций, чем неимперативные побуждения, а значит легче отождествляться с тем или другим вариантом. Вероятно, это связано с тем, что просодические средства имеют однонаправленное с лексико-грамматическими средствами действие в первом случае и разнонаправленное во втором. Концентрация максимального числа оппозиций способствует более легкому распознаванию вариантов. Представляется правомерным говорить о том, что различение вариантов на периферии происходит просодическими признаками самостоятельно, а в центре поля это разграничение возможно только с опорой на лексико-грамматический контекст. Это связано с тем, что в поле побудительности наблюдается два вида синонимии: межуровневая и внутриуровневая. Межуровневая синонимия вызвана многообразием лексико-грамматических структур, выражающих побудительную модальность, и тремя формами взаимодействия с ними просодических средств (однонаправленной, разнонаправленной и компенсаторной). Внутриуровневая просодическая синонимия вызвана экстралингвистическими факторами, разнообразием лексико-грамматических структур и наслоением эмоционально-модальных оттенков. Межуровневая и внутриуровневая синонимия – явления частично перекрещивающиеся. Межуровневая синонимия наблюдается чаще всего, когда побудительная интонация компенсирует неимперативную форму побуждения. Внутриуровневая просодическая синонимия наблюдается, когда при выражении одного из вариантов побуждения, например, приказания наблюдаются различные комбинации просодических средств. Наиболее частыми при выражении приказания взаимозаменяемыми комбинациями оказались: тональный уровень, тональный диапазон и громкость, с одной стороны, и скорость произнесения фразы и объем основного тона в ядре – с другой. При выражении совета наиболее рекуррентными взаимокомпенсирующими просодическими наборами были: тональный уровень и тональный диапазон фразы, с одной стороны, громкость и объем основного тона в ядре – с другой. При реализации просьбы взаимозаменяющимися просодическими параметрами оказались, с одной стороны, тональный уровень и тональный диапазон фразы, с другой – громкость, скорость произнесения фразы, движение и объем основного тона в ядре.

Данные экспериментального исследования позволяют сделать вывод о том, что у побуждений при внутриуровневой синонимии сочетаемость просодических характеристик и степень их спаянности при выражении приказания, совета и просьбы различна. Это, очевидно, в первую очередь, связано с функционально-семантическим характером побуждений. Следовательно, можно заключить, что просодия не только отражает прагматический характер побуждений, но и имеет свои закономерности, позволяющие дифференцировать варианты побуждений. Просодические характеристики, создавая тот или иной вариант побуждения, взаимодействуют, дополняя или компенсируя друг друга, а не являются результатом механического суммирования. Выбранный в конкретной речевой ситуации говорящим из ряда синонимичных разноуровневых средств вариант оформления побуждения призван наиболее сильно воздействовать на партнера, заставить его выполнить установку говорящего. Однако проблема просодической синонимии чрезвычайно сложна и ставит ряд задач, которые могут быть решены только на основе специальных экспериментально-фонетических исследований. Дальнейшего исследования требует также тесно связанное с синонимией явление многозначности языковых единиц. Исследование данных явлений имеет как теоретическое, так и практическое значение. Так, например, большое число комбинаций просодических характеристик при выражении плана трех вариантов побуждений является важным опорным положением при обучении интонации побуждений, а правильный выбор данных комбинаций и наборов является существенным моментом при создании оптимальной эффективности его воздействия. Умелое, корректное использование языковых средств всех уровней дает возможность сделать речь более выразительной и действенной, а адресатам быстро и правильно декодировать типы побуждений. Выявление вариативных и инвариантных просодических средств, которые часто являются причиной просодической синонимии и многозначности, также позволит прогнозировать и разработать формализованные правила для моделирования компьютерными средствами автоматическое распознавание речи.

ЗАКЛЮЧЕНИЕ

Данные экспериментально-фонетического анализа свидетельствуют о том, что побуждение является полисемантичным и разграничение его вариантов происходит как с помощью взаимодействия с грамматическими и лексическими средствами, так и просодическими средствами самостоятельно. Однако комбинированное употребление всех языковых средств не является избыточным и ведет к более легкому распознаванию вариантов побуждений. Побудительный инвариант объединяет не только однородность грамматической структуры (например, императив), или одинаковая лексическая оформленность, но и определенные интегрирующие просодические параметры.

В поле побудительности наблюдаются два вида синонимии: межуровневая и внутриуровневая. Характер побуждения оказывает определенное влияние и детерминирует сочетаемость просодических характеристик и степень их связности как при внутриуровневой синонимии, так и при межуровневой.

Все ведущие и неведущие просодические средства находятся в различной степени зависимости от лексических и грамматических характеристик. Так, при выражении различных типов побуждений они действуют обычно однонаправленно с лексическими и однонаправленно и разнонаправленно с синтаксическими характеристиками. При однонаправленном взаимодействии просодические средства не только дифференцируют типы побуждений, но и усиливают их воздействие, облегчая их правильное декодирование адресатами. При разнонаправленном взаимодействии просодические средства способствуют снятию синтаксической многозначности и дифференциации типов побуждений. Следовательно, принцип замены А.М.Пешковского, формулируемый следующим образом: "чем яснее выражено какое-либо синтаксическое значение чисто грамматическими средствами, тем слабее может быть его интонационное выражение (вплоть до полного исчезновения) ..." [I], на материале побуждений реализуется только частично, а именно, при разнонаправленном взаимодействии синтаксических и просодических средств.

Таким образом, поле побудительной модальности представляет собой общую систему, которая охватывает и упорядочивает всю совокупность вариантов побуждений, при этом прослеживается определенная иерархия отношений "центр – периферия". Каждый вариант, со своей стороны, характеризует это поле, а в совокупности они образуют диалектически единую систему. Взаимодействие и взаимозависимость - вот процессы, которые пронизывают систему поля побудительности. Следовательно, можно утверждать, что вариативные и инвариантные характеристики просодии составляют диалектическое единство и соотносятся как единичное и общее, абстрактное и конкретное и подчиняются таким законам как переход количества в качество, единство и борьба противоположностей.

Литература

А.М.Пешковский. Вопросы методики родного языка, лингвистика и стилистика. Москва, 1930.

PRINCIPLES OF PROSODIC PROMINENCE FORMATION OF WORDS IN RUSSIAN UTTERANCES

Tatiana Nadeina

Department of Philology, Moscow State University
Moscow, USSR, 119899

ABSTRACT

In this paper the necessity to distinguish between two functionally different types of word prominence in an utterance is grounded. The first type is neutral sentence /syntagmatic/ stress which performs constitutive and delimitative functions, and the second type is sentence accent which is related to semantic side of the utterance. Experiments on perception showed that relationship between the sentence accent and neutral sentence stress is not that of complementary distribution and can be realized in one syntagma /sentence/ simultaneously. Among functions of the neutral sentence stress a function of expression of word semantic value is not included, it serves as a means of syntagma phonetic organization and speech rhythmization.

INTRODUCTION

In works on Russian intonation a point of view put forward in works by L.V.Scherba becomes more and more widespread. In conformity with it two functionally different types of word prominence in the utterance are distinguished, namely, neutral sentence stress and sentence sense accent.

The neutral sentence stress is obligatory in a syntagma /or single-syntagma sentence/ and is assigned to its final word thus performing constitutive and delimitative functions. This stress is independent of specific semantic relations in the utterance and serves as a means of intonation segmentation and speech rhythmization.

The second type, the sentence accent, differs from the first one in that it is realized in a sentence under only these conditions, when it is determined by a context, communicative intention of a speaker etc. A place of the sentence accent is not fixed, it can be placed on any word in a sentence. There exist various different terms for this type of word prominence -semantic, logical, contrastive, rhematic etc. Thus it implies that the sentence accent depends on the semantic side of the utterance.

In works by T.M.Nikolaeva the necessity of strict distinguishing between neutral sentence stress /SS/ and sentence accent /SA/ is grounded, since functionally they are heterogeneous phenomena: "The SA is a textual communicative phenomenon and the SS - an intrinsic intonation phenomenon" (I982, p. 9).

This conception is not, however, generally accepted. In works on functional syntax and semantics by soviet and foreign scholars as well as in works on intonation, there is no distinguishing between functionally different types of word prominence in the utterance. It is considered that any word prominence depends on different semantic relations. That's why in these works the term "sentence stress" designates both neutral sentence stress and sentence sense accents.

We consider sentence accent pattern to be the result of the simultaneous realization of functionally different devices of word prominence, namely, the neutral sentence stress /SS/ and the sentence accent /SA/. At the prosodic level the sentence accent pattern is realized in different degrees of prosodic prominence of words which make up the utterance.

To give prove to the proposed point of view the following questions were considered:

I. How do the SS and SA in the utterance correlate? Are they realized simultaneously or does the SA neutralize SS?

2. Is a word, which has the SS in the absence of the SA in the utterance,the point of information focus? In this case, is the degree of prosodic prominence of a word related to its semantic value?

We tried to find answers to these questions by applying to speech competence of native speakers and analysing mechanisms of perception of prosodic prominence of words in an utterance and mechanisms of interpretation of semantic value of words in a text as well.

I

The question of relationship between neutral sentence stress and sentence accent in a Russian utterance is treated differently by scientists that accept functional difference of these types of prominence. Some scholars consider that there exists a possibility of their simultaneous realization in a sentence. For example, in the paper by L.V.Zlatoustova /I963/ it is said that sentence accents "are always realized with the sentence stress and in some cases they overlap the sentence stress but don't neutralize it" /p.I06/. T.M.Nikolaeva also believes that "the presence of a greatly prominent word at the beginning /of a sentence - T.N./ does not mean beginning /of a sentence - T.N./ does not mean that the final part of it lacks corresponding prosodic prominence"/1979, p.I05/. Other scientists, on the contrary, claim that logical stress "neutralizes the sentence stress, makes its realization impossible"/M.Panov,I979,p.88/.

We think that one of the methods of solving this problem is the study of perception mechanisms of word prosodic prominence by native speakers and revealing objective criteria of estimation of a degree of word prominence in the utterance.

We interpret the notion "a degree of word prominence" as its prominence in regard to other words in the utterance. An important aspect of this notion is its consideration from two points of view, namely, objective which presupposes an estimate of value of prosodic word parameters, and subjective which shows their perception by native speakers. In conformity with this a degree of objective prominence and a degree of subjective prominence of words are distinguished in this paper.

While preparing the experiment we proceeded from the notion that when the sentence accent pattern is perceived, a man takes into account different information - segmentic, prosodic, syntactic, semantic as well as extralinguistic context. Since first of all we were interested in the role of prosodic information in word prominence perception, and in order to neutralize the effects of other enumerated factors perception of words cut out from sentences was analysed.

Three-word sentences with identical syntactic pattern /subject + predicate with a dependent word form/, with different word order and with different place of contrastive sentence accent in them /sentence initial, medial and final position/ composed of words: íris, irís, Irina, kupit served as experimental material. Choice of words was stipulated by the desire to achieve maximum phonetic homogenuity of vowels, in order to avoid differences in intensity, duration and fundamental frequency which are characteristic of vowels of different phonetic quality.

Each sentence was read with two kinds of intonation: narrative and interrogative /general question/. The total number of realization of experimental sentences constituted 72 items. Each sentence is characterized by two prosodic variables: I/ type of accent pattern - sentence accent of the first word /designated as SI/, of the second word /S2/, of the third word /S3/; 2/ type of intonation - narrative or interrogative.

The basic experimental sentences were segmented into single words which were used to make up 4 tables containing 50 different realizations of one and the same word /out of this sample only 36 realizations were examined/ singled out from all possible sentence positions.

I7 subjects without special phonetic training /students of philologic department/ took part in the experiment. Their answers were regarded as speech behaviour of native speakers.

Technique of determining the degree of word subjective prominence

In the process of perception test the subjects were first given all the basic experimental sentences and they were informed about the principles of their construction. Then they were asked to listen to the experimental word tables and to decide for each word whether it was accented in the basic sentence or not. It was assumed that the degree of coincidence of subjects' answers /according to which all the words could be divided into two sets - accented and non-accented/ could be used as the estimate for word subjective prominence.

On the basis of the obtained results a coefficient of word subjective prominence /K/ was calculated as a relative number of subjects considered the word to be accented /in %/. Then according to K-values the degree of subjective prominence /P/ was assigned to words in the following way: P=I, if $0 \leqslant K < 30\%$; P=2, if $30 \leqslant \leqslant K < 70\%$; P=3, if $K \geqslant 70\%$.

Results

The results of calculation of word subjective prominence /P/ in sentences of considered types are presented in the table I /mean values for each word in every sentence position/.

Table I

Types of sentences	Narrative sent.			Interrogat. sent.		
	Ist w.	2nd w.	3rd w.	Ist w.	2nd w.	3rd w.
SI	2,67	I,I7	I,67	3,0	I,33	2,0
S2	I,33	2,5	I,83	I,0	3,0	2,33
S3	I,33	I,I7	I,67	I,33	I,0	2,83

It is clearly seen that in sentences with sentence accent on the first or second word /SI and S2 types/ the third word has a greater degree of prominence than an unaccented one. It testifies to the fact that in sentences under consideration the final word is prosodically marked and it can be regarded as a result of the neutral sentence stress.

Thus, values of word subjective prominence in the sentences show that at the level of perception a distinction between an accented word and a word having the neutral sentence stress may exit within one sentence. In other words, the SA and SS are realized simultaneously and the former does not neutralize the latter. The conclusion is proved as well by the results of the experiment which consisted in estimating the degree of word objective prominence in the analysed sentences.

Technique of determining the degree of word objective prominence

The technique of determining the degree of word objective prominence in a sentence is based on the results of the study in the course of which the relationship between values of prosodic parameters of a word and the degree of its subjective prominence were analysed.

The study was preceded by acoustic analysis of basic experimental sentences. Intonograms and wideband spectrograms were analysed and the following parameters were determined: I/ word total duration /T/; 2/ stressed vowel duration /t/; 3/ word maximum intensity /peak value/; 4/ F_o-maximum of a stressed vowel /F_o/; 5/ difference in maximum and minimum F_o-values within a word, we'll call it further bandwidth of F_o /ΔF_o/.

We studied correlations between the values of above numerated prosodic parameters and word subjective prominence /K-coefficients/. The results of the analysis showed that the K-coefficients correlate only with duration values and fundamental frequency. In our test values of word intensity do not correlate with coefficients of word subjective prominence.

On the basis of the results there are reasons to believe that perception of the prominence is based on the estimate of absolute values of prosodic parameters which are compared with some threshold values. If we assume that the threshold value of a prosodic parameter, in regard to which a word is considered to be highlighted, is its mean value in all realizations, then we can sigle out three gradations of a parameter's /Q/ objective value. They are as follows: Q=I if the observed value of the parameter is less than the mean one minus threshold value ε; Q=2 if the observed value of the parameter is in the interval of the mean one plus-minus ε; Q=3 if the observed value of the prosodic parameter is more than the mean one plus ε. Critical values of ε for duration parameter are about I5 centiseconds and 20 Hz for F_o /it is in conformity with earlier published experimental data/.

Comparison between Q-values of different parameters and P-values showed that what is important for prominence perception is not a fixed set of Q-values of prosodic parameters but a complex estimate of objective prominence which takes into account their summary value in a word. Close correlation between summary value of prosodic parameters in a word and the degree of its subjective prominence is revealed.

On the basis of the obtained results we believe that the degree of word objective prominence can be calculated as a sum of values of its prosodic parameters /Qs/:

$$Qs = Q_t + Q_T + Q_{F_o} + Q_{\Delta F_o}$$

In this experiment four prosodic parameters were taken into account. Each prosodic parameter may acquire Q=I,2 and 3, that's why Qs varies from 4 to I2.

Results

In conformity with the proposed technique degrees of word objective prominence in the sentences were calculated. Table 2 presents mean values of word objective prominence /Qs/ in types of sentences under consideration /mean values for each word in every sentence position/.

Let's take a look at table 2. One can see that the degree of objective prominence of the final word in sentences of types SI and S2/both narrative and interrogative/ is greater than that of the other unaccented word. This testifies to prosodic marking of the final word that is caused by the neutral sentence stress.

Table 2

Types of sentences	Narrative sent.			Interrogat. sent.		
	Ist w.	2nd w.	3rd w.	Ist w.	2nd w.	3rd w.
SI	8,3	4,5	6,7	II,5	6,7	7,8
S2	6,2	8,0	7,5	7,5	II,0	9,0
S3	5,8	4,8	I0,2	6,8	6,0	II,5

Thus, the results of the experiments carried out to estimate the degree of subjective and objective prominence of words in three-word sentences with contrastive sentence accent demonstrated that the SA doesn't neutralize the SS: the final word in the sentences with the SA in a non-final position is prosodically marked, highlighted. This prominence is perceived by auditors /under specific conditions of the experiment/, and is proved by objective values of prosodic parameters.

The neutral sentence stress and sentence /sense/ accent are not in complementary distribution, and thus they can be realized simultaneously in one syntagma /sentence/.

2.

The other important question is whether a prosodically marked, highlighted word is a point of information focus and whether the prosodic prominence of a word is related to its semantic value.

It is generally accepted that by means of the sentence accent in the utterance the most important words which have a definite semantic value are highlighted. Often such words are predicted by the preceding context, by word order and accompanied by expressive particles. However, in the case of utterances with no SA /with neutral content/ one can't deny the fact that words which make up the utterance have different semantic values and different degrees of prosodic prominence. The question of what factors determine the prosodic prominence of words in the utterance in the absence of the SA was in the focus of our study, in the course of which the correspondence between the prosodic prominence of words and their semantic value in a text was examined.

5 short newspaper texts with an average volume of about I00 words where the sentence accent occured rarely were used for the experiment. The texts were read by 8 speakers-members of the staff of the philologic department /4 men and 4 women/.

Technique of determining semantic value of words in text

The semantic value of a word is understood as its role in conveying the information comprised in a text. In order to determine the semantic value of words an approach which presupposes an appeal to speech competence of native speakers was chosen. The following procedure was used.

20 subjects were given written equivalents of texts and three successive tasks were set: I/ to reduce the volume of a text by crossing

out some words, maintaining the number of sentences and not violating coherence of the text so that the main /from the standpoint of a subject/ information of the text is left intact; 2/ to underline in the text words and word combinations which should be included into summary in order to reproduce its content in detail some time later; 3/ to give the summary of the text in one's own words. According to the results of answers of the subjects each word of the text can be characterized by a set of three features arbitrarily called "redundance" /A/, "importance" /B/, "richness of content" /C/.

Estimate of the semantic value of words was carried out in two stages. At the first stage a coefficient of word semantic value /S/ was calculated by the following formula: $S = A + B + C$, where A - a relative number of the subjects considered the word to be "redundant", B - of the subjects considered the word to be "important", C - of the subjects considered the word to be "rich of content". The calculated S-values can vary from -I to +2.

At the second stage according to numerical S-values and a combination of A, B and C features the semantic value /R/ was assigned to words in the following way: $R=0$ if $S \leqslant 0$; $R=I$ if $S > 0$ and if $B=0$, if $C=0$; $R=2$ if $0 < S \leqslant I$ and $A=0$ but $B,C \neq 0$; $R=3$ if $S > I$ but $B,C \neq 0$.

Technique of determining the degree of word prominence in texts

The degree of word prominence in texts was analysed in its subjective aspect, i.e. from the point of view of its perception by native speakers. Records of texts read by 8 speakers were presented to the subjects /II students of philologic department/. In the process of audition of the texts they were asked to divide sentences into syntagmas and to highlight the most prominent word in each sentence.

Based on the results of the audition test the degrees of prominence calculated from the data of all speakers and auditors in conformity with the technique reported in the first part of the paper were determined. It should be noted that for the convenience of comparison of the semantic value indices with word prominence estimates we introduced the index P=0 in case K=0.

Results

The study of correlation between different indices of semantic value of words and degrees of subjective prominence of these words demonstrated that words with maximum semantic value /R=3/ can be characterized by the following degrees of prominence: P=I, on the average, 35,2% of cases, P=2 - II,8% of cases, P=3 - 53% of cases and can't have a degree of prominence P=0. At the same time words characterized by the maximum degree of prominence /P=3/ can have any semantic values with approximately equal probability: R=0, on the average, I5,6% of cases, R=I - 30,7% of cases, R=2 - 22% of cases and R=3 - 3I,7% of cases.

The data show that in the texts the relationship between the degree of prominence and semantic value of words is one-sided: words of maximum semantic value have a tendency to be prosodically marked. However, the reverse is not true: maximum prominence of a word does not necessarily indicate maximum semantic value.

The main reason for this asymmetry lies in the fact that in texts the degree of word prominence is mainly determined by the mechanism of neutral sentence stress which is realized on the final word of a syntagma irrespective of its semantic value. On this account one ought to expect that the degree of prosodic prominence of a word depends on its position in regard to syntagma boundaries.

Our results show that for words of minimum semantic value, if they are placed at a syntagma boundary, the frequency of prominence values P=3 increases up to 45%. For words of maximum semantic values the frequency of values P=3 constitutes 95% if these words are at a boundary and 35% if they are placed in the middle of a sentence.

The results prove that the most essential factor which affects the degree of prosodic prominence of a word is its position in relation to syntagma boundaries. In this case word prominence is determined by the effect of the neutral sentence stress. Thus, words of maximum semantic value in a text are not always prosodically highlighted but only in cases when they are under "favourable" sentence conditions. Hence, the expression of semantic value of words in a text is not the function of the SS.

SUMMARY

The obtained results give an answer to the questions that were considered in the first and second parts of our study. It was showed that the relationship between the SA and SS is not that of complementary distribution and that they can be realized simultaneously within one sentence. If the SA is a means of expression of word semantic value, then the neutral sentence stress serves as a means of syntagma phonetic organization, intonational segmentation and rhythmization of speech.

The sentence accent pattern which is realized in different degrees of word prosodic prominence is determined by two functionally different mechanisms - the neutral sentence stress /which highlights the final word of a syntagma/ and the sentence /sense/ accent, the place of which is not fixed.

REFERENCES

Златоустова Л.В, О фонологических функциях фразовых акцентов. // Конференция по структурной лингвистике. Тез.докладов. М., I963.

Николаева Т.М. Функции акцентного выделения и семантико-синтаксическая структура высказывания.// Фонология. Фонетика. Интонология. Материалы к IX Международному конгрессу фонетических наук. М., I979.

Николаева Т.М. Семантика акцентного выделения. М., I982.

Панов М.В. Современный русский язык. Фонетика. М., I979.

FUNCTIONS OF ACCENT PROMINENCE IN SPEECH

TATIANA SKORIKOVA

The Pushkin Institute
of Russian Language
Moscow, USSR, 117485

ABSTRACT

A new method of intonational analysis is presented with two communicative directions in view: text → intonation (a word's accent prominence as effect of text parameters) and intonation → text (accent prominence as a text-formation factor).

INTRODUCTION

Nowadays linguists have evidently ceased to limit accent prominence (AP) to some single language phenomena - be it an intonation pattern, a category of definiteness/indefiniteness, significance, contrast or emphasis of the utterance /1/. Now one is inclined to consider AP as a multi-aspect object of study, which has to do with the semantic framework of a text rather than with semantic meaning of isolated utterances. Hence, AP may be ranked together with general categories of covert grammar of text thus helping to reveal interdependacy of text constituents, i.e. text → intonation (AP as effect of certain text parameters) or intonation → text (AP as a text-formation factor).

The analysis of the intonation of immediate oral speech seems worth-while for both studying the speech-production mechanism and context-bound regularities of AP.

In the present paper special attention is paid to the functions of AP in spontaneous monologue of a scientific character (lectures, reports, etc.). Oral scientific text is characterized by specific accent markers, i.e. words whose AP serves to intensify the communicative message of the utterance /2/.

His public verbal event being limited in time, the speaker has to use a specific text-production program which should provide (besides some cognitively relevant information) argumentation (to convince the audience) and modality (revealing the speaker's attitudes). The study of AP's role in pragmatic constituents of a text may thus involve a number of particular tasks, such as analysis of AP in notional or synsemantic syntactic constructions; syntactic position of AP in the utterance, text-formative functions of AP depending on the kind of texts and so on.

Our approach to AP problems calls forth the role of the speaker himself. In the course of speech the attitudes of the speaker are not inflexible nor intentionally static: "...the speaker's point of view is constatly sliding like the camera man's objective thus exposing the object from different angles" /3,p.155/. Can we relate the changing semantic role of a word in speech continuum to the word's AP? To answer the question we'll first delineate our theoretical departure position.

L.A.Gogotishvili /3/ distinguishes three subsidiary pragmatic sense components which are usually fixed: point of view, attention focus and discourse pivot.

Point of view is determined by the directness (speaker → hearer); hence, we have two positions that of the speaker (=I), and that of the hearer (=You).

Attention focus is some semantic component of the situation chosen as centre for all the other components. Various authors connect the focus with the theme/theme notion having the prosodic correlate on the surface structure of the sentence: "...rougly speaking, the presence of a pitch accent correlates with a focus (and thus with "new information"), while the absence of a pitch accent indicates the lack of focus (or "old information")" /4,p.200/. We assume that AP should have

more ingenuous semantic interpretation: it may highlight, in particular, within one and the same utterance, old as well as new components of sense which become significant for either the speaker or the hearer at a given moment. Hence, using the term "attention focus" we stress the role of AP as a text component /5/.
Discourse pivot is a certain intentional position commited by either the speaker or some generalized social opinion. Our main postulate comes as follows: all the pragmatic sense components mentioned above are prosodically manifested on the acoustic level in the form of AP of some lexical elements. Now we'll discuss the idea in detail using Russian adjectives and adverbs: as we showed elsewhere /6/ attributive words and word combination prove to be good lexical material to study the role of AP in a nexus of meaning, purposes, motivations and other components of a text.

Point of view

Speaking of AP's function to indicate the direction of message orientation (onto the speaker or onto the hearer) we should note that it is posessed by different classes of lexemes in quite different degrees. That's why it seems fruitful to analyze general and particular properties of various lexical classes with the view of their different ability to AP.
Utterance formation is allegedly a multi-stage process beginning with the emergence of some vague semantic concept, then going to the creation of the logical and structural plan and finally producing lexical and phonemic contour. If this model holds true then there should be three groups of words which bear the greatest semantic load and hence are prone to AP:
a) prop words carrying the message of the utterance;
b)deictic words supplying structural and content support to the spoken chain;
c) modal and quantifying words.
AP words can accordingly have the functions of nomination, communicative message (or text formation proper) and modality.
Distributive analysis shows clearly that different classes of words differ both in the intensity of AP and in the frequency of appearing in functions (a-c): accented nowns are mostly nominative; verbs, adjectives and adverbs fall mainly into the second (communicative message) function whereas modal function is performed by predicative words (i.e. verbs, adjectives and adverbs of specific semantics).
To evaluate the ability for AP we introduced index $i = \frac{AP}{NAP}$, where AP is the number of cases when the word (group of words) was found in accent position and NAP - the number of cases of non-accent positions in the texts of special type /7/.
Unlike terminological adjectives whose i lie near to 0 (0,37) words denoting attitudinal meanings have the highest values of i: adjectives denoting negative attitude have i = 3,3; those denoting the highest degree of some property have i = 2,2; adjectives denoting positive attitude, intensifiers and modal words have medium i - values (0,69, 0,59 and 0,57 accordingly).
As might be expected adverbs are very much similar with adjectives: they have a strictly limited number of high i - value words; there exists positive correlation between frequency of a word's AP position and its total text frequency; adverbs with high i - values show much common in semantic features with their adjective counterparts. Within both adjectives and adverbs there are groups of words which are especially apt to AP. These include: 1) words with high text-forming potential (pervyj, vtoroj, odin, drugoj, etc.); 2) words having attitudinal or intensifying function in the text (osnovnoj, sušestvennyj, boľshoj, iskluciteľnyj, neobyknovennyj, praviľno, prosto, veśma, oceń, sovershenno, vsegda, voobše, etc.); 3) words denoting casual/ resultative and some other relations (novyj, poslednij, očerednoj, bližajshij, sledujušij, protivopoložnyj, obraťnyj; togda, teper', sejčas, poetomu, zdeś, etc.).
Accented adjectives and adverbs are often characterized by extremely wide range of semantic values and by high degree of semantic generalization which enables their use in deictic function (i.e. function of direct or inderect reference to the preceding or following context).
To sum up, the high i - value words usually possess presuppositional and deictic properties. The more distinct are the word's attitudinal semes (orientation towards the speaker) or its deictic semes (orientation towards the hearer), the greater would be its AP ability.

Attention focus

By changing the attention focus the speaker can direct the hearer's "view" to various aspects of his message thus bringing home sense pivots. In dealing with the attention focus one should coincide AP with the word's semantics and context.
Different meanings of a word have different AP abilities. The degree of AP of a given meaning depends on the presence (or lack) of elements that make this meaning explicit. Let's consider a group of possessive adjective prounouns as an illustration. This class of words is frequently used to express the notion of I (author) in the text /8/.
In Russian scientific style tradition my 'we' often stands for ja 'I'. Nash 'our' has the highest value of i in the group (i_{nash}= 0,44; $i_{vash'your'}$= 0,22; $i_{moj'my'}$= 0,11). It's worth mentioning that nash has the widest range of meaning among the words of the group in question:
1) 'moj' (my, the speaker's);
2) 'moj i vash' (my and the hearer's);
3) 'nash' (belonging to us). The first meaning usually has no AP whatsoever; the other two have much higher values (i_{nash_2}= 0,41; i_{nash_3}= 0,50).
Meaning 2 ('moj i vash') is coincident with the moment of the speech act. In the course of public speech the speaker draws attention of the audience to this or that item trying to get them share his medidations and to make them part and parcel of the verbal event. The meaning 'moj i vash' is introduced onto a given audience by additional aids of the context, such as: 1) tautological word combination nash s vami (e.g. s točki zrenija náshej s vami raboty...; náshe s vami zasedanije posviašeno voprosu...) 2) inderect appeal to the hearer to follow the speaker's way of thinking (e.g. nu vot esli vozvraš'atsa /k náshej klassifikatsii tipov norm/...); adverbial modifiers of time which refer to the moment of speech act (e.g. Ia dumaju čto tselesoobrazno v náshih diskussijah / segodn'a / i zavtra obsudit'...).
Meaning $nash_3$ has no additional support from the context (e.g. takogo roda rabota / ne provodilaś / ni v náshej rusistike / ni / voobše v tselom). That's why AP of $nash_3$ is not so forcible in speech continuum.
Analysis of a number of meanings in AP position indicates that there exists inverse correlation between frequency of meaning's AP and the degree of its contextual markedness: less text-bound meanings have higher values of i. As to the degree of AP forcibility it is directly dependent on whether the context contain elements reinforcing the meaning in some way or another: the more is the meaning in need of contextual support the more forcible it is marked on the prosodic level.

Discourse pivot

The change of discourse pivots in a scientific oral monologue does not coincide with the change of the speaker's communicative behaviour (as is usually the case with dialogues). Hence, the text of a monologue calls for greater amount of specific markers both to distinguish stages in the speaker's communicative strategy and to outline content and structural parts of the text in the course of speech. It is here that AP undertakes its text-forming functions. For example, a group of adverbs with markedly forcible AP (daľshe, teper', zdes', sejčas, ocen', ves'ma, sovershenno, vsegda, voobše, etc.) have proved to possess the following text loads: 1) promotion of the information; 2) reference to the message that follows; 3) pointing to the previously mentioned fact; 4) stressing the actual moment of speech; 5) intensification of expressive means of the utterance; 6) logical emphasis on some fact.
Within these one can easily distinguish three groups of adverbs with different function loads: a) text-forming elements; b) intensifiers; c) agents of time actualization.
In the adverbs mentioned above the word's semantics is fused with the semantics of AP and their contextual load is assumed in the form of a few typical function and sense combinations: 1) promotion of the information + reference to the forthcoming message or stressing the actual moment (e.g. daľshe, teper'); 2) pointing to the previously mentioned fact + reference to the message that follows (e.g. poetomu, togda); 3) intensification + logical emphasis on some fact (e.g. ocen', ves'ma, sovershenno, vsegda); 4) logical emphasis + intensification (e.g. voobš'e); 5) pointing to the previously mentioned fact + time actualization (e.g. zdes'); 6) time actualization + logical emphasis (e.g. sejčas).
The sense load of the element in AP position attains the greatest functional significance at the moment when a discourse pivot is changed. It seems worthwhile to enumerate the possible lexical variants of most frequent AP words within different discourse pivots and to try to correspond the AP words to a set of means which serve to carry out the sense in oral scientific speech.

CONCLUSIONS

We've discussed some important functions of AP in oral speech stressing the bidirection character of the text-forming process (text ⟷ intonation).
Pragmatic components of the utterance are singled out particularly when the change of points of view, attention focus and discourse pivots take place.
The discourse pivot (making explicit the functional goal of the message) is responsible for interdependence between AP and

various stages of the utterance-formation process. It also determines spatial and temporal limits of the message and contributes to text formation. In this case the text is regarded of paramount importance, the task being to find textual loads of speech elements in AP position (text → AP).

The attention focus corresponds AP with a word's semantic and elements of the immediate context rather than with text parameters. Here we have the reverse direction (AP → text), whose function is to orient the hearer in the speech continuum. Further analysis of AP in this respect may contribute to reveal semantic potential of words in AP position.

The point of view involves speaker/hearer parameters. It is responsible for a general accent contour of the utterance, including AP which helps explicate the speaker's position and AP which takes account of the hearer's needs. Hence, this component functions as a bidirectional text-forming mechanism (text ↔ intonation).

NOTES AND REFERENCES

1. Nikolajeva T.M. Semantika aktsentnogo vydelenija. Moscow, Nauka, 1982.
2. On the role of AP in oral scientific text see Nikolajeva T.M.Funkt3ii aktsentnogo vydelenija v ustnoj naučnoj reči. In: Sovremennaja russkaja ustnaja naučnaja reč. Vol.I. Obš'ije svojstva i fonetičeskije osobennosti. Krasnojarsk, 1987.

3. Gogotishvili L.A. Khronotopičeskij aspekt smysla vyskazyvanija. In: Rečevoje obšenije; tseli, motivy, sredstva. Moscow, 1985.

4. Selkirk E.O. Phonology and Syntax: the Relation between Sound and Structure. Cambridge (Mass): London: the Mit press. Cop., 1984.

5. There are several other terms for this communicatively oriented marking of a semantic component: orientation (see Nooman M. On Subjects and Topics. In: Proceedings of the Third Annual Meating of Berkely Linguistic Society. Berkeley, 1976); empathy (see Kuno S. Theme and Speaker's Empathy.- A Reexamination of Relativization Phenomena. In: Ch.N. Li(ed.) Subject and Topic. Academic Press. N .y., 1976); Point of view (see Chafe W. Giveness, Constrastiveness, Definiteness, Subjects, Topics and Points of View. In: Ch.N.Li(ed.), op. cit., pp.22-55); topic (see Li Ch.N. and Thompson S.A. Subject and Topic: a new Typology of Language. In: Ch.N.Li (ed.), op.cit. pp. 457-489).

6. Skorikova T.P. Funktsional'nyje vozmožnosti intonatsionnogo oformlenija slovosočetanija v potoke reči. Avtoreferat kand. diss. Moscow, 1982.

7. Skorikova T.P. Aktsentno-semantičeskije vozmožnosti prilagatel'nogo v ustnoj naučnoj reči. In: Naučnaja literatura: Jazyk, stil', žanry. Moscow, Nauka, 1985.

8. Skorikova T.P. Intonatsionnyje i strukturno-smyslovyje kharakteristiki atributivnogo slovosočetanija v ustnoj naučnoj reči (na materiale russkogo jazyka). In: Obšije i častnyje problemy funktsionalnyh stilej, Moscow, Nauka, 1986.

ИНТОНАЦИОННЫЕ СРЕДСТВА АКТУАЛЬНОГО ЧЛЕНЕНИЯ РУССКИХ ПРОСТЫХ ПРЕДЛОЖЕНИЙ

ВАН ЧАОЧЭНЬ

II Пекинский ИИЯ
Бэйцзин, КНР

ПАНЬ ХУНХАЙ

Факультет РЯ
II Пекинский ИИЯ
Бэйцзин, КНР

РЕЗЮМЕ

Материал исследовония—русская речь, запись советских дикторов. Задача: установить влияние синтагматического состава на актуальнос членение предложения (АЧП), влияние АЧП на их интонационный контур; характерные черты интонации для выражения АЧП; соотнесенность интонационного выражения темы и ремы с типами интонационых конструкций. Выявлена зависимость интонации от синтагматического состава, порядка темы и ремы, ситуации, семантической и грамматической структуры предложений. Эксперимент выполнен на сонографе//кау-7800.

Актуальное члснение речи представляет собой семантический аспект анализа речи. Оно является также коммуникативно-функциональным анализом. При помощи актуального членения речевое высказывание делится на части: тему и рему, или данное и новое. Мы пользуемся принятыми терминами: тема, рема.

Актуальное членение выражается прежде всего порядком слов и частицами. В устной речи оно всегда получает свое интонационное оформление. Однако между интонационными средствами и актуальным членением нет тождества и точного соответствия.

Экспериментальные данные подтверждают: актуальное членение односинтагмных предложений получает явное интонационное выражение, многосинтагмные предложения также находят просодические соответствия при актуальном членении.

Эксперименты показали: при членении русских предложений на тему и рему долгота звуков среди интонационных компонентов на первый план не выступает. Существует мнение: долгода звуков ремы всега дольше, чем темы. Этот взгляд страдает упрощенностью. Долгота звуков всей части ремы в среднем не больше, чем долгота звуков всей части темы; к тому же наблюдаются случаи, когда среднестатистический параметр величины слогов из части ремы еще короче (см. табл.).

Эксперимент показал, что главным компонентом интонационных средств при актуальном членении предложений на тему и рему служит высота звуков. Между мелодикой двух частей наблюдается заметное падение: Следует различать независимый вариант предложеиия и зависимый.

Независимый вариант обозначает предложение, которое сравнительно самостоятельно семантически и структурно в отрыве от контекста. Например, Конференция/продолжается. На улице/прохладно. Первые части из данных выше предложений являются темой, а вторые — ремой. Тон на ударном слоге последнего слова из темы (-рен-, -у-) поднимается. Тон на заударных слогах того же слова из темы(-ция, -лице) начинает ниспадать. И мелодика из ремы продолжает это нисходяшее движение(см. схемы 1, 2). Таким образом создается видимое противопоставление мелодики темы мелодике ремы. На схеме в мелодике темы и ремы наблюдается резкое падение. Именно это противопоставление дает понимание с помощью аудиального восприятия, что представляет собой ядро высказывания и что служит новой информацией. Примеры можно умножить.

ТЕМПОРАЛЬНОЕ ОТНОШЕНИЕ РЕМЫ И ТЕМЫ

	количество слогов	время звучания фразы	темп фразы	время звучания темы	темп темы	время звучания ремы	темп ремы	время звучания центра	темп ремы без центра	Отношение		
										темп Р : темп Т	темп Р без центра : темп Т	время центра : темп фразы
Я встаю в ссмь часов утра.	8	1070	133,75	450	150	620	124	200		0,827		1,333
В два часа дня я обедаю.	9	1080	120	620	124	460	115	180		0,927		1,5
Мы отвсчасм правильно.	8	990	123,75	500	100	490	163,333	200	96,667	1,633	0,967	1,616
Капустник сочиняют сами студенты.	12	1310	109,167	690	98,75	620	124	250	92,5	1,256	0,939	2,290
Потом начинается урок русского языка.	14	1380	125,455	160	80	1220	135,556	150	133,7	1,694	1,672	1,196

Единица длительности звучания: миллисекунда

Существует суждение, что мелодика в части темы принимает восходящую форму, а в части ремы — нисходящую.

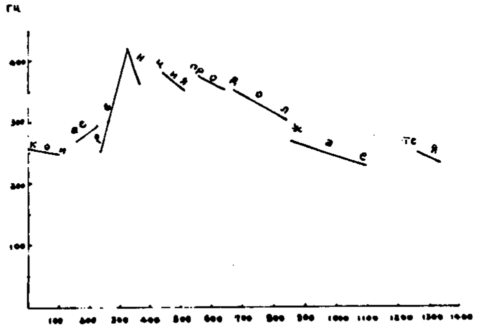


схема 1 Конференция продолжается.

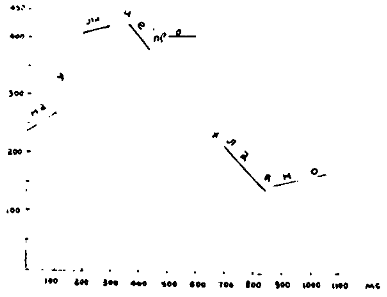


схема 2 На улице прохладно.

В речевой практике дело обстоит гораздо сложнее: встречаются разнообразные интонационные контуры. Пример: Молодой человек/окончил среднюю школу (см. схему 3). Тон

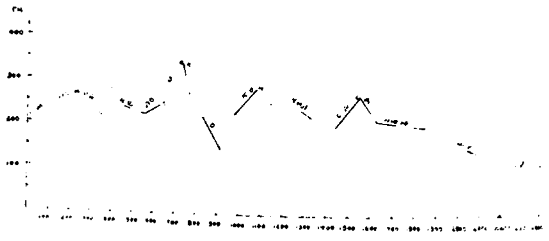


схема 3 Молодой человек окончил среднюю шкслу.

в части ремы с первого слога резко падает, обозначая границу между темой и ремой. Однако, мелодика всей части ремы получает непоследовательно нисходящую форму. Это отличается от данных выше примеров. В примере (см. схему 3) слова из ремы "окончил" и "среднюю" имеют свои восходящий и нисходящий тоны. Лишь в части интонационного центра мелодика становится окончательно нисходящей. Следует отметить, что и в таком случае высота звуков в ударных слогах первых двух слов из части ремы стоит все-таки ниже в сравнении с ударным слогом последнего слова из части темы. Примеры разнообразных интонационных тонов мелодики темы, ремы можно продолжить (см. схему 4).

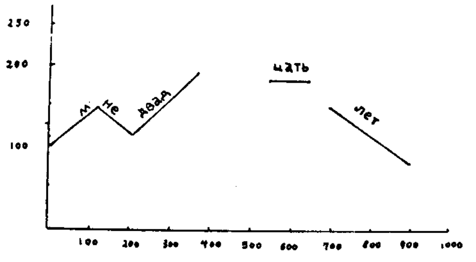
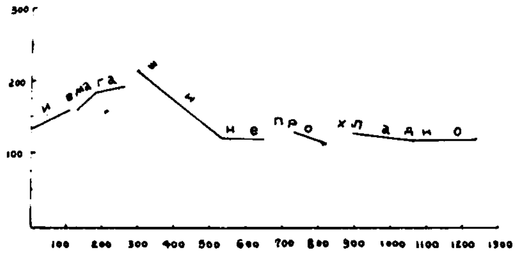


схема 4 Мне двадцать лет.

Среди зависимых вариантов, которые семантически и структурно не самостоятельны и не мотивированы в отрыве от контекста, различаются два случая: инвертированная конструкция и нормативная конструкция. В инвертированной конструкции рема предшествует теме, и интонационный центр находится в первой части предложения. В таком случае интонационный контур приобретает совершенно иной характер. Тон в предцентровой части восходит вверх, а в самом интонационном центре тон падает вниз, и в части темы сохраняется это нисходящее направление. Например: И в магазинс / прохладно. И собрание / продолжается. (см. схемы 5, 6). Мелодика в части темы приобретает ровную форму нисходящего движения. Итак в аудиальном восприятии образуется четкая, но резко отличающаяся от предыдущих предложений граница между ремами и темами. (ср. схемы 5 и 2, 6 и 1).

Нужно особо подчеркнуть отличие таких инвертированных конструкций предложений от нормативных конструкций зависимого типа односинтагмных предложений. Ср. пример: И тогда/будет капустник (см. схему 7).



смеха 5 И в магазине прохладно.

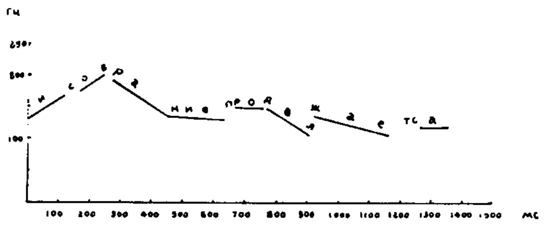


схема 6 И собрание продолжается.

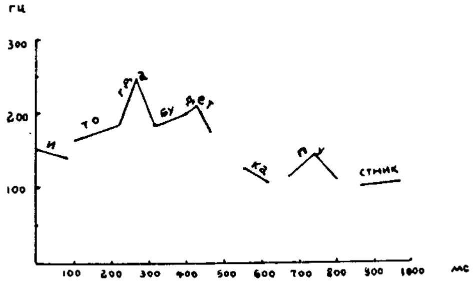


схема 7 И тогда будет капустник.

Анализируя все случаи актуального членения многосинтагмных предложений мы пришли к выводу, что актуальное членение обычно соответствует членению синтагматическому; Граница между темой и ремой именно и есть рубеж двух синтагм.

Что касается интонационного контура, то дело обстоит также как для односинтагмных предложений. То есть мелодика в части темы восходит вверх, а характерной чертой мелодики в части ремы является нисходящее направление (см. схемы 9, 10). Резкое падение между двумя синтагмами создает у слушателей впечатление актуального членения речи. Следует обратить внимание на начало интонационного контура ремы, в котором образуется своя дугообразная форма. Она обычно начинается с первого ударного слога слова ремы (ср. схемы 8, 9). Когда ударному слогу первого слова ремы предшествуют неударные слоги, то они произносятся с нисходящим направлением тона(см. схему 9).

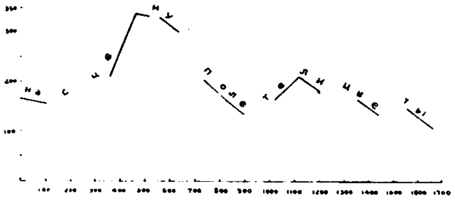


схема 8 На $\overset{6}{\text{сцену}}$/полетели $\overset{1}{\text{цветы}}$.

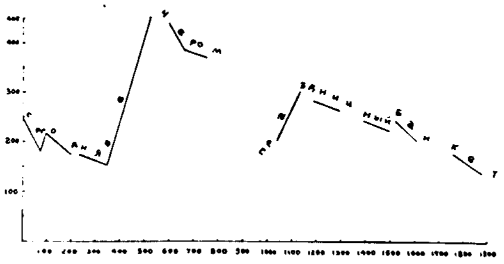


схема 9 Сегодня $\overset{3}{\text{вечерм}}$ /прздничный $\overset{1}{\text{банкет}}$.

Обычно в русской речи центр интонационной конструкции стоит на последнем слове ремы. Однако, в зависимости от разных семантических и структурных факторов центр ИК переходит на предыдущюю позицию. Тогда и интонационный контур после центровой части ремы изменяется, хотя общее направление мелодики все еще сохраняется.

Из сказанного выше мы можем прийти к заключению, что тема выражается исключительно в восходящем направлении мелодики, которая может реализоваться как ИК-3, ИК-4, или ИК 6 (см. схему 10), что рема проявляется в принципе с нисходящим направлением мелодики, если после ремы не следуют другие синтагмы. Рема лишь под влиянием семантических, эмоционально-экспрессивных факторов приобретают восходящюю форму ИК.

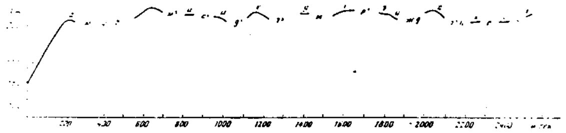


схема 10 Нам с вами сидеть у моря и ждать погоды.

Мы выдвигаем следующие положения:

——Актуальное членение всегда имеет свое интонационное выражение, которое создает аудиальное восприятие и таким образом способствует успешному процессу коммуникации.

——Граница между темой и ремой представлена в резком падении первой части мелодики. Здесь различаются два случая. В нейтральных конструкциях мелодика в части темы обычно восходит, а в части ремы нисходит. В инвертированных конструкциях с центровой части мелодика начинает уже нисходить, и в части ремы продолжается это нисходящее направление.

——Если тема представлена в односложном слове, которое притом не приобретает семантического веса, то она может терять свойственный ей интонационный контур и присоединяться к последующей за ней реме, образуя общее восходящее направление мелодики.

——В многосинтагмных предложениях актуальное членение соответствует членению интонационному. Граница между темой и ремой и есть грань между двумя синтагмами.

——В многосинтагмных предложениях мелодика в части темы всегда реализуется в восходящем направлении,представленном в ИК -3, -4, -6. Мелодика в части ремы обычно проявляется в нисходящем направлении. В зависимости от семантических и эмоционально-экспрессивных факторов интонация может приобрести и восходящую форму.

ВЛИЯНИЕ АКТУАЛЬНОГО ЧЛЕНЕНИЯ НА ИНТОНАЦИОННОЕ ОФОРМЛЕНИЕ РУССКОГО ЧАСТНОГО ВОПРОСА

REINHARD WENK

Sektion Slawistik
Humboldt-Universität zu Berlin
Berlin, DDR 1086

РЕЗЮМЕ

Актуальное членение определяет центрированность мелодического контура русского частного вопроса.
В зависимости от наличия новых элементов /кроме вопросительного слова/ двухцентровый контур проявляется также в вопросах без эмоциональной окраски.

АКТУАЛЬНОЕ ЧЛЕНЕНИЕ ВОПРОСОВ

Прежде чем подойти к изучению влияния актуального членения на интонационное оформление высказывания, нам надо определить, что мы будем понимать под " актуальным членением ", ведь уже I5 лет тому назад Т. М. Николаева /I, 54/ писала: " Актуальное членение, представленное в современной лингвистической теории - это конгломерат различных подходов к гетерогенным языковым явлениям, разъединить и описать которые, может быть, настало время."
Две концепции актуального членения, идущие еще от Матезиуса /ср. 2, 23/, а именно дистинкция " тема " /о чем говорится/ - " рема " /что говорится о теме/ и дистинкция " известное " /" данное "/ - " новое ", большинством лингвистов сегодня уже различаются, и требуется, чтобы изучалось соотношение между ними, напр. /3, 45/, /4, 67/.
Но для этого нужно будет определить все компоненты актуального членения. В то время как " известное " и " новое " можно определить по лексическому составу и конситуации /спрашивается, что известно или ново для адресата/, определение " темы " или " ремы " не так уж просто. Предлагаются разные тесты с " ингерентными " вопросами или с отрицанием. Но трудности начинаются там, где пробуется применять такие тесты к коммуникативному типу " вопрос.". Если актуальное членение признается универсальным явлением /5, 44/, то средства для определения его компонентов должны функционировать в каждом языке и в каждом коммуникативном типе. Поэтому мы думаем, что как универсальное актуальное членение надо признать именно дистинкцию " известное " - " новое ". Именно в этом различении содействует интонация /5, 53/. Но этой дихотомии мало, потому что ее компоненты не исключают друг друга: и " новое " может быть уже " известным ". Поэтому мы приняли и применяем концепцию актуального членения как коммуникативно-прагматического членения, предложенную лингвисткой ГДР Б. Хафтка /6, I60/, и различаем следующие компоненты:

I. " Известное " / в смысле " известно " и одновременно " не ново "/
Сюда относятся
- элементы, которые известны только несредственным предупоминанием,
- элементы, которые однозначно имплицированы.

2. " Известное новое " /в смысле " известно " и одновременно " ново "/
Сюда включаем
- общеизвестные элементы /известные принадлежностью к классам или уникаты/,
- элементы, которые известны только из ситуации,
- элементы, которые неоднозначно имплицированы,
- элементы, которые снова упоминаются после ввода других элементов.

3. " Новое " /в смысле " неизвестно " и одновременно " ново "/. Сюда относятся все элементы, упоминающиеся впервые и которых нельзя вывести ни из контекста, ни из ситуации.

Как уже говорилось, все эти элементы можно определить по конситуации и, может быть, они отражают те три разные степени смысловой важности, которые легли в основу исследований И. Г. Торсуевой /ср. 7, 37-38/.
Но в отличие от Б. Хафтки и большинства других лингвистов мы - на основе исследований Т. М. Николаевой /8/ - не включаем в актуальное членение явления акцентного выделения, потому что их семантическое различие - надо полагать - должно отражаться и в интонационном контуре: для выражения акцентного выделения применяются другие интонационные средства, чем для выражения актуального членения. Это сразу же видно /вернее: слышно/ из примеров Т. М. Николаевой: " Я прошу дать мне пальто." - " Я прошу дать мне пальто"/а не../ /см. 8, 39/. В этом сообщении мы ограничиваемся проблематикой частных вопросов

без акцентного выделения.

Если наш подход к актуальному членению является коммуникативно-прагматическим, мы должны таким же образом определить и суть частных вопросов. Задавая частный вопрос, говорящий располагает определенными сведениями, только одной информации у него нет, которую он заменяет вопросительным словом. Этим отличается модальное исходное положение говорящего от того при постановке общего вопроса, когда он выражает свою неуверенность в том, о чем он спрашивает, см. /9/. Таким функциональным определением частного вопроса мы отграничиваем его от других видов " вопроса с вопросительным словом ", как переспрос, повторение вопроса при ответе и т. п.

По предложенной еще Ш. Балли классификации /I0, 48/ надо различать общие и частичные частные вопросы: " В чем дело? ", " Что случилось? ", но " Кто вышел? ". Исходное положение говорящего при постановке этих двух разновидностей частного вопроса Балли, указывая на свои примеры, описывает так /I0, 48/:

для общих частных вопросов - "... знают, что что-то произошло, но не знают, что именно..",

для частичных частных вопросов - "...знают, что кто-то вышел, но не знают, кто именно..". Оказывается, что Ш. Балли своей классификацией различает разное исходное положение говорящего /спрашивающего/, а не разное актуальное членение, при помощи которого говорящий создает информационные предпосылки для ответа у адресата, т. е. спрашивающий предполагает, что в своем вопросе в данной конситуации является известным или новым или известным новым для адресата, и соответственным образом оформляет его интонационными средствами. В этом смысле надо отвергать мнение Т. П. Ломтева, что в частном вопросе отсутствует актуальное членение, потому что якобы актуализированным элементом является только вопросительное слово /II, 2II/. Одновременно можно установить, что классификация Балли не отражает актуального членения и поэтому не годится для анализа интонационных средств актуального членения, см. /9/. Понимая и принимая актуальное членение как коммуникативно-прагматическое членение в известное, известное новое и новое, мы приходим к другому критерию для различения разновидностей частного вопроса, который имеет значение для их интонационного оформления, а именно: Содержатся ли в частном вопросе, кроме вопросительного слова, еще известные новые и/или новые для адресата элементы? Частные вопросы, для которых надо отвечать на этот вопрос положительно, мы называем " сообщающими ", а остальные - " простыми ". Таким образом, простые частные вопросы содержат известные элементы и только один новый элемент /вопросительное слово/, а сообщающие, кроме него и известных элементов/ - минимум еще один новый или известный новый элемент. Так как новые и известные новые элементы могут различаться по своему интонационному оформлению, надо для наших целей ввести еще одну разновидность частного вопроса, которая отличается тем, что, кроме /нового/ вопросительного слова и известных элементов, в нем содержатся только известные новые элементы /или минимум один/. Такой тип стоит на границе между сообщающими и простыми частными вопросами, и поэтому мы его называем " медиальным ". Вопросительные слова, которые произносятся не в начальной позиции, мы считаем выделяемыми - они находятся вне автоматизированной позиции, ср. /I2, 29/, /I3, I84/ - и поэтому исключаем такие частные вопросы из данного сообщения. В медиальных частных вопросах /ЧВ/ интонационное оформление зависит в высокой степени только от того, как говорящий оценивает предпосылки для ответа у адресата, т. е. от того, помнит ли он еще ту старую информацию, которую он раньше получил. Поэтому такие ЧВ проявляются в речи или в форме простого или в форме сообщающего.

ЦЕЛИ И МЕТОДЫ НАШЕГО ИССЛЕДОВАНИЯ

В рамках исследования интонационных средств для выражения актуального членения в славянских частных и общих вопросах мы выбрали из драмы А. П. Чехова " Иванов" 22 ЧВ в больших контекстах. Они были записаны на ленту в лабораторных условиях 4-мя дикторами-носителями русского языка в конситуациях, а потом подвергнуты аудиторскому анализу. 7-II аудиторов прослушивали все реализации ЧВ, устанавливая

- без контекста: коммуникативное намерение и эмоциональную окраску, главное ударение, ритмическое членение, ядро и другие важные слова вопроса,
- с контекстом: правильность мелодики, позиции главного ударения, интонационного оформления других важных слов и выражение эмоции.

Данные этого анализа были обработаны статистческими методами, в результате чего мы получили интересные наблюдения, которыми мы ограничиваемся в этом сообщении.

РЕЗУЛЬТАТЫ АУДИТОРСКОГО АНАЛИЗА

I. В принятых аудиторами, т. е. правильно интонированных, реализациях испытуемые указывали на анкетах, что они воспринимали " главное ударение "

- в простых ЧВ на вопросительном слове,
- в сообщающих ЧВ не на вопросительном слове./см. таб. I/

2. В сообщающих ЧВ аудиторы в большинстве случаев слышали " главное ударение " на последнем новом элементе в высказывании. В простых ЧВ такой тенденции нет./таб. 2/

3. Слуховой параметр " особенное выделение " в аудиторских ответах дает нам аналогичную картину. Но здесь испытуемые нередко указывают на два /или даже три/ таких выделения в ЧВ, а именно

- меньше всего в простых ЧВ,
- знаменательно чаще в сообщающих ЧВ/таб.3/

Подавляющее большинство сообщающих ЧВ, которые аудиторы характеризуют двумя центрами, имеет их на вопросительном слове в начальной позиции и на последнем новом элементе.

Еще более однозначно эта тенденция к двухцентровости проявляется, если соединяем все аудиторские оценки, которые могут указывать на интонационный центр, а именно оценки по слуховым параметрам " особенное выделение ", " главное ударение " и " ядро вопроса ", и рассматриваем их в отношении к вопросительному слову и к последнему новому элементу или - в простых ЧВ - к последнему известному элементу. /При этом указанные параметры всех реализаций оценивались одним и тем же числом аудиторов./таб. 4/

4. На два интонационных центра указывают и мелодические характеристики сообщающих ЧВ /пока только по слуховым наблюдениям автора, который прослушивал все реализации с половинной скоростью/. В сообщающих ЧВ ударный слог вопросительного слова, т. е. первого нового элемента, произносится на среднем уровне /мы различаем низкий, нейтральный, средний и высокий уровни/. Следующие слоги остаются на этом уровне до предударного слога последнего нового элемента, откуда начинается нисходящая мелодика. Значит, на втором центре употребляется интонема, которую Е. А. Брызгунова называет " ИК-I ". Такую мелодику мы здесь условно называем " первым типом ".

В редких случаях некоторые аудиторы приняли и нисходяще-восходящую мелодику сообщающих ЧВ /тип второй/. Тогда второй центр характеризуется употреблением интонемой по образцу " ИК-4 ", а слоги между центрами - падением тона. Но число отвергающих оценок гораздо выше, чем у первого типа.

Чисто нисходящая мелодика с одним центром на вопросительном слове /тип третий/ у сообщающих ЧВ проявляется и принимается аудиторами довольно редко, тогда как этот тип предпочитается и чаще всего принимается для простых ЧВ. /таб. 5/

Распределение мелодических типов по простым и сообщающим ЧВ очень знаменательно разлиается.

5. Известные элементы мелодически вообще не характеризуются, т. е. никаких предударных или заударных /хорошо слышимых/ интервалов не наблюдается. Наоборот, если даже говорящий допускал такие мелодические характеристики для известных элементов, аудиторы это считали неправильной мелодикой.

Известные новые элементы при употреблении первого типа оформлялись по усмотрению говорящего: если он считал их известными для адресата, не было никаких интервалов, а если он их считал новыми, проявлялись маленькие нисходящие, иногда и восходящие предударные интервалы.

6. Интересную картину мелодического оформления можно было наблюдать у медиальных ЧВ: Если говорящий известные новые элементы считал известными для адресата, он оформлял медиальный ЧВ как простой, т. е. элементы известные новые мелодически никак не характеризовались. А если он их считал новыми, такой вопрос получил мелодическую форму с двумя центрами по мелодике первого типа.

7. Все реализации были оценены аудиторами и по эмоциональности. В их ответах проявляются две тенденции:

- у простых ЧВ двухцентровое оформление указывает на эмоциональную окраску,
- у сообщающих и медиальных ЧВ такое оформление больше всего воспринимается как эмоционально не окрашено./таб. 6/

ВЫВОДЫ

Результаты и наблюдения нашего /предварительного/ аудиторского анализа говорят в пользу того, что актуальное членение частного вопроса /при исключении акцентного выделения/ как в других языках, так и в русском языке влияет на интонационное оформление такого высказывания.

Простые частные вопросы показывают тенденцию к одноцентровому контуру, причем этот интонационный центр находится на единственном новом элементе, т. е. на вопросительном слове. В этой разновидности частного вопроса второй центр /на известном элементе/ проявляется только тогда, когда выражается эмоциональная окраска.

Сообщающие частные вопросы претедуют на двухцентровый контур, который характеризуется первым центром на вопросительном слове в начальной позиции и вторым центром на последнем новом элементе, причем только на этом втором центре проявляется интонема, а именно интонема коммуникативного типа " повествование " с чисто нисходящей мелодикой. Таким образом, двухцентровый контур является характеристикой не только частных вопросов с эмоциональной окраской, ср. /I4, 32/. Нейтральные и эмоционально окрашенные частные вопросы с двумя центрами различаются, по-видимому, употреблением разных интонем именно на втором центре.

Медиальные частные вопросы могут реализоваться или как простые, или как сообщающие. Это зависит от того, как говорящий оценивает осведомленность адресата: известные новые элементы могут быть еще актуальными в его памяти или нет. Употребление двухцентрового контура тогда, когда известные новые элементы считаются новыми, и одноцентрового контура, когда эти элементы считаются известными, подкрепляет наши наблюдения.

Использованная литература

I. Т. М. Николаева
Актуальное членение – категория грамматики текста. In: Вопросы языкознания /I972/ 2, стр. 48-54.

2. J. Firbas
Some Aspects of the Czechoslovak Approach to Problems of Functional Sentence Perspective. In: Papers on Functional Sentence Perspective. Prague 1974, p. 11-37.

3. И. И. Ковтунова
Современный русский язык: Порядок слов и актуальное членение предложения. Москва I976.

4. J. Silić
Od rečenice do teksta. Zagreb 1984.

5. M. A. K. Halliday
The Place of FSP in Lingustic Description. In: Papers on Functional Sentence Perspective. Prague 1974, p. 43-53.

6. B. Haftka
Bekanntheit und Neuheit als Kriterien für die Anordnung von Satzgliedern. In: Deutsch als Fremdsprache 15 (1978) S. 157-164.

7. И. Г. Торсуева
Интонация и смысл высказывания. Москва I979.

8. Т. М. Николаева
Семантика акцентного выделения. Москва I982.

9. R. Wenk
Intonatorische Mittel zum Ausdruck der aktuellen Gliederung in slawischen Entscheidungsfragen. In: Zeitschrift für Slawistik 33 (1988) (im Druck)

10. Ch. Bally (Ш. Балли)
Общая лингвистика и вопросы французского языка. Москва I955.

II. Т. П. Ломтев
Грамматическое и логическое в предложении. In: Исследования по славянской филологии. Москва I974.

12. F. Daneš
Intonace a věta ve spisovné češtině. Praha 1957.

13. R. Wenk
Die Intonation. In: Die russische Sprache der Gegenwart, Band 1. Hrsg. K. Gabka. Leipzig 1984, S. 175-198.

I4. И. Л. Муханов
Пособие по интонации для филологов старших курсов. Москва I983.

Статистика

таб. I

главное удар. на вопр. слове	+	-
сообщ. ЧВ	I20	2I6
прост. ЧВ	96	8

таб. 2

главное удар. в сообщ. ЧВ	+	-
на вопр. сл.	I20	2I6
на посл. нов.	202	I34
на друг. нов.	26	3I0

таб. 3

особ. выдел.	одно	больше
в сообщ. ЧВ	293	96
в прост. ЧВ	II2	I4

таб. 4

особ. выдел. + главн. удар. + ядро вопроса	на вопр. слове	на посл. нов./изв.
в сообщ. ЧВ	394	609
в прост. ЧВ	276	40

таб. 5

число реализ.	тип I	тип 2	тип 3
в прост.	2	0	I0
в сообщ.	2I	6	6

таб. 6

Оценки для реализ. с 2 центрами	эмоц. окраш.	эмоц. не окраш.
прост. ЧВ	I3	I
сообщ. ЧВ	83	I4I

THE PHONOLOGICAL COMPONENT OF LANGUAGE COMPETENCE: SPEECH PRODUCTION AND PERCEPTION IN ONTOGENESIS

A.M. Shakhnarovich

Institute of Linguistics,
Academy of Sciences, USSR
103009, Moscow, K-9,
Semashko st., 1/12

Galina Ivanova

Moscow State Institute of
Foreign Languages
119034, st. Ostozhenka, 38.

The phonological component of linguistic competence is considered. Experimental results concerning the comprehension of sound text are shown.

In psycholinguistic research of speech ontogenesis, quite a few aspects of language competence have been studied so far. The competence in question is understood as the hierarchically ordered system (construction) which represents, in generalized, reduced and specific way, the overall language system. Since in the language competence components and rules are distinguished, the former correspond to the linguistic levels, while the latter constitute a system of "commands". The "commands" prescribe that the speaker should use the given, functionally significant element, for meeting the communicative requirement in the respective communication environment. So far the linguists' attention has been focussed mainly at the lexical and grammatical components, the semantic component being insufficiently studied. The latter is, however, the basic, hierarchically predominant part of the structure. Finally, the phonological (phonetic) component has been almost ignored.

Speech in children comprises specific system of elements accompanied by particular rules of using these elements. Mere phenotypical resemblance of speech elements in children and in adults will be insufficient for claiming the respective genotypical similarity or relationship.

In the peculiar system which is the starting point for the formation and development of man's linguistic capacity, much importance is given to the phonetic component. This is natural, for the system mentioned seems to comprise, at the primary stages, phonetic and semantic components only. As for the expressions resembling the grammatically marked elements, these are actually grammatical from the formal viewpoint, for they are related to the denotational reality and are characterized by a specific combination of meanings. Speech ontogeny starts with the child's primary orientation in the external speech, and therefore, first developmental stages are bound, psycholinguistically, to the phonetic word perception (subtle morphonological distinctions being initially ignored). The constancy of phonetic word perception is ensured by the formation of certain standards, these making up the basis for the phonetic component system. The standards get enriched through the child's work at the phonemes in word, for, as D.B.El'konin put in, he manipulates the phonemic just as he acts upon material objects. Initially, distinct pronunciation can be reduced in the distinct articulation of inflexions only, which testifies to the child's orientation to the sound word. Alongside with the orientation in speech perception medium, the system of standards is enruched due to the development of child's cognitive activity. Thus, filled with new cognitive content, perception standards emerge as part of the system designed to quality the child's environment. Meanwhile, in the speech production domain, the relation of the sound word with the real world semantics is tested and refined.

The major route that speech activity develops by in the formation of generalized ideas of the language reality facts. Phonetic generalizations being the goal, the first stage in distinguishing the sound form of the inflexion (and later of morpheme) and relating its meaning to some aspect of material world. The relation established, both perception and cognitive standards are enriched; therefore, the orientation is stepwise transferred to other parts of the word.

Hence, the second stage in the phonetic component formation is initiated, that is the formation of generalized conceptions correlated to the sounding morphemes which are viewed as derivative word components. This stage proceeds mainly in the speech production domain. Here the acquistion of grammatical meaning starts, accompanied by the semantic development of the derived component and of the overall lexical item.

However, speech ontogeny researchers note that certain words acquire quite early the utteranca status and, therefore, obtain specific intonation marking. Intonation parameters are "tested" rather early and are also due to perception and then to production. Generalized intonation regularities are included further in the phonetic component of language competence.

Prosodic development is most evident in two-word utterances; the transition to these characterizes a specific stage in the overall process of speech development. Early intonation standards are rather sophisticated and advanced; sometimes it is the intonation contour that outranks the semantics of the components. The transition to bi- and multicomponent utterances is followed by the acquisition of logical stress. This phenomenon seems to be beyond the phonetic component limits, for, included in the general set of communicative rules it should belong to the semantic component.

The development of speech perception and production is directly related to the formation, advancement and enrichment of the phonetic component of language competence. This component is realized through a set (system) of specific phonetic perception standards and rules selecting the appropriate elements for meeting the communicative requirement. The development and enrichment of the lexical meaning is ensured both by the child's practical/communicative activity and by the acquisition of phonological generalizations to the delimitation, recognition and distinction of the word. This process is essentially bound to the development and enrichment of the morphological meaning, too. The sound shape of the morpheme is related, directly or indirectly, to the respective aspect of the real world. Thus the basis for presxriptive formational and derivational rules is laid, these rules belonging to the grammatical component (or, to be more precise, to morphological subcomponent) of language competence.

At the later stages of the child's development. speech perception is determined by a richer and more complex system of standards; the standards are further formed and enriched in practical and communicative activities. The phonetic component can be said to acquire its relatively full shape at the stage of adequate discourse perception.

The problem of discourse (cohesive text) perception is related to the issue of acoustic signal determination principle. Some scholars view perception as the stepwise successive process. Another view, which we adhere to, is that man memorizes relatively much information at once, processing it parallelly, quite in accordance with the text/discourse hierarchical structure. Psycholinguistic experiments show that, in speech understanding, the delayed decision strategy is commonly used. At the lower level which is that of phonology, zero decision is made and means for making the final decision are reserved. The final decision is thus postponed to the end of the sounding reflexion /1/.

Establishing meaningful elements in the specch flow segmentation, and minimal meaningful elements in particular, is a most challenging task. Let us assume that speech information is processed by the "dissecting" and intergrating techniques, with "key meaningful points" playing crucial role here. Then the main problem is eliciting major sense fragments and/or blocks functioning in the perception of oral text/discourse.

At the perception level, the major units of meaningful segmentation are taken to be syllable and phonetic word (rhythmic structure). The syllable is defined as the structural sequence of sound-types formed by the vowel sound'types /2/. "The data available show that the phoneme, being the necessary element in word recognition and construction, is "deduced", in some cases, due to the processing of more or less extensive context" /3/. In our opinion, such context can be rendered by the syllable and by the phonetic word, the latter defined as a group of syllables united by one word-stress. In the phonetic word the kernel and the periphery are distinguished. The kernel is related to the stressed syllable, including sometimes pretonic syllables too (especially the first prestressed). In the rhythmic structure strong and weak positions are often recognized; the stressed syllable can be considered the nucleus of the strong position /1/, which is highly relevant for the points made in this study.

Thus, the hypothesis was advanced that the perception of sounding text/discourse should presuppose discrete segmentation in meaningful units. The hypothesis predetermined both the technique and planing of the experiment in which text perception and understanding in children were studied.

A number of experimental series were undertaken, the subjects being preschoolers in a Moscow kinder-garten. The study of speech in children allowed us to follow speech processes in their evolution.

Numerous works show that the acquisition of Russian morphology in children is based on the development pf orientation in the sound shape of words. Primarily the child is guided by the general sound properties of the morpheme, and it is only later that he begins distinguishing separate phonemes in it. Thus the child's original vocabulary comprises root words rather than items consisting of separate sounds. This implies that the child segments out syllables both in hearing and in speaking. Therefore, in experiments with children we take syllables and rhythmic structures as perception pivots. In an experiment made by our colleagues /3/, intuitive syllabation was employed. This technique seems most promising since it allows the research to take into account the specific speaker's peoperties, say his age, etc. In our experiment intuituve segmentation of the ongoing speech by meaningful units is employed.

Other experiments have shown that syllabation in 4,5- and 6-year-old children differs from the respective process in adults. A suggestion has been made that it could be explained by some ontogenetic interference of property phonological and morphological factors. It can be also proposed that syllabation skills are unstable, changing with the child's acquisition of his first language.

Our experiments also test the child's ability of intuituve segmentation of the sound text by phonetic words (rhythmic structures), and the rules of this segmentation. The above mentioned facts taken into account, the hypothesis is formed, according to which the rules of text segmentation by rhythmic structures in children differ from the respective rules in adults. Children's perception and understanding is tested on texts differing in the functional style and structure; besides, the dependency of the understanding on structure; perception and understanding developmental skills are investigated.

The experimental data have been processed and analyzed by comparing the actual content and intonation structure of the sample text on the one hand and the subjective content and intonation structure modelled by the subject in the reproduction, on the other. Content segmentation as made by a professional announcer has been compared to that in children, who acted intuitively. The degree of similarity between the respective structures testifies to the acquired level of the text understanding. Besides, it helps us reveal some intuitive rules of text segmentation by meaningful components.

References

1. Zlatoustova L.V. 1981. Foneticeskije jedinitsy russkoj reci. /Phonetic units of Russian speech/. Moskva: MGU.
2. Bondarko L.V. 1975. Fonemnoje opisanije vyskazyvanija - uslovije i rezul'tat ponimanija teksta./The phonological description of utterance as condition and result of text understanding/. - In: Materialy V Vsesojuznogo simpoziuma po psixolingvistike i teorii kommunikacii. Cast' 1. Moskva.
3. Vinarskaja E.N., Lepskaja N.I., Bogomazov G.M. 1977. Pravila slogodelenija det'mi i slogovyje modeli (na materiale detskoj reci). /Syllabation rules in children and syllabic patterns (speech in children)/. - In: Problemy teoreticeskoj i eksperimental'noj lingvistiki. Moskva: MGU.

THE SENSOR-MOTOR THEORY OF CONSTANT SPEECH PERCEPTION

GEORGE LOSIK

Dept. of Experimental Phonetics
Institute of Linguistics,BSSR Academy of Sciences
Minsk, Byelorussia, USSR.

ABSTRACT

New aspects of sensory and motor theories of speech perception are discussed. The possible mechanisms of constant speech perception by grown-ups as well as mechanisms of stuttering are proposed. In addition, the role of babble-speech for ear-development is analysed. Finally, the results of experimental verification and computer testing of a new sensor-motor theory are reported.

There exists a number of theories of speech perception. The well-known of them are sensory (G.Fant, M.Halle, R.Jakobson) and motor (A.Liberman, F.Cooper, L.Chistovich) theories. But the interest to the verification of these theories is dicreasing recently.

We suppose that there exists a false opinion that the speech signal informativity in these models was taken into account (N.Zagoruiko). Here we offer a new sensor-motor theory of speech perception which explains the mechanism of speech perception by grown-ups and by children. It also explains the formation of samples perception of speech units by a child during the latent interaction of audio-speech and motor systems.

Sensor-motor theory differs from other theories known recently according to the following theses.

Thesis 1. The invariant characteristics of speech units are not present in a common acoustic signal. They appear only on the first stage of speech processing after the signal normalization in audio-speech system.

Nevertheless a new informational characteristic is present in this acoustic signal, making the speech recognition invariant. This characteristic reflects the speech unit variations. A man analyses and memorizes these characteristics at an early age. Thanks to this new information a grown-up reconstructs an initial signal from distorted one during the recognition. Only after this process the reconstructed signal is compared with the sample. We confirm that speech units differ in manner of variation (Δy-description), but not only by their motor program (y-description). Δy-information forms an additional characteristic axis. Recognition in audio-speech system is realized with the help of Δy-description but without the y-description itself.

Thesis 2. The information about pronounciation variation of speech units is memorized in the audio-speech system, but not in a motor one according to a motor theory. Finally, the distortion model of articulation system is formed in the audio-speech system.

Thesis 3. A child can not reveal the variation rules of a speech signal listening to the speech of grown-ups. The acoustic signal is an indefinite function, which is formed by a countless multitude of a small number of argument combinations, i.e. speech commands.

Therefore speech of grown-ups is not used for learning the speech unit variation rules. Only elementary signal variations not the combinational ones are used by children for memorizing. Only directions of variations not the zone of variations (S.Dzhaparidze, I.Zimnyaya) are informational for memorizing.

Hence a child must reveal these rules indirectly. To learn the articulation distortions in the speech of grown-ups he imitates these distortions in his own motor-speech system.

Thesis 4. The repetitions of sounds in babble-speech (e.g. va-va-va,ba-ba-ba, ma-ma, pa-pa) are the acoustic signals, which reflects child's imitation of articulation distortions. These repetitions in babble-speech (i.e. iterations) are initiated by child's audio-speech system, but are accomplished in the motor system according to its degrees of freedom. In the state of motor system jumps are initiated. We think the iteration of the type "va-va" reflects such jump on the acoustic level. Child's auditory system perceives and analyses the multitude of iterations in his own babble. Finally this system accumulates the information about the direction of all elementary jumps. In this process a pair of acoustic signals, not a single one, becomes informative.

Thesis 5. The information about articulation distortions is mastered by a child in his early childhood only, but later it is used by him constantly.During the recognition this information enables him in the sensory system to modulate the multitude of motor work divergences and to realize the selective fitting of the input signal to the sample. During this fitting the statistical information about signal variations enables a child to discover the most probable track in which ree alizing distortion took place. The degree of additional information depends on the probability of track-fitting of input signal to sample, but not on the fact of their fitting. We confirm, that after this sensory learning the motor system is not necessary for recognition.

According to the sensor-motor theory a child forms sound image of a speech unit in his audio-speech system in three stages. During the first stage a child perceives the pronunciation of a word or a syllable articulated by grown-ups. He listens to many separate realizations of a speech unit and forms its average sound sample. In the speech of grown-ups the child hears distorted to a small degree realizations of a speech unit. Therefore they are grouped with little dispersion about an average value. During the second stage a child adopts the skill of pronunciation of speech units, the samples of which were formed in the auditory system. As a result, in the speech-motor system the motor samples of speech units are formed. The thesis about the existence of the third stage is new. At this stage the sensory system receives information about the modifications of those samples, which were formed in it. A child listens to the iterations of his own babble-speech. The development of iteration mechanism in the babble period consists of several substages. At the first one the child exercises the articulation of samples. Here the development of babble is going through without ear participation. Therefore this substage is presented in the babble-speech of deaf children. During this period the inborn program of articulation exercises under cinesthetic control is realized (V.Beltyukov). Therefore at first the child's babbles are realized without iterations in any acoustic situation and not only in stillness. At the next substage the child's ear begins to control the sounds of his own babble. The auditory system adopts a single babble syllable which is pronounced at this moment. Next moment the system stimulates the repeated pronunciation of this syllable. This substage is absent in the babble-speech of deaf children. On the contrary, by the children with ear perception the phenomenon of autoecholaly is developed. This phenomenon is realized in stillness mainly. The babble sounds are combined into iterational chains (va-va-va). However the autonomous mechanism of program repeated triggering is gradually formed in the motor system. After that this mechanism generates the iterations itself without ear participation.

The variation imitation in the motor system becomes possible just at this period. The direction of possible sample variations starts to be coded in double iterations. To our mind, the iterations' mechanism, that is formed with child during his babble, is kept then with grown-up in the blocked state for the whole life. This mechanism can manifest itself again. One of these manifestations is the clonical form of stuttering and Lee effect. Besides the iterations' mechanism promotes indirectly that word-iterations (mamma, papa) take the main place among the first child's words in many languages.

There are some confirmations of sensor-motor theory. One of them is the fact of infringement of ear development by a child caused by the blockade of the babble stage during the speech development. The blockade of listening to iteration in this period is manifested in the underdevelopment of the phonematic ear of child (V.Beltyukov).

We have conducted some special research of child's babble-speech, the results of which prove sensor-motor theory. Firstly, the results,received, that babble iterations are realized during stillness mainly. Secondly, they reveal that double iteration of a syllable is most widely spread. Thirdly, it is shown that sound iterations diads are double triggers of one and the same motor-program. And at last, our results illustrate that iterations in the form of diade exist in the babble of children speaking different languages.

On the basis of sensor-motor theory a new explanation of babble function in speech ontogenesis can be given. The theory explains also sound ontogenesis dissociation in the child from the point of view not only of his speech development (V.Beltyukov), but of his phonetic ear too.

A corresponding mathematical model has been built for the presented theory. This model allows to test new algorithms of teaching and speech recognition in computer experiments on natural speech, to use them in the existing automatic speech recognition system.

A TRANSPORT GLOBULIN, SERUM HORMONE BINDING GLOBULIN, AS A PREDICTING FACTOR OF VOICE CHANGE IN PUBERTY?

Pedersen, M.F.

Vocal Fold Physiology lab.
The Danish Boys' Choir

Moeller, S.

Biostatistical Dpt.
Statens Seruminstitut
2300 Copenhagen S, Denmark

ABSTRACT

In an earlier study we found that serum hormone binding globulin was the most significant predictive factor for pubertal voice change among andrenal hormonal factors in a puberty group of boys aged 13-16 years (p<0.05). In this study we have compared the results with the ones of a group of girls in puberty. The aim was to get a possible understanding of the central biological phenomenons for the regulation of voice in puberty in a better way. The voice parameters were phonetograms and fundamental frequency measured with 2000 electroglottographic circles in continous speech. They were compared with puberty stages, adrenal hormones and sex hormonal changes in boys and girls from 8-19 years of age.

INTRODUCTION

We know very little about the central regulating proteins for voice change in puberty. The transport mechanisms for sex hormones may be involved (1,2). We have tried to examine the serum hormone transport globulin to find out whether it can predict the voice change in puberty not only in boys but also in girls.
We know that this transport globulin does fall in both sexes at time of puberty and a better understanding of the globulin might illucidate the central regulation of voice at a whole. (3,4).
Children at a singing school were analysed with voice phenomena and normal pubertal development together with androgens and oestrogens. Thereafter a statistical analysis was carried out to confirm the function of the transport globulin.

MATERIAL AND METHOD

97 children, 47 girls and 48 boys with trained singing voices from 8-19 years of age in a singing school were included in the study with randomized selection of an equal group in each school class.
The voice parameters included fundamental frequency with a computer program based on analyzing 2000 consequtive electroglottographic circles (5) in a reading situation of a balanced text, IPA book, the northwind and the sun and phonetograms (6) with areas extracted in cm^2 on a standard paper. A coversion factor to dB x Herz was 1 cm^2=32 dB x Herz.
Blood examinations for androgens and oestrogens together with somatic examination were carried out on the same day in each child, before noon and 3-6 days after 1. menstruation day where menarche had taken place.
The measurings of androgens and oestrogens were made at The Hormone Dpt. of Statens Seruminstitut. Logarithmic transformations of observations were required to obtain normal distribution. Data were investigated by one-way analysis of variance and correlation coefficients were calculated comparing all parameters. Multiple regression analysis of fundamental frequency in a reading situation and the lowest tone in the phonetograms with hormone values age and stage of puberty as independent values was carried out.

RESULTS

The change of the fundamental frequency with age is seen at fig. 1. - together with the tone range in semitones in the phonetograms from where also the lowest tones and the areas were extracted.
In table 1. the geometrical mean values for some voice parameters, puberty phenomena and measurings of hormones are seen devided in three age groups.
In table 2. coefficients in boys estimated from multiple regression of fundamental frequency depending on hormone values, age and stage of puberty after reduction of independent parameters are seen.
We have found a correlation coefficient for serum hormone binding globulin in girls in relation to menarche of -.93, which means that serum hormone binding globulin in this study has a predictive value for menarche. In table 3. the best sets of describing variables for the logarithm to fundamental frequency in running speech in girls are shown taking into account that relations are different before and after menarche.

DISCUSSION

We have made an analysis of voice (fundamantal frequency in continuous speech and phonetograms), pubertal stages and androgens together with oestrogens and found that the transport globulin, serum hormone transport globulin, was a significant predicting factor of change of fundamental frequency in puberty in boys in an puberty stage group 2-4 with a significant difference from zero by multiple regression of p<0.05. We have found the change of boys voices to happen at 14,5 years age at the same time as the serum hormone transport globulin is reduced.
In girls in puberty the change of fundamental frequency was not significantly related to the globulin, but serum hormone transport globulin showed a correlation coefficient of: r-0.93. to menarche. When the girls were devided in two groups before and after menarche, several parameters had significant relation to the change of fundamental frequency in running speech in puberty, before puberty: Hight, log (Elso4) p<0.001. and puberty stage p<0.05. - after menarche: log(variation of fundamental frequency in running speech) p<0.001., time after menarche: p<0.01. and age p<0.05.
Of course it has been difficult to set fundamental frequency in speech in relation to traditional pubertal biological changes. Taking time of beginning of menstruation onto account together with serum hormone binding globulin the fundamental frequency change in puberty possibly could be predicted in puberty. One advantage out of many might be to be able to predict to singing teachers in the famous boys choirs the time of sopraneos loosing hight, or changing timbre from child to adult. Much information of biological central regulating factors of voice can be found in studies of puberty also because the psyco-social factors do not influence this time of life to the same extend as later on.

Fig. 1.

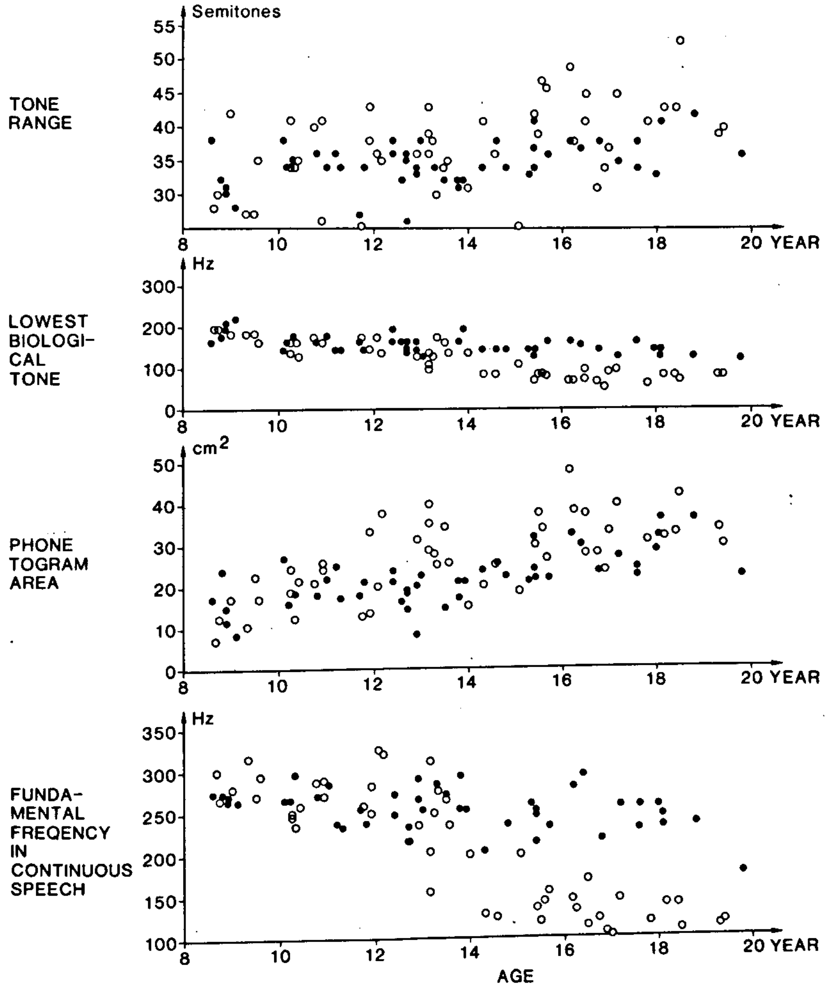


Table 1.

Geometric means of voice parameters, pediatrical measures and puberty hormone differences of 3 groups of ages, (boys and girls).

Age		8.6–12.9	13–15.9	16–19.8
Numbers of boys/girls		19/18	15/12	14/11
Fundamental frequency in speech (Hz)		273/256	184/248	125/241
Variation of fund. freq. (semitones).		3.7/3.2	4.8/4.2	5/5.2
Total tone transf. (semitones)		34.4/23	37.5/30	41.4/38
Lower tone (Hz)		158/166	104/156	72/145
Middle tone (Hz)		435/429	321/409	254/413
Phonetogram area (cm^2) (1 cm^2 = 32 dB x semitone)		19/17.3	28/21.8	34/28.3)
Hight (cm)		143/144.5	157/160	181/165
Weight (kg)		34.4/37.8	56.9/53.0	68.6/64.4
Pubic hair (stage)		1–3/1–4	1–5.5/2–5	5–6/4–6
Testis volume (mltr.)		2.3	13	20
Mamma development (stage)		1–4	2–5	5
SHBG	n mol/1	134/160	66/132.5	45/122.7
DHEAS	n mol/1	1400/3210	4100/3700	5900/7200
Delta 4 androsten dione	n mol/1	1.44/0.59	3.28/1.7	3.43/2.5
Total testosterone	n mol/1	0.54/0.50	10.5/0.76	18.9/0.94
Free testosterone	n mol/1	0.007/0.006	0.14/0.008	0.33/0.009
Dihydro testosterone	n mol/1	0.18/	1.21/	1.57/
Oestrone	p mol/1	/57	/104	/123
Oestradiol	p mol/1	/73	/135	/108
Oestrone sulphate	p mol/1	/732	/1924	/2343

Table 2.

Coefficients estimated from multiple regression of F_o depending on six hormone values, age and stage of puberty after reduction of independent parameters.

Number of boys	Stage of puberty	Geometrical mean values			Coefficient	
		$\bar{x}$ F_o Hz	age	$\bar{x}$ SHBG nmol	age	log SHBG
18	1	274	10,5	141	0.0002	0.010
11	2–4	219	13,5	91	-0.0016	0.501*
19	5–6	129	16,9	42	-0.0014	0.005
48	Total				-0.0033*	0.171*

Mean values of the remaining parameters according to grouping.
* Coefficient is signigficantly different from zero ($p<0.05$).

Table 3.

The best sets of describing variables for the logarithm of fundamental frequency in continuous speech (F_o) calculated for the whole group and for the two subgroups classified by menarche.

All girls*		Pre-menarche		Post-menarche	
	Variable P-value of t-test		Variable P-value of t-test		Variable P-value of t-test
Weight	0,066	Height	0,001	Age	0,033
Log (Tone range in speech)	0,042	Pubic hair (stage)	0,022	Time after menarche	0,008
Log (E_I)	0,054	Log ($E_I So_4$)	0,001	Log (Tone range in speech)	0,001
Log ($E_I So_4$)	0,043			Log (androst)	0,068
SE of estimation	0,034		0,0166		0,0288
SD of log F_o	0,037		0,0300		0,0409
F-test P-value	0,0443		0,0006		0,0036

Correlation coefficient SHBG, r-0.93 to menarche
* n=37 with all relevants measurings.

REFERENCES

1) M.F. Pedersen, E. Munk, P. Bennett, S. Moeller The Change of Voice during Puberty in Choir Singers Measured with Phonetograms and Compared to Andreogen Status together with other Phenomena of Puberty. Proc. X^th Congr. Phonetics Utrics 1984 pp 604-609.

2) M.F. Pedersen, S. Moeller, S. Krabbe, E. Munk, P. Bennett. A Multivariate Statistical Analysis of Voice Phenomena related to Puberty in choir boys. Folia Phoniatr. 37:1985, 271-278.

3) Sean K. Cunningham, Therese Loughlin, Marie Culliton and T. Joseph McKenna. Plasma Sex Hormone Binding Globulin Levels decrease during the Second Decade of Life Irrespective of Pubertal Status. J. Clinical Endocrinology and Metabolism vol 58, 1984, 913-918.

4) Charles E. Larson. The Midbrain Periaquaductal gray: A Brain stem Structure involved in vocalization. J. Speech Hear Research 28, 1985. 241-249.

5) Kitzing, P. Glottoghrafisk Frekvensindikering thesis. University of Lund, Malmoe, Sweden 1979.

6) Seidner, W., Shutte, H.K. Standardisierungsvorschlag. Stimmfeldmessung/Phonographie Proc. IX^th Congr. Union eur Phoniatr, Amsterdam 1981, 83-87.

ANALYSE SPECTRALE DES VOYELLES CHEZ LES SUJETS ATTEINTS DE MALADIE DE PARKINSON

ARABIA-GUIDET C

INSERM U3 "Physiologie et
Pathologie Cerebrale"
47 Bd de l'hopital
75651 PARIS Cedex 13

MANTOU A

Laboratoire "Image et Parole" et
Laboratoire de Phonetique (DRL)
Universite Paris 7
2 Place Jussieu 75005 PARIS

CHEVRIE-MULLER C

INSERM U3 "Physiologie et
Pathologie Cerebrale"
47 Bd de l'hopital
75651 PARIS Cedex 13

RESUME

Cette etude se place dans le cadre plus général d'un programme, en cours d'elaboration, d'aide au diagnostic automatise des troubles de la phonation

L'analyse du signal electroglottographique ainsi que l'analyse spectrale du signal acoustique permettent en effet, en procédant a l'etude quantitative et descriptive du timbre vocalique, d'aider a la detection de troubles, tant au niveau de la source sonore qu'a celui de l'articulation

L'etude presentee ici concerne uniquement l'analyse spectrale de voyelles /a/ du français prononcees par une population temoin en voix "normale" et en voix "faible" et par des sujets atteints de maladie de Parkinson.

Les resultats font apparaître que la fréquence des formants varie tres peu d'un groupe a l'autre. Par contre, la répartition spectrale de la puissance est significativement différente d'une population à l'autre et semble être un indice pertinent pour la detection et eventuellement la classification d'une voix pathologique.

INTRODUCTION

L'etude du timbre vocalique de sujets atteints de troubles phonatoires a deja éte entreprise par plusieurs auteurs, essentiellement dans le cas de pathologies du larynx.

Nous nous intéressons plus particulierement aux troubles lies a un disfonctionnement du système nerveux central tels qu'on peut les rencontrer chez les sujets parkinsoniens

La voix des malades parkinsoniens a été décrite par la plupart des auteurs de la littérature comme une voix de faible intensite. L'alteration du timbre a aussi ete frequemment signalee et l'élevation du fondamental est également classique ([3],[6]) L'analyse, comportant une comparaison sujets temoins/sujets malades, a donc éte complétée par une autre etude comparative avec les mêmes sujets temoins mais avant la consigne de parler d'une voix "faible"

CORPUS-SUJETS

Le corpus est constitue de deux /a/ tenus et de dix /a/ situes a l'interieur des mots suivants : /kaRe/, /fam/, /lega/, /banan/, /banan/, /ʒaze/, /gaRgarism/, /lasœR/, /kuppapje/ et /maʒik/

L'ensemble des locuteurs, tous de sexe masculin, est constitue de dix sujets témoins et dix sujets parkinsoniens.

Tous ces sujets ont prononce les dix mots. Pour les voyelles tenues, l'analyse porte sur six locuteurs parkinsoniens (PK) et sept locuteurs témoins en voix "normale" (NX). En voix "faible" (FA), quatre personnes ont prononcé les deux /a/ tenus ainsi que les six premiers mots de la liste.

METHODE

Rappelons brievement la methode employee déja décrite dans un précédent article ([1]) les enregistrements sont numerises à 16 kHz et un calcul de transformée de Fourier rapide (FFT) est effectue avec une fenetre de Hanning de 512 points (32 ms) dans la partie centrale des voyelles. Les spectres sont ensuite normalises par rapport à leur puissance totale.

Ces spectres, presentant des "pics" et des "vallées" dans les mêmes regions, ont ete divises en six classes dont les bornes correspondent a la fréquence moyenne des minima, soit

Classes	C1	C2	C3	C4	C5	C6
Fréquence en Hz	60 à 375	375 à 1125	1125 à 1875	1875 à 3000	3000 à 4500	4500 à 7000

Nous avons egalement realise des spectrogrammes en bande large d'une partie du corpus (cf fig 1) : leur observation confirme bien la rapidite et l'existence paradoxale de transitions brusques perçues a l'oreille chez les Parkinsoniens ([5])

De fait, la partie stable des voyelles est souvent tres breve et les formants difficiles a localiser De même, l'articulation des consonnes parait assez imprecise.

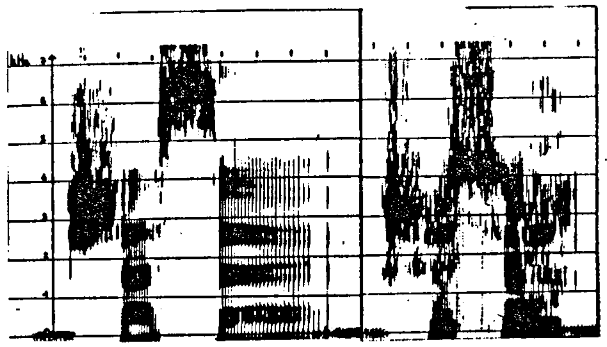


Fig. 1 Spectrogrammes du mot /lasœR/ à g. : sujet temoin, à d. Parkinsonien

Ces spectrogrammes, faute de pouvoir fournir des mesures sûres, n'ont pas ete exploites et toutes les mesures indiquees ci-dessous ont donc ete determinees a partir des spectres

RESULTATS

Une premiere analyse a consiste a chercher si la cible articulatoire (et donc le timbre de la voyelle /a/) etait bien atteinte ou non Le tableau ci-dessous indique qu'elle l'est puisqu'on ne decele aucune difference importante dans la frequence des formants parmi les trois groupes de sujets, aussi bien pour les voyelles tenues que pour les /a/ situés a l'interieur des mots (cf tableau 1)

	NX		FA		PK	
/a/ (mots)	μ	σ	μ	σ	μ	σ
F1	560	95	590	60	545	105
F2	1480	140	1455	65	1480	80
F3	2420	135	2505	120	2445	205
F4	3520	135			3505	195
/a/ tenus	μ	σ	μ	σ	μ	σ
F1	665	65	655	85	625	115
F2	1140	105	1210	170	1230	130
F3	2500	170	2575	180	2555	205
F4	3530	295			3350	330

Tableau 1 Moyenne et ecart-type des frequences des formants F1 à F4 en voix normale et faible et chez les Parkinsoniens dans les mots (en haut) et les voyelles tenues (en bas)

Les variations observées entre les /a/ des mots et les /a/ tenus, en particulier sur les formants 1 et 2, se retrouvent de façon identique dans les trois populations et le test t de Student effectue sur ces mesures montrent que les differences entre voix "normale" d'une part et voix "faible" ou pathologique d'autre part ne sont jamais significatives

Le tableau 2 montre au contraire que la puissance de ces formants varie selon les voix:

	NX		PK	
/a/(mots)	μ	σ	μ	σ
F0	100	14	112	8
F1	91	15	79	20
F2	70	18	54	16
F3	52	17	36	15
F4	39	19	27	13

Tableau 2 Moyenne et écart-type des puissances en dB du pitch et des formants chez les témoins et les Parkinsoniens (normalisées sur celle du F0 temoin)

Même si les ecarts-types sont importants et donc les moyennes à prendre avec precaution, on constate de façon assez nette une puissance plus élevée chez les Parkinsoniens que chez les témoins dans le pic du fondamental et par contre plus basse dans les pics formantiques (cf fig. 2).

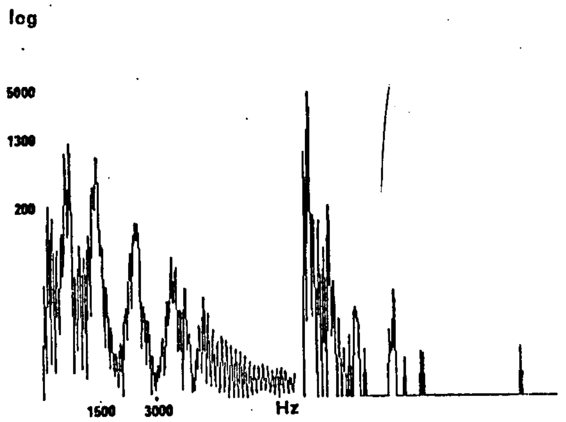


Fig. 2 Spectres du mot /banan/ à g. : sujet témoin, à d. Parkinsonien

Ces résultats vont être confirmes par l'etude de la répartition de la puissance des classes C1 à C6 définies ci-dessus

Les tableaux 3 et 4 synthétisent toutes les mesures prises et le resultat du test t de Student (apparie dans la mesure du possible) effectuées sur les paires voix "normale"/voix "faible" (NX/FA), voix "faible"/Parkinsonien (FA/PK) et temoin/Parkinsonien (NX/PK)

/a/	NX1		NX2		FA1		PK1		PK2	
(mots)	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
C1	2920	1520	2740	1330	4800	1620	4210	1300	4560	2760
C2	2897	825	2942	746	2018	785	2578	898	2529	925
C3	1649	616	1617	546	1014	424	1067	512	1000	522
C4	1244	381	1314	530	948	480	879	316	907	413
C5	974	480	971	644	664	297	711	485	741	531
C6	309	282	411	465	471	126	510	122	496	238
/a/	NX3		NX4		FA2		PK3			
tenus	μ	σ	μ	σ	μ	σ	μ	σ		
C1	3410	1470	2450	1590	5920	1710	4980	900		
C2	3168	1007	3909	1161	1640	1252	2625	771		
C3	1106	380	1389	477	782	405	991	382		
C4	962	326	1110	449	572	275	532	148		
C5	1142	709	977	688	521	291	435	222		
C6	204	155	161	128	502	279	431	208		

Tableau 3 Moyenne et ecart-type de la puissance par classe.
/a/ des mots : NX1: 6 mots x 4 loc., NX2: 10 x 10, FA1: 6 x 4, PK1: 6 x 10, PK2: 10 x 10
/a/ tenus : NX3 : 4 loc., NX4 : 7 loc., FA2 : 4 loc., PK3 : 6 loc.

/a/(mots)	NX1/FA1		FA1/PK1		NX2/PK2	
ddl	21		77		181	
	t	S	t	S	t	S
C1	5,851	***	-1,695	NS	5,591	***
C2	-4,796	***	2,568	*	-3,305	***
C3	-4,777	***	0,431	NS	-7,801	***
C4	-2,377	*	-0,750	NS	-5,824	***
C5	-2,903	**	0,423	NS	-2,643	**
C6	2,420	*	0,766	NS	1,581	NS
/a/tenus	NX3/FA2		FA2/PK3		NX4/PK3	
ddl	7		17		23	
	t	S	t	S	t	S
C1	5,611	***	-1,558	NS	4,713	***
C2	-3,241	*	2,124	*	-3,156	**(*)
C3	-1,502	NS	1,149	NS	-2,253	*
C4	-4,011	**	-0,400	NS	-4,083	***
C5	-2,305	NS	-0,738	NS	-2,503	*
C6	4,520	**	-0,633	NS	4,003	***

Tableau 4 Résultats des tests de comparaison des populations voix "normale"/voix "faible", voix "faible"/Parkinsoniens, voix "normale"/Parkinsoniens.

COMMENTAIRE

Chaque classe C1 a C5 contient respectivement le pitch F0 et les formants F1 a F4 La puissance de ces classes reflete donc l'intensite globale des harmoniques situes dans le voisinage de chacun des pics principaux du spectre

La comparaison entre témoins en voix normale et Parkinsoniens a permis de mettre en évidence que la puissance des classes C2 à C5 chez les premiers est toujours superieure a celle des voix de Parkinsoniens, au contraire de celle des classes C1 et C6, toujours inferieure. De plus, il y a une inversion de pente dans la partie basse du spectre puisque la puissance en C2 est plus élevée que celle de C1 (cf fig 3 et 4)

Par contre, l'hypothèse qu'on a faite d'une modification de la repartition de la puissance du spectre liee chez les Parkinsoniens à une diminution de l'intensite vocale est confirmee par la similitude entre voix "faible" et voix de Parkinsonien. En effet, qu'il s'agisse de mots ou de voyelles tenues, il n'y a pas de difference significative entre ces deux populations Cependant la puissance en C1 est sensiblement plus faible chez les seconds et donc la pente du spectre plus raide. La raison de cette dissemblance pourrait être trouvee dans une légère difference d'intensite vocale entre les voix "faibles" des sujets normaux et les voix des Parkinsoniens puisque l on sait par ailleurs, que le spectre du signal glottique est d'autant plus raide que la voix est moins intense ([2],[4])

On a évidemment confirme qu'il existait une, disparite importante entre voix "normale" et voix "faible" les differences sont tres significatives par exemple dans les mots Les classes C1 a C3 qui contiennent l'essentiel de l information tant sur la source sonore que sur le timbre du phone realise n' ont rien de comparable : chez les temoins, peu d'energie en C1 et beaucoup en C2; situation inverse chez les voix "faibles" Les classes suivantes presentent des divergences moins importantes

Ceci est moins vrai dans les voyelles tenues puisque, si la différence reste importante pour la premiere classe, elle s'estompe dans tout le reste du spectre Sans doute, la cause doit être recherchee dans le fait que parler "faiblement" pour les sujets temoins n'est pas naturel donc difficile a maitriser et ce d'autant plus que la voyelle est accentuee Cependant, le nombre peu élevé d'items doit inciter a la prudence quant a une quelconque interprétation

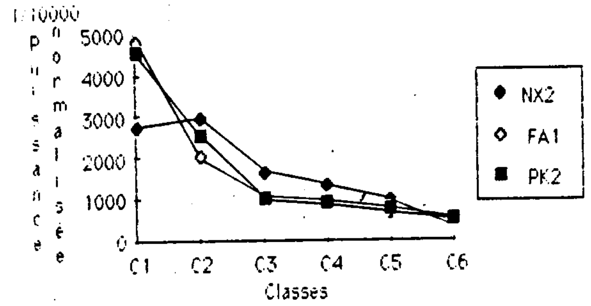


Fig. 3 Puissance par classe dans les /a/ des mots pour les trois groupes de sujets.

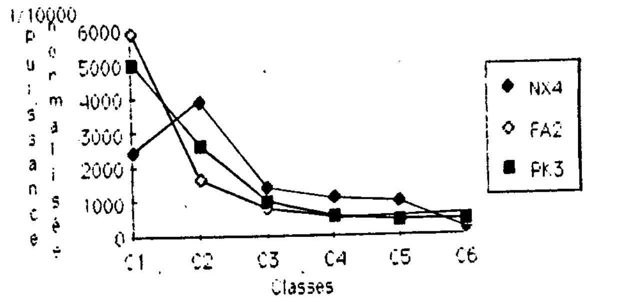


Fig 4 Puissance par classe dans les /a/ tenus pour les trois groupes de sujets

Si la pente du spectre est à peu prés parallele de C2 à C5 entre les divers sujets, elle est par contre très différente d'un groupe à l'autre entre C1 et C2 d'une part, et C5 et C6 d'autre part. Comme le montre le tableau 5, ce phénoméne est amplifié encore sur les voyelles tenues.

	NX	FA	PK
/a/(mots)			
C2/C1	1,07	0.42	0.59
C6/C5	0,42	0,71	0,67
/a/ tenus			
C2/C1	1.60	0,28	0,53
C6/C5	0,16	0,96	0,99

Tableau 5 Rapport des puissances moyennes des classes adjacentes C2/C1 et C6/C5 pour les trois groupes de sujets.

CONCLUSION

Le calcul par classes, c'est-à-dire par bandes de frequence fixes, semble fournir un bon indice de description du timbre des sujets parkinsoniens Cependant, il faut l'utiliser avec precaution compte tenu des variations non negligeables dans la population temoin (voir les ecarts-types).

Il est en tous cas plus fiable que la puissance des formants qui est souvent trop faible pour en faire une mesure correcte et qui necessite la detection prealable, pas toujours aisee et de plus inutile puisque non pertinente, de ces formants

On a pu confirmer que certaines caracteristiques spectrales de la voix parlee des Parkinsoniens etaient liees a la diminution d'intensité de l'emission vocale. On peut discuter du mecanisme de cette diminution d'intensité · insuffisance respiratoire (signalee par certains auteurs) et/ou rigidite (raideur) des cordes vocales et des muscles larynges (en faveur d'une telle rigidité plaide aussi l'elevation du fondamental)

Les donnees actuellement recueillies sont insuffisantes pour evaluer les consequences possibles d'une rigidite de l'appareil articulatoire supra-glottique. On notera cependant la realisation formantique normale de la voyelle /a/ Une etude ulterieure devrait envisager les mêmes mesures sur d'autres voyelles

REFERENCES

[1] ARABIA-GUIDET C, MANTOY A. Diagnostic automatise des troubles de la phonation · analyse spectrale des voyelles ITBM vol 7, n° 6, 1986 p 728-736.

[2] FANT G. Preliminaries to analysis of the human voice source STL QPSR n° 4, 1982 p 1-27.

[3] GUIDET C., CHEVRIE-MULLER C Methode de traitement du signal electroglottographique · application au diagnostic automatise des troubles de la phonation. ITBM vol 4, n°6, 1983 p 617-635.

[4] LONGCHAMP F. Les sons du français : analyse acoustique descriptive. 1985.

[5] RONTAL M., ROLNICK M Objective evaluation of vocal pathology using voice spectyrography . Ann Otol. n° 84, 1975 p 662-671

[6] SEGUIER N, SPIRA A., DORDAIN M., LAZAR P, CHEVRIE-MULLER C Etude des relations entre les troubles de la parole et les autres manifestations cliniques de la maladie de Parkinson Folia phoniatrica 26, 1974 p 108-126.

ATAXIC DISARTHRIA

E.N. VINARSKAYA, T.I. BABKINA

Maurice Torez Moscow State Pedagogical Institute of Foreign Languages
Ostozhenka 38, Moscow, USSR, 119034

ABSTRACT

On the basis of phonetic and experimental-phonetic study of cerebellum or ataxic disarthria three syndrome variants related to local lesion of (1) cerebellum vermis, (2) intermediate and lateral cerebellum zones, (3) right cerebellum hemisphere, have been singled out.

It is well known that local lesion of the cerebellum and its conductive systems cause disorders in the fluency of speech and disprosody. Furthermore, speech becomes slurred |1,2,3,4,5,6,7|. Instrumental phonetic studies of disarthria which began in the middle of the 20th century have dealt mostly with cerebellum or ataxic disarthria. According to the data obtained |5|, it is characterized by the following symptoms (in order of their diminishing significance): inaccurate consonant articulation; excessive or weak prominence of stressed syllables; irregular slurred speech, vowel distortion, a shrill voice, increased sound and pause duration, voice monotony, even voice loudness and slow tempo of speech.

Our study of cerebellum disarthria was carried out with patients of three nosological groups. They comprised 5 sclerosis disseminata patients, 6 patients with local neurosurgical cerebellum diseases and 6 patients with degenerative cerebellum diseases. The total number of defective speech reactions studied was 8.039. Speech investigation included the date obtained using the auditory method of phonetic analysis as well as the results of experimental phonetic study of speech using oscillography and intonography (The instrumental part of the investigation was carried out in the Laboratory of Phonetics, headed by A.P. Belikov in Maurice Torez Institute of Foreign Languages.). The patients were given the following tasks: to prolong vowels, to reproduce string of CV syllables, rhythmic syllabic structures with different stressed syllable positions, sensegroups of different communicative types and sensegroups with different logical stress positions.

Our observations confirm the data, according to which cerebellum disarthria is caused by the same phenomena of adiadochokinesia, dismetrics, asinergia and intentional tremor that are observed in extremities motor disorders. Speech discoordination is one of the most perceptible symptoms of local cerebellum deficiency.

First of all note should be made of speech tension which is an outcome of discoordinate work of certain muscles, when sumultaneous innervation of agonists and antagonists takes place and when on the contrary, functional sinergists don't work simultaneously. Typical of discoordination causing speech tension are cases of syllable prolongation and simultaneous unchanged or even increased syllable reduction.

Motor speech discoordination involving time delay in switching the innervation of certain muscles over to the innervation of the antagonistic ones (adiadochokinesia) brings about slowing down of speech tempo. It is easily percieved by ear and becomes more obvious in experimental phonetic studies. There are also regular instances of speech monotony when all speech segments are pronounced in averaged voice register; insufficient loudness variations, when all the segments are equally loud or equally low; averaged vowel tambre accompanied by low degree of contrastivity between A-, И- and У- vowels in word stressed positions; absence of, or on the contrary, excessive reduction of unstressed syllables and absence of tempo variations.

Defective prosody make the speech of cerebellum patients slurred, not articulate enough, i.e. disarthric. Disprosodic slurred speech is must characteristic of sclerosis disseminata patients and those with systemic degerative cerebellum diseases. Sometimes it was registered in spontaneous speech, sometimes in reading texts, but more often in special phonetic tasks.

The results of our research make us doubt if slurred and disarthric speech in local cerebellum lesion necessarily involves brain stem structure lesion.

A characteristic symptom of cerebellum disarthria is failure in speech fluency which is generally termed as "scansion". Scansion (Lat. "scando" - measured speech) denotes metric recitation with verse emphasized rhythmic structure |8,9|. But can we say that cerebellum patient speech resembles recitation and is measures? Their similarity seems to be rather vague: abnormal speech has no metric organization, its hypermetric prominant elements are irregular, they are marked by discord and a variable set of the means used(sometimes it is duration, sometimes laudness, sometimes pitch, voice quality or a combination of several means). Therefore, defining cerebellum speech as "scansion" is not correct.

Prosodic disorders make cerebellum speech phonetically non-normative. Non-normative features are first of all traced in quantitative characteristics of prosodic parameters that are not appropriate for the given situation: increased duration of stressed and unstressed syllables in rhythmic structures, increased vowel and consonant duration, a higher degree of loudness and melodic expressiveness of speech, etc. These disprosodic features, especially when in discord with each other (e.g., syllable duration increases with a simultaneous increase in its qualitative reduction) make the speech not only tense and not distinct enough, but also abnormal in view of the given speech situation and context.

Secondly, cerebellum discoordination can lead to a situation when prosodic parameters of phonetic units no longer correspond to the aim of the speech act. When asked to reproduce a string of equally stressed syllables in a maximum fast tempo the patient arranges them in rhythmic structures; trying to reproduce an affirmative sensegroup the patient uses a wrong melodic structure and changes it into an affirmative one; using melodic means of sense emphasis instead of dynamic ones the patient distorts the logical structure of the utterance.

Taking into account the prosodic disorders described above, we can single out three variants of cerebellum disarthria. Most significant for the first variant was tense and slurred speech accompanied by voice tremor in pronouncing continuant vowels, a low degree of loudness though without scansion. The patients found it more difficult to repeat phonetic tasks after the investigator rather than to perform the task with the help of a speech instruction. The degree of disarthria worsened considerably in ortho-clinostatic test: the results of all the tasks were worse in sitting position and even more so in standing position as compared to lying position. This first variant of disarthria was most vivid with a female patient with cerebellum vermis lesion (the group of patients with degenerative cerebellum and conductive systems diseases).

The second variant of cerebellum disarthria was registered in the majority of degenerative celebellum cases and sclerosis disseminata cases that are caused by a bilateral dificiency in intermediate and lateral cerebellum zones. First and foremost the syndrome brought about disorders in speech fluency (syllable - to syllable and scandent speech) which were accompanied by slow tempo of speech, frequent increase in voice loudness and inaccurate production of Russian prosodic norms.The patients were worse at repeating neurophonetic tasks, made more errors than in performing the tasks when assisted with a speech instruction. Disarthric syndrome worsened under ortho-clinostatic test conditions.

The third variant of cerebellum disarthria observed in our material, was typical of the neurosurgical group of patients with unilateral lesion of the right cerebellum hemisphere which is functionally related to dominant left hemisphere of the cerebrum. It was revealed in slow, slurred, tense, syllable-to-syllable, monotonous speech. No intention in the prolongation of vowels was observed. Speech loudness was often normal. The characteristic feature was the use of prosodic speech norms. The patients were better at repeating the phonetic tasks than at their realization according to the speech instruction, which made these patients significantly different from those of variants I and II. Ortho-clinostatic tests yielded negative results.

There are grounds to believe that this variant of cerebellum disarthria is caused by difficiency in the lateral zone of the right cerebellum hemisphere. Affecting the motor programmes of the secondary associative praxic cortex, generalization of which are formed under the influence of language phonetic norms, pathalogical cerebellum affects of this kind cause particularly severe prosodic disorders in spontaneous speech. These disorders diminish when the patient produced an utterance trying to imitate the investigator's speech patterns.

In all the three groups of patients task complication aggrevâted disarthric disorders. In utterances of automatic character as in counting to 20, week days ennumerating and so on, motor speech ataxia was observed only with 3 patients; in reciting, characterized by regular prosodic structure., motor speech ataxia was more frequent (8 patients) though had a mild form; In reading and retelling prosaic texts where the prosodic structure is less regular, motor speech ataxia had a higher degree; in experimental-phonetic tasks performing motor speech ataxia had the most severe form and was registered with 15 patients out of the 17 examined.

The variants of cerebellum disarthria outlined in our investigation and the structural - functional considerations discussed call forth a further accumulation and analysis of the factual material. However, we can say even at present that phonetic and experimental studies contribute to tne diagnosis of cerebellum local lesion and widen our knowledge about the cerebellum functions.

REFERENCES

|1| М.Б. Кроль. "Невропатологические синдромы". Гос. мед. издательство УССР, 1933.

|2| Л.Б. Литвак "Локально-диагностические особенности дизартрии и дисфонии в неврологической клинике". Сборник: Вопросы патологии и речи, Харьков, 1959.

|3| И.М. Иргер "Клиника и хирургическое лечение опухолей мозжечка". Медгиз, 1959.

|4| В.Н. Винарская, А.М. Пулатов. "Дизартрия и её топико-диагностическое значение в клинике очаговых поражений мозга" Медицина, Уз.ССР, 1973.

|5| F. Darley, A. Aronson, J. Brown. "Motor speech disorders". Saunders Co., 1975.

|6| A. Lecours. F. Lhermitte. "L'aphasie", Flammarion, 1979.

|7| A. Aronson. Motor speech signs of neurologic disease. In: Darby J. (Ed.) Speech evaluation in medicine. Grune & Stratton, 1981.

|8| В. Даль. "Толковый словарь живого великорусского языка", т.IV, Гос. изд. иностр. и нац. словарей, 1956.

|9| А. Квятковский. "Поэтический словарь". Сов. энциклопедия, 1926.

PHONETICS OF STUTTERING

GLORIA J. BORDEN

Dept. of Speech
Temple University
Philadelphia, Pa. 19122
USA

Haskins Laboratories
270 Crown St.
New Haven, Ct. 06511
USA

ABSTRACT

Articulatory and acoustic records of the speech of severe stutterers, mild stutterers, and nonstutterers were measured. Although severe stutterers spoke more slowly when fluent than other subjects, they did not significantly differ in the first few glottal pulses of voicing, their coordination of lip/jaw movements with vocal fold positioning for voicing, in kinematic relationships between displacement and velocity, nor in proportional segment durations as measured from sound spectrograms. Stuttered samples were aberrant in all measures, but the normal phasing of lip/jaw and vocal fold movements remained evident during tremors. Findings from this study fail to support a temporal motor deficit theory of stuttering.

INTRODUCTION

Since no one understands exactly what is happening when someone stutters, much less what originally caused it, theories of stuttering appear, disappear, and reappear with the passing of time. At times, the psycho-social aspects of stuttering are emphasized, at other times the biological motoric aspects of it are emphasized, and at times it is viewed primarily as learned behavior. Some theorists view stuttering as a problem in self perception (Harrington, 1987), while others view it as a deficient timing mechanism for speech (Van Riper, 1973; Perkins et al,1976; and Kent,1984). One view holds that stuttering is part of a continuum of fluency ranging from a high degree of fluency to the high degree of disfluency evident in the speech of severe stutterers (Starkweather, 1987). Another possible view is that normal disfluency and stuttering are discontinuous representing an abrupt change in speech mode.

What can we learn of these things by examining the phonetics of stuttering, the articulatory and acoustic correlates of speech ? A popular view of stuttering at the present time is that stutterers exhibit motor deficiencies of the speech production systems even when they are perceived to be fluent by listeners. Evidence to support the motor deficiency view includes slower speech rate, slower speech reaction times, and slower articulatory movements in the fluent speech of stutterers than in the speech of nonstutterers (See Bloodstein, 1983 and Starkweather, 1987 for reviews).

METHOD

We have collected and analyzed a large amount of data on severe stutterers, mild stutterers, and normal speakers performing a task of repeating numbers 4253 and 3425 until speech was fluent. Articulatory and acoustic analyses were performed. To perform the articulatory analysis, respiratory, laryngeal, and supralaryngeal (lip/jaw) movements were inferred from recordings made from a pneumograph, an electroglottograph, and an optical tracking system. Velocity changes were derived from the movement waveforms. The articulatory analysis included temporal measures, kinematic measures, and qualitative inspection of voice initiation indices. Temporal measures included speech rate, duration of movements, times from onset of movement to peak velocity of the movement and to voice onset, and cross-system (laryngeal-lip/jaw) intervals between corresponding onsets and peak velocities. Kinematic analysis involved plotting relative velocity by displacement measures. Qualitative inspection of electroglottographic waveforms of voice initiation was performed on both fluent and stuttered samples.

For the acoustic analysis, sound spectrograms were measured for voice onset time (VOT), duration of the stop-gap, and duration of the vowel in the utterance ´two´ /tu/ in the context ´425´. Consonant/vowel ratios and stop-gap to VOT ratios were computed as well as the proportions of time taken within the mean total utterance duration for the stop-gap, VOT, and vowel segments.

RESULTS

Our results generally support a view of stuttering that goes against the currently popular notion of a motor timing deficit underlying even the fluent speech of stutterers. Our data suggest that although the speech motor system is vulnerable to breakdown, that breakdown is an abrupt change in the mode of speaking, discontinuous with the fluent speech of the same speaker. This is not to deny the presence of covert stuttering. We see evidence of stuttering in some samples perceived to be fluent.

When speakers are truly fluent, however, they do not significantly differ from nonstutterers on several critical articulatory dimensions:(1.) initiation of voicing as determined from analysis of the first few glottal pulses according to the EGG signal, (2.) coordination of lip/jaw movement with vocal fold positioning; for all subjects there is close coordination between lip/jaw opening and vocal fold adduction for the vowel, and (3.) relationship between the kinematic features of displacement and velocity (Figure 1); normally, increased displacement is correlaated with increased velocity. Further, there is no significant difference between groups for three acoustic characteristics: VOT, consonant to vowel ratios, and segment durations as a percentage of total utterances times. Although severe stutterers are slower than normal in speech rate aand thus exhibit longer vowel and stop-gap durations, normal acoustic relationships among segments are maintained (Figure 2)

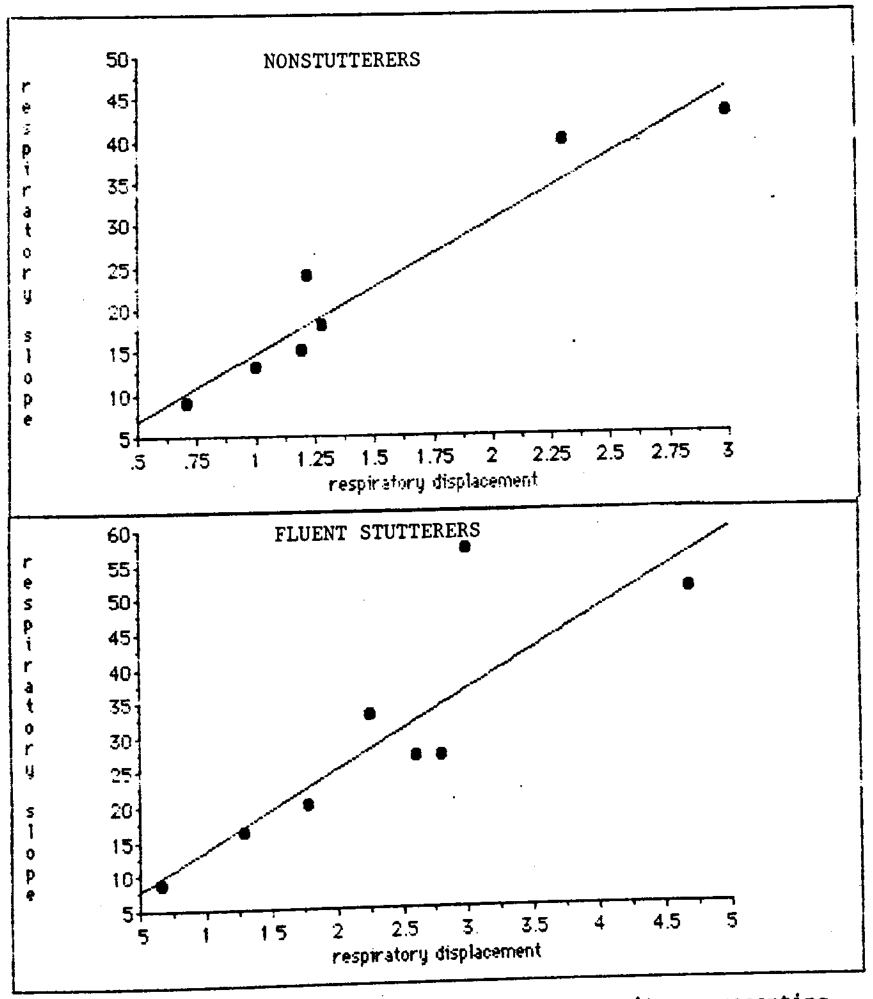


Figure 1. Peak to peak displacement in arbitrary units representing exhalation are plotted according to slope. Stutterers like nonstutterers exhibit steeper slopes for larger displacements.

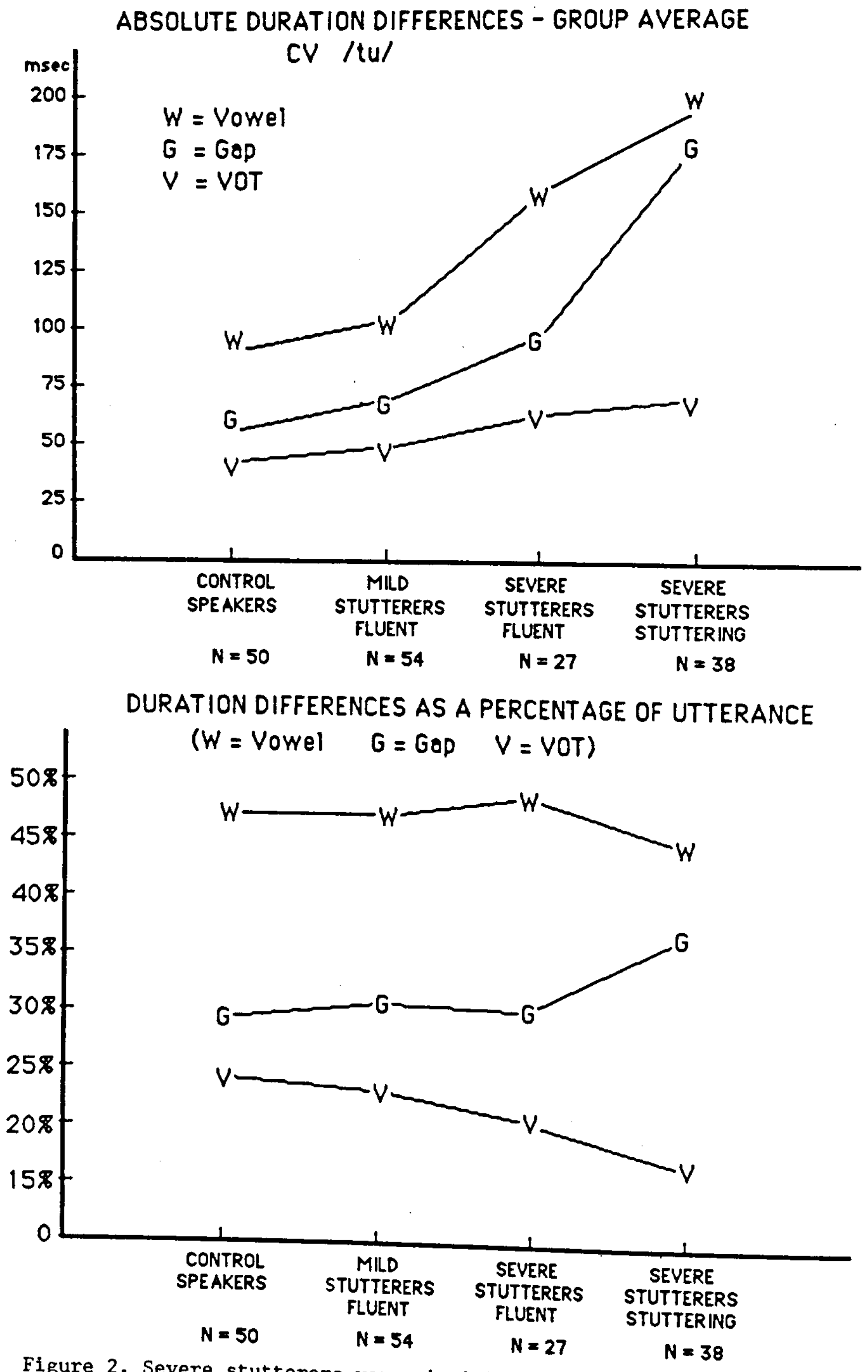


Figure 2. Severe stutterers were significantly slower for the stop-gap and vowel segments than nonstutterers. Proportionally, there was no difference.

There are some significant differences between the fluent utterances of the severe stutterers and those of the nonstutterers. All of the articulatory and acoustic measures that significantly differ from normal are positively correlated with speech rate. Severe stutterers, unlike the mild stutterers, are slower in speech rate than the nonstutterers. The slower rate is reflected in certain articulatory measures: longer times between onset of exhalation and voice onset, between lip/jaw opening for the vowel and voice onset, and in the duration of lip/jaw opening for the utterance ´five´. Slower rate for the severe stutterers is also reflected in certain acoustic measures: significantly longer stop-gap durations and vowel durations.

We were impressed, however, with the evidence that despite the slowed speech of some stutterers, their fluent speech maintains its proportional relationships in the acoustic signal and preserves the coordination between the movements of the lip/jaw system and the positioning of the vocal folds for voicing.

For stuttered samples, of course, all indices, temporal, kinematic, EGG traces during voice initiation, and spectrographic measures are highly aberrant. The kinematic relationships between displacement and velocity changed during stuttering tremors with higher velocities per unit of displacement (indicating a stiffer system) than for fluent samples. Voice initiation according to EGG patterns reveal an abnormally gradual rise in amplitude to effect continued vibration upon release of a stuttering block. Other qualitative differences between stuttered and fluent samples were that whereas fluent voicing gives evidence of a relatively stable open phase and more gradual opening, voicing upon release of a stuttering episode reveals a sharper opening and a brief, less stable open phase. This may indicate a stiffer than normal system. Highly ritualized patterns used to break the blocks were observed. An indication of the cross-system coordination that can occur during the most ´uncoordinated´ moments of stuttering, two severe stutterers who demonstrate simultaneous and phase related tremors of the lip/jaw system and of the vocal folds also show that throughout the tremors, the lip/jaw opening phase of the tremor is coordinated with the vocal fold adductory phase of the laryngeal tremor. These actions are appropriate for vowel initiation, although voicing failed to occur or was aborted upon each trial (Borden et al,1985). Finally, spectrographic measures of stuttered samples show significantly longer stop-gap durations, vowel durations, and voice onset times (VOT) than normal (especially when the block occurred upon the release of the stop).

When we inspected data for evidence of continuity between the initial disfluent utterances and the fluent utterances across the 10 or more repetitions, we found instead evidence of discontinuity, a step function that separated fluent from disfluent samples. Nor did we find a fluency continuum going from normal speakers to mild stutterers to severe stutterers. Rather, the mild stutterers, when fluent, were indistinguishable from normal, while the severe stutterers, when fluent, were notably slower in their speech, although they preserved proportionally normal acoustic segments and normal articulatory coordination across speech motor systems.

DISCUSSION

These findings may not generalize to all stutterers; they await further data for verification. The significantly slower rate that we found for severe stutterers, as well as the articulatory and acoustic indices found to be correlated with the slow rate might be interpreted by some theorists to support the idea of a temporal motor deficit that is hard-wired into the speech motor system or possibly a fault in the temporal programming of the systems. The evidence, however, of normalized acoustic relationships in the speech and of normal articulatory patterns, especially the high degree of lip/jaw coordination with the larynx force us to reject a temporal motor deficit explanation. The slower speech of the severe stutterers may simply reflect a technique acquired for avoiding increased tension in the speech mechanisms. The problem may not lie in timing mechanisms but in the tension settings of the muscles. During the fluent speaking mode, the settings may be appropriate, but during stuttering episodes, the settings may be out of balance across muscle groups cooperating for a certain function. Especially vulnerable to disruption are the settings required to position and tense the vocal folds appropriately for voicing.

REFERENCES

Bloodstein,O. A Handbook on Stuttering Chicago:National Easter Seal Society, 1983.

Borden,G.J., Baer,T., and Kenney,M.K. "Onset of voicing in stuttered and fluent utterances" Journal of Speech and Hearing Research 28, 1985, 363-372.

Harrington,J."Stuttering,delayed auditory feedback and linguistic rhythm" Journal of Speech and Hearing Research(in press)

Kent,R.D. "Stuttering as a temporal programming disorder" In R.F.Curlee & W.H.Perkins (Eds.) Nature and Treatment of Stuttering: New Directions San Diego: College-Hill Press, 1984,283-301.

Perkins,W., Rudas,J.,Johnson,L.,& Bell,J. "Stuttering: discoordination of phonation with articulation and respiration"Journal of Speech and Hearing Research19, 1976, 509-522.

Starkweather,C.W. Fluency and Stuttering Englewood Cliffs,NJ:Prentice-Hall, 1987.

Van Riper, C.G. The Nature of Stuttering Englewood Cliffs,NJ:Prentice-Hall, 1973.

PHONOLOGY OF STUTTERING

LEONID V. IVANOV

Russian Language Institute
Academy of Sciences
Moscow,USSR 129019

ABSTRACT

Here follows the discussion of the most interesting problems wich are indicated by the results of phonological subtests of the Multilevel Linguistic Test proposed to 90 stutterers being treated at the Speech Pathology Center in Moscow.Stutter can be interpreted as a new distinctive feature inherent to stutterers' "dialect".Its existence is connected with the changes in the syntactic and the semantic levels.Stutter can be compared with the slip of tongue in the frame of time-space planning model of stuttering.Certain features of the disorder are similiar to child language and aphasia.

0.In the last years the problem of definition of the stuttering phenomenon has been widely discussed in the literature(see,e.g. /1/).It seems however that the linguistical 'diagnosis' for stuttering(i.e. the study of the functions of levels of stutterers' language system) is more important.

As it was shown in /2/,/3/ and in some other reports,stutterers evidenced serious disturbances in the scene analysis/synthesis and its description,in the functional description of objects,in composing,storing and producing complex phrases,in completing long phrases and other semantic and syntactic difficulties.Bilingualism and language interference are to be considered as causing particular troubles in the patients.

1.As the above listed difficulties seem to affect the principal levels of language structure,the study of morphological and phonological means in stutterers turns out to be decisive in testing the hypothesis of the particular dialect(or few ones) that's created in the process of the developement of the disorder.

The hypothesis implicates the following problems:

a) One can (preliminary) observe that,at least partialy,the stuttered sounds are likely to be compared with the sounds of languages that are 'exotic' for the given speaker.From that point of view in stutterers' speech the new 'phonetical' features appear(such as 'aspiration','emphatisation', 'prolongation' and some others - cf./4/). The sounds,when stuttered,are 'marked'(and 'unmarked' in the fluent pronounciation - cf. the data of /5/).These features could be described(because of different degrees of laryngeal participation - cf./4/;another aproach in /6/) as Trubezkoy's "correlations of second degree"/7/.

b) At the same time the stutterers' active vocabulary is narrowed(among others my patients could not recall/use such words as NEW,HEAT,COMFORT,SUCCESSFULL,TO UNITE,TO COMPLETE,SALESMAN,TRUNC,etc.).The smaller size of active vocabulary can be one of the reasons that the patients fail to produce examples of the minimal pairs for certain distinctive features(Ž-Š,G-K and others) even if the investigation continues for a rather long period(till two days; the subtest often ends up with the patient's refuse to continue any kind of linguistic testig - the difficulty compared only with the subtest for making up the long phrase).

c) Phonetically different performance,on one hand,and the failure to produce examples for minimal pairs,on the other,get stuttering phonology close to the child phonology (cf."cortical immaturity in stutterers" discussed in /8/).To some extent, stutterers are to be considered as being in the phase of language developement that was described by Jakobson as "oubli des phonations" /9/: children confound the sound pairs (in stutterers' case the difference is not 'phonological') but can distinguish them in audition.

A case study can illustrate the thesis. Nastya,4:5(yrs:mos),Russian speaking young girl,began to stutter(according to her mother) at 2:2.Stutterings were mainly blocks and prolongations.At the same period,she began to learn by heart the passages from the poems that were read to her and recite them without stuttering.The intensive stuttering lasted for 5 months,then,suddenly, the amount of stuttering drammatically decreased.At about3:10 she started telling herself rather long nonsense texts.At 4 her texts became meaningfull,and interjections(SO TO SPEAK,LORD, etc) appeared.

The investigation started when she was 4:3.The phonetical subtests designed personally for her were the following:

Q.Instead of saying /DOM/('home') I say /TOM/('volume').That's the language the fox speaks.How do you think he will say /BAR/('bar')?
A. ----
Q.(the first question repeated)How will he say /TAM/('there')?
A./ZABOR/('fence').
Q.Instead of /DOM/ the fox says /TOM/,instead of /BAR/ he says /PAR/('steam'),instead of /TAM/ he says /DAM/('I'll give'). What will he say for /POL/('floor')?
A.'I don't know ... yet '(exactly 'after' Jakobson!)
Q.How will the fox say /SAD/('garden')?
A.'/SAD/ will be /SKAF/'(pronounced as /ɬKAF/ as she had sigmatismus lateralis;it can be preliminary supposed that the tongue position,just like in aphasics,is connected here with certain brain processes).

...

Q.Intstead of /SOL'/('salt') the fox says /SOL/.How will he say /BOL'/('pain')?
A.'/BOL'/ - it's when you fall and something is painfull on the asphalt'.
Q.(the question repeated) -
A.'/BOL'/ - it's when it's /BOL'NO/('painfull').
Q.(the question repeated) What will he say for /ROL'/('role')?
A.'/ROL'IK/... /KROL'IK/('roller' - 'rabbit')(the answer shows that the kid understood that she was supposed to change only the first sound of the word;she tried to do it later in her /NAMPA/ for /LAMPA/;the only idea she could not get was how the sounds were related;is her favour for semantic associations instead of phonetic changing a general tendency in stutterers?).

She confounded /R/ and /L/.At first she did not correct me when asked her about /LAK/ 'laquer' instead of /RAK/'crab'.But then she asked me what will I say for /KRASKI/ ('paints').I answered /KLASKI/.She reacted: '/KRAS/ - ha-ha - it's like /KRAS/(normally /KLAS/'grade') - where you study in the school!'.

The case is interesting because she must be considered as the high-risk infant,though actually the number of her stutterings is rather small.Anyway,the comparison between stutterers and child language must be the problem for further investigation.

2.Though the question of the loci of stuttering had been put up long ago /10/,it still remains not quite clear(see,e.g./11/)

As a working hypothesis one can assume that:

a)The stuttered word differes from the non stuttered one not only phonetically,but with its value on the other levels of language structure(especially in the semantic aspects).The supposition leads to the conclusion that the stutter <u>per se</u> is the value of the distinctive feature(or few ones) in the stutterers dialect.The study of the question is complicated by the fact that stutterers sort of speak the 'normal' language;that's one of the reasons for their speech changing under different conditions (cf/12/;cf.,also,the orientation of the patients on various clichées and standards, such as the speech of TV announcer ,etc.).

b)The stuttered word differes from the non stuttered one because of its different position in the speech sequence,that is,there are certain positions that tend to be stuttered more than the others (apart from the classical "first three words"/11/ my clients tend to stutter on every 4th,8th and so forth,word of syntagm and on the conjunctions of complex phrases).Various relations in which the stuttered word takes part are arising the problem of time-space planning of stutterers speech(cf./13/). The problem along with the slow,slurred speech of stutterers,the changes in the intonation patterens and often phonetical errors ressemble very much the features,described by Alajouanine in the patients with the damages in the frontal lobes of brain /14/.

3.The problem of time-space planning may help to link the stutter in its various appearances with the slip of tongue(that,as the stutter,appears under certain conditions in the fluent speech as well as in aphasia).The two principal kinds of the slip - perseveration (/NE NEDO/ instead of /NE NADO/ - from the patient V119) and anticipation(/POSLE SLUŽBY V ARMIJU JA POPAL NA LEČENIE V MOSKVU/ - instead of /... V ARMII .../ - from the patient K12) may be interpreted in this model as the stutter,extended in time and space symmetrically around some "nucleus"(it is interesting,that , according to /8/,stutterers perseverate less than the normal speakers - maybe,because some part of their perseverations converts into stutters).It seems that in the normal speech sequence there are certain places for pauses,corresponding to the "nucleus"(as it was stated in /15/,based on the different kind of testing,"the'true'relation is between natural pauses and stuttering").In the cases of stuttering certain restrictions for the distances between pauses seem to appear.Not contradictory to the model seem be the linguistic analysis of slips /16/ as well as the phenomena of word and syllable repetitions(where,"nucleus" being stable,the sphere of its influence is extended).The time restrictions data are also indicated by the results of syntactic subtests.Maybe,the interjections(LEMME SEE) appear exactly on the margins of these time intervals.That is supported by the fact that the "embolus" can be meaningless(/PI/ in the Cheremiss patient) or can consist of the words from foregn languages(i.e. Russian /TAK/'so' in the Armenian patient), thus carrying no but temporal function.

REFERENCES

/1/M.E.Wingate,Definition <u>Is</u> the Problem, Journal of Speech and hearing Disorders, 1984,49,430.

/2/L.V.Ivanov,Speech disorders,Mentality and Bilingualism:towards a formal description.Tallin,Symp.Formalis.Hist.Lingu.,1986.

/3/L.V.Ivanov,Stuttering as the Disorder in the Syntactic Hierarchy. Tallin,1987(in press).

/4/M.R.Adams,R.Reis,The influence of the onset of phonation on the frequency of stuttering,Journ.Speech Hearing Res.,1971,14, 639-644.

/5/D.E.Metz,E.G.Conture,A.Caruso,Voice onset time,frication and aspiration during stutterers' fluent speech,J.Speech Hearing Res.,1979,22, 649-656.

/6/M.E.Wingate,"Vocalisation"≠ Phonation, J.Speech Hearing Res.,1979,22, 657-658.

/7/Н.С.Трубецкой,Основы Фонологии,Москва, 1960.

/8/D.G.Sayles,Cortical excitability,perseveration and stuttering,J.Speech Hearing Res.,1971,14 ,462-475.

/9/R.O.Jakobson,Les lois phoniques du language enfantin et leur place dans la phologie generale.In:Selected Writings,vol.1, The Hague:Mouton&Co,1962.

/10/S.F.Brown,The loci of stuttering in the speech sequence,Journal of Speech Disorders 1945,10,181-192.

/11/M.E.Wingate,The first three words,J. Speech Hearing Res.,1979,22,604-612.

/12/M.A.Young,Comparison of stuttering frequencies during reading and speaking,J.Speech Hearing Res.,1980,23,216-217.

/13/W.H.Perkins,J.Bell,L.Johnson,J.Stocks, Phone Rate and Effective Time Hypothesis of Stuttering,J.Speech Hearing Res.,1979, 22,747-755.

/14/T.Alajouanine,A.Ombredane,M.Durand, Le Syndrome de la Dêsintegration Phonêtique dans l'Aphasie,Paris,1939.

/15/S.Griggs,A.W.Still,An Analysis of Individual Differences in Words Stuttered,J. Speech Hearing Res.,1979,22,572-580.

/16/V.A.Fromkin,The non-anamalous nature of anamalous utterances,Language,1971,47, 27-52.

DYSPHASIA (SPEECH DISTURBANCES), CAUSED BY THE FUNCTIONAL STAMMERING (Phoniatric aspects)

IRAIDA KRUSHEVSKAYA

Dept. of Oto-Rhino-Laryngology
Research Institute of Capacity to Work
Minsk,Byelorussia, USSR 22008I

Stammering is a result of the functional variations in the central nervous system, influencing the motor mechanisms of respiration, phonation and articulation. The study of biomechanisms of the process of speech- and voiceformation will add the new facts for the correction of the existing methods of the rehabilitation of the patients with functional stammering.

The functional stammering is related to the constitutional disturbances of the speech, and not being independent disease, is considered a symptom in quite a number of diseases of the central nervous system.

Stammering is a result of the functional variations in the central nervous system, influencing the motor mechanisms of respiration, phonation and articulation.

According to the data of Zeeman, up to 30% of stammering children have inherited dysphasia from the parents. The reason of stammering in these cases may be the congenital constitutional deficiency of motor mechanisms. By origin it is customary to distinguish two types of functional stammering: stammering, which appeared in the period of development and posttraumatic stammering.

In view of the fact, that voice-speech process depends on the activity of respiratory, phonator and articulator organs, the convulsive conditions of that or this part of the organs cause the corresponding form of stammering. For example, the convulsion of respiratory muscles determines difficulty of inhalation or exhalation which causes the interruption in the process of voiceformation. Pharyngospasm leads to the intermissions in voting. Spasm of muscles of the articulator system impedes the formation of phonemes. Potention of the respiratory, phonation and articulatory muscles is characteristic for hyperkinetic form of stammering. According to the data of chronaximetry - chronaxia of the buccal muscles achieves 0,I5 m/sec.

Relaxation of muscular tension is typical for the buccal muscles lengthens from 0,4 to 0,5 m/sec.

The function of the closed throat ring may be violated in such cases and air will pass through the nose during the pronunciation of mouth sounds.

Sometimes, the disturbance of breathing of stammerers may be strongly pronounced and noticeable for the people surrounding them. During the observation of the function of external respiration of such patients the usage of the thoracic respiration was typical for childhood and adult age. As a rule, stammerers have asymmetric breathing that is the left and right sides contract asynchronously, as evidenced by the contractions of diaphragma during roentgenoscopy. The number of respiratory movements per unit time is not constant, during paroxysm it becomes so frequent, that sometimes it achieves paradoxical figures. These features are accompanied by acceleration of expiration phase, which may be here interrupted by inhalation. The disturbance of breathing at continuous stammering takes place not only during the phonation process, but during the condition of rest. The isolation of the mouth cavity from the nasal one with the closed throat ring may be incomplete at reduction of muscular tension, air may penetrate into the nose, that creates difficulties during the pronunciation of the explosive sounds.

The absence of air in the mouth cavity negatively affects the articulation. The movement of the articulatory muscles is sharply limited. One can notice that the violation of the possibility to make the simplest movements with the tongûe (as its raising upwards and lowering), the displacement of an angle of the mouth to the right and to the left, it droops. The paralysis of the corresponding muscles is not discovered in this position.

The speech is monotonous, colourless, deprived of melodiousness, the artificial drawl of the vowels only emphasizes these qualities. When examining speech one should pay special attention to stresses, appreciate the words from the point of view of grammar and syntax.

The disturbance of the function is expressed in constantly repeated strong compression of the vocal folds. The data of endoscopy ascertain in such cases the dilation of the blood vessels, stasis of blood flow, and also the parts with varicose vessels on the mucosal internal edge of the vocal folds, the vestibular folds, in the subfolded zone. During the long-term laryngospasms the mucosal membrane becomes stagnetely hyperemic, the dystrophiy changes develop with the deaf of surface layers of epithelium. In such cases the mucosal membrane may thicken, hypertrophy, more often hypertrophic laryngitis, keratosis, pachydermia and others organic diseases of the vocal folds.

According to the preliminary data of electronic laryngostroboscopy it is discovered, that in cases of strong compression of the larynx the vocal folds may come one over another, traumatizing the mucosal one, the rhythm of oscillation of the vocal folds is asynchronous, the amplitude is inconstant. In such cases the larynx moves up, down and forward. The voice becomes firm, explosive, the attack of the sound is hard.

Motor hypertonus leads to the development of hyperkinetic dysphonia, spastic aphonia.

In case of hypotonus the symptoms of hypokinetic dysphonia are developing, that is a reverse symptom. The flabbiness of the vocal folds and the absence of motor movements in them (the data of electronic laryngostroboscopy) create the impossibility of voiceformation.

Sharp tension and compression of the vocal folds as well as their flabbiness are noted only during an attack of stammering. The study of biomechanisms of the process of speech- and voiceformation will add the new facts for the correction of the existing methods of the rehabilitation of the patients with functional stammering.

It is known, that the intellectual people can conceal stammering much better, whereas the mentally deficient and psychopatic persons manifest their ailment in an expressed form.

Our data on the treatment of neurogenic dysphonia by the method of acupuncture are used with regard for an individual corresponding approach, which is typical for hypo- and hyperkinetic forms in the whole complex of rehabilitative measures. Timely successful rehabilitation, implemented, especially in childhood will allow determine without limitation the labour orientation of these patients whereas at adult age it will raise their labour ability, increase the labour potential of the country.

A TECHNIQUE FOR THE PHONETIC TRANSCRIPTION OF STUTTERING

JONATHAN HARRINGTON

The Centre for Speech Technology Research and Department of Linguistics, University of Edinburgh, Scotland.

ABSTRACT

This paper describes a phonetic transcription system that has been developed from an auditory, spectrographic and electrolaryngographic analysis of the production of stuttered speech by 36 adult stutterers and an electropalatographic study of stuttered speech produced by 2 adult subjects. The system enables the transcription of both 'repetitions' and 'prolongations' and provides some guidelines for their distinction. In the final section, a description is given of some of the characteristics to which the production of stuttered speech seems to conform.

1. INTRODUCTION

One of the main advantages of the development of a phonetic transcription system for stuttering is that it enables the classification of stuttered speech which may in the past have been obscured by the usage of ill-defined terminology As Wingate [1] has noted, some of the more common terminology includes: repetitions, prolongations, interjections, part-word repetitions, word repetitions, phrase repetitions, blocks, blocking, blockades, silent blocks, hard contacts, forceful attacks, spasms, broken words, revisions and incomplete phrases. Cutting across this, there is also 'tonic' and 'clonic' stammering and stuttering and variations thereof, including 'primary clonus', 'tonoclonus', 'clonotonus' and 'initial tonus' [2] [3]. In a phonetic transcription, auditory impressions are componentially analysed in terms of a finite set of articulatory parameters with recognisable acoustic correlates: therefore, confusion which may result from the welter of labels referred to above is to a large extent eliminated, since their inclusion is not necessary. At the same time, since a phonetic transcription has articulatory referents, it can provide a convenient bridge between an auditory impression of stuttered speech and an empirical analysis using physiological or acoustic techniques.

The transcription system reported in this paper is based on the production of around 800 stuttering disfluencies produced by 36 stutterers and transcribed by a trained phonetician. A full list of the transcribed material is given in [4].

2. METHOD

36 adult, male and female stutterers all attending speech therapy clinics around Edinburgh and Cambridge were recorded in a sound treated recording studio. Two gold-plated, surface electrodes from a Fourcin electrolaryngograph were secured with a band around the subject's neck at the level of the thyroid cartilage. The electrolaryngographic signal was stored on channel 2 of the Revox A77, channel 1 being used for the audio signal. The subject read the first two paragraphs of the *Rainbow Passage* and avoided, as far as possible, the use of any 'techniques' to improve fluency which may have been learned at speech therapy clinics. Following the *Rainbow Passage*, the subject relaxed for at least five minutes while the next recording was prepared in which the experimental design was the same as above. In addition, connecting wires were fed from a Tektronix TM 504 frequency generator to a pair of headphones worn by the subject in the recording studio. The subject was asked to read a list of 200 monosyllabic words one at a time following the offset of a 1 kHz tone from the frequency generator. The stimulus was activated by the experimenter following the production of each word by the subject and was designed to prevent coarticulation across word boundaries.

For the electropalatographic recording, two subjects, one male, one female, were selected from the population of 36 stutterers. In choosing the subjects, it was necessary to ensure that the majority of their disfluencies were realised as some form of lingual-palatal contact. An upper plaster cast impression was made for each subject and from this an acrylic palate containing 62 silver electrodes extending as far back as the junction of the hard and soft palate. These two subjects read the same list of 200 monosyllables described above; in addition, a Reading University electropalatograph [5] connected to a Commodore 3032 computer was used at a sampling rate of 100 Hz to store the palatograms as a function of time.

3. PHONETIC TRANSCRIPTION SYSTEM

3. 1 Syntagmatic Analysis

A subsequent analysis of the transcription suggested that disfluencies can be classified into four broad types:

(1) $\overset{\{\ \}}{[\text{e}]}\ [\text{min}]$ (2) [m] [m] [min]

(3) [m m min] (4) $\overline{[\text{m} \qquad \text{m}}\text{in}]$

The first type of disfluency, which does not bear any phonetic relationship to the target syllable (*mean*), is sometimes referred to as an *interjection*: the production of pause filling sounds such as 'er', 'um' are included under this category. When such interjections are produced, they have been transcribed phonetically (as far as this is possible), and a brace notation has been placed above the transcription. In (2), the disfluency is realised as two [m] segments.

Two criteria have been adhered to in this type of transcription. First, the duration of the [m] segments of the disfluency is approximately the same as the duration of [m] in the target syllable. Second, after the production of each [m] of the disfluency, there may be a pause of several seconds, in which the subject might exhale and inhale. If this kind of transcription has been used, the clear segmentation of the separate [m] segments using acoustic techniques should always be possible. In (3), the [m] segments are also of approximately the same duration as [m] of the target syllable but they are transcribed inside the brackets to indicate that the duration between them is negligible. Spectrographic and laryngographic analyses of type (3) disfluencies show a succession of similar 'events' which often cannot easily be segmented. In the laryngogram of the disfluency in Figure 1, realised as [n n n n n n n], it is possible to detect by eye the repetition of a similar pattern 8 times, but the boundaries between successive repetitions cannot be easily determined.Aaother characteristic feature of type (3) disfluencies is that the duration of the repeated segments tends to be approximately equal. In· Figure 1 for example, each 'segment' appears to consist of either 11 or 12 cycles. In the spectrogram in Figures 2 of a disfluency realised as $[\mathrm{p^h\ p^h\ p^h\ p^h\ p^h}]$ (target syllable *piece*) the duration of successive $[\mathrm{p^h}]$ segments varies betweeen 172 ms and 185 ms.

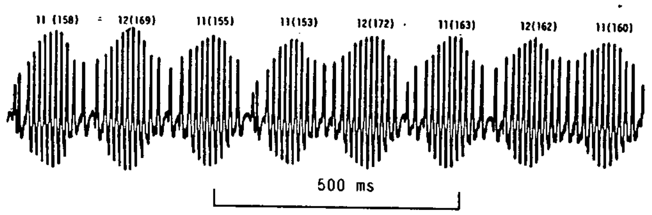


FIGURE 1: Laryngogram of a series of [n] repetitions (target syllable *no*). The estimated number of periods for each repetition is shown above the laryngogram; the number in brackets, adjacent, shows the corresponding duration in milliseconds.

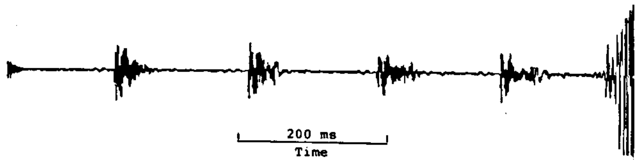


FIGURE 2: Audio wave (digitised at 16 kHz, low-pass filtered to 8 kHz in AUDLAB[1] on a MassComp MC-500) of four $[\mathrm{p^h}]$ segments (target syllable *piece*). The durations of the four $[\mathrm{p^h}]$ segments (closure and aspiration) are 172 ms, 181 ms, 185 ms and 172 ms respectively.

Unlike type (2) disfluencies, the vocal organs do not return to a neutral position between successive segments.

A transcription of type (4) has been used when the disfluency is realised as a prolonged section of the prevocalic consonant(s) of the target syllable and is continuous with the vowel: the duration for which the segment is produced is indicated impressionistically by the length of the bar.

3. 2 Paradigmatic Analysis

In the preceding section, some rules were specified for the realisation of the disfluency as either [X] [X] [X], or [X X X] or $[\overline{X \qquad X}]$ in which [X] is usually phonetically similar to part of the prevocalic consonants of the target syllable. In this section, the discussion will focus on a detailed evaluation of [X] itself and is applicable principally to $[\overline{X \qquad X}]$ disfluencies which were more common than the other two.

In the transcription system to be developed below, the convention is adopted that those attributes which, in auditory terms, are relatively invariant throughout the production of the disfluency are transoribed below, and at either end of, the bar: $[\overline{m \qquad m}]$ designates therefore that an [m] quality pervades the disfluency. Variations away from this constant auditory impression are transcribed as symbols, or diacritics, at various points above, or below, the bar.

3.2.1 Laryngeal Analysis.

The majority of disfluencies in the analysed corpus were either voiceless throughout, fully voiced or produced with a prolonged glottal stop: (5) - (7) show the transcriptions which have been used for these three laryngeal settings in the hypothetical disfluent production of a [w] initial syllable:

(5) $\overline{w \qquad\qquad w}$ (6) $\overline{\underset{\circ}{w} \qquad\qquad \underset{\circ}{w}}$ (7) $\overline{\widehat{wʔ} \qquad\qquad \widehat{wʔ}}$

While it was possible to classify almost all of the disfluencies as either (5), (6) or (7) above, a detailed auditory and laryngographic analysis showed that either of the three phonatory settings could fluctuate to a different short-term setting requiring a transcription with a different symbol of diacritic. If this fluctuation is both audible and very short in duration, then a + symbol is used at a relevant point above the bar, the appropriate symbol or diacritic being transcribed below the + and bar. Thus the production correlate of (8)

(8) $\underset{\sim}{\overset{+}{\overline{w \qquad\qquad w}}}$

is a prolonged voiced labial-velar approximant with a sudden fluctuation to creaky voice at the point marked by the + symbol; furthermore, this fluctuation occurs about half-way into the prolongation, as indicated by the position of the + relative to the onset and offset of the prolonged continuant. By analogy, (9) - (11) are interpreted as follows:

(9) $\underset{\circ\ \ \circ}{\overset{+\ \ +}{\overline{w \qquad\qquad w}}}$ (10) $\underset{\circ \qquad \circ}{\overset{+}{\overline{w\ \ ʔ\ \ w}}}$ (11) $\overset{+}{\overline{w\ \ ʔ\ \ w}}$

in (9)[1], the prolonged continuant is voiced, except for two short intervals of voicelessness at the points marked by +; in (10) and (11), the prolonged continuants are voiceless and voiced respectively except for a brief production of $[\widehat{wʔ}]$ at the points marked by the +.

In the transcription of stuttering disfluencies, it is often necessary to indicate changes in the phonatory setting which span an interval of time longer than the momentary change which corresponds to the + symbol. In this case, the interval is marked by two vertical lines on the bar, as in (12) and (13):

(12) $\overset{|\ |\quad |\ |}{\overline{n \qquad \underset{\circ}{\ } \qquad ʔ \qquad n}}$ (13) $\overset{|\ | \qquad\qquad\qquad |\ |}{\overline{\underset{\sim}{n} \qquad\qquad\qquad \underset{\circ}{n}}}$

In (12), the voiced alveolar nasal is produced continuously, except during two intervals: in the first, a voiceless nasal is produced (for the impressionistic interval indicated by the vertical lines); production of [n] subsequently continues until the next interval marked by the vertical lines in which $[\widehat{nʔ}]$ is produced; the remainder of the disfluency is produced as a fully voiced [n]. In (13), as in (12), a voiced alveolar nasal is produced continuously except for two intervals, one at the beginning (alveolar nasal with creaky voice) and one at the end (voiceless alveolar nasal). It is also possible to include + symbols within an interval of | |. In (14), for example,

(14) $\overset{|+\ \ +|}{\overline{s \qquad \underset{\sim\ \sim}{z} \qquad s}}$

the disfluency is a prolonged voiceless alveolar fricative. For the interval between the vertical lines, the disfluency becomes voiced and, in addition, the onset and offset of the voiced interval are creaky, as indicated by the diacritics beneath the + symbols.

3.2.2 Supralaryngeal Analysis.

The transcriptions in (15) and (16) (target syllable *cap*) are used to describe two different kinds of repetition:

(15) $[\mathrm{k^h} \quad \mathrm{k^hap}]$ (16) $[\overset{+}{\overline{\mathrm{k\ \ x\ \ k}}}\mathrm{{}^hap}]$

In (15), the first $[\mathrm{k^h}]$ has the same phonetic characteristics as $[\mathrm{k^h}]$ of *cap*. The transcription in (15) may correspond to what is habitually referred to in the literature as a stuttering *repetition*. The term 'stuttering repetition' is applicable to disfluencies of type (15) because a section of the prevocalic consonants is repeated once (or several times, usually at approximately equal intervals, as in Figure 2). In (16), the back of the tongue is raised to the velum forming a complete closure for an abnormally long duration. But in addition, this closure is punctuated by a release and turbulence at the point marked by the +. Apart from this momentary release, the closure forms a continuum, making the disfluency inseparable from the target syllable. It is quite possible that the release in (16) is involuntarily caused by the counteracting forces of high intra-oral air-pressure and an abnormally tense tongue-velum contact. If intra-oral air-pressure continually increases during complete occlusion, a point may come when the driving force of the former overcomes the resistance offered by the closure. In this case, the aerodynamic power may force a gap in the sealed tongue-velum contact, and this would result in a brief interval of frication.

It is often the case that the release and frication in disfluencies of type (16) are very short in duration. Figure 3, for example, shows releases which are so short that they are visible on spectrograms as a series of vertical lines that correspond to repeated burst onsets.

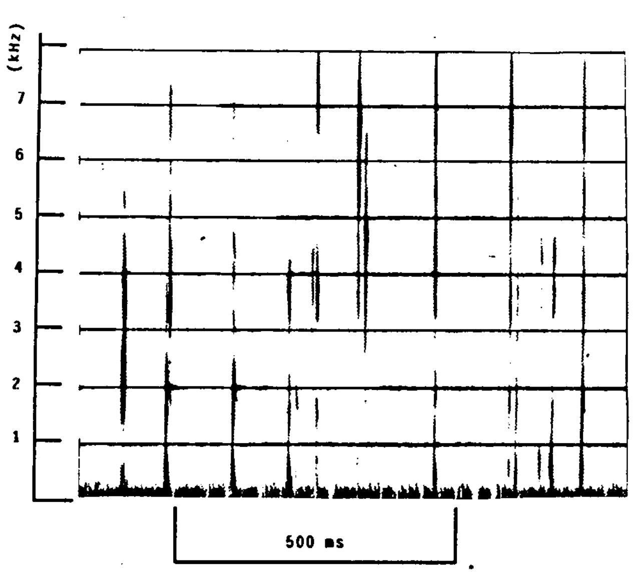


FIGURE 3: Spectrogram (300 Hz analysing filter) of a series of burst releases (target syllable *to*); two such releases occur at an interval of approximately 3 ms.

In order to distinguish these much shorter releases from those in (16), a single point is transcribed which corresponds to each release, as in (17):

(17) $\overset{+\ \ +}{\overline{\mathrm{k\ .\ \ .\ \ k}}}\mathrm{{}^hap}$

Characteristic of many disfluencies is an increase in approximation between active and passive articulators probably caused by excessive tension throughout the vocal tract system, as suggested by Dalton & Hardcastle [6]. The closer approximation of approximants and fricatives resulting in a stop-like production is very noticeable in several subjects and the electropalatographic data often showed a considerable increase in the surface area of lingual-palatal contact in disfluently produced stops and affricates, as shown in Figure 4.

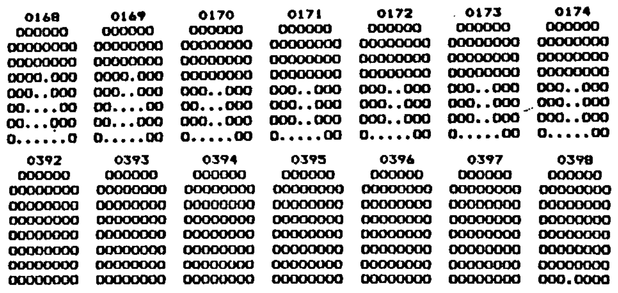


FIGURE 4: palatograms (sampling rate 100 Hz) of a prolongation of an affricate closure (target syllable *Jew*). Palatogram 168 occurs 210 ms after the onset of the closure; the increase in surface area of lingual-palatal contact (0 designates contact) is apparent from palatogram 392 (2.45 ms after the closure onset).

The progressively closer approximation of disfluently produced continuants can be transcribed in the following two ways:

(18) $\overline{\mathrm{s\ \ st\ \ \ t}}$ (19) $\overline{\mathrm{s\ \ t\ \ t}}$

(18) corresponds to the production of a prolonged alveolar fricative which then became a prolonged alveolar stop at the location of the first [t] segment; in (19), the onset of the disfluency is an alveolar

fricative; thereafter, the degree of approximation gradually increases until a stop is produced. In (19), therefore, the closer approximation of the articulators is gradual, whereas in (18) it is comparatively abrupt.

Although the disfluency was most often realised as a section of the prevocalic consonant(s), occasionally there is a 'deflection' in the disfluency towards the vowel target. Thus, medially in a prolonged [n] continuant, for example, a vowel-like production might be audible which is very short in duration. A detailed acoustic and electropalatographic analysis of such vowel-like productions, reported in [7], has shown that they can consist of the entire acoustic vowel onglide of the corresponding fluently produced syllable, but not the acoustic vowel target. A spectrogram illustrating this phenomenon is shown in Figure 5 below.

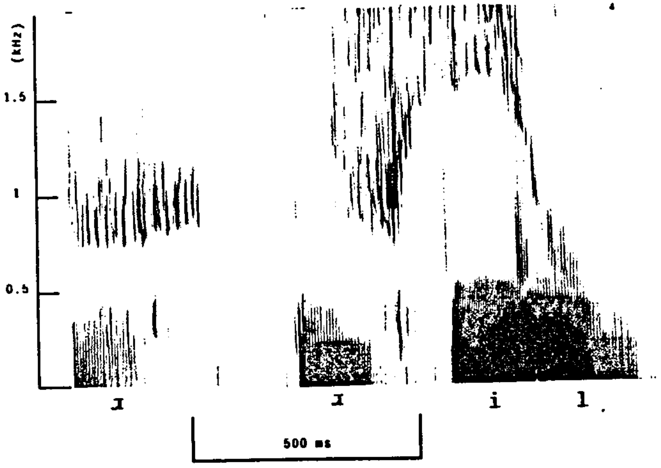


FIGURE 5: Spectrogram (300 Hz filter) of the offset of a prolonged [ɹ] continuant and the onset of *reel*. At the offset of the prolonged continuant, F2 rises less steeply to around 1 kHz compared with the F2 transition from [ɹ] to [i] in *reel* which rises to a target of 1.8 kHz.

The presence of the acoustic vowel onglide, or part of the acoustic vowel onglide, can be indicated by using the relevant vowel as a diacritic, as in (20):

+
e
(20) n n

4. CONCLUSIONS

An examination of a large corpus of stuttered speech [4] has shown that it is possible to make certain generalisations about the phonetic characteristics of disfluencies, at least when stutterers produce monosyllables in isolation; as far as possible, the design of the transcription system has been based on such general phonetic properties.

As discussed in more detail in [7], a stuttering disfluency can consist of any part of the syllable from its acoustic onset to the end of the acoustic vowel onglide (i.e., the acoustic vowel target and post-vocalic consonants are never realised as part of the disfluency). For the great majority of stutterers, part of this section of the syllable is prolonged in time. 'Prolonged in time' means that the dynamic change in shape of the supralaryngeal tract is minimal and that the disfluency is (in the majority of cases) either voiceless or realised as a glottal stop throughout its production. There are some, but less frequent, cases when the disfluency is fully voiced or produced with creaky voice from its onset to its offset. Disfluencies of type (2) and (3), in which an entire section of the prevocalic consonant(s) is repeated, are possible, but they are less frequent than type (4) disfluencies.

Characteristic of very many disfluencies seems to be a boost in the level of tension in the vocal tract system which may cause both the closer approximation of active and passive articulators (examples (18) and (19) and Figure 4) and some stop releases (examples (16), (17) and Figure 3). A more detailed empirical analysis is necessary to enrich this data.

Finally, it is possible, although comparatively rare, for part of the vowel to be realised during the production of a prolonged section of the prevocalic consonants (example 20); in such cases, the audible vowel-like sound is very brief and an acoustic analysis shows that the formant transitions 'bend towards', but never attain, the acoustic vowel target.

REFERENCES

[1] Wingate M.E. (1976) *Stuttering: Theory and Therapy.* Irvington: New York.

[2] Froeschels E. (1943) Survey of the early literature of stuttering, chiefly European. *Nervous Child* 2, 86-95.

[3] Morávek M. & Langová J. (1967) Problems of the development of the initial tonus in stuttering. *Folia Phoniatrica* 19, 109-116.

[4] Harrington J.M. (1986) *The Phonetic Analysis of Stuttering.* Ph.D dissertation, Department of Linguistics, University of Cambridge, England.

[5] Hardcastle W.J. (1972) The use of electropalatography in phonetic research. *Phonetica* 25, 197-215.

[6] Dalton P. & Hardcastle W.J. (1977) *Disorders of Fluency.* London: Arnold.

[7] Harrington J.M. (in press) An acoustic and electropalatographic study of stuttered speech. In Peters H. & Hulstijn W. (eds.) *Speech Motor Dynamics in Stuttering.* Springer Verlag: New York.

NOTES

[1] For clarity, diacritics are transcribed a little further below the segment level than is usual to indicate a phonatory setting that is maintained *throughout* the disfluency (e.g. (10), in which the entire disfluency is voiceless, as opposed to (9), in which two brief intervals of voicelessness occurs medially in the disfluency).

[2] See Terry M., Hiller S., Laver J., & Duncan G. (1986) The AUDLAB interactive speech analysis system. *IEE international conference on speech input/output: techniques and applications. Conference publication* 258, 263-265.

THE AUDITORY MODELLING DILEMMA, AND A PHONETIC RESPONSE

ANTHONY BLADON

Phonetics Laboratory, University of Oxford
41 Wellington Square
Oxford, OX1 2JF, U.K.

A. A DILEMMA IN CURRENT AUDITORY MODELLING

In recent years, results in psychoacoustics and auditory physiology have become routinely available to speech researchers. Numerous computational models of peripheral auditory processing have been published, some being only partial models, but some, including those by the following authors, being rather more complete (see the Blomberg et al. review [5], papers by Cohen, Divenyi, Lyon, Seneff in [14], Dolmazon and Boulogne [9], Cooke [6].)

However, at the time of writing there are many uncertainties about what should go into an auditory model for speech processing. Different models will result, depending on how the investigator views such matters as the following:

(a) which of the many reported psychophysical effects the model incorporates (a partial list could include a tonality scale, frequency masking and resolution, temporal masking and resolution, saturation, equal loudness curves, total loudness, lateral suppression, combination tones, retention of phase information);
(b) which of the physiological findings it seeks to replicate (such as phase-locking, adaptation to a steady-state signal, recovery, probabilistic neural firing, onset/offset asymmetry, efferent intervention, interactions at various stages of the auditory process);
(c) whether it is safe to extrapolate to speech signals, from data of the above types obtained mostly with simpler stimuli; and if not, what modifications to make;
(d) parameters of these models which are intended to be variable (e.g. time windows, bandwidths);
(e) parameters which are empirically variable because we do not yet know what values they should have;
(f) design considerations (more functional versus less so, more data reduction versus less).

In all of these general ways, including the extent to which they have taken specific account of speech, published auditory models reveal considerable differences.

As if this indeterminacy were not itself enough of a nuisance, there is also a multiplicity of answers to the question of methods of evaluation of such models. One method is to use an auditory model to preprocess the signal at the front end of an automatic speech recogniser, and to consider the model to be improved when the recognition rate improves. This brings with it the enormous variable of the recogniser characteristics themselves, for which there is no foreseeable standard. An alternative possibility is to calibrate our auditory models against human perceptual data, such as confusion matrices, perceptual distance judgements, recognition against noise, etc. The main problem here is that there is an acute shortage of such data; but problems of language bias and task differences also add to the difficulty of interpretation. Finally, since almost all auditory models presuppose a calculation of distance between a stored reference pattern and an incoming candidate signal, there is the open question of a distance metric.

Putting together all these uncertainties, the researcher is confronted with a dilemma. It is that, at the current stage of knowledge, we are faced with more variants of auditory models than we can ever possibly test experimentally; and yet, if we do not test the models, there is no way to identify a better model and know when progress has been made. The essence of a modelling exercise is to advance by successive testing and refinement.

Inevitably therefore we need to identify some factors to help limit the search among candidate auditory models. Some expedients which may assist in this task include:
* cost, computability
* best guessing
* functional overlap
* limiting the objective

* improved data on the auditory processes
* improved knowledge in speech perception.

Further elaboration of these possibilities, and of individuals' answers to them, could form a worthwhile discussion issue at this Congress. Our own personal decisions are implicit in the Appendix, in which we briefly sketch the current implementation of an auditory model at Oxford.

B. ONE RESPONSE: SPEECH PERCEPTION

Meanwhile in this paper we concentrate on two strands of research, illustrated mainly from our own work, which can contribute to constraining the search among auditory models. The first, and more familiar, exercise involves trying to refine existing knowledge about speech perception. There is of course nothing new in that, as a research programme. However, the strategy we wish to advocate adds to that position, by suggesting that advances in speech perception research can, when cautiously intepreted, help us to infer (or, more realistically, to state speech-based preferences about) properties of auditory analysis which might underlie the findings. We reason back from these findings, as it were, so as to shape our expectations about an auditory model for speech.

1. Diphthongs

Consider diphthong sounds, for example, as a test case for dynamic auditory modelling of speech. We know that confusions between steady-state vowels and diphthongs are rare; the spectral change in diphthongs is somehow auditorily salient. And it is quite well established that there are auditory mechanisms (e.g. neurons in the cochlear nucleus and inferior colliculus) which respond specifically to a change in a stimulus. However, it is possible to imagine more than one way in which the auditory system might assign importance to this particular kind of changing signal. We can formulate the issue as a speech perception experiment: are diphthongs perceived in terms of their endpoints, or, irrespective of the targets achieved, in terms of a constant rate-of-change? Several authors have addressed this issue experimentally, but in our opinion (see [1]), inconclusively.

In our presentations of diphthong stimuli, which had been artificially cut back in a variety of ways, and when offered a good range of possible transcriptions. of the diphthong quality, our trained listeners consistently responded in terms of the endpoints actually achieved (and not the rate-of-frequency-change). Moreover, when listening to diphthongs whose transitional interval was excised completely, 100% identification was maintained, and the fact that there was an instantaneous spectral jump in these edited diphthongs was hardly noticed at all. At the same time, listening to stimuli consisting of the transition alone (in running speech, but without the early and late steadier-states of the diphthong segment) led to many confusions.

From these studies our first conclusion had to be that "the one thing the ear is not doing, during the transitional part of a diphthong, is estimating the spectral shape change over time" [1, p.152]. Instead, the data were interpreted as suggesting the following role for spectral change in the auditory processing of diphthongs (and perhaps other speech sounds as well). Recall that, whether the spectral change in a diphthong lasts 100 ms or (artificially) 0 ms, it suffices to tell the listener that the sound is diphthongal in quality. The auditory role of spectral change in a diphthong may therefore be, first, as a weighting flag, alerting the system to assign extra distinctiveness to the current stimulus (because it contains spectral change); and second, as a temporal pointer, designating temporal regions of the signal (here, the adjacent endpoints) which the system should inspect more closely for their spectral content.

What do these interpretations mean for a physiologically-based auditory model? Some evidently quite appealing parallels can be drawn - for example, with peaks in neural discharge rate, with adaptation and with recovery in the auditory nerve. Such data show that, when spectral change intervenes, adapted fibres may recover leading to an enhancement of contrast in the adjacent segments (cf. [8]). On the other hand, other prospective model components would fare less well: lateral suppression, for example. We might be justified in inferring that modelling this behaviour would not be productive in the case of a diphthong transition. This is because lateral suppression would predict a frequency-sharpening effect, whereas our findings seem to confirm the other view (cf. [10, 17]) that when listening to rapidly changing signals, the frequency analysis of the ear is much coarser than otherwise.

2. Laterals

Lateral consonants have been another focus of our recent interest. It turns out that, in a limited way which is however reinforced from other speech data, laterals shed light on the question of auditory integration (versus resolution) of frequency. We (Bladon and Burleigh) recently manipulated lateral consonants, both in isolation and in a CV context, in respect of several variables including the width (from 0 to 5 Bark) of an antiformant-like notch in the spectrum. A 1-Bark notch, see Figure 1, which is typical of what regularly occurs in production samples, was essentially undifferentiable from no notch at all. Consistent with other experiments on fricatives, this finding suggests that the auditory filter for speech "smoothes over" a typical lateral consonant's spectral notch. Our experiments were not very sensitive, but a JND for notch width in the region of 2-4 Bark was indicated. The straightforward notion of psychophysical critical band (corresponding to a resolution of 1 Bark) is not, it seems, an appropriate model. A wider-scale integration is operative.

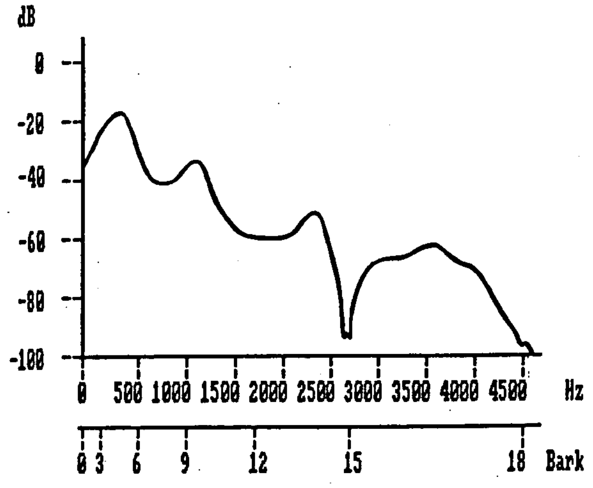


Figure 1. Spectrum of a synthetic lateral consonant used in Experiment 1. The bandwidth of the anti-formant notch was varied from 1 Bark (shown here) to 5 Bark.

Could the lateral's antiformant be detected instead, at least when adjacent to a vowel, by temporal auditory mechanisms, such as enhanced onset/offset of discharge in certain auditory channels? Our results revealed not: notches remained barely detectable, and one can only surmise that the notch is not salient enough (in a word such as "law") to survive temporal masking. Our experiments went on to suggest that the auditory signature of lateralness relies instead on grosser characteristics such as transition duration and overall amplitude envelope.

The concept of a wider-scale (>>1 Bark) auditory integration, for speech sounds, will in due course merit some further attention. We shall return to it at a later stage of the next section, in which we address a second methodological response to the modelling dilemma.

C. A RESPONSE FROM LINGUISTIC PHONETICS

A second way of limiting our testing of auditory models is by virtue of the objective we set. One well established objective, which underlies much of the philosophy of our model given in the Appendix, is to use it for pre-processing the signal supplied to an automatic speech recogniser. But that is not the objective we wish to pursue here. Suppose instead as an interesting objective, that auditory modelling should equip us better to understand the auditory constraints upon language systems and language use. After all, speech is designed not only to be spoken but also to be heard. It turns out that this fact can be inferred to lie at the basis of a whole gamut of properties of sound-systems, their long-term structural trends, distinctive features, and aspects of sound change.

These inferences, and the explanatory value they have for linguistic phonetics, have been fleshed out elsewhere, [2]. Generalising from them to the theme of this paper, it can be said that sound-system properties show evidence of long-term influence from two main kinds of auditory behaviour: one, the asymmetry in auditory representation of energy onsets (which are disproportionately more salient) versus offsets; and two, the wide-scale spectral integration mentioned earlier.

The inclusion of onset/offset asymmetry in an auditory model for speech processing seems well justified by numerous linguistic examples. Summarising [2], there are various instances of unaccounted directionality in phonological behaviour which could be due to the stronger representation of auditory onsets. For instance, phonological nasalisation of vowels spreads very commonly onto a preceding vowel (as it did in the history of French), but only rarely onto a following one. Lateral consonants can vocalise (as in Cockney "field") but commonly do so after, and rarely before, a vowel. The rarity of aspiration after (but not before) a vowel, as in word-final /h/ or in preaspiration, is another often-noted directional asymmetry. In all these cases, a general tendency to spectral energy offset is what characterises the rare occurrence; whereas the common member contains more of an onset. All the cases (as well as others, such as patterns of syllable consonant formation) could well have this auditory foundation.

Now to pick up the earlier reference to the bandwidth of auditory integration. The limited evidence of the lateral consonant notch can be supplemented very considerably, so as to show that much of speech behaviour, especially the long-term organisational properties of sound systems, is consistent with an auditory resolution as wide as some 3.5 Bark. The psychophysical evidence for this idea, it must be said, is still not large; the linguistic evidence, however, is mounting.

If we suppose, then, that two vowel formants are integrated into a single auditory percept when they are less than 3.5 Bark apart, a number of interesting observations follow. Syrdal and Gopal [20] showed how the vowels of American English partition into categories, defined by formant integration versus resolution, which align impressively with the distinctive-feature classification of these vowels into grave/acute, diffuse/compact. The same kind of partition applies if we reanalyse, in Bark-scale integration terms, the Lehiste data [13] for /r,l/ of American English, see Figure 2; and likewise, though not shown here, if we reanalyse in the same way the fricative spectra of Polish reported in [12]. In a nutshell, the identification of certain sounds in language (probably those with a strong spectral pattern), seems to be favoured, on a long-term basis, if they maintain boundaries some 3.5 Bark apart.

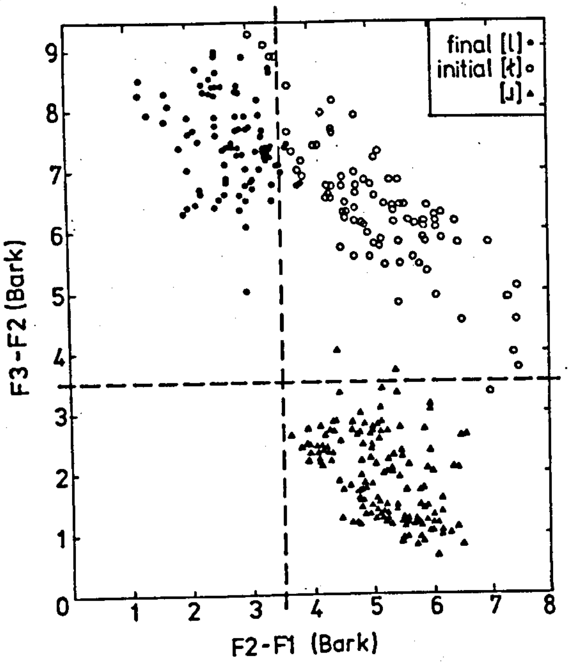


Figure 2. Liquid consonants of American English, their F3-F2 distance plotted against their F2-F1 distance (both in Bark). Each data point is one male (of 6) in one context (of 15).

Space limitations here do not permit the other examples of this kind to be elaborated in detail. In brief, though, the assumption is of a 3.5 Bark band of spectral integration, within which formants will be clearly integrated, outside which they will be clearly resolved, but if falling near the boundary formants will be auditorily less distinct, hence perhaps disfavoured in language and unstable. In these terms, it becomes possible to understand that there could be an auditory motivation for several properties of vowel systems. One such is the under-population, whether in actual languages or in computational simulations of vowel systems, of the "close" region of vowel space. Another is the dimension of "brightness", often noted to be a consistent reality for naive listeners; and a third is the auditory dimension of "rhotacised". We can also understand the strong disfavouring, in languages, of "interior" vowels. Finally, if we imagine vowel space to be a juxtaposition, in (perhaps) three dimensions, of zones of auditory integration/resolution, then we can understand further general properties such as that the number of height distinctions in back vowels is rarely more than the number in front vowels. All of these observations follow from the same basic assumption mentioned at the start of the paragraph; all can be appreciated, with a little patience, from the (rather conjectural) diagramming of cardinal vowels, Figure 3.

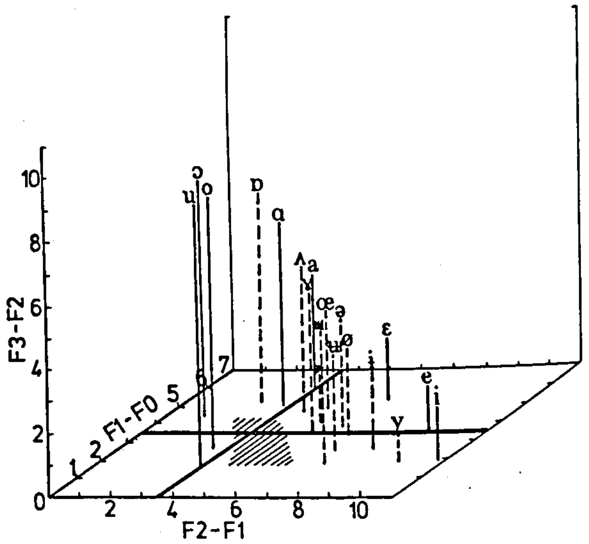


Figure 3. Cardinal vowels visualized in a three-dimensional Bark space defined by whether there is integration or resolution of their various spectral peaks, within a 3.5 Bark band. Primary cardinal vowels as solid lines, secondary ones as dashed lines.

In sum, this demonstration brings home quite forcibly how an auditory model of the identification of vowels in actual languages may well need to incorporate specifics which are not typical of most psycho-physical models on offer today. In addition, some components of existing models may need to be emphasised at the expense of others. To further refine those models, as phoneticians, we may have to resort to the circularity of picking out persuasive trends in the very data we want an auditory model to explain, as a way of focussing the forbiddingly large search space.

APPENDIX: CURRENT OXFORD AUDITORY MODEL

Our current version of auditory modelling routines is built on the foundations of the vowel model used by [4], extending it to include important aspects of the auditory processing of dynamically changing events.

The model is a modular piece of analysis and display software, written in C to run under Unix on a Masscomp 5500 computer with array processor and high-resolution colour graphics system.

The modules which are currently available are outlined below, with skeleton comments: for a fuller description, see [3]. For the honing and encoding of the algorithms in question we are indebted to C. deSilva.

1. Middle ear transfer function

Two alternatives are embodied, following [7], with some simplification. One alternative relates the pressure at the eardrum to the displacement of the stapes, and has essentially the form of a low-pass filter; the other, a stapes velocity function, resembles a pass-band filter with broad skirts.

(a) Displacement function (normalized to unity at 0 Hz):

$$(25.0 / (1.0 + t^{2} (25.0 - 6.0 t^{4})))^{0.5}$$

(b) Velocity function (normalized to unity at its maximum):

$$\sqrt{\frac{20.8925937\ t^{2}}{1.0 + t^{2} (25.0 - 6.0\ t^{4})}}$$

where t = frequency in Hertz/1500.

2. Frequency conversion to the Bark scale

The conversion from the physical Hertz scale of frequency to the Bark scale of tonality or perceived pitch is accomplished by the following formula from Traunmüller (unpublished):

$$\text{Bark} = \left[\frac{26.81\ h}{1960.0 + h} \right] - 0.53$$

where h is the frequency in Hertz.

3. Freq. conversion to the ERB-rate scale

This is intended as an alternative to the Bark scale, or more strictly, a compromise among several different suggested scales of tonality. The conversion from Hertz to ERB-rate is accomplished by the formula from [15]:

$$\text{ERB-rate} = 11.17 \log \left[\frac{k + 0.312}{k + 14.575} \right] + 43.0$$

where k is the frequency in kiloHertz.

4. Frequency smearing

Spectral masking effects are modelled by convolving the spectral values in linear units with a function derived from [18]:

$$10.0 \log \left[15.81 + 7.5\ x - 17.5 \sqrt{1 + x^{2}} \right]$$

where x = 0.474 k (in Bark), and k is a scaling factor which enables a selection of different -3dB bandwidths for this function, according to the relationship:

k=1.429046/required -3dB bandwidth (Bark).

5. Decibel spectra

Spectral values, s, scaled in linear units of pressure are converted to decibels by the familiar relationship:

$$\text{dB} = 20.0 \log(s)$$

6. Equal-loudness curves (phons)

The equal-loudness curves of [16], Table 8, Appendix 4, are used to convert decibel values to phons. These curves are used principally because of the wide range covered, 0 to 15000 Hertz. The phon values are determined by substituting the spectral values in decibels into quadratic functions whose coefficients are functions of the frequency.

7. Total loudness (sones)

Loudness levels in phons are converted to total loudness values in sones by the use of the loudness indices given in [19] Table I, interpolating when necessary. Below 18 phons, loudness is approximated by:

$$\text{sones} = (\text{phons}/40)^{2.884}$$

8. Enhancement of spectral change

Two alternative models (8 and 9 below) are being explored, each of which incorporates some enhancement of the input signal at instants of rapid spectral change - [11], Divenyi in [14]. In the simpler implementation, enhancement of spectral change is carried out by adding, at each frequency point, a multiple of the rate of spectral change at that point. The formula used is:

$$o(h,t) = i(h,t) + a \left[i(h,t+1) - i(h,t-1) \right]$$

where:

$i(h,t)$ is the input spectral value at freq h, time t.

$o(h,t)$ is the output spectral value at freq h, time t.

a is a user-controllable sharpening (= change enhancement) factor.

9. Neural adaptation/recovery effects

As an alternative to the preceding, it is possible to combine some change-related enhancement with other properties of auditory nerve behaviour, specifically neural adaptation/recovery effects. They are modelled by combining a filter which models exponential decay to an equilibrium level with one whose output is related to the derivative of the input. The model has four parameters: (i) equilibrium output level, that is, the steady-state output of the filter when the input is identically zero; (ii) adaptation time constant, which represents the time taken for the output to decay to 1/e of the difference between its initial value and the equilibrium value; (iii) recovery time constant; (iv) input response factor, which determines the amount to which changes in the input are reflected in the output. The formula has:

$$o(h,t) =$$

$$c \cdot o(h,t-1)+(1-c) \cdot y0+r\left[i(h,t) - i(h,t-1)\right]$$

where:

$i(h,t)$ is the input spectral value at frequency h, time t.

$o(h,t)$ is the output spectral value at frequency h, time t.

$y0$ is the equilibrium output level.

r is the input response factor.

c is related to the time constants as follows:

$$c = \exp\left[- \left[\frac{\text{interval between spectra}}{\text{ad/rec time constant}} \right] \right]$$

REFERENCES

[1] A. Bladon, Sp.Comm. 4, 145-154, 1985.

[2] A. Bladon, In G. McGregor, "Language for Hearers", Pergamon, 1-24, 1986.

[3] R.A.W. Bladon, C.J. Clark, C. deSilva, P.F.D. Seitz, Prog.Rep.Oxf.Un.Phonet. Lab. 2, 28-42, 1987.

[4] R.A.W. Bladon, B. Lindblom, J.Acoust. Soc.Am. 69, 1414-1422, 1981.

[5] M. Blomberg, R.Carlson, K. Elenius, B. Granström, In R. Carlson and B. Granström, "The Representation of Speech in the Peripheral Auditory System", Elsevier, 197-201, 1982.

[6] M.P. Cooke, Sp.Comm. 5, 261-281, 1986.

[7] P. Dallos, M.C. Billone, J.D. Durrant, C.-Y. Wang, S. Raynor, Science, 177, 356-359, 1972.

[8] B. Delgutte, N.Y.S. Kiang, J.Acoust. Soc.Am. 75, 897-907, 1984.

[9] J.M. Dolmazon, M. Boulogne, Sp.Comm. 1, 55-73, 1982.

[10] P. Escudier, J.L. Schwartz, Sp.Comm. 4, 189-198, 1985.

[11] S. Furui, M. Akagi, Proc.12th.Int. Cong.Acoust., A2-6, 1986.

[12] W. Jassem, In B. Lindblom, S. Ohman, "Frontiers of Speech Communication Research", Academic, 77-91, 1979.

[13] I. Lehiste, "Acoustical Characteristics of Selected English Consonants", Indiana Univ, 1964.

[14] P. Mermelstein (ed.), "Proc. Montreal Symposium on Speech Recognition", Canadian Acoust. Assn, 1986.

[15] B.C.J. Moore, B.R. Glasberg, J.Acoust. Soc.Am., 74, 750-753, 1983..

[16] D.W. Robinson, R.S. Dadson, Brit.J. App.Phys., 7, 166-181, 1956.

[17] M.R. Schroeder, IEEE Comms.Magazine, 23, 54-61, 1985.

[18] M.R. Schroeder, B.S. Atal, J.L. Hall, In B. Lindblom, S. Ohman, "Frontiers of Speech Communication Research", Academic, 217-229, 1979.

[19] S.S. Stevens, J.Acoust.Soc.Am., 33, 1577-1585, 1961.

[20] A.K. Syrdal, H.S. Gopal, J.Acoust.Soc. Am., 79, 1086-1100, 1986.

AN OPTIMUM PITCH PROCESSING MODEL FOR SIMULTANEOUS COMPLEX TONES

Adrianus. J.M. Houtsma and John. G. Beerends

Institute for Perception Research, P.O. Box 513,
5600 MB Eindhoven, The Netherlands.

ABSTRACT

An extension of Goldstein's Optimum Processing Theory is presented which can account for pitch perception behavior for simultaneous complex tones. The essence of the theory is that all aurally resolved stimulus frequencies are transformed into independent Gaussian random variables with a variance that depends only on the frequency of each partial. A central processor is assumed to use its prior knowledge about the number of simultaneously present tone complexes and the proper parsing of the observed random variables to find the respective fundamentals of the best fitting harmonic templates. In a series of pitch identification experiments for two simultaneous two-tone complexes with diotically and dichotically distributed partials, some model assumptions and their consequences were tested. It was found that (1) the processes of estimating two simultaneous (missing) fundamentals are to a large extent independent, (2) that the central processor tends to group the partial percepts on the basis of common fundamental and not on the basis of ear input, and (3) that pitch identification performance degrades only noticeably if none of the stimulus partials of both tone complexes are aurally resolved.

INTRODUCTION

The problem how we perceive the pitch of complex tones has kept psychoacousticians busy for more than a century. In particular the problem of the so called "missing fundamental", a pitch percept that corresponds with the fundamental frequency of a harmonic tone complex while that complex actually has only overtones, has been the object of many experimental and theoretical studies. Various pieces of important empirical evidence and theories to account for such evidence have been brought forward by Seebeck [1], Ohm [2], Helmholtz [3], Fletcher [4], Schouten [5] and Békésy [6].

More recent experiments by Plomp [7], Ritsma [8] and Houtsma and Goldstein [9] have progressively shown that the real cause of the "missing fundamental" phenomenon must not be sought in the peripheral, but rather in the central part of the auditory system. The new experimental evidence has led to the formulation of some new central pitch theories, of which the Virtual Pitch Theory of Terhardt [10] and the Optimum Processor Theory of Goldstein [11] are the principal variants. These theories were developed and quantified mostly on the basis of pitch perception data obtained with isolated complex tones or short sequences of such tones.

In music, especially in the Western hemisphere, we usually deal with harmonic or polyphonic sound patterns in which either a melody is accompanied with chords or several melodies are played simultaneously against one another. This poses the interesting problem how our auditory system is able to perceive two or more simultaneous pitches when it is acoustically exposed to a cluster of harmonics that belong to several different tone complexes. The same problem actually occurs when one tries to track the prosodic contours of two simultaneously spoken sentences or, more realistically, when one tries to follow the pitch contour of one spoken sentence against a background of other speech. Although both central pitch theories mentioned [10,11] are in principle able to cope with this problem, this has never been worked out specifically or tested against systematic empirical data.

In this study the Optimum Processor Theory of Goldstein will be extended and tested with experimentally obtained pitch identification data for two simultaneous complex tones. The model extension will be treated in Sect. I. Descriptions of the experimental procedure and the results are given in Sects. II and III. Computer simulations of model performance are discussed in Sect. IV, and conclusions of the study are presented in Sect. V.

I. EXTENSION OF THE OPTIMUM PROCESSOR THEORY

In Goldstein's Optimum Processor Theory [11] and in a later extension of that theory [12] it was assumed that :

1. the complex tone input in both ears is spectrally analyzed and only frequency information of sufficiently resolved partials is kept; phase and amplitude information is discarded;
2. independent Gaussian random variables r_i, of zero mean and with variance depending on frequency only, are added to each resolved frequency to form the noisy frequeny codes $x_i = f_i + r_i$;
3. a central processor rank-orders all noisy frequency codes from both ears and performs a maximum-likelihood estimate of the best-fitting harmonic numbers and fundamental of some underlying harmonic complex-tone template.

This model, which was originally formulated to describe perception of a single pitch from a single complex tone, can easily be extended to accomodate identification tasks of pitches from simultaneously sounding complex tones. In this study we will focus on the task of identifying two fundamental pitches in an acoustic stimulus that comprises two simultaneous two-tone complexes, each one having successive harmonics. Extension of the model to other cases, e.g., three or four simultaneous two-tone complexes or two simultaneous multi-tone complexes, is, in principle, not different but may be computationally more complex.

Suppose now that the acoustic stimulus consists of four frequencies: $f_1 = mf_{01}$, $f_2 = (m+1)f_{01}$, $f_3 = nf_{02}$, and $f_4 = (n+1)f_{02}$, and that these frequencies are all peripherally resolved by the auditory system. The frequencies f_1 through f_4 are then transformed into four independent Gaussian random variables x_1 through x_4, having means of f_1 through f_4 respectively, and standard deviations $\sigma(f_1)$ through $\sigma(f_4)$. If we denote $\sigma(f_i)$ simply as σ_i, the likelihood function to be optimized by the processor is given by the expression:

$$\begin{aligned} L(f_1,f_2,f_3,f_4) \;=\; & \frac{1}{4\pi^2\sigma_1\sigma_2\sigma_3\sigma_4}.exp[-\frac{(x_1-f_1)^2}{2\sigma_1^2}]. \qquad (1)\\ & exp[-\frac{(x_2-f_2)^2}{2\sigma_2^2}].exp[-\frac{(x_3-f_3)^2}{2\sigma_3^2}].\\ & exp[-\frac{(x_4-f_4)^2}{2\sigma_4^2}]. \end{aligned}$$

Maximizing Eq.(1) is equivalent to maximizing the log-likelihood function:

$$\begin{aligned} \Lambda(f_1,f_2,f_3,f_4) \;=\; & -\frac{(x_1-f_1)^2}{\sigma_1^2} - \frac{(x_2-f_2)^2}{\sigma_2^2} - \qquad (2)\\ & -\frac{(x_3-f_3)^2}{\sigma_3^2} - \frac{(x_4-f_4)^2}{\sigma_4^2}. \end{aligned}$$

In interpreting this log-likelihood function, the knowledge the processor has about the make-up of the stimulus and the task to be performed becomes very important. We will fist discuss the case (A) in which the processor has full knowledge of the fact that there are two two-tone complexes, and hence two pitches to be found, as well as knowledge of the correct parsing, i.e., of the correct harmonic interpretation of each observed input x_i. We will then discuss another case (B) where the number of complex tones present is known, but the correct parsing is unknown to the processor.

Case A. When the number of fundamental pitches to be identified and also all parsing information is available to the processor, it makes the following substitutions in Eq.(2):

$$\begin{aligned} f_1 &= \hat{m}\hat{f}_{01} \qquad (3)\\ f_2 &= (\hat{m}+1)\hat{f}_{01}\\ f_3 &= \hat{n}\hat{f}_{02}\\ f_4 &= (\hat{n}+1)\hat{f}_{02} \end{aligned}$$

and maximizes the expression with respect to the (lower) harmonic number estimates $\hat{m}$ and $\hat{n}$ and the fundamental pitch estimates $\hat{f}_{01}$ and $\hat{f}_{02}$. Because of the statistical independence of the input variables x_i, the first two terms and the last two terms of Eq. (2) can be maximized separately. The two independent fitting procedures, each one identical to the one described by Goldstein [11], yield the optimum harmonic-number estimates $\hat{m}$ and $\hat{n}$ as well as the fundamental pitch estimates $\hat{f}_{01}$ and $\hat{f}_{02}$. The probabilities $\Pr[\hat{m} = k]$ and $\Pr[\hat{n} = l]$, with k and l being integers, are discrete probabilities which can be computed from the stimulus frequencies f_i and the fixed and known frequency coding noise function $\sigma(f_i)$, and the fundamental pitch estimates are given by the expressions:

$$\begin{aligned} \hat{f}_{01} &= \frac{[x_1/\hat{m}]^2 + [x_2/(\hat{m}+1)]^2}{x_1/\hat{m} + x_2/(\hat{m}+1)} \qquad (4)\\ \hat{f}_{02} &= \frac{[x_3/\hat{n}]^2 + [x_4/(\hat{n}+1)]^2}{x_3/\hat{n} + x_4/(\hat{n}+1)}. \end{aligned}$$

The probability density functions of the estimates $\hat{f}_{01}$ and $\hat{f}_{02}$ are nearly-discrete functions with the main modes at f_{01} and f_{02}, the correct fundamental estimates, with probabilities $\Pr[\hat{\mathrm{m}} = \mathrm{m}]$ and $\Pr[\hat{\mathrm{n}} = \mathrm{n}]$ respectively. Correct identification of the two pitches therefore boils down to two independent correct identifications of the respective lower harmonic numbers m and n.

Case B. When the processor only knows the number of fundamental pitches to be identified, but does not have any information about the proper parsing of the input variables x_i, it tries, in principle, all possible interpretations of the x_is which are, in this case, 24 permutations. In practice, only the following three permutations are relevant in most cases because of simple ordinal properties of the input variables and their possible interpretations:

$$\begin{array}{lll} (a) & (b) & (c)\\ f_1 = \hat{m}\hat{f}_{01} & f_1 = \hat{m}\hat{f}_{01} & f_1 = \hat{m}\hat{f}_{01}\\ f_2 = (\hat{m}+1)\hat{f}_{01} & f_2 = \hat{n}\hat{f}_{02} & f_2 = \hat{n}\hat{f}_{02}\\ f_3 = \hat{n}\hat{f}_{02} & f_3 = (\hat{m}+1)\hat{f}_{01} & f_3 = (\hat{n}+1)\hat{f}_{02}\\ f_4 = (\hat{n}+1)\hat{f}_{02} & f_4 = (\hat{n}+1)\hat{f}_{02} & f_4 = (\hat{m}+1)\hat{f}_{01} \end{array}$$

Group (a), of course, represents the correct parsing, but the interpretations of (b) and (c) may result in a larger likelihood function value and therefore in a better fit on a given trial because of the noise in the variables x_i. Interpretations (b) and (c) will almost always lead to incorrect pitch identifications, however. We will refer to such mistakes as parsing errors.

It is far from clear whether or not the extension of the Optimum Processor Theory as it has been described so far offers a realistic account of human pitch perception for situations of simultaneous complex tones. Some particular questions remain to be answered. Are the fundamental estimation processes for each complex tone in a chord really independent ? Does the processor actually have knowledge of the correct parsing and interpretation of the perceived partials, or can such knowledge be externally supplied ? When two partials of different complex tones have exactly or almost the same frequency, are they both unavailable to the processor because they are peripherally unresolved, or is some frequency information still transmitted to the processor ? These questions are investigated in the following set of experiments.

II. EXPERIMENTS

Musically experienced subjects performed a series of pitch identification experiments with two simultaneously sounding notes, each note made with a harmonic two-tone complex. One complex, representing the lower note, comprised the frequencies $f_1 = mf_{01}$, $f_2 = (m+1)f_{01}$, the other complex representing the higher note the frequencies $f_3 = nf_{02}$, $f_4 = (n+1)f_{02}$. The respective fundamentals f_{01} and f_{02} were both elements of the note set {do, re, mi, fa, so} or, equivalently, the frequency set {200, 225, 250, 267, 300} Hz, and could not be the same on any given trial. Both lower harmonic numbers m and n were independent random integers beteen 2 and 10. Note durations were 600 ms and intensities were 20 dB above threshold, with 30-dB SL broadband noise as a general masking background. Sound stimuli, which were computed and stored on a Philips P857 minicomputer, were played back through a 2-channel, 12-bit D/A converter and presented through headphones to the subject who was seated in a sound-insulated chamber. The task of the subject was to identify both simultaneously perceived (missing) fundamentals f_{01} and f_{02} on each trial by pressing two out of five buttons on a response box in any temporal order. There was unlimited response time, and each response triggered presentation of a new trial after a brief fixed delay.

Four stimulus conditions were investigated. In condition 1, which was diotic, all four stimulus frequencies f_1 through f_4 were presented to both ears. In condition 2, which was dichotic, one note (comprising the frequencies f_1 and f_2) was presented to one ear, while the other note (with the frequencies f_3 and f_4) went to the other ear. In conditions 3 and 4, which were also dichotic, the frequencies of both notes were split up between the ears. In condition 3, one ear received f_1 and f_3, while the other ear received f_2 and f_4. In condition 4, one ear received f_1 and f_4 while the other ear received f_2 and f_3.

Since there are ten different combinations of two notes in a total set of five, and since there were $9 \times 9 = 81$ different harmonic representations of each two-note combination, there was a total of 810 physically different stimuli and a total of 10 different response categories. Each of these stimuli was, on the average, presented six times to each of four subjects, for a total of 4500 identification trials per subject for each stimulus condition.

III. RESULTS

The raw data of all experiments consisted of a record for each trial of the presented fundamentals f_{01} and f_{02}, the lower harmonic numbers m and n, and the subject's two responses R_a and R_b. A response (R_a, R_b) could be an identification of the perceived fundamentals (f_{01}, f_{02}) or (f_{02}, f_{01}), since the order of pressing the response buttons was arbitrary.

To obtain some insight in the perceptual independence of the identification processes for each of the two simultaneous notes, the raw data were processed by two different methods. In the first method all trials were counted for every (m, n) combination where both f_{01} and f_{02} were identified correctly. The results of this way of counting yielded half-matrices of 'percent correct' scores, $\mathrm{Pc}(k,l)$, for each subject, in which k and l are integers representing the harmonic numbers (m,n) or (n,m). They are half-matrices because of the built-in symmetry around the main diagonal, which makes both halves of the matrix mirror images. In the second method only the correct identification of one of the two simultaneous notes was considered as a function of both (lower) harmonic numbers but regardless of the identification response to the other note. The resulting score, designated as $\mathrm{Pc}(k|l)$, represents the percentage correct identifications of f_{01} for $k = m$ and $l = n$, as well as the correct identifications of f_{02} for $k = n$ and $l = m$. The total count for each subject yielded full 9x9 matrices.

Both processed data matrices $\mathrm{Pc}(k,l)$ and $\mathrm{Pc}(k|l)$ can be used to find an underlying processor performance function $\Pr[\hat{k} = k]$, the processor's probability of correctly estimating the harmonic order of any complex tone. This was done with a minimum chi-square fitting procedure which looked for those $\Pr[\hat{k} = k]$ functions that provided the most likely account of the empirically obtained data matrices $\mathrm{Pc}(k,l)$ and $\mathrm{Pc}(k|l)$. The details of this procedure, which also involved some assumptions about the decision process for the particular experimental paradigm that was used, are discussed in a recent publication by the authors [13].

The functions $\Pr_1[\hat{\mathrm{k}} = \mathrm{k}]$ derived from the matrix $\mathrm{Pc}(k|l)$and $\Pr_2[\hat{\mathrm{k}} = \mathrm{k}]$ derived from $\mathrm{Pc}(k,l)$ are shown in Fig. 1a-d as triangles and squares respectively for the experimental conditions 1 through 4. One can show that, if $\Pr_1[\hat{\mathrm{k}} = \mathrm{k}] > \Pr_2[\hat{\mathrm{k}} = \mathrm{k}]$ for low values of k and $\Pr_1[\hat{\mathrm{k}} = \mathrm{k}] < \Pr_2[\hat{\mathrm{k}} = \mathrm{k}]$ for large values of k, the two fundamental pitch identification processes are mutually dependent in the sense that the perception of the more salient pitch, i.e., the one represented by the lowest harmonic numbers, inhibits correct perception of the less salient pitch [13]. Figure 1a-d shows that in condition 2 only subject JH noticeably exhibits this effect of mutual dependence of the two identification processes, but in conditions 1, 3 and 4 all subjects except MZ seem to show a small amount of mutual dependence.

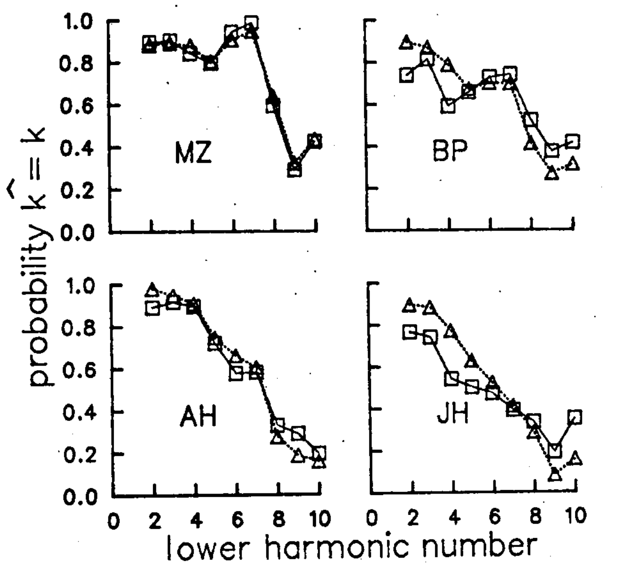


Fig. 1a. The processor's probability of identifying the correct harmonic order of a complex tone. The harmonic order is shown on the abscissa. Triangles designate $\Pr_1[\hat{\mathrm{k}} = \mathrm{k}]$, squares $\Pr_2[\hat{\mathrm{k}} = \mathrm{k}]$. Computed from data of condition 1.

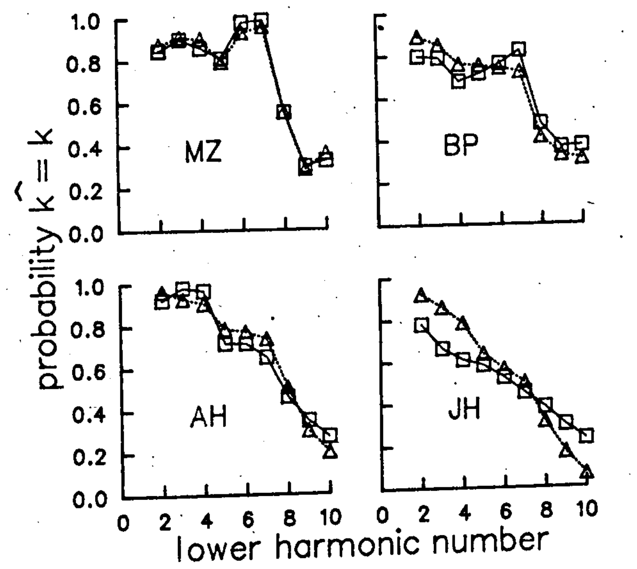


Fig. 1b. Same as Fig 1a, but computed from data of condition 2.

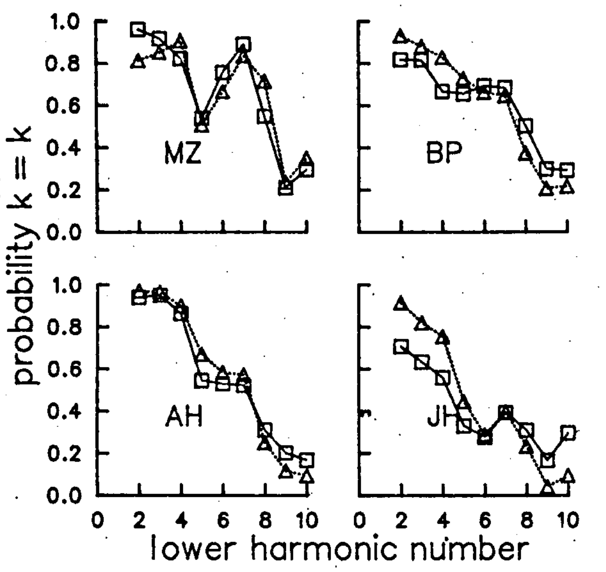


Fig. 1c. Same as Fig 1a, but computed from data of condition 3.

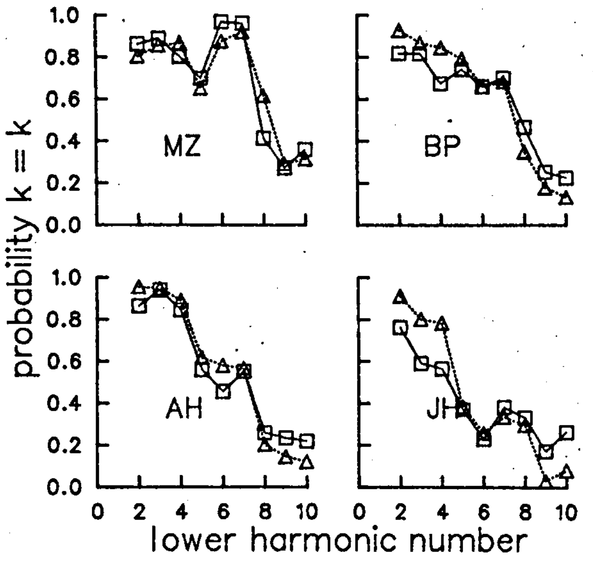


Fig. 1d. Same as Fig 1a, but computed from data of condition 4.

From either probability function $\mathrm{Pr}_1[\hat{\mathrm{k}} = \mathrm{k}]$ or $\mathrm{Pr}_2[\hat{\mathrm{k}} = \mathrm{k}]$ one can now compute the model's variance function $\sigma(f)$ which represents the frequency coding noise and is its only free parameter. A set of those sigma functions is shown in Fig. 2. The functions were computed from the averaged $\mathrm{Pr}_1[\hat{\mathrm{k}} = \mathrm{k}]$ and $\mathrm{Pr}_2[\hat{\mathrm{k}} = \mathrm{k}]$ functions obtained from the experimental data of dichotic condition 2. The $\sigma(f)/f$ functions have the typical U-shape which was also found in an earlier study [11], and have also the same general magnitude. The low-frequency slopes of these functions, however, are much steeper than the average slope found in that earlier study. We think this is caused by an over-estimate of $\sigma(f)$ at low frequencies in the present experiments. Partial frequencies below 1000 Hz, with fundamentals limited between 200 and 300 Hz, could occur only for very low harmonic numbers where identification is close to perfect and occasional mistakes are more made through carelessness or poor attention than through insufficient salience of pitches.

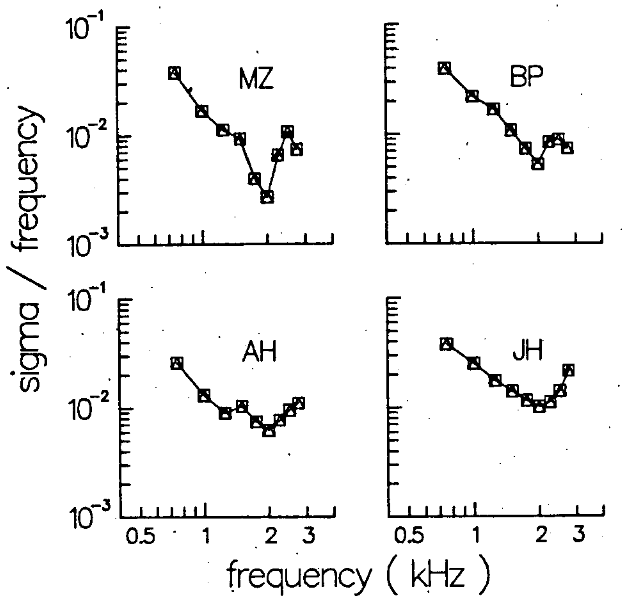


Fig. 2. Model variance or "noise" functions $\sigma(f)/f$ computed from the averaged functions Pr_1 and Pr_2 shown in Fig. 1b.

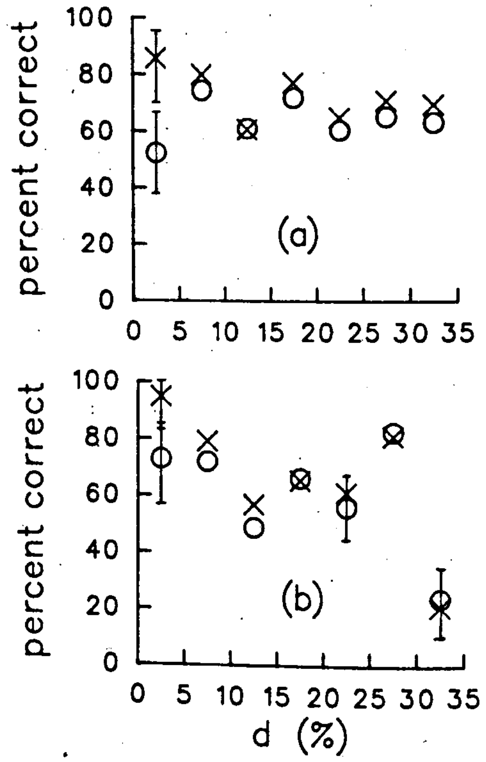


Fig. 3. Correct identification scores for both simultaneous fundamentals as a function of d defined by Eq. (5). (a) Diotic condition 1 (circles) and dichotic condition 2 (crosses). (b) Dichotic conditions 3 (circles) and 4 (crosses).

In order to examine the influence of spectral interference on pitch identification performance, all 810 different stimuli were mapped on a frequency-difference measure d, defined as:

$$d = \sqrt{d_1^2 + d_2^2}, \qquad (5)$$

where d_1 represents the smallest frequency difference between any two harmonics and d_2 the next smallest difference in the total four-tone stimulus. Stimuli were grouped on this d-scale in bins of 5 % and, in order to limit the general degrading effect of high harmonic numbers on pitch identification performance, only stimuli were included having lower harmonic numbers m and n of 5 or less. Percentages correct pitch identification of both fundamentals as a function of d are shown in Fig. 3a for experimental conditions 1 and 2. Under diotic condition 1 values of d below 10 % imply that the partials mf_{01}, nf_{02} as well as the partials $(m+1)f_{01}, (n+1)f_{02}$ must have interfered with one another because of limited frequency resolution in the cochlea. Under dichotic condition 2 such interference was not possible because potentially interfering partials went to different ears. Figure 3a shows that only for the lowest d-values, between 0 and 5 %, there is a noticeable difference between the scores of conditions 1 and 2. The figure also shows, however, that performance for diotic condition 1, although degraded, is still well above the expected chance level of 10 % correct. Similar results were obtained with the data from conditions 3 and 4. They are shown in Fig. 3b.

IV. MODEL SIMULATIONS

The data presented in the previous section show a general performance deterioration with increasing (lower) harmonic numbers m and n, and also a dependence of performance on the the presentation conditions 1 through 4. The data still provide insufficient information, however, about the relative contribution of parsing errors compared with errors caused by interference of partials or mutual dependence of pitch identification processes. To study the influence of parsing errors in more detail, a computer simulation experiment was performed with the model discussed in Sect. I. To simulate each subject's performance, the $\sigma(f)/f$ functions derived from the data of condition 2 were substituted in the model to specify the exact amount of noise to be added to each frequency component of the simulation input. For all 810 stimuli 25 computations were made (with new noise samples added to partials each time) of the maximum log-likelihood function of Eq. (2) without knowledge of the correct parsing, as outlined in Case B of Sect. I. Simulations were made on a Vax 11/780 computer. Those stimuli for which the correct parsing was always obtained were put in a stimulus subset PNS (parsing-non-sensitive). The remaining stimuli, for which (occasionally) the likelihood function came out maximum with the wrong parsing, were put in the subset PS (parsing-sensitive). With the subsets PS, PNS and also with the entire set PS+PNS, the simulation experiment was now repeated for all four stimulus conditions (1 through 4) and with substitution of the appropriate $\sigma(f)/f$ function obtained from the data of each particular subject under that condition with stimulus subset PNS. Simulation was done with the use of Eq. (3), implying knowledge of the correct stimulus parsing, and with substitution of all 24 possible permutations outlined under Case B of Sect. I, implying the absence of this knowledge. The results, expressed as a percentage correct identifications of both fundamentals f_{01} and f_{02} pooled over all values of m and n, are shown in Fig. 4 for stimulus conditions 1 and 2. For each of the three stimulus (sub)sets, the solid line represents the actual performance of the subject, the dashed line the performance level simulated with parsing knowledge, and the dotted line the level simulated without this knowledge. For the PNS-subset one expects all three performance levels to be identical. The fact that this is not exactly the case is not a truncation effect in which the number of experimental and simulated trials was smaller than required by the Law of Large Numbers, but represents a small uncertainty about the details of the simultated decision strategy. One also observes that for the subset PS and for the entire stimulus set PS+PNS the performance level of all subjects (solid lines) is much closer to the performance level simulated with parsing knowledge (dashed lines) than to the level simulated without that knowledge (dotted lines). This is true for dichotic condition 2 as well as for diotic condition 1. Results similar to the ones shown in Fig. 4 were obtained for stimulus conditions 3 and 4. This finding is important because in the diotic condition 1 no explicit parsing information was supplied to the subjects, and in conditions 3 and 4 an explicit attempt was actually made to supply them with false parsing information. If this wrong information had been used by the subjects, their performance would have been at chance level, which was easily shown by simulation. Actual performance was well above chance level for those conditions, however, as is evident from Figs. 1c,d. The empirical and simulated results tell us that subjects somehow do have a fairly accurate knowledge of the proper interpretation of the various stimulus partials in the percept of simultaneous complex tones, but that this knowledge is not obtained on the basis of ear input. It is probably obtained on the basis of experience with the harmonies of the stimulus set and a general tendency to group perceived partials holistically on the basis of common fundamental. Something similar was also found by Deutsch [14] and Butler [15] who used entirely different musical paradigms.

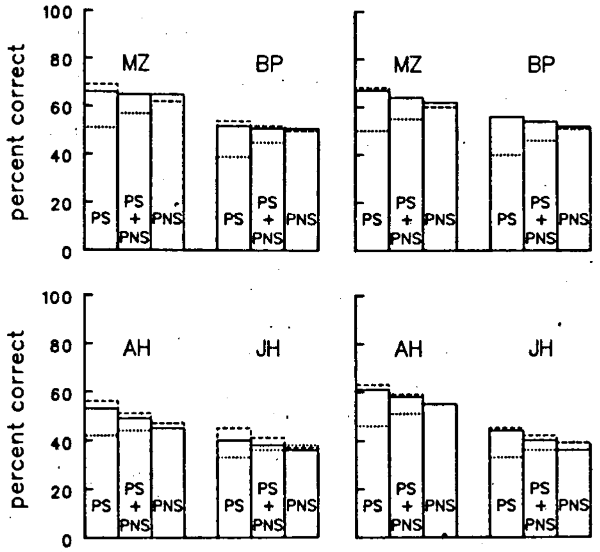


Fig. 4. Measured (solid lines) and simultated performance levels with (dashed lines) and without (dotted lines) knowledge of stimulus parsing. Columns PS are for the stimulus subset that is prone to parsing errors, columns PNS for the subset that does not induce such errors, the central columns for the entire set. Left: diotic condition 1; right: dichotic condition 2.

V. CONCLUSIONS

From the experimental and simulated results of this study the following conclusions are drawn:

1. The task of identifying two pitches when exposed to two simultaneous complex tones is separable into two pitch identification processes which are to a large extent independent. The small amount of mutual dependence that is sometimes observed tends to support the notion that the more salient pitch, represented by lower-order harmonics, is processed first and interferes with the processing of the less salient pitch. This mutual dependence of the two processes, small as it may seem, is largely responsible for the degradation of performance when going from dichotic condition 2 to diotic condition 1 and finally to dichotic conditions 3 and 4.
2. Information of the correct parsing and interpretation of perceived stimulus partials is, to a large extent, available to the central pitch processor. It is independent of the manner in which partials are distributed between the ears. This is consistent with other results on simultaneous-melody perception in the literature [14,15], and with informal observation of ordinary musical practice in which both ears are always exposed to all partials of simultaneously-playing musical instruments.
3. Interference of spectrally close partials has a surprisingly small effect on pitch identification for two complex tones, at least as long as either tone is represented by harmonics of sufficiently low order. Although it is known that high-order harmonics of a single complex tone do not contribute much to fundamental pitch sensation [9] and are as such not available to the central processor [11], it now appears that aurally non-resolved harmonics belonging to different tone complexes are not entirely discarded. They may instead be transformed into a single (noisy) percept that can be used more than once by the processor when filling in the variables of Eqs. (1) or (2). This idea will be investigated further in a future study.
4. The human central pitch processor appears not to be hard-wired or specifically programmed for one particular way of processing stimulus tones. On the contrary, its processing algorithm appears to be quite *cooperative* and *interactive* with the task it has to execute.

VI. ACKNOWLEDGMENTS

This work, initiated under a grant from the U.S.A. National Institutes of Health (NS 11680) was executed with a grant from the Netherlands Foundation for the Advancement of Pure Research (ZWO 560-260-009).

VII. REFERENCES

[1] A. Seebeck, "Beobachtungen über einige Bedingungen der Entstehung von Tönen", Ann. Phys. Chem. 53, 417-436, 1841.

[2] G. Ohm, "Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen", Ann. Phys. Chem. 59, 513-565, 1843.

[3] H.L.F. von Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, F. Vieweg & Sohn, Braunschweig, 1863.

[4] H. Fletcher, "The physical criterion for determining the pitch of musical tone", Phys. Rev. 23, 427-437,1924.

[5] J.F. Schouten, "The residue and the mechanism of hearing", Proc. Ned. Akad. Wetenschap 43, 991-999, 1940.

[6] G. von Békésy, "The missing fundameental and periodicity detection in hearing", J. Acoust. Soc. Am. 51, 631-637, 1972.

[7] R. Plomp, "Pitch of complex tones", J. Acoust. Soc. Am. 41, 1526-1533, 1967.

[8] R.J. Ritsma, "Frequencies dominant in the perception of the pitch of complex sounds", J. Acoust. Soc. Am. 42, 191-198, 1967.

[9] A.J.M. Houtsma and J.L. Goldstein, "The central origin of the pitch of complex tones: evidence from musical interval recognition", J. Acoust. Soc. Am. 51, 520-529, 1972.

[10] E. Terhardt, "Zur Tonhöhewahrnehmung von Klängen II: Ein Funktionsschema", Acustica 26, 187-199, 1972.

[11] J.L. Goldstein, "An optimum processor theory for the central formation of the pitch of complex tones", J. Acoust. Soc. Am. 54, 1496-1516, 1973.

[12] A. Gerson and J.L. Goldstein, "Evidence for a general template in central optimal processing for pitch of complex tones", J. Acoust. Soc. Am. 63, 498-510, 1978.

[13] J.G. Beerends and A.J.M. Houtsma, "Pitch identification of simultaneous dichotic two-tone complexes", J. Acoust. Soc. Am. 80, 1048-1055, 1986.

[14] D. Deutsch, "Two-channel listening to musical scales", J. Acoust. Soc. Am. 57, 1156-1160, 1975.

[15] D. Butler, "A further study of melodic channeling", Percept. Psychophys. 25, 264-268, 1979.

THE INFLUENCE OF INTERAURAL PHASE UNCERTAINTY ON BINAURAL SIGNAL DETECTION

ARMIN KOHLRAUSCH

Drittes Physikalisches Institut, Universität Göttingen
Bürgerstr.42-44, D-3400 Göttingen, FR Germany

ABSTRACT

This study investigates whether binaural signal detection is improved by the listener's a priori knowledge about the interaural phase relations. We measure binaural masked thresholds and vary the interaural phase of masker and test signal randomly within the same measurement. A comparison of the results with experiments applying a fixed binaural configuration shows no significant differences. The results allow an examination of different model predictions in relation to the simultaneous processing of signals with distinct interaural phase relations.

INTRODUCTION

Speech perception in background noise is to a large extent dependent on the function of the binaural hearing system. This fact can be tested qualitatively by occlusion of one ear in a typical "cocktail-party" situation and, more exactly, by listening tests in a defined acoustical condition. A quantitative measure of the noise reducing ability is given by the Binaural Masking Level Difference (BMLD), the threshold difference between monaural and binaural signal presentation. Binaural thresholds depend on the interaural parameters (time and level differences) of the background (masking noise) and the test stimulus. Similar to monaural experiments, only the interaural parameters within a limited frequency range around the test frequency contribute to the masking /1/.

The experiments in this study investigate a specific binaural aspect of signal detection. This aspect shall be explained by a short discussion of two models for binaural signal processing proposed by Durlach /2/ and Colburn /3,4/.

In the Equalization and Cancellation (EC) theory /2/, binaural unmasking is explained by mathematical operations, which are performed on the acoustical inputs to both ears in order to reduce the intensity of the masking signal. In a first "Equalization" step the maskers from the left and the right ear are adjusted to each other by internal transformations of amplitude (by attenuation) and time (by delay). These transformations are accompanied by errors, described as amplitude and time jitter. Therefore, the subtraction of the two adjusted masking signals in the second step does not totally cancel the masker intensity. For most interaural phase relations, however, this binaural processing leads to an increased signal-to-noise ratio, which is directly related to the lower binaural masked thresholds. The transformations are performed on the peripherally bandpass filtered signals within a critical band. In the description of the theory, it remains unclear whether this system is able to apply distinct transformations simultaneously.

The "auditory-nerve-based model" from Colburn /3,4/ differs from the EC-theory by including a detailed description of the peripheral transduction process from acoustical waveforms to neural activity. In the central part of the model, the synchronous neural activity is measured for pairs of fibres from the left and right acoustic pathway having identical best frequency f_i and a specific internal time delay τ_i. This part of the model can be described as a two-dimensional pattern of coincidence detectors with internal delay τ and best frequency f as the two dimensions. For a fixed frequency f_i, the coincidence values along the τ-axis represent an estimate of the cross-correlation function of the input to the right and the left ear within the frequency channel i. From the activity within this two dimensional pattern a decision variable is derived, which can be used to calculate binaural masked thresholds /4/. As all coincidence detectors analyze the input signals simultaneously, different internal delays (corresponding to different Equalization transformations in the EC-theory) can be applied even within the same frequency channel simultaneously.

The experiments described in this paper were performed to test the differences between the two models in this point. We used a binaural masking noise with distinct interaural phase relations in different frequency regions. Thus, according to the EC-theory, the optimal binaural processing strategy had to be different for different test signal frequencies. By introducing uncertainty about the test signal frequency and phase, we could test whether a priori knowledge of the interaural phase relations is advantageous for the listeners, as it would be predicted by the EC-theory.

METHOD

Apparatus

The experimental setup for measuring binaural masked thresholds is shown in Fig.1. The experiments were controlled by a 16 bit microcomputer TI 980A, which also generated the sinusoidal test stimuli. They were converted to analog signals by means of a 2 channel 12 bit D/A-converter at a sampling rate of 5 kHz, low pass filtered at 1 kHz and attenuated. The dichotic noise masker had a steep transition of the interaural phase difference from 0 to π at 500 Hz. This masker was generated digitally and stored on magnetic tape. Computer controlled gate switches were used to turn the noise on and off at the appropriate instants of time. The masker was low pass filtered at 2.5 kHz and presented at an overall level of 75 dB SPL. Masker and test signal were added with the appropriate interaural phase relations and presented to the subject over headphone (Sennheiser HD 44) in a sound insulated booth.

Threshold Procedure

Binaural masked thresholds were determined with an adaptive 3 Interval Forced Choice (3 IFC) procedure. The 500 ms noise masker was presented in three sequential intervals separated by short breaks of 100 ms. In one randomly chosen interval, the test signal was added to the temporal center of the masker. In the main experiment, the test signal had a duration of 20 ms including 5 ms linear ramps. After each trial (a group of three intervals), the subject had to specify the number of the interval containing the probe tone. The level of the test signal was changed adaptively following a two-down-one-up rule /5/. After two subsequent correct responses, the level was decreased by 1 dB, after each incorrect response, it was increased by the same amount. In the beginning of the measurement, the level was lowered after each correct response until the subject first failed to specify the correct interval. The threshold value was finally calculated by averaging the signal level of the 15 trials that followed the second lower turning point of the signal level. Each data point in the figures is based on at least four such measurements. Five subjects aged 23 to 30 years participated in the experiments.

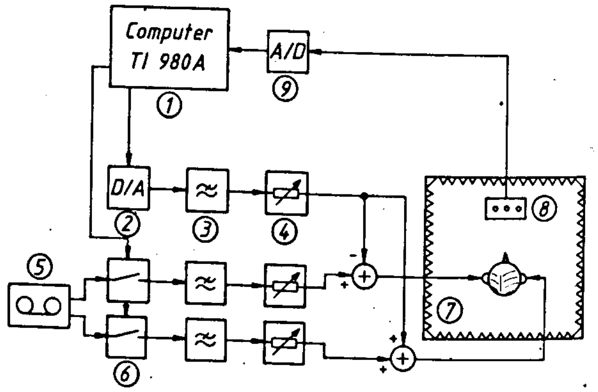


Fig.1: Experimental setup: (1) 16 bit microcomputer TI 980A; (2) 12 bit D/A converter; (3) low pass filter Krohn + Hite, 48 dB/octave; (4) manual attenuators; (5) two channel white noise, stored on magnetic tape; (6) computer controlled gates; (7) sound insulated booth; (8) response box; (9) 12 bit A/D converter.

Frequency and phase uncertainty

To introduce uncertainty about the test signal frequency and the test signal phase, the threshold procedure was modified in the following way: Within one measurement, the thresholds for two different test signals were determined using the adaptive 3 IFC procedure. For each trial, one of the two signals was randomly chosen with probability 1/2. The two signals differed in frequency and interaural phase difference. The frequencies of each pair of signals were chosen symmetrically around 500 Hz, e.g. 450 and 550 Hz or 200 and 800 Hz. The phase difference was always opposite to the noise phase difference at that frequency, e.g. π for the lower test frequency and 0 for the higher test frequency. Thus, the subject had no prior knowledge about the frequency and interaural phase of the test signal in the next trial. The level adjustment for either test frequency followed the algorithm described above and the measurement was completed if for both signals the number of 15 trials was reached.

EXPERIMENTS

Influence of test signal duration on the BMLD

In the first experiment we investigate the BMLD pattern for the masker with frequency varying interaural phase difference. To define the interaural conditions of our experiments, we use the notation common in binaural psychoacoustics: N and S describe noise masker and (test) signal respectively, the interaural phase differences are given by indices (0 indicates in-phase, π antiphase presentation). In addition, we introduce the notation $N_{0\pi}$ for the masker with phase difference 0 below 500 Hz and phase difference π above 500 Hz. By inverting one channel of this masker, the components below 500 Hz are in antiphase and the components above 500 Hz in phase ($N_{\pi 0}$).

In Fig.2, we demonstrate the effect of the interaural phase step of the masker for a 250 ms S_π test signal. Open and closed symbols respresent the BMLD values for $N_{0\pi}$ and $N_{\pi 0}$ masker respectively. The continuous line gives the values for a masker with fixed phase difference of 0 at all frequencies (N_0). The step of the interaural masker phase strongly influences the binaural thresholds between 400 and 650 Hz. A detailed analysis of this BMLD pattern leads to the conclusion that the masker cross-correlation averaged over the critical band at the test frequency is the crucial factor in this experiment /1/. As this correlation variies between +1 and -1 for test frequencies close to 500 Hz, the BMLD of the test signal also variies by about 15 dB. At test frequencies well apart from 500 Hz, no

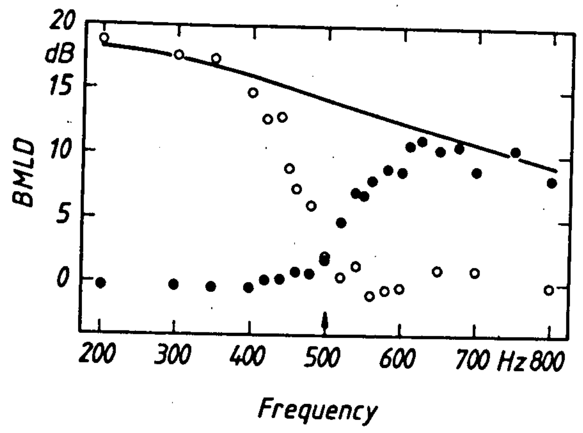


Fig.2: BMLD of a 250 ms test signal in the configurations $N_{0\pi}S_\pi$ (o) and $N_{\pi 0}S_\pi$ (●). The continuous line shows the N_0S_π BMLD. The arrow marks the transition of the interaural phase difference of the masker at 500 Hz. One subject.

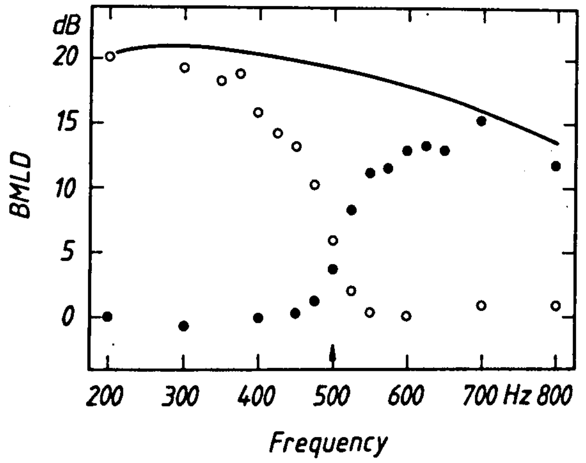


Fig.3: Same as Fig.2 for a 20 ms test signal.

influence of the phase transition is observed and the $N_{0\pi}$ and $N_{\pi 0}$ maskers have the same effect as the N_0 masker.

In order to test the influence of test signal duration in this detection task, we repeated the same experiment with a 20 ms tone (Fig.3). The slope of the data in this figure is the same as for the 250 ms test signal in Fig.2. The slight increase of the BMLD for the shorter test signal confirms the observations in other binaural configurations /6,7/. The broadening of the transition range may be due to the widening of the test signal spectrum.

Generally, the presence of two spectral masker ranges with different interaural phase relations which would require different processing strategies in the view of the EC theory, does not hamper the binaural system. Even for a short test signal of 20 ms, the maximal amount of binaural unmasking is reached at test frequencies well apart from 500 Hz.

Influence of frequency uncertainty on diotic masked thresholds

In the previous experiment, the test signal always was presented at a constant frequency within one measurement block. To determine the importance of this a priori information, we introduced uncertainty concerning the test frequency by the algorithm described above. Both test signals and the masker were presented in phase (N_0S_0 condition).

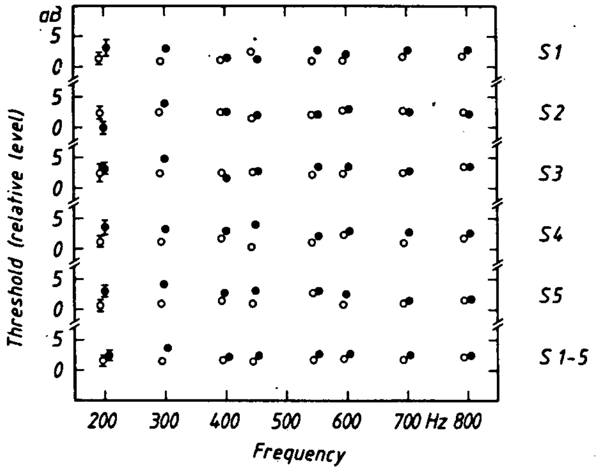


Fig.4: Binaural masked thresholds of a 20 ms test signal in the diotic configuration N_0S_0. o fixed test signal frequency, • uncertain (one out of two) test signal frequency. Values for 5 listeners and their means.

In Fig.4, the results for uncertain frequency presentation (•) are compared to threshold values for fixed frequeny (o). The uncertainty has only a slight influence on the masked thresholds, the averaged difference between the two experimental conditions amounts to 1 dB.

In this experiment, the subjects are obviously able to concentrate on different frequency regions simultaneously without strong reduction in sensitivity. If the number of alternative frequencies is further increased, a monotonous rise of the thresholds is observed /8/.

Frequency and phase uncertainty in a dichotic detection task

In the following experiment, we apply the uncertain frequency algorithm to a dichotic condition. In this case, the frequency uncertainty is accompanied by an uncertainty about the optimal binaural processing strategy. The masking noise is in phase at frequencies below 500 Hz and in antiphase at frequencies above 500 Hz ($N_{0\pi}$). The low-frequency test signal was interaurally inverted (S_π), the high-frequency signal was in phase (S_0). Thus, the two test frequencies correspond to the two different binaural conditions N_0S_π (low-frequency stimulus) and $N_\pi S_0$ (high-frequency stimulus). For comparison, we determined the binaural thresholds for fixed test frequency in the same binaural conditions.

Fig.5 shows the results of three listeners for fixed (o,□) and randomly chosen (•,■) test signals. For all subjects, there is no significant difference in the threshold values of the two measurements. Additional experiments in other binaural conditions confirmed this result: The binaural unmasking process is not influenced by uncertainty about the interaural phase of masker and test signal.

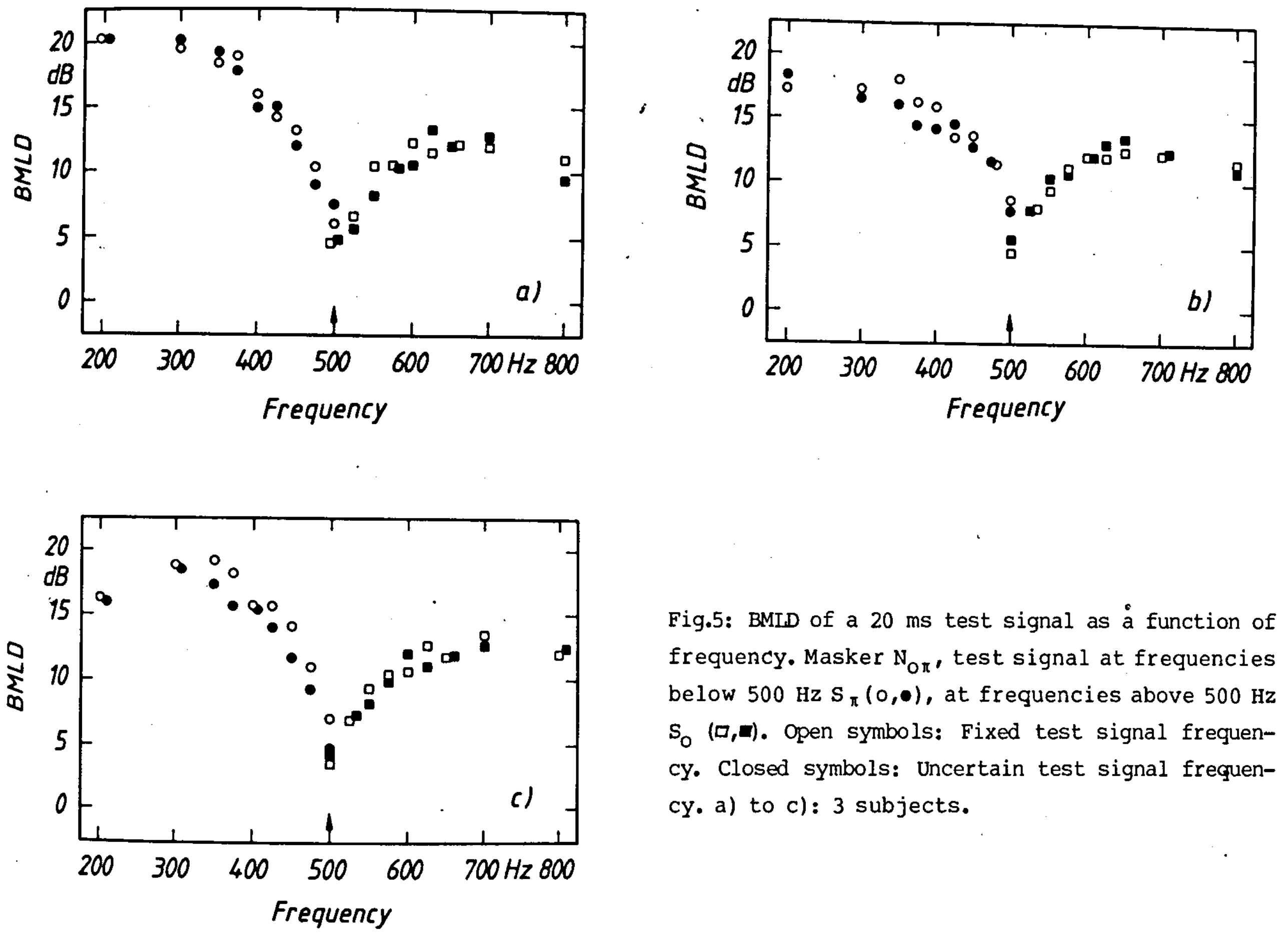


Fig.5: BMLD of a 20 ms test signal as a function of frequency. Masker $N_{0\pi}$, test signal at frequencies below 500 Hz S_π (o,•), at frequencies above 500 Hz S_0 (□,■). Open symbols: Fixed test signal frequency. Closed symbols: Uncertain test signal frequency. a) to c): 3 subjects.

DISCUSSION

The results presented in Fig.5 emphasize the following conclusions: The detectability of short test signals presented randomly in different critical bands with different interaural phase relations is as good as for fixed signal parameters. As we assume that an adjustment of the optimal processing strategy is not possible within the short test signal duration, one could argue that the ear uses a priori information about the different binaural conditions. This a priori knowledge could be aquired in the beginning of the measurement, as the two test signals are presented clearly audible. It could be stored as different processing strategies for the two frequency regions of interest.

However, this way of reconciling our experimental results with the ideas of the EC theory does not hold if the two signals are presented within the same critical band. For the test signal pair 500 Hz S_π/ 500 Hz S_0, the optimal strategy has to be subtraction of the two channels (for the S_π signal) and addition for the S_0 signal. Thus, different strategies are necessary according to the test signal phase. As the binaural system reaches a significant BMLD in this condition, it must be able to apply different transformations instantaneously within the same critical band.

This outcome of the experiments is much more compatible with cross-correlation models of binaural interaction /3,4,9-11/. In these models each binaural stimulus leads to a specific two-dimensional excitation pattern (cf. introduction). Test signals with different interaural phase relations excite different places along the τ-axis. Uncertainty about test signal frequency and phase results in uncertainty about the exact place of excitation within this cross-correlation pattern. Our results emphasize that uncertainty about interaural phase (one dimension within the cross-correlation pattern) has the same (negligible) effect as uncertainty about test signal frequency (the other dimension, cf. Fig.4). In the same way as in monaural hearing the listener can concentrate on different frequencies simultaneously, he seems to be able to concentrate on different interaural delays in binaural hearing. Therefore, the a priori information about the interaural parameters does not further improve the detection of binaurally presented signals.

ACKNOWLEDGEMENTS

I would like to thank M.R. Schroeder, U. Sieben and B. Kollmeier for many helpful discussions during this study and my colleagues for the participation in the listening tests. Technical support by L. Liebe and G. Kirschmann-Schröder is gratefully acknowledged. This project was supported by the Deutsche Forschungsgemeinschaft (Schr 38/20).

REFERENCES

/1/ A. Kohlrausch, "Auditory filter shape derived from binaural masking experiments", submitted for publication.

/2/ N.I. Durlach, "Binaural signal detection: Equalization and cancellation theory", in: Foundations of Modern Auditory Theory, Vol.II., J.V. Tobias (ed.), Academic Press, New York, 1972.

/3/ H.S. Colburn, "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination." J. Acoust. Soc. Am. 54(1973), 1458-1470.

/4/ H.S. Colburn, "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise." J. Acoust. Soc. Am. 61(1977), 525-533.

/5/ H. Levitt, "Transformed up-down methods in psychoacoustics." J. Acoust. Soc. Am. 49(1971), 467-477.

/6/ D.M. Green, "Interaural phase effects in the masking of signals of different durations." J. Acoust. Soc. Am. 39(1966), 720-724.

/7/ A. Kohlrausch, "The influence of signal duration, signal frequency and masker duration on binaural masking level differences." Hearing Research 23(1986), 267-273.

/8/ D.M. Green, "Detection of auditory sinusoids of uncertain frequency." J. Acoust. Soc. Am. 33(1961), 897-903.

/9/ L.A. Jeffress, "A place theory of sound localization." J. Comp. Physiol. Psychol. 61(1948), 468-486.

/10/ J. Raatgever, F.A. Bilsen, "A central spectrum theory of binaural processing. Evidence from dichotic pitch." J. Acoust. Soc. Am. 80(1986), 429-441.

/11/ W. Lindemann, " Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals." J. Acoust. Soc. Am. 80(1986), 1608-1622.

THE INTERFACE BETWEEN ACOUSTIC-PHONETIC AND LEXICAL PROCESSES

William D. Marslen-Wilson Uli H.Frauenfelder

Max-Planck Institute for Psycholinguistics,
Nijmegen, The Netherlands

Medical Research Council Applied Psychology Unit,
Cambridge, England

Abstract

Speech research and psycholingustic research into spoken language comprehension have a common interest in the processes of acoustic-phonetic analysis. This paper argues that this common goal should be reflected in a common research programme, integrating together the questions and the techniques of the two disciplines. Without such an integration, neither discipline can expect to achieve adequate answers to its characteristic questions.

Introduction

A fundamental goal of the phonetic sciences is to characterise the ways in which the acoustic signal is mapped onto an acoustic-phonetic level of mental representation. This objective is subsumed within one of the major goals of experimental psycholinguistics -- namely, to characterise the mapping from the speech signal onto a level of meaning representation. But despite this intimate inclusion relation, there has been surprisingly little direct contact between the two disciplines. Research in phonetics has paid only limited attention to the wider context within which the processes of acoustic-phonetic analysis presumably operate. By the same token, psycholinguistic research into spoken language comprehension has tended to neglect the complexities of the acoustic-phonetic input and its analysis.

This bidirectional indifference is doubly surprising when we consider just how strong the interdependence must be between the two disciplines at the point where their interests directly converge: that is, at the interface between acoustic-phonetic and lexical processes. From the phonetic perspective, the emphasis, naturally enough, is on the computation of acoustic-phonetic representations from the speech input. From the psycholinguistic perspective, the extraction of meaning depends upon access to the mental lexicon, and this in turn depends upon the ability of the system to map the speech input onto mental representations of lexical form via an acoustic- phonetic representation.

In other words, both disciplines are closely concerned with one and the same representation -- what we label here as the **input representation** -- mediating between the speech signal and the mental representations of lexical form. The goal of this paper is to demonstrate that it is both necessary and possible to investigate the properties of this representation from both perspectives. We want to show, on the one hand, how the input representation, and the processes mapping it onto the lexicon, are constrained by the signal and its acoustic-phonetic analysis, and, on the other, how the input representation and its construction are influenced and constrained by the target lexical representations, and by the properties of the language in general.

Research in our laboratories over the past three years has been guided by this dual perspective, aiming at the development of a unified picture of the early stages of the speech understanding process. The following sections give an overview of some of this research. In the first part, we will focus on some ways in which the properties of the speech signal constrain the input representation and lexical access, and on the consequences of this for the theoretical assumptions that have been used to justify the separation of acoustic-phonetic issues from the lexical level. In the second, we will discuss research into the processing structure of the interface, focussing on the directionality of information flow within the system. In the concluding section of the paper we will turn to some research into the role of the listener's system of phonological knowledge in mediating the relationship between the signal and the lexicon, and the consequences of this for the input representation.

The speech signal and lexical access

The conventional division between psycholinguistic research into spoken word recognition and phonetic research into speech analysis is based on the assumption that lexical access is largely insulated from the detailed properties of the speech signal and the way it carries information over time. This assumption in turn depends on a number of further assumptions about the properties of the speech processing system. The most important of these -- as we argued here four years ago (9) -- seem to be the following.

First, one must assume that there are two distinct levels of perceptual representation computed during speech analysis. These correspond, respectively, to an acoustic-phonetic level of analysis and to a lexical level. Secondly, one must assume that the properties of the acoustic-phonetic level, and of the processes that map from the speech signal onto this level, can be determined solely with reference to phenomena internal to this level, and without reference to the role of these processes in providing the basis for a further mapping onto the mental lexicon. Thirdly -- and most crucial for the psycholinguistic neglect of the speech signal -- there is the assumption that the representation generated at the acoustic-phonetic level (the input representation) is highly abstracted from the detailed properties of the input to the acoustic-phonetic processor. In fact, psycholinguistic research into lexical access has standardly been conducted on the assumption that the input to the lexicon is a string of phonemic labels, and, indeed, that this is also an adequate characterisation of the properties of lexical form representations.

In this part of the paper we will argue that this cluster of assumptions is false. The detailed properties of the speech signal, and of the way it carries discriminating information over time, are tracked faithfully and continuously at the lexical level. The psycholinguistic problems of lexical access and selection cannot be isolated from the problems of acoustic-phonetic analysis.

The salient feature of the speech signal, considered as an information channel, is that it is based on a continuous sequence of articulatory gestures, which result in a continuous modulation of the signal. Cues to any individual phonetic segment are distributed across time, and, in particular, they overlap with cues to adjacent segments. This means that the speech signal is rich in what we can call **partial information** -- that is, anticipatory cues to the identity of an upcoming segment. As the listener hears one segment, he will also hear partial cues to the identity of the next.

An example of this is the presence of cues to the place of articulation of a word-final plosive in the formant structure of the preceding vowel. Thus, in the word **scoop**, the lips may move towards closure for /p/ during the vowel, while in **scoot** the tip and body of the tongue are brought forward to form closure for the /t/. Both movements, conditioned by the place feature of the consonants, produce differences in the formant frequency patterns towards the end of the vowel.

The question we have asked in recent research is whether this type of partial information is made available at the lexical level. How far is the on-line process of lexical access and selection sensitive to the continuous nature of information transmission in the speech signal, and to the availability of partial information as it accumulates over time? To the extent that such sensitivity can be demonstrated, then the separatist assumptions we listed above seem to fail. If word candidates can be accessed and identified on the basis of partial information about the identity of a sound segment, then this causes fundamental problems for the claim that the speech input is mapped onto representations of word-forms in the mental lexicon in terms of complete phonemes (or units of a similar or larger size)

We have investigated this question in a number of studies,carried out in English, Bengali, and Dutch (4, 7, 11, 12), which have used the speech gating task to trace the temporal microstructure of acoustic-phonetic uptake during spoken word-recognition. We focus here on the English studies, looking at the uptake at the lexical level of partial cues to word-final place and voice in CVC monosyllables (11,12).

These were experiments in which listeners heard gated fragments of CVC's, drawn from pairs contrasting in place (like **scoop/scoot**) or in voice (like **log/lock**). The words were presented in increments of 25 msec, focussing on the 125 milliseconds leading up to the closure of the vowel. Gate 0 in Figure 1 represents the gate at which the vowel terminated. The subjects were required at each increment to say what they thought the word was, or was going to become.

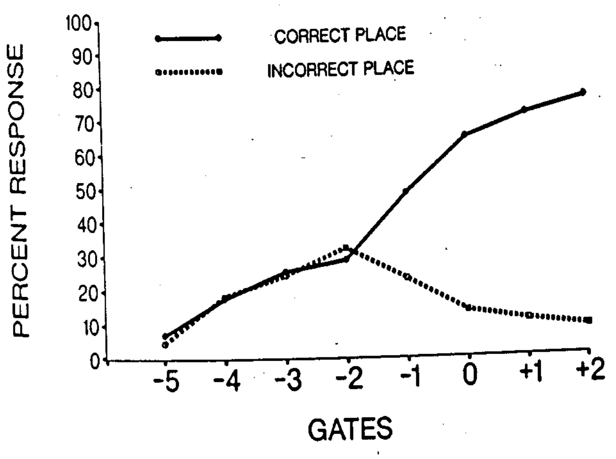


Figure 1: Lexical responses to pairs of CVC's contrasting in place. The correct responses are cases where the subjects respond with the member of the pair with the correct place (e.g., scoop); incorrect responses are cases where they respond with other member of the pair (e.g., scoot).

For the place contrasts, which here involved CVC's ending in voiceless plosives and matched for frequency, partial information as to place of articulation is conveyed by the changing spectral properties of the vowel as it approaches closure. The question at issue was whether this would affect lexical access and selection, as reflected in the subjects' responses at each gate. If so, then their responses should start to diverge before vowel closure (i.e., before Gate 0), and certainly before they hear the plosive release, falling some 80-100 msec after closure. The results, summarised in Figure 1, clearly show this early divergence, with a strong preference at Gate 0 for the word with the correct place of articulation.

For the voicing contrasts (involving pairs like **rip/rib** and **dog/dock**) we were asking similar questions, but looking now at a durational cue -- vowel length is a powerful cue to voicing in English. In the gating task, listeners hear the vowel slowly increasing in length over successive gates. Our question was whether they could exploit this information as it became available.

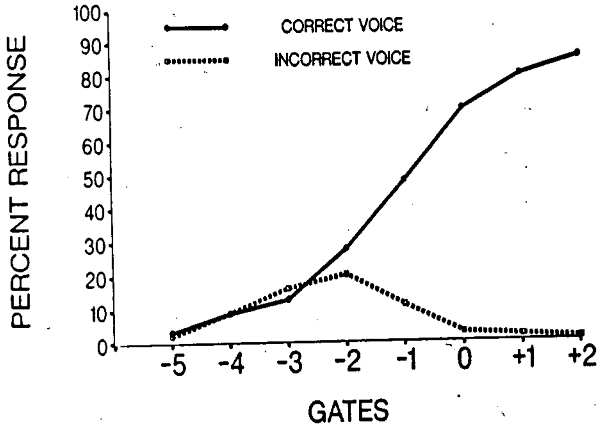


Figure 2: Lexical responses to pairs of CVC's contrasting in voice. The correct responses are cases where the subjects respond with the member of the pair with the correct voice (e.g., dock); incorrect responses are cases where they respond with other member of the pair (e.g., dog).

Turning to Figure 2, where Gate 0 again represents vowel closure, we see that after an initial period in which voiceless responses predominate, listeners start to successfully discriminate voiced from voiceless words as soon as the length of the vowel starts to exceed the durational criterion (around

135 msec from vowel onset for this particular stimulus set).

What we find, then, in Figures 1 and 2, is clear evidence for the immediate uptake of accumulating acoustic information. There do not appear to be any discontinuities in the projection of the speech input onto the lexical level. The speech signal is continuously modulated as the utterance is produced, and this continuous modulation is faithfully tracked by the processes responsible for lexical access and selection. As the spectrum of a vowel starts to shifts towards the place of articulation of a subsequent consonant, this is reflected in a shift in listeners' lexical choices, which becomes apparent about 25-50 msec before closure. As the duration of a vowel increases, the listener produces lexical choices that reflect these changes in duration, shifting from voiceless to voiced as the durational criterion is reached and surpassed, at about 50-75 msec before closure. There is immediate use of partial durational cues, just as there is immediate use of partial spectral cues.

Lexical processing is clearly not insulated from the detailed properties of the speech signal. Psycholinguists interested in the temporal structure of the speech understanding process cannot ignore the variations in information flow that stem from the continuous modulation of the incoming speech signal. By the same token, research into the properties of acoustic-phonetic processing will have to acknowledge the direct relevance of its subject-matter for processes at the lexical level.

The processing structure of the interface

We turn now a different perspective on the properties of the interface between acoustic-phonetic and lexical processes. This involves the structure of the interface viewed as an information-processing system. If we are talking about different levels of analysis during speech processing, and discussing the flow of information between these levels, then we need to specify the directionality of this flow, and to determine the contraints on how information at one level can affect processes at another level. Is information-flow strictly "bottom-up", in the sense that information flows in one direction only, from the speech signal, via an acoustic-phonetic processor, up to the lexical level. Or does the system allow for "top-down" effects as well, in the sense that information at a higher level can feed back to lower levels and directly affect the outcome of these lower level processes.

We have seen in the preceding section how the time-course of lexical access and selection is determined by the properties of the signal and its on-line acoustic-phonetic analysis. Information originating in the sensory input flows continuously in a bottom-up fashion to drive lexical processing. The further results of these gating studies (11, 12) looking at the effects of at least one lexical variable (the frequency of occurrence of the word being heard), suggest that the bottom-up (sensory) input has the priority in determining the outcome of lexical access and selection. Although word frequency did have an effect, with subjects initially tending to respond with more frequent words (for pairs of words where we explicitly contrasted frequency), the scope of these effects was severely limited. In particular, frequency only affected lexical processing under conditions where the available sensory information was sufficiently ambiguous or indeterminate to allow a choice between one or more alternatives. But these effects dissipate immediately as soon as more determinate bottom-up information became available.

This processing asymmetry between the importance of bottom-up and top-down processes is in apparent conflict with a strong current trend to assign an important role to top-down information-flow during speech processing. On this type of account, decisions about the content of the sensory input (i.e. the identity of segments) are affected by expectations coming from higher levels of processing. In effect, the perceptual output of the mechanisms of speech perception are assumed to vary as a function of the lexical context in which the speech input occurs.

Evidence for top-down information flow comes from a variety of sources, including studies of phonetic categorization and phoneme restoration.

Phonetic categorization (for example, the identification of stimuli falling at different points along some acoustic-phonetic continuum) has been claimed to be influenced by the lexical status of the item bearing the phonetic segment. In one such study (6) the voice onset time (VOT) of stimulus initial stop consonants was manipulated to construct a voicing continuum. These stimuli were chosen such that the lexical status of each stimulus changed from a word to a non-word, as in dash/tash, as a function of the voiced or unvoiced character of this initial phoneme. Subjects were asked to make a forced phonetic choice (i.e. between /d/ and /t/). It was found that the lexical status of the item led to a shift in the location of the phoneme boundary along the VOT continuum, in the direction of word rather than non-word responses. This result was interpreted as showing that the perception of the identical speech sound can differ depending on its status at the lexical level -- whether it forms a word or a non-word.

The phenomenon of phoneme restoration has also been taken as evidence for a contribution of the lexical level to phonetic processing. Listeners typically report that an utterance sounds intact even when a part of it has been replaced by an extraneous noise. According to Warren (13), this ability to restore the missing speech sound shows that the perception of speech is mediated by higher levels, via a top-down processing link.

A major weakness in this type of evidence for top-down perceptual processes is that it ignores the temporal properties of the proposed top-down information flow. In fact, it is critical to determine when top-down information first becomes available to influence processing, and whether this influence operates quickly enough to affect the continuous and immediate bottom-up analysis. The importance of temporal variables in controlling information flow in lexical processing can be seen in a more recent study (2) of phonetic categorization, where time constraints were introduced. When subjects were required to make speeded phonetic decisions, a lexical effect was found only for slow responses, but disappeared for fast ones. This suggests that a certain amount of time, or more precisely - a substantial amount of acoustic information - needs to accumulate before the lexicon can exert its influence on the bottom-up analysis. The critical question arises whether this top-down influence comes into play before the bottom-up analysis has been completed.

We have conducted a number of experiments investigating the temporal properties of these proposed top-down lexical effects, in order to determine whether lexical constraints do in fact influence on-line processes at lower levels of analysis. We will present here a sample of this research (for more details, see Frauenfelder, Segui, and Dijkstra, this volume).

To trace the time-course of lexical effects, we selected words containing phoneme targets to be detected in different positions in the words. These targets occupied four different positions with respect to the words' uniqueness points (word onset, before uniqueness point, after uniqueness point, word offset) -- for an example set, see Table 1.

Table 1

Examples of Monitoring Stimuli

	WORD	NONWORD
ITEM ONSET	Pagina	Pafima
BEFORE UP	jaPanner	joPammel
AFTER UP	olymPiade	arimPiako
ITEM OFFSET	bioscooP	deoftooP

The Uniqueness Point (UP) is defined as the point at which a spoken word becomes uniquely identifiable, going from word-onset. Nonwords were created by changing one or more segments in the original word, while keeping the target's local phonetic environment as constant as possible. The dependent variable that we measured was the subject's latency to detect a previously specified phoneme target. The difference in these detection latencies to phoneme targets in the same position in matched words and nonwords was taken to provide a measure of the lexical contribution to the phoneme detection process. Figure 3 shows these differences between words and nonwords as a function of target position.

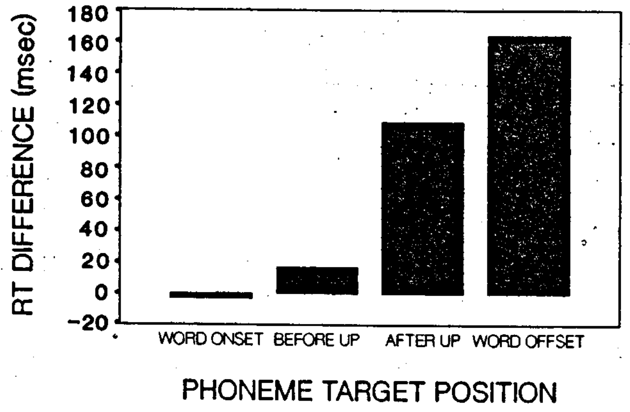


Figure 3: Mean differences between targets in words and non-words as a function of target position.

There are significant differences between words and nonwords, but only after the UP. This suggests that the lexicon exerts its effect on phoneme detection responses only at a point in processing where just a single candidate is still compatible with the sensory input. Listeners have already accessed full lexical information and have available the entire phonological description of the word (10). This severely limits the potential role of top-down lexical effects in speech processing. If there is a top-down influence on lower level processes -- and there has been some dispute as to whether phoneme detection tasks tap a sub-phonemic lexical level at all (e.g., 1)-- then these effects seem to come into play only after the bulk of the bottom-up processing has been completed. This defeats the potential processing function of top-down effects in theories like TRACE (8), where top-down information-flow to the phoneme level has the effect of tuning the responses of phoneme nodes as a function of their lexical environment.

Phonological aspects of the interface

We have spoken so far about the relationship between the acoustic-phonetic analysis of the speech signal, and the processes of form-based lexical access and selection. We now turn briefly to the potential role of phonological factors in determining the character of this interface.

Research in acoustic-phonetics and in lexical access typically assumes a fully transparent relationship between the signal and the lexicon - that the speech signal makes information available and that this information is straightforwardly mapped onto representations of lexical form. In fact, the listener's system of phonological knowledge may mediate this relationship (3,5), with consequences for the interpretation of the signal which are not deducible from the properties of the signal alone.

We see this, for example, in recent research that reveals phonologically based assymetries in the lexical interpretation of cues in the speech signal. These are studies looking at the perceptual consequences of vowel nasalisation (4,7,11), contrasting languages like Bengali, where nasal is distinctive for vowels, with languages like English, where it is not.

For English listeners, the presence of nasalisation in a vowel is an unambiguous signal that they are hearing an oral consonant followed by a nasal vowel. But for Bengali listeners, where phonetically equivalent vowel nasalisation holds both for nasal vowels preceding oral consonants and for oral vowels preceding nasal consonants, the presence of nasalisation is ambiguous. What one sees, however, in a gating task carried out with Bengali listeners (7), is a very strong bias to interpret nasality as signalling the underlying marked value [+nasal] for the language. They interpret nasalisation as signalling the presence of a nasal vowel followed by an oral consonant -- the exact opposite of the interpretation of the same acoustic feature in a language like English. This choice of the Bengali listeners is only explicable if one takes into account the structure of their phonological systems. It is not explicable just in terms either of the signal, or of the representations of lexical form, taken on their own.

A different kind of asymmetry in the interpretation of nasalisation (7,11,12), is a difference in the signal value of the presence as opposed to the absence of nasalisation. When a vowel is nasalised in English, this has a strong effect on lexical choice, ruling out word-candidates where oral vowels are followed by oral consonants. The absence of nasalisation, in contrast, seems to have weaker effects, and does not prevent listeners from selecting CVC's ending with nasal consonants.

This asymmetry may reflect the status of the nasal feature for vowels in English. Because English has no nasal vowels, it is likely that the abstract specification of English vowels does not include the feature [nasal]. This means that when an unnasalised vowel is being heard, there is nothing in the abstract representation of lexical items ending in nasals that could exclude these as possible responses. If a vowel has no nasality feature, then the absence of nasalisation cannot be a discriminant property of the input. In contrast, when the vowel is nasalised, this is a positive cue to the status of the following consonant, and is treated as such by the listener.

These are only preliminary investigations of some asymmetries in the interpretation of acoustic cues at the lexical level. But if we are correct in suggesting that the ' formal properties of phonological representations can help determine the lexical interpretation of the speech input, then this has important implications for how we should investigate the properties of the acoustic-phonetic processing space within which the listener accesses the mental lexicon. We are arguing, then, not just for an integration of the questions and the techniques of speech research and of psycholinguistics, in studying the interface between acoustic-phonetic and lexical processing, but also for the full engagement in this enterprise of the associated linguistic disciplines.

REFERENCES

1) Cutler, A., Mehler, J., Norris, D. & Segui, J. Phoneme identification and the lexicon. *Cognitive Psychology*, 1987, (in press)

2) Fox, R.A. Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 1984, 10, 526-540.

3) Frauenfelder U. H. & Lahiri, A. Understanding words and word recognition: Can phonology help? Paper presented at the MPI Conference on lexical processing and representation, June, 1986.

4) Frauenfelder, U.H., Wessels, S., & Marcus, S.M. Manuscript in preparation, Max-Planck Institute for Psycholinguistics, Nijmegen.

5) Frazier, L. Structure in auditory word-recognition. *Cognition*, in press, 1987.

6) Ganong, W.F. III. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, 6, 110-125.

7) Lahiri, A., & Marslen-Wilson, W.D. Manuscript in preparation, Max-Planck Institute for Psycholinguistics, Nijmegen.

8) McClelland, J.L. & Elman, J.L. The TRACE model of speech perception. *Cognitive Psychology*, 1986, 18, 1-86.

9) Marslen-Wilson, W.D. Perceiving speech and perceiving words. In M.P.R. v. d. Broecke & A. Cohen (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences*. Dordrecht: Foris, 1984.

10) Marslen-Wilson, W.D. Function and process in spoken word recognition. In H. Bouma & D.G. Bouwhuis (Eds.) , *Attention and Performance X: Control of Language Processes*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1984.

11) Warren, P., & Marslen-Wilson, W.D. Continuous uptake of acoustic cues in spoken word-recognition. *Perception and Psychophysics*, in press, 1987 (a)

12) Warren, P., & Marslen-Wilson, W.D. Cues to lexical choice: Discriminating place and voice. Manuscript, Department of Experimental Psychology, University of Cambridge, 1987 (b)

13) Warren, R.M. Perceptual restoration of missing speech sounds. *Science*, 1970, 167, 392-393.

ON THE PREDICTION OF PHONEME RECOGNITION BY THE HEARING IMPAIRED

REVOILE, S.G., BUNNELL, H. T., and PICKETT, J.M.

Center for Auditory and Speech Sciences
Gallaudet Research Institute
Gallaudet University

ABSTRACT

A review of recent research on psychoacoustic correlates of subnormal speech recognition by hearing-impaired listeners reveals a confusing diversity of results. This may be due partly to the use of recognition measures that derive only an overall score of percent phonemes correct or a signal level for a criterion percent correct phonemes. A set of studies is described in which this problem is approached by scoring the correctness of perception of specific consonant features, such as voicing or manner, for correlation with related psychoacoustic capacities, such as discrimination of vowel duration, discrimination of the rate of formant transition, and modulation detection. The availability of cues to consonant features is controlled by processing to degrade selected cues in the test syllables and/or by selection of listeners with limited cue reception. Fairly useful correlations were found from the point of view of predicting feature perception differences among individuals.

INTRODUCTION

The past decade has witnessed increasing interest in the relation between impaired speech recognition and basic psychoacoustic performance, such as auditory resolution of intensity, frequency, and time. Knowledge of such relations could contribute to our understanding of speech processing at the cochlear level; when substantial speech/psychoacoustic relations are found, the similarities between the psychoacoustic assessments and analogous physiological measures of audition enable inferences to be drawn about the auditory analysis of speech. A more practical intent of these studies is to identify measures that might predict speech recognition performance more effectively than traditional and current audiological techniques. However, whether for predictive purposes or to gain more basic knowledge, relations between speech/psychoacoustic performances have yet to be firmly established due in part to the diversity of findings throughout investigations thus far. The variability in results is apparent among some of the more recent studies.

Haggard et al. (1986)[1] found that the predictiveness of psychoacoustic tuning curves (PTC) and pure tone sensitivity for speech performance in noise was affected by the frequency response used for speech presentation. For a flat response, tone thresholds and abbreviated PTC measurements (Lutman and Wood, 1985[2]) at 2 kHz were each similarly effective in predicting speech performance. For a rising response, adequate predictions for speech performance required both tone thresholds and PTCs. The speech measure was a speech identification test from which an overall score was obtained for correct consonant recognition.

Of the audiometric and psychoacoustic variables measured by Lutman and Clark (1986)[3], threshold sensitivity at 2 kHz was the best predictor of, although only moderately related to, speech performance in noise of 23 hearing-impaired subjects ranging in age from 44 to 72. Subject age also showed some predictive value for speech performance. While frequency resolution and gap detection at 2 kHz were moderately related to speech performance via simple correlations, they lacked uniqueness as predictors of speech performance via multiple regression. The index of speech performance was the signal-to-noise ratio, determined adaptively, that yielded about 70% correct word identification.

Using a similar speech performance index Stelmachowitz et al. (1985)[4] measured recognition of monosyllabic words in noise and correlated the S/N ratio for 75% correct with impaired subjects' frequency selectivity, as indicated from psychoacoustic tuning curves. (These curves plot the levels, as a function of frequency, of narrow-band remote maskers that just mask a probe tone, of 2000 Hz in this study). Frequency selectivity values derived from the curves were found to predict the speech recognition level with correlations on the order of .65 in the broad-band noise and .70 in the low-pass noise; combining various selectivity parameters accounted for 68% and 54% of the variance, respectively.

Lamore et al. (1985)[5] tested 32 severely/profoundly hearing-impaired adolescents for audiometric and psychoacoustic performances, which were analyzed relative to the speech reception threshold (SRT) and maximum word discrimination scores in quiet. (SRT is the lowest speech level or speech-to-noise ratio at which the listener judges meaningful speech to be at least 50% correct.) Between the two speech measures, the SRT showed somewhat better relations to the auditory measures than did the word discrimination scores. Of course, the SRT showed the highest relations to pure tone threshold sensitivity (.83 to .92). At least moderate relations were seen between the SRT vs DL for frequency and for critical ratios measured with low and mid frequency tones (correlations were .57 to .76). These same auditory measures yielded the highest correlations to word discrimination scores in quiet. Measures of amplitude modulation for white noise and temporal integration generally manifested poor relations to the speech performances.

Dreschler and Plomp (1985)[6] related various audiometric and psychoacoustic variables to speech perception in quiet and in noise for 21 hearing-impaired subjects from 13 to 20 years of age. The measure of speech perception was the SRT for sentences (Plomp and Mimpen, 1979)[7]. Speech perception in quiet was best predicted by the amount of loss, as represented generally by the mean audiometric tone thresholds, and specifically by the 500-Hz threshold. In contrast, audiometric variables had little power for predicting speech perception in noise, which was best predicted by measures of gap detection and the critical ratio--believed to reflect indirectly the frequency resolution capabilities of the ear. These psychoacoustic measures accounted for about 70% of the variance among the subjects' hearing for speech in noise.

Preminger and Wiley (1985)[8] tested consonant recognition in quiet and PTCs for 3 pairs of two subjects each matched for audiograms showing either flat, low-, or high-frequency losses. Within either subject pair of low- or high-frequency losses, the subject with poorer consonant recognition manifested a more abnormal PTC in the frequency region where the loss was greatest. Hence, pure tone threshold sensitivity appeared to show less association with consonant recognition than did frequency resolution.

Dorman et al. (1985)[9] related identification for synthetic, burstless /ga/ in a /ba, da, ga/ continuum to 2 kHz tone thresholds and frequency selectivity of aged listeners who generally showed reduced perception for /ga/. The optimal tokens of /ga/ contained F2 starting frequencies near 2 kHz. While a moderate relation to /ga/ identification was found for 2 kHz tone thresholds, a lower relation occurred for the measure of 2 kHz frequency selectivity versus /ga/ identification.

With few exceptions (e.g., Dorman, 1985[9]; Preminger and Wiley, 1985[8]), a common element among these studies of psychoacoustic versus speech performance is the representation of speech recognition by a total performance score, which is then analyzed for its relation to psychoacoustic performance. Such global scores are gross measures of speech recognition and do not reveal the particular acoustic patterns in speech that may be imperceptible for a hearing-impaired listener. Among the temporal and spectral cues available for use in recognition of phonemes, the hearing-impaired listener may rely exclusively on one type of cue, because other cues are imperceptible. For example, in distinctions of final consonant voicing, some of our listeners seemed to depend primarily on the vowel duration cue since their reception was reduced for consonant constriction cues or spectral cues in the vowel offset (Revoile et al., 1982[10], 1985[11]). If a listener's speech perception is limited to use of only certain cues, such information would seem relevant to correlational analyses of speech and psychoacoustic performances.

Underlying the effort to relate psychoacoustic performance and speech recognition is the assumption that psychoacoustic discrimination for a particular type of stimulus is associated with a similar mode of auditory processing for speech recognition. Yet few attempts have been made to examine given auditory abilities in relation to the use of particular types of acoustic patterns involved in speech recognition. The speech/psychoacoustic relations may be more easily determined if speech recognition is measured according to each of several classes of cues that contribute to phoneme distinctions, for example, temporal cues, dynamic spectral cues, or static spectral cues. Relative to these classes, then psychoacoustic performances for analogous stimuli could be examined (as well as for the more traditional types of psychoacoustic stimuli used for this purpose). The psychoacoustic performances found to relate most highly with the speech measures may be those that used acoustic stimuli resembling the critical speech-cue patterns.

A study modeled after this approach would test consonant recognition according to articulatory features using stimuli altered to contain only one type of cue per feature category. Thus, recognition of a test consonant would depend only on the perceptibility for the listener of the single available cue per stimulus. The stimuli used to test psychoacoustic performance would be selected according to the dominant acoustical domain of the critical cue. For example, in a study of VOT use for voicing perception of initial stop consonants, a temporal measure of psychoacoustic performance would be indicated.

In our investigations of speech perception by the hearing impaired we have employed cue-degraded spoken stimuli to determine the cues used by hearing-impaired listeners for consonant recognition. Listeners are tested with different conditions of cue degraded nonsense syllables, among which the different redundant cues have been progressively eliminated. When a listener's performance declines significantly for a given condition of cue deletion, then it is apparent that the absent cue is important to perception.

The following text describes three experiments in which the relations were examined between psychoacoustic performance and consonant recognition, where both were measured by stimuli thought to require similar types of auditory processing.

VOWEL DURATION DISCRIMINATION VS USE FOR CONSONANT VOICING IDENTIFICATION

Some hearing-impaired listeners with severe/profound losses may rely predominantly on the vowel duration cue to distinguish voicing perception for word-final consonants (Revoile et al., 1982[10], 1985[11]). In an experiment on enhancement of the vowel duration cue (Revoile et al., 1986[12]), we used stimuli representing a range of vowel durations to test recognition of final fricative voicing for severely/profoundly hearing-impaired listeners (N=25).

Method. For each of the syllables /bæs/, /bæz/, /bæ f/, /bæv/, ten different spoken utterances were selected from a larger pool of stimuli to vary systematically in vowel duration. The vowels of the /bæs/-/bæf/ utterances ranged from 214 to 298 ms, and of /bæ z/-/bæv/, from 277 to 412 ms.

Identification of the utterances was tested via single-interval trials in which a listener chose a response from among the 4 syllables plus /bæ/, displayed on an answer box. No feedback of correct responses was given. Voicing perception was scored across all four fricatives (errors for place of articulation were ignored), yielding a percent correct score for each of two blocks of 40 utterances. These tests were administered at the end of the experiment on vowel duration cue enhancement.

The stimuli used to test vowel duration discrimination ranged from 200 to 475 ms, in steps averaging 10 ms. The stimuli were composed of iterations of a single pitch period that had been extracted from a typical test syllable. Duration discrimination thresholds were measured adaptively via 2A/3IFC trials in which listeners chose the longer stimulus relative to two identical reference stimuli of 197 ms, each. Correct-answer feedback followed each trial. About 42 trials were presented to reach threshold, which was taken as the smallest duration difference yielding 70% correct responses. About 5 threshold measurements of vowel duration discrimination were obtained per listener throughout the experiment.

All stimuli were presented to each listener's better ear at a comfortable level, established via an adaptive procedure at the beginning of each test session. A computer controlled all procedures. Stimulus events during trials were signalled to the listener by flashing lights on an answer box.

Results. Final fricative voicing perception for the listener group averaged 68% (S.D.=17.3) and ranged from 35% to 98% among the listeners. For discrimination of vowel duration, the listeners' mean jnd was 36 ms (S.D. = 24.5) relative to a 197-ms reference vowel. Among the listeners the jnds ranged from 17 to 89 ms, however about 3/4 of the group obtained jnds smaller than the mean.

A multiple regression analysis was carried out to examine the relations of various auditory measures, including vowel duration discrimination, to final fricative voicing perception. While the zero-order correlation (simple *r*) of vowel duration discrimination versus fricative voicing was modest (-.41) relative to the other variables analyzed, the third-order partial correlation (-.37) re vowel duration discrimination equaled that of the 250 Hz tone threshold, which showed the highest zero-order relation to the fricative voicing scores (-.87).

The similarity among listeners for discrimination of vowel duration may account for the barely moderate relation of this measure to final fricative voicing perception, for which a fairly continuous distribution of scores was found. This continuous distribution was partly due to the inclusion of some listeners who could use cues in the vowel onset to distinguish final consonant voicing, in the absence of the vowel duration cue.

It is possible that the relation between vowel duration discrimination and use of the vowel duration cue for final consonant voicing perception is too categorical to be well described by a regression model. That is, some criterion level of vowel duration discrimination may be necessary to support the use of the vowel duration cue. Howevèr, discrimination ability beyond that criterion level may be unrelated to cue use. Supporting this idea is our finding that 4 of the 5 listeners who performed at chance level [<55%] for voicing perception, were also those who showed the poorest vowel duration discrimination (jnd >57 ms) among the listeners overall.

TRANSITION DISCRIMINATION VERSUS USE FOR GLIDE MANNER PERCEPTION

For many persons with hearing impairments, the difficulties they experience in speech recognition may be related to deficient reception for formant transitions in speech. There seems to be no research that has attempted to test a given group of subjects for their ability to discriminate transitions, as well as for their use of transitions for consonant perception. In this study we examined hearing-impaired listeners both for their use of transitions for consonant perception and for their discrimination of transitions that were generated to simulate those present in the spoken syllables containing the test consonants.

Method. Perception was assessed for the initial phonemes of the syllables /wæk, jæk, bæk, gæk, dæk, æk/. For each syllable, 6 different utterances (male talker) were tested. A block of 36 utterances was tested at least 5 times for each listener, according to the procedure described above for the study of vowel duration cue use.

The listeners were 21 impaired- and 6 normal-hearing students at Gallaudet. The mean .5, 1, 2 kHz threshold average was 84 dB HL for the impaired group, and ranged from 64 to 105 dB HL. All listening was done monaurally (better ear) with the stimuli presented at the listeners' comfortable levels. The normal-hearing listeners were presented the stimuli at 75 dB SPL.

Among the other measures obtained were discrimination thresholds for different continua of synthetic transitions that varied in either frequency extent or duration. The stimuli were two-formant signals (parallel resonance synthesis) designed to resemble the transition characteristics of the velar glide and stop consonants in the /jæk/ and /gæk/ syllables. Thresholds were assessed using 2A/3IFC trials, with subsequent feedback, in an adaptive procedure. In tests for discrimination of transition duration, the transition frequency extent was held constant; the duration threshold obtained corresponded to the just-longer duration of transition that could be discriminated relative to a reference stimulus with a stop-like duration of the transition. For discrimination of transition frequency extent, the threshold was the minimum transition frequency extent that could be discriminated relative to a steady state formant.

Results. Results are reported here for one measure each of transition duration discrimination and of transition frequency extent discrimination. For both transition frequency extent and transition duration discrimination, performance by the hearing-impaired group was poorer than that seen for the normal group. In the test for transition duration discrimination, the hearing-impaired group required a transition that was 123 ms in duration in order to distinguish the difference relative to the 70 ms reference stimulus. In comparison, the normal group could discriminate a shorter duration transition of about 100 ms relative to the 70 ms reference.

Transition frequency extent was discriminated relative to a steady state reference. The hearing-impaired group required an F1 transition of over 100 Hz to discriminate its presence relative to a steady state stimulus. The normal-hearing listeners discriminated an F1 transition of less than 50 Hz relative to the steady state.

Percent correct perception is reported only for the glides of each repetition of a block of utterances, per listener. Mean perception of glide manner (place errors ignored) for the hearing-impaired group averaged 73%, while mean phoneme perception was about 60%. (Note that the normal-hearing group perceived /wæk/ and /jæk/ syllables with 100% accuracy.)

The results for glide manner perception for the /jæ k/ and /wæk/ syllables were analyzed relative to the discrimination results for transition duration and transition frequency extent. Pearson product moment correlations for /wæ k/ versus discrimination of frequency extent and of transition duration, respectively, were -.27 and -.23, and for /jæk/, -.46 and -.59 (p<.05). The direction of the correlations was negative for each analysis indicating generally that the higher the glide manner score, the smaller the frequency extent or duration of transition that could be discriminated.

The size of the coefficients obtained between transition discrimination and /jæk/ perception, both greater than -.40, reveals that transition discrimination was moderately related to manner perception for the velar glides. In other words, the discrimination ability for the transition stimuli tended to indicate how well a listener would perceive manner perception for the velar glides. This suggests that the listeners' abilities to use transitions for velar glide manner perception appears somewhat related to their ability to discriminate glide-like synthetic transitions.

AMPLITUDE MODULATION VS INTERVOCALIC CONSONANT IDENTIFICATION

Finally, as an adjunct to several speech perception studies, we have obtained some data on the detection of amplitude modulation at very low rates by our hearing-impaired listeners, an interest stimulated in part by the work of Festen and Plomp (1983)[13]. Within the context of a continuous utterance, relatively rapid spectral and temporal variations must presumably be processed to support consonant perception. For instance, static spectral shape cues to stop consonant identity (Blumstein and Stevens, 1979)[14] which are usually present for initial consonants in CV or CVC syllables spoken in isolation, can be absent or less salient for intervocalic stops in longer utterances. This places a greater burden on the listener to process transition and/or spectral change cues to consonant identity. Thus, the rationale for these studies was the possibility that a measure like the amplitude modulation threshold, because it involves processing amplitude variations in time, might be more representative of a listener's ability to perceive intervocalic consonants than more static sensitivity measures.

While prior work has used modulated broadband noise (e.g., Bacon and Viemeister, 1985[15]; Formby, 1985[16]; Lamore, et al., 1985[5]), we have used AM tones, sometimes in a tone complex, and have concentrated on modulation rates between 4 and 12 Hz. These very low rates are of interest both because they fall within the area of maximal sensitivity to sinusoidal amplitude modulation, and because they correspond to the region of temporal variation most associated with dynamic speech events (Plomp, 1984[17]). Four Hertz is a typical syllable rate for conversational speech; a rate of twelve Hertz falls in the range of more rapid fluctuations that may be associated with the detection of formant transitions. That is, the rapid rise and fall of amplitude within a single auditory "channel" as a formant passes across that channel is similar in wavelength to a 12-Hz modulation rate.

Method. These data were obtained from 9 listeners who, with other subjects, were involved in a study of speech enhancement that required identification of stop consonants in LPC vocoded speech. The speech stimuli were three-syllable nonsense utterances of the form /əCVlə/ and /əlVCə/ in which V = (/i/, /a/, /u/), and C = (/p/, /t/, /k/, /b/, /d/, /g/). The tokens, originally produced by a male talker, were analyzed and then resynthesized with an LPC vocoder program. They were identified in separate listening tests for the voiced and voiceless stops. In addition to the standard identification tasks, the listeners' modulation thresholds were measured for detecting several AM stimuli. Four AM conditions were used, all with a carrier frequency of 1024 Hz. For three conditions, the carrier was modulated at a rate of 4 Hz, and for the fourth condition the modulation rate was 12 Hz. Two of the 4-Hz modulated tones were modulated sinusoidally, but differed in duration (500 ms versus 1000 ms). For the third 4-Hz modulation condition, the carrier was modulated by a square wave. In the following, these conditions are labeled SN4 (4-Hz sinusoidal modulation), SQ4 (4-Hz square wave modulation), SN4X (4-Hz sine modulation extended in duration from 500 to 1000 ms), and SN12 (12-Hz sinusoidal modulation).

The procedure was a 3-interval task, similar to those described above, in which depth of modulation was varied adaptively in discrete logarithmic steps to locate the listener's threshold for detecting the presence of modulation.

Results. Table I presents the thresholds obtained from each listener for the various stimulus conditions, three-frequency pure tone threshold averages (PTA), and average percent correct stop consonant identification (ID) scores for consonants in either stressed syllable Initial, or Final position, and Combined.

The modulation thresholds varied over a range of about 25 dB and were generally well correlated with both the pure tone thresholds and with the identification scores. For example, the correlation between PTA and SN4X was -.636 while the correlation between PTA and combined ID score was .572. The measure most strongly related to consonant ID score (r=.840) for these 9 listeners was their threshold for detecting 4-Hz sinusoidal modulation in the extended-duration AM stimulus (SN4X). Under regression analysis, the relationship between SN4X threshold and Combined identification score was significant [F(1,7) = 16.81; p=.0046]. Once this variable was included in the regression equation, no other variable produced a significant improvement in the prediction.

These results suggest a moderately strong relationship between AM detection thresholds and intervocalic consonant identification, however, they must be considered with caution. First, it is clear from the regression analysis that variance associated with AM threshold is not independent of that associated with the three frequency averages. Secondly, the correlations here benefit from the small number of listeners and relatively wide range of hearing loss they exhibit: had our population of listeners been less diverse, or our population of sentences more diverse, weaker correlations would likely have obtained. Finally, it is worth mentioning that the LPC vocoding process used in this experiment may be thought of as adding noise to the speech. Consequently, the identification measure here is actually similar to a speech-in-noise test.

SUMMARY

Prediction of speech recognition abilities on the basis of auditory discrimination capacities should be tested using stimuli that represent the speech features which are cued by the particular acoustic differentials of the psychoacoustic discrimination tests. The potential application of this approach was tested in three studies using psychoacoustic measures of 1) duration discrimination, 2) rate of formant transition, and 3) detection of amplitude modulation to be correlated with recognition, respectively, of stop voicing, glide consonants, and intervocalic stops. The results yielded generally modest correlations suggesting that much additional research is required prior to the successful development of a predictive relational model between psychoacoustic characteristics of impaired hearing and phoneme perception.

Table I. Three-frequency (.5, 1.0, 2.0 kHz) Pure Tone Averages (PTA), Modulation Detection Thresholds (dB re 100 percent modulation), and percent correct consonant identification scores in syllable Initial position, syllable Final position, and Combined. See text for modulation conditions.

Subject	PTA	SN12	SN4	SQ4	SN4X	Initial	Final	Combined
BT	72	-28.0	-16.9	-39.9	-22.1	43.68	39.17	41.42
BK	65	-25.3	-23.8	-29.0	-27.8	50.99	62.63	56.81
CD	70	-41.1	-28.1	-33.2	-34.0	78.56	73.05	75.81
HJ	60	-28.4	-25.7	-28.6	-29.4	63.52	74.13	68.82
HK	57	-25.6	-20.4	-34.3	-24.8	72.43	65.63	69.03
HE	60	-28.3	-24.5	-29.6	-25.7	45.08	35.01	40.05
MM	75	-16.2	-12.7	-23.8	-15.4	36.06	34.82	35.44
PC	43	-32.7	-22.1	-32.8	-32.9	73.04	69.55	71.30
RD	76	-20.8	-16.9	-19.6	-18.8	45.78	43.89	44.80
Mean	64.2	-27.4	-21.2	-30.1	-25.7	56.57	55.32	55.95
S.D.	10.5	7.0	5.0	6.0	6.2	15.50	16.79	15.72

REFERENCES

[1] Haggard, M., Lindblad, A., and Foster, J. (1986). "Psychoacoustical and audiometric prediction of auditory disability for different frequency responses at listener-adjusted presentation levels," Audiology, 25, 277-298.

[2] Lutman, M., and Wood, E. (1985). "A simple clinical measure of frequency resolution," Br.J.Aud., 19, 1-8.

[3] Lutman, M., and Clark, J. (1986). "Speech identification under simulated hearing-aid frequency response characteristics in relation to sensitivity, frequency resolution, and temporal resolution," J. Acoust. Soc. Am., 80, 1030-1040.

[4] Stelmachowitz, P.G., Jesteadt, W., Gorga, M.P., and Mott, J. (1985). "Speech perception ability and psychophysical tuning curves in hearing-impaired listeners. J. Acoust. Soc. Am., 77, 620-627.

[5] Lamore, P., Verwey, C., and Brocaar, M. (1985). "Investigations of the residual hearing capacity of severely hearing-impaired and profoundly deaf subjects," Audiology, 24, 343-361.

[6] Dreschler, W., and Plomp, R. (1985). "Relations between psychophysical data and speech perception for hearing-impaired subjects," J. Acoust. Soc. Am., 78, 1261-1270.

[7] Plomp, R., and Mimpen, A. (1979). "Improving the reliability of testing the speech-reception threshold for sentences," Audiology, 18, 43-52.

[8] Preminger, J., and Wiley, T. (1985). "Frequency selectivity and consonant intelligibility in sensorineural hearing loss," J. Sp. Hear. Res., 28, 197-206.

[9] Dorman, M., Marton, J., and Hannley, M. (1985). "Phonetic identification by elderly normal and hearing-impaired listeners," J. Acoust. Soc. Am., 77, 664-670.

[10] Revoile, S., Pickett, J. M., Holden, L.D.,and Talkin, D. (1982). "Acoustic cues to final stop voicing for impaired- and normal-hearing listeners." J. Acoust. Soc. Am., 72, 1145-1154.

[11] Revoile, S., Holden-Pitt, L., and Pickett, J. M. (1985). "Perceptual cues to the voiced-voiceless distinction of final fricatives for listeners with impaired or normal hearing." J. Acoust. Soc. Am., 77, 1263-1265.

[12] Revoile, S., Holden-Pitt, L., Pickett, J., and Brandt, F. (1986). "Speech cue enhancement for the hearing impaired: I. Altered vowel durations for perception of final fricative voicing." J. Speech Hear. Res., 29, 240-255.

[13] Festen, J., and Plomp, R. (1983). "Relations between auditory functions in impaired hearing," J. Acoust. Soc. Am., 73, 652-662.

[14] Blumstein, S.E. and Stevens, K.N. (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. America, 66, 1001-1017.

[15] Bacon, S.P. and Viemeister, N.F. (1985). "Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners," Audiology, 24, 117-134.

[16] Formby, C. (1986). "Frequency and rate discrimination by Meniere patients," Audiology, 25, 10-18.

[17] Plomp, R. (1984). "Perception of speech as a modulated signal," In Van den Broecke, M.P.R., and Cohen, A. (eds.) Proceedings of the Tenth International Congress of Phonetic Sciences. Foris Publications, Cinnaminson USA. 29-40.

COOPERATIVE AND INTERACTIVE MODELS OF HEARING
AN INTRODUCTION

MANFRED R. SCHROEDER

Drittes Physikalisches Institut, Universität Göttingen
Bürgerstr. 42-44, D-3400 Göttingen, West Germany, and
AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

Since the days of Ohm, Seebeck and Helmholtz, we have gained great insight into the functioning of the peripheral auditory system of man (and many other species). We finally begin to understand even Tartini's (1740) "terzi suoni" as a nonlinear byproduct of active mechanisms in the inner ear.[1]

Highly sensitive *physical* measurements led to the discovery of "Kemp echoes" and oto-acoustic emissions. Combined with intracellular (hair cell) recordings and other *physiological* triumphs, the resulting insights have taught us what we have suspected all along: the mechanical resonators in the inner ear (von Békésy's traveling waves as seen in post mortem specimens) are but *pre*filters for delicate active processes to build on. Only thus can we understand the astounding sensitivity of our ears (close to the detection of random molecular motion) and their great frequency resolution (almost--but not quite--violating Heisenberg's Uncertainty Principle).

Psychoacoustic research has likewise made great strides in decades past and has contributed enormouly to our understanding of hearing. *Mathematical models* of auditory functions have had another heyday--and some models have even clarified our thinking about the ear.

While most stimuli used in hearing research have been of the nonspeech kind--the familiar tones, clicks and hisses in a veritable carnival of post-, pre-, and simultaneous masking paradigms, pulsation thresholds, loudness estimates etc.--a considerably corpus of knowledge using speech or *speech-like* stimuli has also been built up.

These experiments have taught us much about subtle interactions between the time and frequency domains and much else that is important for speech perception.

However, there remains a large white area of unexplored territory, located somewhere between these two continents explored by past research using speech- and nonspeech-like stimuli. Numerous experiments beckon us to enter this noman's land to discover properties of the auditory system that are *pre*-speech and yet transcend previous nonspeech work.

Why is it that a vibrato ("frequency modulation") imparted coherently to the individual components of a tone complex will tend to fuse these tones into a perceptual whole? Indeed, what *are* the physical parameters in complex stimuli that will lead to perceptual fusion on the one hand and, on the other hand, permit the separation of different "voices" (including those of musical instruments and other nonspeech signals).

Why is masking of a tone pulse influenced by the coherence ("phase stability") of the masker *hundreds* of milliseconds before or after the maskee [1]? What kind of temporal integration is at work here?

Considerable integrative capabilities of the human ear for nonspeech stimuli have also been observed in the *frequency* domain, both regarding gross spectral features ("formants") and spectral fine structure ("pitch"). A recent model [2] addresses the latter problem, but much remains to be explored on spectral shape integration and timbre perception.

In this context, we might consider also the subnormal speech of hearing-impaired speakers as another kind of prespeech. Certainly, the results of research on recognition of such "speech" are confusing, to say the least, in terms of present psychoacoustic understanding [3].

[1] Not surprisingly, these "amplifiers" involve bio-*molecular* processes--what *else* could go nonlinear at basilar membrane "excursions" smaller than 1 Angstrom, the diameter of the hydrogen atom.

The perceptual correlates of phonetic features and their *relational* properties [4] are another important subject that needs to be explored further in the arena of cooperative models of hearing.

Another area of concern is the interface between acoustic-phonetic analysis and lexical processes. W. D. Marslen-Wilson and U. H. Frauenfelder, in their contribution to this Symposium [5] argue strongly for integrating the methodologies of acoustic-speech research and psycholinguistics.

One of the strongholds of cooperative auditory processing is the human binaural system and we welcome a contribution with exciting new results to this important topic [6]

It is hoped that the present Symposium, in addition to auditing exciting new results from its participants, will provide a forum for focusing on those cooperative and interactive aspects of hearing that have been hitherto largely ignored. Wish that Tallinn will provide a strong stimulus in our search for new and revealing auditory stimuli!

REFERENCES

[1] Work by S. Mehrgardt, cited by M. R. Schroeder, "Speech and Hearing: Some Important Interaction," in M. P. R. Van den Broeke and A. Cohen (eds.), Proc. 10th International Congress of Phon. Sci. pp. 41-52 (1983).

[2] A. J. M. Houtsma and J. G. Beerends: "An Optimum Pitch Processing Model for Simultaneous Complex Tones," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).

[3] S. G. Revoile, H. T. Bunnell, and J. M. Pickett, "On the Prediction of Phoneme Recognition by the Hearing Impaired," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).

[4] K. N. Stevens, "Relational Properties as Perceptual Correlates of Phonetic Features," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).

[5] W. D. Marslen-Wilson and U. H. Frauenfelder, "The Interface Between Acoustic-Phonetic and Lexical Processes," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).

[6] A. Kohlrausch, "The Influence of Interaural Phase Uncertainty on Binaural Signal Detection," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).

RELATIONAL PROPERTIES AS PERCEPTUAL CORRELATES OF PHONETIC FEATURES

KENNETH N. STEVENS

Research Laboratory of Electronics and
Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge MA 02139 USA

Abstract

The view taken in this paper is that the apparent variability of the acoustic correlates of phonetic features is reduced if these correlates are described in relational terms. That is, the acoustic properties are specified in relation to the spectral and temporal context in which they occur. We present a number of examples of vocalic and consonantal features for which there appear to be advantages in specifying the acoustic correlates in relational terms.

1. Introduction

Some acoustic attributes of speech sounds produced by different speakers and in different contexts show a great deal of variability. This variability is especially evident if the properties are specified in terms of absolute measurements such as the frequencies of spectral prominences, times between acoustic events, or amplitudes of particular regions in the frequency-time representation of speech. The view taken in this paper is that much of this variability tends to disappear if the properties that constitute acoustic correlates of the phonetic features are defined in a relational sense. The term *relational* is taken to mean that an attribute at a particular frequency and time in the speech stream is specified in relation to the context in frequency and in time in which this attribute occurs.

2. Three frequency regions

In what follows, we shall consider examples of acoustic properties of classes of speech sounds that appear to be described most naturally in relational terms. The descriptions are usually in terms of relations between spectral prominences or periodicities at a particular point in time, or relations between spectrum amplitudes in particular frequency ranges in adjacent time regions. These relational properties will refer to spectral characteristics that are observed within broad regions of the audible frequency spectrum. In identifying the properties in different broad frequency regions, we are implying that the sound can be processed in the auditory system in different ways in these frequency regions. That is, we are assuming that the capabilities of the auditory system for processing sound in these frequency regions may be different, although we recognize that some aspects of the processing are common to all frequency regions. The edges of these frequency regions are not well defined and there may be some overlap between the regions. Before discussing in detail examples of the various relational properties, we will review briefly what the different basic frequency regions are, and what are some of the bases for selecting these particular regions.

The lowest frequency region extends up to about 800 Hz, and usually encompasses the frequency range of the first formant for adults. Within this frequency range, the frequency resolution as defined by the critical bandwidths, or by the bandwidths of the psychophysical or physiological tuning curves, tends to be independent of frequency, and is less than 100 Hz. The shape of the tuning curves is more or less symmetrical in this range and there are no low-frequency tails on the tuning curves [1]. The time resolution is poorer than it is at higher frequencies, as expected on the basis of the narrower frequency resolution.

Above the low-frequency region, the critical bandwidths increase with frequency, and the tuning curves are characterized by low-frequency tails. That is, the auditory filters in this frequency range can respond to low-frequency energy in the sound if it is sufficiently large. The time resolution at these higher frequencies is good, and the auditory filters respond within a millisecond or two to an abrupt increase in amplitude.

This high-frequency region is divided into two parts. The lower part, which we call the midfrequency region, encompasses the normal range of variation of the second and third formants for vowels. The bandwidths of these formants for vowels tend to be narrower than the bandwidths of the auditory tuning curves in this midfequency region. Within both the low- and midfrequency ranges, there tends to be synchrony of firing of individual auditory-nerve fibers to pure tones and to the frequencies of the first three formants for fibers whose characteristic frequencies are in the vicinity of one of these frequencies [2, 3, 4].

At much higher frequencies, probably above 3000 or 3500 Hz, the frequency resolution is much poorer than in the mid- and low-frequency range. The synchrony of firings of auditory-nerve fibers, either at their own characteristic frequencies or at the frequencies of nearby spectral prominences, is much less evident in this frequency range [2]. The frequencies of the formants for vowels do not contribute significantly to vowel identification at these frequencies. Spectral energy in this frequency range contributes primarily to the identification of particular consonant features, and the important aspect of the auditory-nerve response is probably its strength rather than the temporal characteristics of the nerve firings [5]. There are only about 5 critical bands relevant to speech processing in this high-frequency range, out of a total of about 22 such bands when they are spaced in such a way that one critical band separates adjacent filters.

3. Some relational properties for vowels

3.1 *F*0 contours

In describing the fundamental frequency ($F0$) variations in tone languages, it is common to use labels such as high, mid, and low tones. These labels are generally considered, however, to apply in a relational sense. A vowel with a high tone, for example, is regarded as having, at some point within the vowel, a maximum in $F0$ in relation to the $F0$ in nearby regions, either within the same vowel or within adjacent sonorant regions. Thus in a monosyllabic word with a high tone, the $F0$ is higher near the midpoint of the vocalic nucleus than it is near the beginning and end of the vowel. The value of $F0$ depends, however, on the individual talker, as well as the position of the vowel in the sentence unless the word is spoken in isolation. A listener presumably interprets this concave downward contour as being qualitatively different from a contour that is falling or is concave upward. When interpreting $F0$ contours, a listener seems to be examining $F0$ at one point in time in relation to $F0$ at other points—that is, the listener is making use of relational properties. Experiments have shown that the range of frequency variation of individual tones in a tone language (Mandarin) can be reduced drastically without modifying their perceptual integrity as long as the contour shape is preserved, i.e., as long as the relational aspects of the contour are maintained [6]. Thus $F0$ contours provide examples of acoustic properties that need to be defined in relational terms.

3.2 Relations between formant frequencies

Experiments of Chistovich and her colleagues [7] have suggested that some aspects of the auditory processing of a vowellike sound with two spectral prominences are qualitatively different depending on the frequency spacing between the prominences. This conclusion is based on experiments in which subjects are asked to match a vowellike sound with an adjustable single spectral prominence to be similar in quality to a two-prominence test stimulus. When the spacing between the two prominences is less than about 3-4 Bark, listeners tend to adjust the frequency of the matching stimulus to be between the frequencies of the two prominences of the test stimulus. For a greater spacing, a best match is obtained when the frequency of the matching prominence is equal to one or other of the prominences of the test stimulus. Our own (unpublished) experiments, using slightly different stimulus characteristics, have led to results that are consistent with those of Chistovich et al. Syrdal [8] has applied this concept of a critical formant spacing to an examination of the analysis and perception of vowels in English. She has shown that back vowels tend to have an $F2-F1$ difference that is within or close to this critical range, whereas for front vowels it is the $F3-F2$ spacing that is less than 3-4 Bark. One can conclude that vowel perception is based on a relational property among the spectral prominences for the vowel: the second formant in relation to the first, or the second formant in relation to higher spectral prominences—usually $F3$.

3.3 Breathy vowels

A different kind of relational property has been shown to distinguish between breathy and nonbreathy vowels in languages that use this feature contrastively. The amplitude of the fundamental component of the vowel in relation to the spectrum amplitude of the first formant is greater for breathy than for nonbreathy vowels [9, 10]. The perceptual relevance of property has been shown by Bickley [10]. While a quantitative specification of this property, valid across all vowel heights, has yet to be developed, it is clear that it is the relation between the amplitude of the fundamental component and that of higher components that contributes to the identification of the breathiness feature.

3.4 Formant contours

When a vowel is produced in the context of consonants, the formant frequencies usually vary with time throughout the vowel, and there may not be a time interval in which the formants remain relatively fixed. Several experiments have compared the identification of vowels characterized by time-varying formant frequencies and those with steady formants [11, 12, 13]. A general outcome of these experiments is that the identification of a vowel with time-varying formant frequencies cannot be predicted from the formant frequencies sampled at a particular point within the vowel such as the midpoint or the point where one of the formants reaches a maximum or a minimum value. Identification of the vowel is dependent on the entire contour. Preliminary data of Huang [12] and of Di Benedetto [13] suggest that the processing may be different for the first formant (i.e., within the low-frequency region defined above) and for the second formant. When the first formant ($F1$) has a concave-downward shape, the equivalent vowel height is that of a steady vowel with a lower $F1$ than the maximum $F1$ on the contour, i.e., the listener extracts some kind of average frequency from the contour. The $F2$ contour tends to be identified with a steady vowel for which $F2$ is beyond the maximum or minimum $F2$ on the contour, particularly when the contour is concave upward. Again we can conclude that the listener interprets the extreme values of the formants in relation to the formant contour preceding and following these maximum or minimum values.

4. Some relational properties for consonants

4.1 Acoustic correlate of sonorancy

A consonantal segment with the feature [+sonorant] is characterized by continuity of the spectrum amplitude at low frequencies in the region of the first and second harmonics—a continuity of amplitude that extends into an adjacent vowel without substantial change. This property is a consequence of the fact that there is essentially no obstruction to the airflow in the airways above the larynx. The vocal folds can therefore continue to vibrate in a normal manner, so that the low-frequency amplitude in the radiated sound remains unchanged. The perceptual salience of this low-frequency continuity has been examined in a limited way through experiments in which the spectrum amplitude at low frequencies was manipulated in the consonant region of a synthetic consonant-vowel stimulus (unpublished research of S.E. Blumstein and K.N. Stevens). This manipulation resulted in a continuum in which the consonant was heard as a prevoiced stop at one end and a nasal consonant (either [m] or [n] in two different continua) at the other. The [+sonorant] nasal consonant was heard when the amplitude of the lowest spectral prominence in the consonant was equal to or greater than the amplitude in the same frequency region in in the adjacent vowel. When the low-frequency energy in the consonant region was weaker, listeners tended to hear the consonant as a stop rather than a nasal. The correlate of sonorancy appears to involve a relation between the low-frequency spectrum amplitudes in the consonant and vowel regions.

4.2 Acoustic correlate of anteriority

A consonant with the feature [-anterior] is produced with a constriction at some point along the vocal tract posterior to the alveolar ridge. When the constriction is located in this region, the lowest front-cavity resonance of the vocal tract will usually be either $F2$ or $F3$. Such a consonant will have a spectrum that contains one or more spectral prominences in the midfrequency region as defined in section 2 above. The acoustic correlate of [-anterior] is that the spectrum amplitude of at least one of these prominences is approximately the same as the amplitude of the spectral prominence at the same frequency in the adjacent vowel. That is, there is continuity of the spectrum amplitude for one or more spectral prominences at the release of the consonant into the adjacent vowel. The evidence for this property relating midfrequency spectrum amplitudes before and after the consonant release is derived from acoustic analysis data and from perceptual experiments with synthetic consonant-vowel stimuli [14, 15, 16].

In separate perceptual experiments with stop and fricative consonants, the amplitude of a spectral prominence in the consonant region was manipulated and listeners were asked to identify the consonant. Results for the fricative experiment have been described previously [17]. For the fricatives, the stimuli at the ends of the continuum were /š/ ([-anterior]) and /s/ ([+anterior]), whereas for the stops they were /g/ and /d/ or /k/ and /t/ [16]. In all cases, the [-anterior] consonant was heard when the amplitude of the midfrequency spectral prominence in the frication noise was equal to or greater than the corresponding peak at the vocal onset, whereas the [+anterior] cognate was heard when the noise peak was weaker. It is suggested, then, that the acoustic correlate of the feature [-anterior] can be described as a property indicating the relation between spectrum amplitudes in the frication noise and in the vowel.

4.3 Acoustic correlate of coronality

Consonants that are classified as [+coronal] are often described as having a spectrum that rises with increasing frequency, so that there is substantial energy in the spectrum in the higher-frequency range (as defined in section 2 above) [18, 19]. Acoustic analysis of a variety of stop, nasal, and fricative consonants, together with some limited perceptual experiments [15, 17, 20], have led to a revised formulation of this property in relational terms. The property specifies that the high-frequency spectrum amplitude in the vicinity of the consonantal release should exceed the high-frequency amplitude observed after the onset of the following vowel. The spectrum at the release is regarded as a spectrum that would be represented in the peripheral auditory system in the sense that it is governed by the frequency resolution and adaptation characteristics of the peripheral auditory system.

This specification of the acoustic correlate of [+coronal] as a relational property involving changes in high-frequency energy is supported in part by the results of perceptual experiments in which the high-frequency spectrum amplitude of a burst is manipulated in synthetic consonant-vowel syllables [20]. Except for this high-frequency energy, the spectrum of the burst had no major prominences. The consonant was identified as an alveolar stop (i.e., having the feature [+coronal]) when the high-frequency amplitude of the burst was within about 5 dB of the high-frequency amplitude at the onset of the vowel (assuming neutral formant transitions). If one takes into account the overshoot in the response at the auditory nerve following an abrupt onset, this burst amplitude might be expected to yield a high-frequency response at onset that equals or exceeds the response in the following vowel. When the high-frequency spectrum amplitude of the burst was weaker, the consonant was heard as a labial. We conclude that the feature [+coronal] is based on a relational property, this time involving the high-frequency spectrum amplitude in the consonant region in relation to that in the adjacent vowel.

5. Concluding remarks

The examples that have been presented here could be expanded to include a variety of other acoustic properties. These examples provide support for a view that some advantage might be gained if the acoustic correlates of phonetic features were expressed in relational terms, particularly relations between acoustic events at nearby points in time. For the most part, these relational properties are probably manifested in the listener's auditory pathway at a level somewhat higher than the level of the auditory nerve. The representation of the speech stream at the level of the auditory nerve, however, provides a first step in the series of transformations that ultimately lead to the relational properties from which phonetic features can be derived.

6. Acknowledgements

The preparation of this paper was supported in part by grants from the National Institute of Neurological and Communicative Disorders and Stroke and the National Science Foundation.

7. References

1. N.Y.S. Kiang, T. Watanabe, E.C. Thomas, and L.F. Clark. **Discharge patterns of single fibers in the cat's auditory nerve.** Cambridge MA: MIT Press (1965).

2. D.H. Johnson. *The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones.* **J. Acoust. Soc. Am., 68**, 1115-1122 (1980).

3. E.D. Young and M.B. Sachs. *Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers.* **J. Acoust. Soc. Am, 66**, 1381-1403 (1979).

4. B. Delgutte and N.Y.S. Kiang. *Speech coding in the auditory nerve: I. Vowel-like sounds.* **J. Acoust. Soc. Am., 75**, 866-878 (1984).

5. B. Delgutte and N.Y.S. Kiang. *Speech coding in the auditory nerve: III. Voiceless fricative consonants.* **J. Acoust. Soc. Am., 75**, 887-896 (1984).

6. V.W. Zue. Unpublished research.

7. L.A. Chistovich and V.V. Lublinskaya. *The "center of gravity" effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli.* **Hearing Research, 1**, 185-195 (1979).

8. A.K. Syrdal. *Aspects of a model of the auditory representation of American English vowels.* **Speech Communication, 4**, 121-135 (1985).

10

9. P. Ladefoged. *The linguistic use of different phonation types.* In **Vocal Fold Physiology: Contemporary Research and Clinical Issues.** D. Bless and J. Abbs, eds. San Diego: College-Hill Press, 351-360 (1983).

10. C. Bickley. *Acoustic analysis and perception of breathy vowels.* **Speech Communication Group Working Papers I.** MIT, Cambridge MA , 71-81 (1982).

11. B. Lindblom and M. Studdert-Kennedy. *On the role of formant transitions in vowel identification.* **J. Acoust. Soc.Am., 42**, 830-843 (1967).

12. C.B. Huang. **Perceptual correlates of the tense/-lax distinction in General American English.** SM thesis, MIT, Cambridge MA (1985).

13. M.G. Di Benedetto. **An acoustical and perceptual study on vowel height.** Ph.D. thesis, University of Rome (1987).

V.W. Zue. **Acoustic characteristics of stop consonants: A controlled study.** Ph.D. thesis, MIT, Cambridge MA (1976).

15. D. Kewley-Port. *Time-varying features as correlates of place of articulation in stop consonants.* **J. Acoust. Soc. Am., 73**, 322-335 (1983).

16. S. Hawkins and K.N. Stevens. *Perceptual basis for the compact-diffuse distinction for consonants.* **J. Acoust. Soc. Am., Suppl. 1, 80**, S124 (1986).

17. K.N. Stevens. *Evidence for the role of acoustic boundaries in the perception of speech sounds.* In **Phonetic Linguistics**, V. Fromkin, ed. New York: Academic Press, 243-255 (1985).

18. G. Fant. **Acoustic theory of speech production.** The Hague: Mouton (1960).

19. S.E. Blumstein and K.N. Stevens. *Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants.* **J. Acoust. Soc. Am., 66**, 1001-1017 (1979).

20. R.N. Ohde and K.N. Stevens. *Effect of burst amplitude on the perception of stop consonant place of articulation.* **J. Acoust. Soc. Am., 74**, 706-714 (1983).

ROMAN JAKOBSON AND THE SEMIOTIC FOUNDATIONS OF PHONOLOGY

Henning Andersen

University of Copenhagen

This paper focuses on aspects of Roman Jakobson's theory of phonology which have had very little impact on contemporary phonology, but hold promise for its future development. Four specific contributions are mentioned: (1) the semiotic analysis of distinctive features; (2) the delineation of the content categories of phonic signs; (3) the understanding of language structure as a system of diagrammatization; and (4) the relation between code structure and phonetic regularities in messages. These contributions form the semiotic foundations of phonology.

0. The title of this panel discussion holds an implicit, but obvious invitation to look back over the development of modern phonology and to assess the relationship between Roman Jakobson's work and the course this development followed.

But the topic of this session does not have to be understood as an invitation only to look back. It would indeed be more in keeping with Jakobson's undaunted forward-looking spirit to examine the relationship between his work and contemporary phonology with a view to the future. One might ask whether there are still important lessons we can learn from this great teacher and look to see if there are elements of his understanding of phonology or insights of his which, for whatever reasons, were left unexplored and remain to be exploited, and which might enrich the development of phonology in the years to come.

This is the tack I have chosen for this brief presentation.

The phonological enterprise, in which Jakobson was the prime mover, from its very beginnings in the 1920's was a three-pronged research program aimed at describing the phonological patterns of the languages of the world, establishing a typology of known phonological systems, and uncovering the universal laws that underlie all such systems. Trubetzkoy's Grundzüge der Phonologie can be read as an interim report on this project, for it sums up results achieved by the mid-thirties along all these three lines of research.

Jakobson early saw that the question of phonological universals was the most important task on the agenda. And from his programmatic statement at the First International Congress of Linguists (SW I:3ff.) to The Sound Shape of Language (1979), he made this task a central topic in his scholarly work.

It is perhaps not surprising, in view of this research emphasis of his, that he came to be understood as an advocate of the description of all sound patterns in terms of universals. This understanding, though, is a misunderstanding in several respects.

In the first place, universals and description are categorially distinct. Universal phonology is concerned with the species general laws that constitute the phonological aspects of our faculté de langage and explain the relative uniformity of all phonological systems. Descriptive phonology, on the other hand, is concerned with the given language particular sound pattern, which must be grasped as a synchronically constituted part of a cultural tradition, fulfilling a multitude of functions in a given community. A sound pattern is like all other parts of a linguistic code a product of history. In the individual, it is a technique which has been acquired through an interplay of nature and nurture. As a social code, however, it is a system of historically transmitted, voluntarily imposed semiotic conventions. The description of such patterns is an entirely different undertaking from the explanation of their relative uniformity.

In the second place, even though Jakobson can perhaps be faulted for permitting this essential distinction to be blurred in some of his writings, there is no doubt that he himself had a constant and clear conception of the object of description, of what a sound pattern is, as distinct from the laws that govern phonological systems (cf. SW I:3).

To Jakobson a sound pattern was first and foremost a sort of ideology, that is, a system of signs constituting social values (SW I:9). As such it forms part of a cultural pattern, that all-encompassing ideology through which members of a community cognize their world. This conception was the basis of the parallels he drew between phonological markedness and the asymmetrical relations observable in other paradigms of cultural values, which may be equally culture specific, be superficially similar, but organized differently in different cultures, and be equally subject to revaluations in the course of history (cf. Trubetzkoy 1985: 161f.).

It is not difficult to see here a continuity of thought between Saussure and Jakobson. Jakobson acknowledged this continuity many times (e.g. SW I:312 and especially SW I:743ff.).

The structuralist conception of phonology as a system of signs constituting social values served as the major premiss for Jakobson's consistent efforts to grasp the sounds of language in semiotic terms, to illuminate the phonic side of speech as a system of semiotic systems embedded in culture. It is reflected in his writings on all the fields he contributed to in which the role of phonic elements is central, from descriptive linguistics to language acquisition and aphasia, to poetic language, and to the relations between linguistics and other sciences.

The Saussurean inspiration led Jakobson to two important advances. One was the semiotic analysis of distinctive oppositions, first presented in lectures in 1939 (SW I:280ff.) and in 1942 (1978, especially p. 59ff.). The other was a gradually elaborated identification of the categories of content expressed by phonic signs. His acquaintance with the ideas of C.S. Peirce led him to an explicit understanding of what it means for a system to be structured. His work in poetics, finally, led him to an understanding of the relation between code and messages which is crucially important for phonology. Let me pass these four points in review.

1. In 1939 Jakobson contrasted the minimal units of phonology with other linguistic signs and determined that the distinctive features, or, as I will call them here, the diacritic opposites, that serve to distinguish and identify the signifiants of segmental morphemes are signs in the Saussurean sense, that is, comprise a signifiant and a signifié, but differ radically from all other linguistic signs: (1) they have no specific signifiés, but signify mere otherness of their referents, the signifiants of morphemes; thus they are all systematically synonymous; (2) unlike all other linguistic signs, which are defined and form oppositions in terms of their signifiés, the diacritic opposites are defined and form oppositions on the basis of their signifiants.

The first of these characteristics is essentially related to their function: without such purely diacritic signs, the rich inventories of content signs of human languages would be impossible. Their synonymy, again, facilitates the maintenance of communication through sound change. By their second characteristic they are--in Saussure's well-known phrase--oppositive, relative and negative. Hence they can remain invariant despite differences among speakers in vocal tract configurations. This relational character also makes it possible for them to serve as a conceptual carrier wave for a variety of categories of other phonic signs.

The fact that diacritic opposites are defined and form oppositions on the basis of their signifiants entails that a phonological theory that treats them as mere names, without regard for the substance in which they are manifested, is essentially meaningless, as Chomsky & Halle very succinctly showed (1968:400f.).

Also, since the diacritic opposites have no positively defined phonetic signifiants, the language specific implementation rules that specify their realization cannot be captured by rewrite rules of the now customary kind. To conform to their nature, the linguist's rules will have to transform diacritic oppositions into sound differences.

Jakobson repeatedly spoke about the systems of diacritic oppositions as structured, but had little to say about the variability of such structures (cf. SW I: 709, 1979:166). However, there are enough indications of variability in the ranking of diacritic oppositions in different languages to make this a fruitful basis of a systematic phonological typology (cf. Andersen 1975), and to hypothesize variability in ranking to account for synchronic variation (cf. Andersen 1974, Gvozdanović 1985) and to explain major types of phonological drift (cf. Andersen 1978).

2. In 1929 Jakobson defined three classes of phonic signs (SW I:20).

The analysis of 1939 determined that these differ by the character of their signifiés. (1) The "stylistic" signs belong to the category of "content signs", which includes lexical and grammatical morphemes; Jakobson exemplified them with some of the kinds of pragmatic signs to be mentioned below. (2) The signifiés of the diacritic opposites, as mentioned, refer to the signifiants of segmental morphemes. (3) Contextual variants, finally, are phonic signs whose signifiés refer to the signifiants of diacritic opposites.

Such a semiotically based classification yields a principled basis for delimiting and subdividing the field of phonology. It could be defined as the study of systems of diacritic oppositions only; thus SW I:20. It could study the systems of phonic signs under (2) and (3) above; thus SW I:297. But it could also encompass all phonic signs with social value.

In his later works Jakobson tended towards this, the widest possible understanding as he repeatedly emphasized the interdependence of the all the parts of the over-all code, the consequent impossibility of describing any part of it in isolation, and in particular, the importance of consistently analysing speech sounds in regard to meaning.

From the point of view of their signifiants, all three classes of phonic signs are clearly inseparable. One can recognize the primacy of the diacritic opposites, which are constitutive of articulate speech. They can be conceived of as a carrier wave, as I suggested above, and the other kinds of phonic signs as "episigns" which deform or refract ideal projections of the diacritic oppositions into sound differences. But in speech none of these signs occur separately. And already in the intended audible output, which determines a speaker's articulatory implementation of his messages, and against which he monitors his proprioceptive feedback, signifiants of all three classes are superposed.

The analysis of the signifiés of the "stylistic" phonic signs remained very sketchy in Jakobson's writings. His famous paper on "Linguistics and poetics" (1960) defined a number of generic categories of pragmatic content, which are obviously in part expressed by phonic signs and which encode language particular categories of emotive, aesthetic, conative, phatic, and referential content (cf. SW I:295). They refer to elements of the communicative situation and thus are shifters (cf. SW II:131f.).

In 1979 Jakobson pointed to some generic categories of societal phonic signs, with which a speaker adjusts his pronunciation, in accordance with the norms of his speech community, to give expression to socially defined categories of status and role (1979:42). Such phonic signs symbolize the speaker's relation to his speech community and/or his addressee(s). Jakobson insisted on the distinction between these symbolic signs, which refer to cultural categories, and the pure indices properly termed physiognomic (Bühler 1978: 286), which facilitate speaker identification (cf. Andersen 1979).

The extent to which pragmatic and societal signs may be encoded in the use of diacritic oppositions and allophonic variation is obvious, and Jakobson early emphasized the necessity of examining all three classes of signs in their interrelations in explications of phonological change (e.g., SW I:19, 216).

3. One of Jakobson's lasting contributions to the study of morphology was the demonstration that when morphological analysis is carried through to an adequate depth, the incidence of combinations and/or concatenations of diacritic opposites in grammatical morphemes can be seen to mirror relations of meaning in the grammatical pattern of the language.

The tendency for vowels to be specialized for grammatical content in the Semitic languages was mentioned in 1929 (SW I:9) as an example of what Jakobson later called the "sense-determining function" of diacritic opposites. Later studies of the interaction between the phonemic and grammatical aspects of language (cf. SW II: 103ff.) cited numerous further examples of such specialized utilization of complexes of diacritic opposites for features of grammatical content in various languages, most convincingly in the detailed analyses of the inflectional patterns of Russian and other Slavic languages (e.g., SW II: 115ff., 119ff., 143ff., 148ff., 154ff., 184ff., 198ff.).

The semiotic basis for this is the "inverted" character of the diacritic signs (SW I:286ff.). Their signifiés ('otherness') serve to distinguish morpheme shapes; but by the substance of their signifiants they serve to identify morpheme shapes. Relations among complexes of diacritic opposites can consequently be used to reflect relations among the signifiants of morphemes.

The theoretical generalization implicit in the above-mentioned morphological studies was made explicit only in 1965 (SW II:345ff.), when Jakobson arrived at a Peircean characterization of language structure as essentially a "'system of diagrammatization', patent and compulsory in the entire syntactic and morphological pattern of language" (SW II:357).

Several kinds of diagrammatization can be distinguished. (1) By their perceptual dimensions diacritic oppositions are associated with, and diagram, other perceptual dimensions. This is synaesthesia (cf. 1979:188ff.). (2) Thanks to similarities between perceptual dimensions and other experiential dimensions, complexes of diacritic opposites can be used to form iconic (in Peircean terms: imaginal or metaphoric) lexical signifiants such as onomatopoea and ideophones (cf. 1979:179ff.). (3) Their direct association with lexical content in "word affinities" (cf. 1979: 195ff.) may be partly imaginal, but is in any case essentially diagrammatic, partial identity of signifiés being reflected by a partial identity of signifiants. (4) Most of the diagrammatic relations between grammatical meaning and sound which were revealed in Jakobson's morphological studies are of this last-mentioned type; but to these must be added (5) the morphophonemic alternations, which form diagrams indexing signifiants and/or signifiés of contiguous morphemes.

The great variety of diagrammatic relations in language awaits a thorough analysis. But Jakobson's contributions offer us a key to an understanding of language structure, which can be used to disclose in explicit terms the integration of phonology with the content systems of individual languages, the interplay between meaning and sound in historical phonology, and eventually the universal determinants of these interrelations between linguistically formed sound and sense.

4. The analysis of contextual variation forms a lacuna in Jakobson's investigations of sound patterns. He did not go beyond his understanding of 1939 that allophonic variants refer to the significants of diacritic opposites (cf. SW I:469 and 1979:42).

The reason for this neglect is consistent with his Saussurean understanding of the linguistic sign, which does not acknowledge the need the specify the syntactics of each linguistic sign. At the same time this neglect may be related to the absence in Jakobson's thinking of the fundamental distinction between language system and language norms (cf. Coseriu 1952).

It is in fact a question of phonological norms whether, say, flatted consonants in a given language are realized as labialized, retroflex, or pharyngealized, and whether a given five-vowel system is normally realized in one or another of several ways in a given community (cf. Vysotskij 1967). Similarly it is a question of norms whether and to what extent a given diacritic opposition is contextually suspended (the distinctive opposites being deleted in specified environments), and with what non-diacritic features complexes of diacritic opposites must be expanded before they are realizable in speech.

It follows that any individual sound pattern must be described as a system of diacritic oppositions conjoined with a historically established set of phonetic norms, expressible as rules of implementation. Phonological typology must consider not just systems of diacritic oppositions, but also the diverse norms of realization attested in different languages for each system type. Universal phonology must determine the freedom with which one and the same system type can be conjoined with different phonetic norms and the universal limits of this freedom.

From a semiotic point of view, however, Jakobson's theoretical advances offer a good basis for the systematic investigation of contextual variation.

First, variation rules expand complexes of diacritic opposites with subsidiary phonic signs that act as indexes, pointing to the context to which they have been assigned. The importance of these "auxiliary-sociative" signs for communication was recognized and repeatedly emphasized by Jakobson (cf. SW I:469, 1979:42). They are the semiotic counterpart of morphophonemic alternants.

Secondly, variation rules produce distributional patterns which carry information about the system of diacritic oppositions. Note that the mere fact that different phonetic values are in complementary distribution is a sign that they do not form a diacritic opposition: complementation diagrams the absence of opposition. Furthermore, by assigning different non-diacritic values to different contexts variation rules correlate non-diacritic values with (complexes of) diacritic opposites according to an apparently universal principle, viz in such a way that marked values are assigned to marked contexts, and unmarked to unmarked. This principle, by which equivalences in markedness are diagrammed by contiguity relations, was first discovered by Jakobson (1960) and by him held to be characteristic of the poetic function of language. However, it has been shown to be a much more general principle, in evidence whenever a value system is manifested syntagmatically (Andersen 1987). The phonetic co-occurrence relations thus codified in variation rules are the phonological counterpart of the lexical and morphological systems of diagrammatization which were mentioned above.

Finally, since the phonetic norms as a whole are established by convention, they serve to symbolize the speaker's allegiance to the socially or geographically defined community for which the given set of norms holds (1979:42).

5. To most phonologists today, probably, Jakobson is known primarily as the initiator of the search for phonological universals, the scholar who more than anyone else contributed to the modern understanding of the dependence of phonology on the innate capacities which are man's by nature. There seems to be a growing awareness among phonologists that there is a great deal more in language particular sound patterns than is accounted for by such universals. In turning our attention to these idiosyncratic aspects of phonology, we need not turn our backs on Jakobson. On the contrary, his view of a sound pattern as first and foremost a system of signs with social value and his substantive contributions to the elucidation of the character of these signs and of their interrelations offer a fruitful orientation and effective conceptual tools for future work in phonology.

References

Andersen, Henning. 1974. "Markedness in vowel systems", Proceedings of the 11th International Congress of Linguists, ed. by L. Heilmann, 891-7. Bologna: Il Mulino.

_______. 1975. "Variance and invariance in phonological typology", Phonologica 1972, ed. by W. U. Dressler & F. V. Mares, 67-78. München: Wilhelm Fink.

_______. 1978. "Vocalic and consonantal languages", Studia Linguistica A. V. Issatschenko a Collegis Amicisque oblata, ed. by L. Durovic et al., 1-12. Lisse: Peter de Ridder.

_______. 1979. "Phonology as semiotic", A semiotic landscape, ed. by S. Chatman et al., 377-81. The Hague: Mouton.

_______. 1987. "On the projection of equivalence relations into syntagms", New vistas in grammar, ed. by S. Rudy & L. Waugh. In press.

Bühler, Karl. 1978. Sprachtheorie. Die Darstellungsfunktion der Sprache. Stuttgart: Fischer.

Chomsky, Noam & Morris Halle. 1968. The sound pattern of English, New York: Harper & Row.

Coseriu, Eugenio. 1952. "Sistema, norma y habla", in his Teoría del lenguaje y lingüística general, 11-114. Madrid: Editorial Gredos.

Gvozdanovic, Jadranka. Language system and its change. On theory and testability. Berlin: Mouton De Gruyter.

Jakobson, Roman. 1960. "Linguistics and poetics", Style in language, ed. by Thomas A. Sebeok, 350-77. Cambridge MA: MIT Press.

_______. 1971a. Selected writings, I. Phonological studies. 2nd, expanded edition. The Hague: Mouton.

_______. 1971b. Selected writings, II. Word and language. The Hague: Mouton.

_______. 1978. Six lectures on sound and meaning. Trsl. by John Mephan. Hassocks, Sussex: The Harvester Press.

_______ & Linda Waugh. 1979. The sound shape of language. Bloomington: Indiana UP.

Trubetzkoy, Nikolaj S. 1958. Grundzüge der Phonologie. Göttingen: Vandenhoeck & Ruprecht.

_______. 1985. N. S. Trubetzkoy's letters and notes, prepared for publ. by Roman Jakobson. Berlin: Mouton De Gruyter.

Vysotskij, S. S. 1967. "Opredelenie sostava glasnyx fonem v svjazi s kacestvom zvukov v severnorusskix govorax", Ocerki po fonetike severnorusskix govorov, ed. by L. L. Kasatkin, 5-82. Moscow: Nauka.

TOWARDS A CALCULUS OF INTONATION CONTOURS FOR SENTENCES OF ARBITRARY SYNTACTIC COMPLEXITY

E.V.Paducheva

Dept. of theoretical foundations of informatics,Institute
of scientific and technical information
Moscow, USSR, 125219

ABSTRACT

Rules are proposed which describe cooccurrence restrictions for tone units in acceptable intonation contours of sentences and thus generate intonation contours for sentences of infinite length and syntactic complexity.

INTRODUCTION

Roman Jakobson more than anybody else contributed to formation of modern phonology in its present shape, having characterized the phoneme as a bundle of distinctive features; having raised the problem of the laws of combination of phonemes; and introducing the notion of protracted distinctive feature as a supersegment characteristics of a word overcoming phoneme boundaries, cf. protracted feature "voised" in the initial combination of consonants in Russian vzdrognut' or protracted feature "voiseless" in the initial consonant combination in vstreča. Though Roman Osipovich payed relatively less attention to prosody than to segmental phonology, the subsequent development of phonology proved his ideas in prosody to be stimulating for researchers of different schools and trends.

Those who study prosodic aspect of a sentence usually presume that it is sufficient to describe main functional "blocks" of intonation - tone units, - which characterize relatively simple and short sentences: the problem of synthesis of these blocks into sequences which really occur in linguistic activity, when the speakers generate utterances of considerable complexity, is left aside. Meanwhile the utterance in a natural language can have theoretically infinite length and cooccurrence of tone units in grammatically correct sentences is governed by sufficiently sophisticated rules. The problem arises - to delimitate these rules.

INTONATION CONTOUR OF A SENTENCE

We accept an assumption - perhaps, a bit simplifying, that intonation structure of a sentence, i.e., its intonation contour, can be represented as a well formed sequence of tone units. This amounts to saying that a tone unit is supposed to be an elementary segmental unit of intonation. There is no generally accepted intonational transcription for Russian. In our transcription we shall use the following symbols. Raising tone (IK-3 according to [1]) is symbolized by / . Symbol ! denotes accent - in the sense of [3], i.e. a fall of the base tone on the stressed syllable of the word, but not a glide; cf. the accent on the word kogda in the sentence Kogda! on priedet? uttered in a context where the fact of his arrival has already been mentioned. Falling tone \ (IK-1 according to [1]) marks the end of an indicative statement independent of the its textual context. This tone unambiguously expresses the main phrasal stress of the sentence. Falling tone, at least phonologically, is not a primitive - it is rather a conflation of two features: accent and the indicator of the completion of the utterance. Thus, (1) presents a minimal pair: the question (a) contains a pure accent, while the statement (b) - an accent in combination with the indicator of completion:

(1) a. On napišet/ ¦ ili pozvonit!?
    b. On napišet/ ¦ ili pozvonit\.

The accent can be conflated with the indicator of noncompletion; this combination is symbolized as ∨ (in terms of [1] it is IK-4), and is used to mark either a constituent or a syntactically completed sentence which alludes to a specific textual context of such a type as can be introduced by an adversative conjunction, cf. A vaš ∨ gde bilet? ⟨With the tickets of all the rest everything is in order⟩; Pokatal-s'ja, i xvatit ∨ ⟨the other should also have an opportunity⟩.

There are contrastive tones corresponding to the raising and to the falling tone; they are symbolized as // and \\ . Exact phonetic characterization of the symbols used in transcription is here irrelevant; for concrete details cf. [1], [2].

The borders between tone groups are indicated by a dotted vertical line ¦ . The symbol of the tone is placed after the word which constitutes the intonation centre of the tone group. More exact indication (e.g., marking the stressed syllable) is not necessary for our purposes.

The analysis of a sentence into its tone groups is slightly obscured by the so called binding of accents, described in [2], cf.:

(2) Kakuju⌜ knigu⌝ ty citaeš?
(3) Kuda⌜ ty položil Bulgakova⌝?

We assume that sentences like (2) consist of a single tone group with two centers, while in sentences like (3) two tone groups can be delimitated.

It would be natural to suppose that each tone group must have its own phrase accent. But for convenience of marking correlations between intonational and syntactic structure of a sentence it is preferable to make also use of unaccented tone groups. Thus in sentence Posredstvennyj\\ poet ¦ byl Šaxovskoj its second main constituent can be considered a separate tone group. Besides, if we cannot afford of unaccented tone groups then the theme of a sentence (i.e. a component of its topic-comment articulation) usually will not constitute a separate tone group, as in a sentence On xudožnik\. Some by no means relevant intonational distinctions will not be reflected in our transcription, just because they are besides the point - such as tempo and register distinctions which allow to express the accent characterizing a constituent as a whole in contradistinction to the accent of a word, see [2]; different degrees of reduction of unaccented words described in [3] are also ignored.

It must be underlined that many types of sentences allow for variation of intonation contours, and it is not claimed that the transcription proposed is the only one possible.

A BRIEF SURVEY OF LITERATURE

The idea of a generative grammar for intonation contours was first suggested by N.Chomsky and M.Halle [3], but with several simplifying assumptions, cf. in this connection [4] with references to some earlier publications. Thus, of all parameters characterizing intonation only the strength of stress (alternatively, the degrees of reduction) was taken into consideration. Meanwhile intonation contour is defined by a far more rich complex of prosodic features (the pitch; tone fallings and tone raisings; the accent; tempo etc.), and it is reasonable to try to generate the complete configuration of such features. As for the strength of the stress, there are no more than two degrees of it which are meaning-differentiating: the original mechanism of word stress reduction depending on the depth of syntactic embedding (the so called Nuclear Stress Rule) proposed in [3], presupposes theoretically unlimited scale of such degrees, but these distinctions are rather phonetical than phonological by nature (i.e. they are automatical and not specifically meaning-differentiating). Besides, it was assumed in [3] that intonation, let alone contrasts, is predicted by the surface syntactic structure of a sentence. But this hypothesis must now be rejected, - indeed, there are at least three groups of arguments that contradict it.

Firstly. The resulting intonation contour is influenced not only by the surface structure but also by the deep structure of a sentence.

Secondly. Intonation contour of a sentence is to a certain extent predicted by prosodic properties of its words, i.e. by its lexical composition and not by its syntax. And the effect of the lexical composition does not always amount to contrasts. Thus, personal pronouns in Russian are normally unstressed, no matter what their syntactic function is, and acquire stress only in a marked context, as, e.g., the following: Čašče/ vsego ¦ menja\ vidjat v Oblomove¦, govorja, čto ja èto lico/ ¦ napisal s sebja\. Among particles one are always unstressed (like i, -taki, me), the other always stressed, at least in one of their meanings. Besides, many particles are capable of predicting the stress placement in words with which they are connected, cf. [5]. Quasi-synonyms razve and neuželi are known for their prosodic differences [6]: the former induces the intonation of a question, and the latter - of a statement. There are also nonauxiliary words with idiosyncratic prosodic properties. Thus, Russian adjective redkij (in one of its uses) calls for emphatic stress inside the noun phrase that contains it; usually this noun phrase must occupy the initial position in a sentence: Redkaja\\ ptica ¦ doletit do ego serediny!; Redkaja para sapog\\ ne proxodila čerez ego ruki!. Analogously for the adjective raznyj: Raznyje\\ ljudi živut na ostrove!; Raznyje\\ byvajut slučajnosti!. Limited capacities to occupy thematic position in the topic-comment structure usually characterize words with negative and quantifitory meaning. Example (1) shows that thematic position rejects words with negative meaning:

(1) Mne dostatočno/ ¦ vašego raskajanija\.
    *Mne nedostatočno/ ¦ vašego raskajanija\.

Nedostatočno must bear main stress, i.e. it must have a falling and not a raising tone. Here are some examples with quantifiers (see also [7], p. 127):

(2) a. Častaja pričina bolezni/ ¦ - prostuda.
    b. *Redkaja pričina bolezni/ ¦ - prostuda.
(3) a. Často/ ¦ on prixodil zapolnoč\.
    b. *Redko/ ¦ on prixodil zapolnoč \.

(though Izredka/ ¦ on prixodil zapolnoč).
(4) a. Nedavno/ ¦ on vernulsja\.
b. *Davno/ ¦ on vernulsja\.
(5) a. Inogda/ ¦ on šel peškom\.
b. *Vsegda/ ¦ on šel peškom\.
(6) a. Bol'šinstvo/ učastnikov ¦ bylo\ zanjato v pervom akte.
b. *Men'šinstvo/ učastnikov ¦ bylo\ zanjato v pervom akte.

Cf. also an example from [8], p. 289: in sentence Snegu malo vypalo the verb is obligatory unstressed.

Many adverbs bear obligatory contrastive stress, cf. Naprasno\\ ty staralsja (*Naprasno/ ¦ ty staralsja\); Tščetno\\ ¦ stal by ja èto skryvat'! (*Tščetno/ ¦ stal by ja èto skryvat'\). The word ran'še (in one of its uses), is on the contrary, invariably thematic: Ran'še / ¦ on byl vsegda mračen\ (*Ran'še\\ on byl vsegda mračen). The main stress on ran'še is possible only in the presence of a particle: Ëto ran'še\\ on byl vsegda mračen.

And thirdly: linear-intonational structure has its own semantics, which is in many respects independent both of syntax and of lexical composition of a sentence; hence the notion of a communicative paradigm constituted by a set of sentences with the same array of lexemes and syntactic structure, which differ from each other in their linear-intonational structure.

As the semantics of the topic-comment structure is closely connected with a pragmatic context, usual semantic distinctions caused by variation of tone placement (cf. Ivan poedet v Kiev\\; Ivan poedet\\ v Kiev; Ivan\\ poedet v Kiev) can be accompanied by quite specific ones. Thus, in (a) U menja ostalos'\ nemnogo vremeni dlja osmotra goroda and (b) U menja ostalos' nemnogo\ vremeni dlja osmotra goroda semantic opposition is nearly that of antonymy. See also example (7):
(7) a. Doma/ ¦ Ivana net\.
b. Doma xleba\ net.

For (b) the contour Doma/ ¦ xleba net\ is practically excluded, but by purely pragmatic reasons. In fact, the semantic contribution of a contrastive theme to the sentence meaning is very rich: it enriches the meaning of a sentence by an implicature which, in this case, can be worded as follows: 'And for other places it is not so' or 'And about other places nothing is known' [7], which is quite sound for (a) but trivial and meaningless for (b).

And the last example of a pragmatic component in the intonation meaning. Cf. sentence Sledujuščaja/ stancija ¦ - Zvenigorod∨ pronounced with this intonation by an announcer in a suburban train. This intonation must be treated as a flagrant error. Indeed, in the situation described the adversative context, which is alluded to by IK-4, makes one think that this announcement will be followed, in due time, by an analogous announcement concerning the station which goes after Zvenigorod. But this expectation fails, because Zvenigorod is the last station of the railway.

AN OUTLINE OF THE CALCULUS

The set of all intonation contours can be enumerated with the help of a calculus of the following shape. Suppose that all elementary contours are given, which characterize syntactically non-extended sentences or constituents consisting of one or two tone groups, cf. contour (\) in Nastupila vesna\; raising-falling (RF) contour (/¦\) in Keramika/ ¦ - èto krasivo\; contour (\\ ¦ ◡) in Posredstvennyj\\ poet ¦ byl Šaxovskoj. More complex contours owe their existence to the fact that elementary contours (or, more exactly, syntactic constituents with elementary contours) are inserted in a context with a given tone characteristics. Thus we arrive at rules of substitution which generate more and more complex contours from the simplest ones. Rules of substitution may, certainly, refer to the context, as is natural for a context sensitive phrase structure grammar.

When we try to substitute some intonation contour for a tone group in a certain context the following situations may arise: 1) substitution is impossible, i.e. the sequence of tone groups which is the result of the substitution does not correspond to any acceptable intonation contour; 2) substitution is possible but a) the substituted contour must be transformed in a special way; b) the context must be transformed in a special way. Presumably, only such transformations are allowed which do not change the topic-comment semantics conveyed by intonational means. We hope that a finite set of rewriting rules of the type described above will constitute a model enumerating well formed intonation contours. Though only substitution rules have generative power, prohibitions deserve attention as well, for they explicate useful cooccurrence restrictions.

EXAMPLES OF SUBSTITUTION RULES

Rule 1. RF-contour (/¦\) when inserted in a position with inherent raising tone, must be transformed into a subcontour (' ¦/); in other words, the former main stress after substitution becomes a secondary stress and the former secondary stress becomes a mere accent. In examples below a is an independent sentence; in b this very sentence is put into an embedded position:
(1) a. Zavtra/ ¦ položenie izmenitsja\.
b. Esli zavtra' ¦ položenie izmenitsja/ ¦ ja vam soobšču\.
(2) a. Derevjannyj lubok počti isčez/ ¦ , a mednyj prišel v upadok\.
b. Kogda derevjannyj lubok počti isčez', a mednyj prisel v upadok/ ¦ , voznik lubok iz medi staroobrjadcev\.
(3) a. Čtoby pereexat' v stolicu/ ¦ , on soglasen na žertvy\.
b. Čtoby pereexat v stolicu' ¦ on soglasen/ na žertvy/?
(4) a. Keramika/ ¦ - èto krasivo\.
b. Keramika' ¦ - èto krasivo/?
(5) a. Muzyku/ ¦ on ljubit\.
b. Muzyku' ¦ on ljubit/?
(6) a. Posle spektaklja/ ¦ pozvoni\ mne.
b. Posle spektaklja' ¦ pozvoni/ mne/! (request)
(7) a. Na nej byla belaja šuba/ ¦ i šljapka\.
b. Na nej byla belaja šuba' ¦ i šljapka/?
(8) a. On priedet v ijule/ ¦ ili v avguste\.
b. Esli on priedet v ijule' ¦ ili v avguste/, on ešče uspeet.
(9) a. Ivan/ ¦ živet v Kazani\.
b. Ivan' ¦ živet v Kazani/ ¦ , a ja v Syzrani\.
(10)a. V novom zdanii/ ¦ budet biblioteka\.
b. V novom zdanii' ¦ budet ta biblioteka/, ¦ kotoraja ran'še byla v podvale\.

It is clear from these examples that the transformation which RF-contour undergoes is independent of syntactic construction it belongs to - it may be a coordinative group, subject-predicate combination, adverbial modifier + extended sentence etc. Equally inessential is the nature of the construction which delivers a position with an inherent raising tone - it can be a prepositive modifying sentence; a general question; the first component of a compound sentence, an utterance with the illocutionary force of request; a sentence with a restrictive modifier, as in (10b), etc. Thus it is clear that tones may interact directly with each other without such intermediaries as syntax or meaning.

Substituted RF-contour may belong not to an independent sentence, but to a constituent, cf. Dom otdyxa / ¦ stoit na beregu reki\ and Dom otdyxa' ¦ , kuda my poedem/ ¦, stoit na beregu reki.

Rule 1 shows that raising tone cannot stand syntactic embedding.

Rule 2. RF-contour when embedded into a position with inherent raising tone can be simplified into a contour consisting of one tone group with a raising tone - the intonation centre corresponding to the place of the former falling tone. In other words, two tone groups can merge into one, the former secondary stress having faded away and the former main stress having been transformed into a raising tone:
(11) a. Voprositel'naja/ intonacija ¦ otličaetsja ot predupreditel'noj\.
b. Voprositel'naja intonacija otličaetsja ot predupreditel'noj/ ¦ tol'ko bolee vysokim registrom golosa\.
(12) a. Ivan/ ¦ živet v Kazani\.
b. Ivan živet v Kazani/ ¦ , a ja v Syzrani\.

This rule does not generate any new contours but it captures one of the most important regularities of behavior of intonational constituents in such conditions when their hierarchy becomes more complicated.

Rule 3. If in a RF-contour the raising component is not obligatory (as, e.g., is the case in a coordinative group), it can be transformed, in a position with inherent raising tone, into a subcontour (/¦'). Thus in (13) utterance a is intonated in accordance with Rule 1, and b - in accordance with Rule 3:
(13) a. Kon"junkcija P&Q istinna' ¦ , esli P i Q istinny/ ¦ , i ložna/, esli P i Q ložny\.
b. Kon"junkcija P&Q istinna/ ¦ , esli P i Q istinny', i ložna/ ¦ , esli P i Q ložny\.

In (14) the end of the first of the two conjoined clauses is marked by tone ∨, as if it were a separate sentence:
(14) Idet napravo/ ¦ - pesn'∨ zavodit ¦ , nalevo/ ¦ - skazku\ govorit.

Rule 4. RF-contour can be inserted into a position with an inherent falling tone (non contrastive); if it gets into the context of a raising tone on the left, this raising tone can be replaced by tone ∨:
(15) a. Rano utrom/ ¦ Petja otkryl kalitku\.
b. Rano utrom∨ ¦ Petja otkryl kalitku/ ¦ i vyšel na lužajku\.

Replacement of a raising tone by an accent is also possible: Teatr byl zakryt' ¦ , tak kak truppa/ ¦ uexala na gastroli\.

Rules 2 - 4 reveal a general tendency of language to avoid sequences of identical tones; cf. the impossibility of *Ja uveren/ ¦ , čto Pavel/ ¦ nam pomožet\. Example (from [1]) where such sequence is allowed, requires explanation: Pojmannyx ptic/ ¦ vyderživajut na karantine/ ¦ i očen' xorošo kormjat\. Even accent, which is the most neutral of all types of phrasal stress, allows for repetition only under very special conditions, cf. an example from [2]: Vy v Telavi' ¦ kogda' poedete? where the sequence of identical tone groups is conditioned by a split word order.

There are, though, clear cut syntactic exceptions to this regularity; thus, appositive construction, on the contrary, is based intonationally on tonal repetition of the preceding stress:
(16) a. Ostal'nye razmestilis' nemnogo podal'še\ ¦ , na drugom beregu reki\.

b. Nemnogo podal'še/ ¦, na drugom beregu reki/ ¦, razmestilis' ostal'nye\.

Yet contrastive tone is not repeated by an apposed phrase: Takoj// on ¦ čudak\\ ¦, vaš Vanja!.

PROHIBITED SUBSTITUTIONS

1. RF-contour cannot be inserted into an unstressed position after a contrastive tone:

(1) a. Ja sčitaju, čto Whorf/ ¦ byl lingvistom osobogo \ roda.
b. *Ja znaju\\ ¦, čto Whorf/ ¦ byl lingvistom osobogo\ roda.

In such position RF-contour is simplified into a contour with two accents; thus, (1b) ⇒ Ja znaju\\, čto Whorf! byl lingvistom osobogo! roda. Cf. also: Tol'ko to/ prekrasno ¦, čto ser'ezno\ and Ne tol'ko\\ to! prekrasno, čto ser'ezno!.

2. Contrastive tone groups allow of no syntactic embedding; they are confined to independent sentences, thus belonging to the so called main clause phenomena:

(2) a. Da, ja kot\; no ljudi inogda tak// nevnimatel'ny.
b. Ja ne znal/, čto ljudi tak\ nevnimatel'ny.

Specific tone on tak disappears in the embedded position. Sentence Vernulis'\\ naši guljaki, being embedded, sounds unnatural: *Nesmotrja na to, čto vernulis'\\ naši guljaki...; cf. also: Kogo// on tol'ko ne sprašival\ but *Ja dumaju, čto kogo// on tol'ko ne sprasival\.

Prohibitions 1 and 2 taken together provide an explanation to the fact that a sentence cannot contain more than one contrastive tone (of the same direction). Thus, sentence (3) is not well formed because it combines two contrasts:

(3) *Jasno, čto imenno ètot\ smysl ¦ peredaetsja predloženiem (a)// ¦, a ne (b)\\.

Moreover, contrastive falling tone functions as the main sentence stress; thus, if there is another candidate for the post of the main stress bearer in a sentence, a conflict is bound to arise:

(4) *On ni s togo ni s sego/ ¦ vzjal da i rasskazal\\ mne ¦ dovol'no zamečatel'nyj slučaj\ (Turgenev).

(5) *Už on dostaval\\ ¦ - dostaval! iz-za pazuxi ¦ skomkannoe pis'mo na imja Oblomova\.

Syntactic and lexical peculiarities of these sentences create conditions for preposed, and thus contrastive, accent; while final noun phrase, being indefinite, also longs for the position of the main stress bearer. Though contrastive stress is usually treated as a phonetic phenomenon, it seems that contrast cannot be identified on purely phonetic grounds: means of expression for contrast are scarce and disparate (e.g., in [1] it is mentioned that contrast can be conveyed by strengthened word stress or by more distinct pronounciation of phonemes). Contrast is definitely opposed to its absence only when it is confirmed by structural or semantic factors. On the other hand, phonetic means must be quite explicit if without intonation the intended meaning will be lost, cf. ⟨- A sbežavšij byl vaš dvorovyj čelovek? - Kakoe dvorovyj čelovek? Èto by ešče ne takoe bol'šoe mošenničestvo.⟩ Sbežal// ot menja ¦... nos\\ : it is contrast that transforms this sentence into an identity statement. Tone group bearing the main stress (in particular, contrastive stress) can easily change its place in a sentence if there is no other contrasts in the same sentence. If there are, then removel of a contrastive tone group from its final position destroys the communicative structure. Example: Lož'// ¦ - religija rabov! ¦ i xozjaev\\. Pravda\\ ¦ bog svobodnogo čeloveka!. For the second sentence underlying word order and intonation are as follows: Bog svobodnogo// čeloveka ¦ - pravda\\. Indeed, svobodnyj čelovek is the theme of this sentence, and a contrastive one, because it is opposed to raby i xozjaeva from the first sentence: pravda is the rheme, also contrastive, for it is opposed to lož' in the first sentence. Emphatic preposition of the contrastive rheme destroys this structure: communicative meanings which were expressed explicitly by the underlying word order in the resulting sentence are only guessed due to lexical associations.

We may add now that Roman Jakobson's favorite idea about iconic character of language works in the sphere of word order much better than with intonation: falling and raising tones give way to one another in intonation contours without any direct relation to the meaning of the utterance.

References

[1] Russkaja grammatika. Moscow: Nauka, 1980.

[2] Kodzasov S.V. Intonacija voprositel'nyx predloženij: forma i funkcija. - In: Dialogovoe vzaimodejstvie i predstavlenie znanij. Novosibirsk, 1985.

[3] Chomsky N., Halle M. The sound pattern of English. N.Y. etc.: Harper & Row, 1968.

[4] Selkirk E.O. Phonology and syntax: the relation between sound and structure. Cambridge, Mass.: MIT press, 1984, p. 285.

[5] Nikolaeva T.M. Semantika akcentnogo vydelenija. Moscow: Nauka, 1982.

[6] Apresjan Ju.D. Tipy informacii dlja poverxnostno-semantičeskogo komponenta modeli smysl - tekst. - Wiener slavistischer Almanach, Sonderband I, 1980, p. 51.

[7] Paducheva E.V. Vyskazyvanie i ego sootnesennost' s dejstvitel'nost'ju. - Moscow: Nauka, 1985.

[8] Zolotova G.A. Kommunikativnye aspekty russkogo sintaksisa. - Moscow: Nauka, 1982, p. 289.

ON THE TONAL AND PROSODIC ACCENTS OF THE BALTIC SEA AREA

Kalevi Wiik

University of Turku
Finland

The language area that can be called the Baltic Sea Area includes three main groups of languages: Scandinavian, Baltic, and Finnic. One phonetic feature that seems to be more or less common to many of these languages is the "singing" quality of speech; four of the languages are generally considered to be tone languages with the opposition between two tonal accents (Norwegian, Swedish, Latvian, and Lithuanian), while two of them (Danish and Livonian) are stød languages. Even in those languages that are not tone or stød languages (e.g. Estonian and Finnish) there are dialect differences that can be explained by referreing to tonal curves.

I first give an inventory of the main types of tonal accents of this area. Next I give the geographic distribution of the main tonal and prosodic types. Thirdly, I consider the origin of the various types. It is my purpose to present a more or less uniform explanation for the rise of the støds in Danish and Livonian and the tonal accents in Swedish and Norwegian. I also give an explanation for the rise of the Estonian and Livonian syllable gradation with the various tonal curves associated with the strong and weak grades. Finally I ponder on the influence that the languages of this area may have exerted on each other. Many of the dialect differences can be explained by language contacts.