ASSIMILATION AS A CONTINUOUS COARTICULATORY PROCESS: FIRST ARTICULOGRAPHIC RESULTS

B. Pomplno-Marschall

Institut für Phonetik und Sprachliche Kommunikation
der Universität München, Germany.

ABSTRACT
Alveolar to velar assimilation in stop place of articulation was studied in German utterances by spectrographic analysis and by recording tongue movements with an electromagnetic articulograph. Spectral and positional differences between the stop consonants and between different speaking rates were analyzed. Increased speaking rate for the alveolar stops clearly resulted in positional changes of the tongue towards a position appropriate for velar stops but still significantly different from the latter. This seems to be in accordance with the view that assimilation is a continuous coartulatory process.

1. INTRODUCTION
In 1988 Nolan [2] reported an EPG study on English alveolar to velar assimilation of stop place of articulation in utterances of the form "...bed girls..." (vs all velar "...beg girls..."). Assimilation resulted in EPG patterns with reduced as well as with a total loss of alveolar contact. This finding seemed to be in accordance with the view that assimilation is rather a continuous coartulatory process than a process of featural change (cf. also [1]). Since despite the lack of alveolar contact even those 'weakly' assimilated items could be discriminated from the all velar utterances better than chance with the following pilot experiments we wanted to study the differences in tongue movement in alveolar to velar assimilation.

2. PROCEDURE
Two male German speakers read ten utterances of the type "Wir wollen zu Bett gehen" (vs all velar "Wir wollen zu Beck gehen") ten times at two different speaking rates (normal/fast) in randomized order. Besides the audio signal tongue movements were recorded with the help of an electromagnetic articulograph (AG100, Carstens Medizinelektronik; cf. [3]) via three coils placed on the midsagittal line of the tongue: (1) as far back as possible (back coil), (2) ca. 0.5 cm behind the tip of the tongue (front coil), and (3) midway between the others (mid coil).

Besides spectrographic analysis of formant frequencies in the middle of the preceding vowel and at implosion of the alveolar/velar stop, and of acoustic segment duration, the position of the three coils (as their x/y-coordinates on the midsagittal plane) were determined at the following points in time: (1) in the middle of the preceding vowel, (2) at the beginning of stop closure, (3) at the first stop release (if present1), and (4) at the second stop release.

3. RESULTS
In contrast to the English study the auditory analysis of our data revealed that there is only a weak tendency for assimilation in this German material constructed in parallel. One of the speakers almost never produced perceivable assimilations. For the numerical analysis we therefore chose the item with the most occurrences of assimilation, i.e. "Er wird es bald kriegen" (vs all velar "Er wird das Balg kriegen"). The mean coil position at the four points in time for both subjects and both places of articulation are seen in Figure 1. The differences in coil placement for the two subjects are clearly seen besides the differences in stop place of articulation.

For the statistical the analysis the positional data as well as the measured frequencies for the second and third formants at implosion were subjected to separate two-factorial analyses of variance (with the factors speech rate and place of articulation). There was no influence of speech rate on the formant frequencies at implosion for either speaker. The only significant effects were a higher F2 (p < .001) and a lower F3 (p < .001) for velars. The positional data showed significant effects only for one speaker (S1). His mean coil positions and standard deviations are shown in Table I and II. The differences in tongue contour are also shown in Figure 2.

The analyses of variance showed significant interactions between the factors speech rate and place of articulation only for the y-position of the mid and back coil at the first stop release: while for alveolar/velar utterances at fast rate of speech coil 2 is on the average 4.7 mm higher (p < .001) than in the case of normal rate of speech, no such rate effects are seen for the all velar utterances. A parallel effect is seen for the y-position of the back coil at the first stop release: the value for the fast alveolar utterance is on the average 3.9 mm higher than for the utterance at normal rate. At the same time the differences between fast alveolar/velar and all velar utterances always remain significant (p < .01; .001).

4. DISCUSSION
These results (cf. Figure 2) can be summarized as follows: Whereas tongue position
at implosion and the first stop release does not change dramatically with speaking rate for all velar utterances, the back of the tongue clearly adopts a higher position in the fast alveolar/velar utterances. But on the other hand this higher position does not reach the configuration of the all velar utterances. This clearly seems to be in accordance with the view that assimilation rather is due to a continuous coarticulatory process than a process of featural change.

5. FURTHER EXPERIMENTS
In a second pilot experiment these effects were studied in more detail with another male German subject. Here, additionally, we wanted to study the influence of context on alveolar/velar assimilation: Besides preceding /_ald/g/ as in the experiments above ("bald/Balg") simple /_ad/g/–endings ("Tag/Tag") were used with following accented vs unaccented /ge/ ("geben" vs "gestanden" or "gehalten"; accented syllables bold).

6. REFERENCES

Table I:
Mean coil position at implosion (a) and at stop release (b) in cm (1st line: x; 3rd line: y; 2nd/4th line: standard deviations) at slow rate for S1 (a: alveolar, v: velar)
(a) coil
1 2 3
7.36 7.34 9.84 9.87 10.74 10.90
.13 .10 .10 .15 .08 .15
18.23 17.87 18.99 19.30 18.17 18.57
.21 .26 .15 .10 .10 .16
(b) coil
1 2 3
7.24 7.24 9.67 9.60 10.61 10.76
.10 .09 .08 .15 .07 .11
17.91 17.47 18.81 19.37 18.26 19.00
.16 .22 .13 .10 .13 .12

Table II:
Mean coil position at implosion (a) and at stop release (b) in cm (1st line: x; 3rd line: y; 2nd/4th line: standard deviations) at fast rate for S1 (a: alveolar, v: velar)
(a) coil
1 2 3
7.36 7.40 9.81 9.76 10.71 10.84
.05 .16 .11 .13 .12 .11
18.29 17.87 19.10 19.46 18.23 18.83
.13 .24 .12 .10 .11 .11
(b) coil
1 2 3
7.17 7.37 9.53 9.64 10.59 10.81
.08 .16 .03 .13 .07 .16
18.06 17.49 19.29 19.49 18.64 19.06
.19 .26 .12 .07 .05 .11
AUTOMATIC PROCEDURE FOR LARYNGOGRAPHIC (Lx) ANALYSIS OF PHONATION CONTRASTS

J. H. Estling**, B. C. Dickson*** and J. R. Woolsey*

**University of Victoria, Canada
***Speech Technology Research, Ltd., Victoria, Canada

ABSTRACT

This report describes an automated, microcomputer-based procedure for comparing laryngographic (Lx) waveforms. The program enables researchers to analyze the cycle-by-cycle changes in vocal fold vibratory characteristics that may signal linguistic contrasts or differences in long-term voice quality. The first differential of the original Lx signal, captured digitally, is marked to indicate beginning, ending, upper and lower limits. A set of ratios is then obtained relating increasing voltage to decreasing voltage for each period of the signal. Considerations such as the inherent variability of the Lx signal, techniques of Lx recording, and applications of the algorithm are discussed.

1. LARYNGOGRAPHIC ANALYSIS

Electrical impedance laryngography has been used in phonetic research to quantify differences between contrasting types of phonation [4]. Such contrasts appear linguistically in languages like Korean at the syllable level in conjunction with phonologically distinct manners of consonantal articulation [1] [6] [7] [8] [9]. Phonatory contrasts also appear as long-term postures in voice quality with largely indexical significance [3, 10]. One problem in the analysis of the larynx waveform (Lx) has been the highly variable data that it yields. The signal is obtained by means of superficial throat electrodes which measure decreasing impedance as the vocal fold mass comes together, and increasing impedance as these structures separate [5]. Different models of laryngograph and differing recording procedures result in Lx signals with varying phase characteristics. This makes it difficult to analyze characteristics of individual waveform periods to distinguish, for example, a breathy voice from a harsh voice. Another problem is the DC float that characterizes many Lx signals and which makes establishing a baseline for reliable measurement of individual period characteristics particularly difficult. Aspects of obtaining an initial, workable Lx signal are dealt with in section 2. A solution to the baseline problem is presented in section 3. The method of segmenting the waveform to obtain a ratio is presented in section 4.

2. RECORDING Lx

Recordings of Lx signals made on a standard AM tape recorder tend to be distorted by phase shift. For this reason, it is valuable to develop procedures for direct digitization of the Lx signal, using an adequate (16-bit) data acquisition system. However, if this is not possible, the signal may be recorded using a system that does not introduce phase distortion, such as a Sony PCM digital audio processor and recorder, as has been used for these experiments.

Attempts at controlling DC float in the Lx signal include utilizing that proper coupling with the input preamplifier or similar analog conditioner is maintained. However, low-frequency oscillations can be expected as a result of laryngeal movement around the axis of the electrodes. This is more apparent during continuous speech than during examples taken from sustained vowels, owing to the natural raising and lowering of the larynx in the less controlled situations. The polarity of the Lx signal must also be considered. When the signal is taken directly from the laryngograph, the high impedance component of the signal is converted to the maximum positive voltage in the waveform, while the low impedance component results in a negative voltage. As the low impedance component is a result of maximum current flow across the electrodes, this will occur when the glottis is in its maximally closed phase of the voicing cycle. Because it is more representative of laryngeal behaviour to display the closed phase on the positive side of the waveform, we prefer to invert the polarity of a signal that has been taken directly from the laryngograph. However, if the signal is passed through a preamplifier at any stage, this will result in the polarity being reversed.

3. DATA ACQUISITION

Data acquisition is carried out using the CSL digital signal processing system [2], operating on an IBM-AT workalike. Data acquisition is performed at a rate of between 10K and 40K samples/second and the resulting sampled data files are passed to the EDIT320 software package. In that package, the waveform is displayed graphically and manipulated to enhance the laryngeal characteristics of interest (see FIG. 1).

The first differential of the waveform emphasizes the change in voltage over time, thus providing a representation of the high signal that closely models significant changes in current (as the impedance changes from, e.g., high impedance during the open phase of the laryngeal cycle to low impedance during the closed phase) as in equation (1).

\[ \frac{dy}{dt} \sim y_i - y_{i-1} \]

where \( i = \{1, 2, 3 \ldots n \} \) sampled data points

A side effect of taking the first differential is that low frequency oscillations attributed to larynx movement, as well as DC float, are eliminated.
4. PROCESSING ALGORITHM
To derive the ratios for each Lx period of a range of voiced speech, the original Lx signal is loaded using EDIT320 and flipped if necessary (the sign is changed on each amplitude value) depending on recording conditions, and the first different threshold is calculated. A minimum positive threshold and a minimum negative threshold are then selected, using horizontal cursors as illustrated in FIG. 2, to eliminate the effect of arbitrary zero crossings. For each period in the marked range, the greatest negative excursion (i1) that is less than the negative threshold and the greatest positive excursion of the period (i2) that is greater than the positive threshold are identified. A third value (i3) is defined as the subsequent greatest negative excursion below threshold, beginning the following period. A ratio is then calculated for each Lx period as shown in equation (2).

\[
\frac{(i_2 - i_3)}{(i_3 - i_2)} \]

where i = the selected sampled data point

For each succeeding period, the previous i3 becomes the new i1, until the end of the range is reached. The resulting ratios are stored in a file; as shown in FIG. 3, computed for a portion of the differentiated signal for harsh voice.

5. APPLICATIONS
Applications of this analysis algorithm focus on the identification of phonatory differences at the segmental, CV, or long-term level. The hypothesis that a distinctive 'breathy' phonatory quality is associated prosodically with the liens (vs. aspirated or foris) consonant series in Korean as a principal cue in identifying meaning in CV sequences, for example, can now be tested. Lx rise-time to fall-time ratios of sets of controlled phogetic models can also be compared with specific language data or with examples of phonation in pathological speech.

Initial examination of phonation types using this procedure illustrates that harshness and creakiness, which have low Lx ratios, differ from modal voice, and from whisperiness and breathiness, which increase progressively in Lx ratio range, as predicted in prior research [3].

REFERENCES
ELECTROPALATOGRAPHY OF CONVERSATIONAL SPEECH

L. Shockey

University of Reading, U.K.

ABSTRACT

Electropalatography was used to sample natural conversational English. A tabulation was made of cases where alveolar obstruents could occur and of how these underlying consonants were realised. The results reflect large scale reduction of alveolars in conversational speech, some of which (e.g. reduced lateral contact) seem to be common to all members of the set and some of which are more particular to the class of speech sounds involved (lateralls, nasals, stops, fricatives).

1. INTRODUCTION

Until recently, little or no research using electropalatography was focused on tongue-palate contact during relaxed, unselfconscious speech such as that which we use in everyday discourse. The reason for this is presumably the unease which besets phoneticians when they think about doing research on non-laboratory speech: in collecting free conversation, one cannot control for any of the variables known to influence articulation, among them segmental environment, stress, place in utterance, and word class. In addition, one never knows how many tokens of a given type will appear on any particular occasion, thereby making it hard to apply standard statistical measures to the results. Yet, surely if our goal as linguists is to model speech as it is used by ordinary people in daily life, it is vital to develop techniques for collecting and analysing data about this type of speech. Electropalatography provides an indirect but dynamic picture of articulator movement and as such is an invaluable adjunct to auditory and acoustic analysis of natural speech.

2. EXPERIMENTAL METHOD

In this study, acoustc and EPG data were collected from two subjects involved in conversation. The subjects were both longterm EPG users, having been on the team which developed the system currently in use at Reading University. They reported feeling very comfortable wearing the palate and experiencing no interference with articulation. Each of the subjects was seated comfortably in a small room and asked to talk to another member of the research team whose speech was not being monitored. The experimenter was in an adjoining room, listening to the conversation. After an initial period during which the conversants seemed to have become involved in discussion and to be producing unselfconscious output, the experimenter collected three-second samples of acoustic and EPG data. The acoustic signal was sampled at 10KHz and the EPG output at 100Hz. One minute of speech was collected from subject WJ, a West Midlands speaker with considerable Standard Southern overlay and 1.5 minutes collected from subject FG, a Standard British Southern speaker.

An impressionistic phonetic transcription of the collected corpus was done as well as a phonemic transcription. A tabulation was then made of cases of /t,d,s,z,n,1/ (the alveolar consonants involving contact in English), and each phonemic form related to both its phonetic transcription and the span of 10-millisecond EPG patterns which corresponded to it. The phonemic category provided a list of places where it would in theory be possible to find a maximally-articulated alveolar consonant; the phonetic realisations were divided into three categories: complete closure, incomplete closure, and deletion. These are very crude divisions. Complete closure was defined as the case in which every column of the palatogram indicated contact in at least one of the first four rows. Many kinds of complete closure were noted. For example, several degrees of lateral contact could be seen for everything except [l]: some showed a great deal of lateral contact, presumably indicating a high tongue position. Less side contact was visible in others, suggesting a laxer closure. The tokens with weak lateral contact were very common: this may prove to be a predictable feature of English conversational speech.

3. RESULTS

Not all underlying alveolars were fully realised, and in a pattern which was relatively similar from speaker to speaker. Table 1 shows summary data averaged over all consonants for each speaker and for both speakers combined.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>all alveolars</td>
</tr>
<tr>
<td>complete</td>
</tr>
<tr>
<td>incomplete</td>
</tr>
<tr>
<td>deleted</td>
</tr>
<tr>
<td>glottailed</td>
</tr>
<tr>
<td>total</td>
</tr>
</tbody>
</table>
Characteristic realisation patterns emerged for each manner of articulation:

1. /n/ -- Reduction of /n/ can be attributed to two main factors, a) a Vn sequence is often reduced to a nasalised vowel before another alveolar consonant, and b) [n] often shows incomplete closure intervocally.

In addition, [n] shows, in common with most of the other consonants investigated, a tendency to be articulated with a central groove before a fricative. It is a well-established tenet of phonetics that the production of the near-closure for a fricative involves finer motor control than the (theoretically) complete closures found for stops and nasals. Electropalatograms show that preparation for the groove configuration begins in preceding alveolar consonants and can sometimes be detected in vowels preceding such clusters.

2. /l/ -- In these subjects, there were two distinct realisations of /l/. One involved contact with the palate and was found syllable-initially, at the trailing end of a cluster, and intervocally. The other involved no contact and was found at the leading end of a cluster and finally. The light or "semivocalised" closure which was noted by Hardcastle and Barry [1] in some environments was not found to be characteristic of these subjects: subject FG showed four anomalous cases, but these were a very small proportion of the total.

3. /s/ and /z/ -- These sounds tend to be preserved in some form, but (as mentioned above) often get a very wide channel in these data, implying (in agreement with the lateral contact discussed above) less raising of the tongue toward the palate than is found in citation forms.

4. /d/ -- A fully closed [d] is normally found after another non-nasal alveolar, especially word-finally when the next word begins with a vowel. The closure tends to weaken intervocally, even if the [d] is word-final. (The resulting segment does not sound like a fricative or look like one on an acoustic display. This is presumably because there is little or no airflow through the constriction). [d] is especially prone to deletion in the environment n\_C.

5 /u/ -- Fully-articulated tokens tend to be found syllable-initially, especially word-initially and especially in stressed syllables. After the alveolar nasal or fricative and intervocally, [l] can be either fully closed or incomplete. No closure is normally found in the environment C\_C.

For both speakers, /l/ was usually realised as a glottal stop in the environment V\_C and in absolute final position.

4. DISCUSSION

Let us return briefly to the notion of a normal or target articulation. While it is clearly desirable for all speakers to be able to produce a maximally-differentiated set of alveolars in citation-form words in a laboratory situation, it seems obvious from the above that less fully realised tokens are very much a part of conversational speech and are in themselves normal. The implication for those using EPG didactically is obvious: it would be excessively demanding and in some sense even incorrect to expect maximally differentiated tokens of most alveolar consonants (in some environments) in unselfconscious speech. Variation in production which comes about not only through coarticulation with surrounding segments but also through position in the linguistic unit (syllable, word) and position with respect to stress must be taken into account. There might also be a generally lower longterm jaw/tongue setting in conversational speech, which leads to less side contact and bigger fricative grooves, and may be one of the reasons for the observed incomplete closures. (See [2] for further discussion of this question).

The latter point must be reiterated with respect to general phonetic theory: these data provide further evidence for the assertion that the physical properties of the vocal tract alone cannot account for the patterns of reduction we find in conversational speech. An /nu/\# sequence behaves very differently from an /nu/\#u/ sequence with respect to reduction: it is the higher-level linguistic construct which determines the possibility of phonetic variation, though the construction of the vocal tract is one of several factors which determine the nature of the variation.

BIBLIOGRAPHY


PHONOLOGICAL DISRUPTION IN WORD PRODUCTION

Gregory V. Jones

University of Warwick, Coventry, England.

ABSTRACT
In naturally occurring speech, people occasionally find that they have a word "on the tip of the tongue". In this state, they may produce other words related in either sound or meaning to the targets. Are these other words instrumental in causing the TOT states, or are they merely by-products of the TOT states?

1. INTRODUCTION
In spontaneous utterances, most people occasionally experience difficulty in producing an intended word. In this state, a person may be confident that the word he or she wishes to generate is within his or her mental lexicon. The word nevertheless remains temporarily unavailable, seemingly "on the tip of the tongue". While people are in this tip-of-the-tongue (TOT) state, they often do not remain mute but instead produce words other than the target-word at which they are aiming. Such words have been termed "interlopers" [5,6]. An early example was reported by the writer George Lewes, partner of the novelist Mary Ann Evans (George Eliot), as follows.

I was one day relating a visit to the Epileptic Hospital, and intending to name the friend, Dr. Bastian, who accompanied me, I said, "Dr. Brinton;" then immediately corrected this with, "Dr. Bridges," - this also was rejected, and "Dr. Bastian" was pronounced. I was under no confusion whatever as to the persons, but having imperfectly adjusted the group of muscles necessary for the articulation of the one name, the one element which was common to that group and to the others, namely B, served to recall all three [7, p. 128].

Lewes's observation was discussed widely, for example in France by Ribot [11, p. 19] and by Binet [1, pp. 113-114]. However, greater generality was clearly to be obtained by the collection of a corpus of such observations. Early corpora were assembled by Woodworth [14] and Wenzl [12,13]. More recent corpora have been described by Reason and his colleagues [9,10], Cohen and Faulkner [4] and Burke, MacKay, Worthley, and Wade [3]. In all of these studies, a considerable number of TOT states were found to be characterised by the occurrence of interlopers that were related to their respective targets in either their sound or their meaning. The nature of the empirical stochastic contingency between relatedness in sound and relatedness in meaning of an interloper to its target is still, however, unclear. For this reason I have in a recent unpublished study collected a small corpus of naturally occurring TOT states.

2. TOT CORPUS
TOT experiences were collected from undergraduates at the University of Warwick over a period of several weeks. In this sample, the number of interlopers generated by the participants themselves (as opposed to those generated by bystanders) was 100. The interlopers were classified as being related both phonologically and semantically (PS), phonologically
alone (Ps), semantically alone (pS), or neither phonologically nor semantically (ps). The observed incidences were PS = 29, Ps = 3, pS = 67, and ps = 1.

A striking and unexpected aspect of the preceding results was that almost all (96%) of the interlopers were semantically related to their targets. At first sight, this result appears to conflict with the previous observation of many interlopers categorised as phonologically related to their targets [10, p. 124]. However, closer examination of the examples provided by Reason and Mycielska indicates that in each case their "most phonological pathways" display semantic relatedness also (e.g., target = pomander, interloper = pot-pourri).

3. INTERLOPER ORIGINS

What are the origins of the interlopers that commonly occur in TOT states? Two logical possibilities may be distinguished. The interlopers may arise either before or after the disruption in target word generation. In particular, the interlopers may either be instrumental in causing the disruption or be merely a consequence of the disruption. To use medical terminology, the interloper could be considered either as a pathogen (i.e., cause of disruption) or as a sequela (i.e., consequence of disruption).

In the case of words related in meaning, the Sequela hypothesis seems a priori plausible. Words produced in normal utterances are presumably selected largely on the basis of their meaning. Thus after a target word becomes unavailable, it might be expected that a person's attempts at word generation will yield other words which are related in meaning to the target. In contrast, the Pathogen hypothesis (that the interlopers themselves cause the disruption) seems implausible. It is obvious that other words related in meaning to intended target words are routinely generated in many normal meaningful utterances (e.g., consider the target word "water" in the sentence "The swimming pool water was chlorinated"). Since we generally have no difficulty in speaking such sentences, we may infer that target generation is not likely to be prevented by the activation of words related in meaning that act in a pathogenic manner.

In the case of interloper words related in sound to the target, it is in contrast difficult to establish their role by a priori reasoning. On the Sequela hypothesis, such interlopers might arise if it is the case that, for some unrelated reason, only a partial phonological specification of the target word becomes activated. Subsequently, other words sharing this partial specification might be generated as sequelae. Most would be expected to be also related in meaning to the target, since semantic factors would presumably remain important in guiding word production. On the Pathogen hypothesis, it might be possible for phonological interlopers to act as blockers. Perhaps a word which is similar in sound to the target word receives activation by chance shortly before generation of the target is completed, and acts as a phonological decoy receiving in sum more activation than the target itself. Again, this is clearly more likely to occur if the interloper and target are related in meaning as well as in sound.

4 SOME INTERLOPER EXPERIMENTS

How can one distinguish between the Pathogen and Sequela hypotheses for the origins of phonological interlopers in TOT states? Two recent studies [5,6; see also 8] developed further an experimental method of investigating the TOT state introduced by Brown and McNeil [2]. Brown and McNeil showed that reading people definitions of moderately rare words induces TOT states on the order of 10% of occasions.

In the new studies, people were again presented with definitions of moderately rare words, such as
"Something out of keeping with the times in which it exists". But now the definition was followed immediately by an interloper word also presented by the experimenter. Equal numbers of the four types of interlopers distinguished earlier (PS, Ps, pS, and ps) were used - that is, the interloper was either related both phonologically and semantically to the target, phonologically alone, semantically alone, or neither phonologically nor semantically, respectively. For the present example definition, the interloper was "abnormality". This was of the PS type since it was related in both sound (initial phoneme and number of syllables) and meaning to the target "anachronism".

It was found that interlopers which were related in sound to their targets were more likely to lead to TOT states, irrespective of whether they were related in meaning. This result is consistent with the Pathogen hypothesis since that hypothesis asserts that phonological similarity between interloper and target is instrumental in engendering TOT states, in contrast to the Sequela Hypothesis's assertion that the interloper is merely a by-product of naturally occurring TOT states. Nevertheless, considerable empirical work remains to be done to examine further the effects of artificially supplied interlopers, and in particular more extensive work with a wider range of experimental materials is needed.

5 REFERENCES
ARTICULATORY-ACOUSTIC CORRELATIONS IN THE PRODUCTION OF FRICATIVES

N. Nguyen-Trong, P. Hoole, & A. Marchal

1. CNRS, URA 261, Univ. Provence, Aix-en-Provence, France
2. Institut für Phonetik, München Univ., FRG

ABSTRACT
This work is aimed at exploring the relationships between a set of articulatory parameters, and the acoustic output, in the production of French fricatives /s/ and /f/. More specifically, we attempt to find out whether the dimensions of maximal contrast among the fricative spectra, are correlated with movements of lingual transducers monitored by means of an electromagnetic (EMA) system. Our results show that the EMA measurements can be considered to be very reliable. It appears that the spectra can be regenerated with a good accuracy from these measurements, with the help of a statistical method the advantages of which are pointed out. In conclusion, implications of this work in the domain of articulatory modelling are discussed.

1. INTRODUCTION
In a recent work [3], Hoole et al. have shown that the EPG tongue-contact patterns in the production of the fricatives /s/ and /f/ in English, were strongly correlated with a set of acoustic parameters extracted from the corresponding spectra by means of a factorial analysis. It has appeared that this relationship was close enough to allow a prediction of the acoustic data from the EPG data, with the help of a multiple linear regression. The results supported the conclusion that an empirical investigation of this kind, was suitable for providing information on the articulatory-acoustic correlations, which could be fruitfully incorporated into a model of fricative production [2,5]. The present experiment was based on the same methodological principles, and was aimed at investigating in a more extensive way two specific points. First, the question could be raised to know whether it is possible to relate the articulatory parameters with the spectra themselves by regenerating these spectra from the acoustic factors. Second, it seemed important to compare the results of the multiple linear regression, with those of a method giving the possibility to detect non-linear articulatory-acoustic relationships similar to the ones which are described in Stevens’ quantal theory [6].

2. MATERIAL
The experiment has been carried out at the Institute of Phonetics of Munich University. It consisted of an audio recording synchronized with a parallel EMA tracking. The electromagnetic system used is a commercially available device (Articulograph AG 100, Cawana, Göttingen, FRG) which has been recently described in [4,7], and which allows monitoring of articulatory movements with the help of five electromagnetic transducers (coils). In the present experiment, three coils were attached to the mid-line of the tongue, one was attached to the lower incisors, and one reference coil to the upper incisors. The tongue board was placed so far back as possible, the frontmost coil about 1 cm back from the tip, and the third coil in between these two. The output of the EMA system was digitized (sampling rate 200 Hz) and transferred online to a PC AT-386 computer where software compensation for the effects of possible tilt of the receiver coils was applied. The digital signal, which represented the displacement of each of the five coils in the x-y plane, was finally stored on a hard disk. The audio signal was recorded by means of a B&K microphone on a DAT recorder, digitized on a LSI 11/33 compact (sampling rate 16000 Hz, LP filtered at 7500 Hz), and aligned with the EMA signal thanks to a set of synchronization pulses recorded on the second track of the DAT tape. The estimated accuracy of the alignment was +/- 3 ms. The speech material consisted of the following combinations: /s,f,s/s,f,s/, /s,s/, /s/s/s/, /s/s/s/ embedded in 9 sentences which have been pronounced from 8 to 9 times by two male native speakers of French (AM, NN). In this paper, results will be presented for speaker AM.

3. ANALYTICAL PROCEDURES
3.1. EMA measurements
To minimize the variations in the articulatory signal which could have been generated by any head movements, the coordinates of the coil attached to the upper incisors have been subtracted from those of the other coils. Moreover, for each repetition, the whole cloud of data has been rotated around the origin in the x-y plane, so as to achieve a vertical orientation of the first principal axis of the jaw movement. Figure 1 displays the positions, averaged over all the repetitions, of the tongue-base, tongue-mid, tongue-tip and jaw coils, at the mid-point of the fricative (which has a coordinate on the time axis determined with the help of the acoustic signal) for each item.

3.1. Acoustic analysis
The acoustic analysis consisted in calculating an FFT spectrum within a 32 ms Hamming window centered at the fricative mid-point. This spectrum was next reduced to 21 components by averaging the spectral energy over 1 Bark intervals from 0 to 8 kHz. Moreover, the information below 8 Barks has been ignored, in such a way that the acoustic data were finally made up in the present experiment of a set of 13-dimensional vectors. For reasons that are given below, we have chosen to proceed to a new data reduction by means of a principal-components analysis, which proved that 4 linear combinations of the 13 original parameters could account for more than 90% of the variance among the spectra. It has appeared that the 1st 2 factors were sufficient to differentiate the fricatives /s/ and /f/ from each other. Factor 1 is interpretable as a dimension of average energy; factor 2 can be considered as underlining an opposition between the spectra which show a local maximum within the 12-15 Barks range, and those in which the energy is relatively higher above 15 Barks.
4. RESULTS

4.1. Empirical regression of the acoustic factors from the EMA data

One approach to exploring EMA-acoustic relationships consists of analyzing separately the way in which the data are distributed in the EMA space, and in the acoustic space, to examine whether the sources of variation can be considered as being the same in the two cases (pronounced constant, immediate context, carrier sentence, etc.). But it is also possible to check the existence of such relationships, by attempting to predict the acoustic parameters from the articulatory ones. At the present time, the information obtained with EMA on the mid sagittal section of the vocal tract, doesn’t allow calculation of the area function required by a standard acoustic model to synthesize the output signal. In this experiment, the predictions have been based on the so-called statistical regression, which has been performed in two different ways, since we have compared the results of a classical, multiple linear regression, with those of an empirical, non-linear variant [1]. In the second case, the predicted value of a given acoustic parameter \(y\) for a given articulatory input (a tongue profile composed of 3 points) was simply defined as the \(y\) mean value associated with the input k-nearest neighbors in the articulatory space (k being determined by the user). It can be easily shown that the regression achieved by means of a local approximation, is suitable for modeling non-linear relationships, between any number of independent variables and the to-be-predicted one.

The calculations have involved the articulatory and acoustic data relative to all the VC(W) sequences recorded by speaker AM. The prediction quality has then been assessed on a test set which was composed of the tongue «profiles» averaged over all the repetitions for each of the two consonants, in each possible context. In Table 1 are given the \(r^2\)’s corresponding to the correlation between the measured and the predicted values for factors 1, 2, 3 and 4. It appears clearly that the \(r^2\)’s are high whatever regression technique is used, and reflect a close relationship between the EMA parameters and the dimensions of maximal variance among the spectra, while the empirical regression (referred to as REG.2 in Table 1) produces results which are systematically superior to those of the multiple linear regression (REG.1). The number of neighbors, on the basis of which a value has been given to each tongue profile for each of the 4 factors in REG.2, was fixed to 5.

4.2. Regeneration of the original spectra from the acoustic factors

The fricative spectra have been finally regenerated from the output parameters of the empirical regression, through the usual operation based in the present case on the eigenvectors of the covariance matrix relative to the 14 original acoustic variables [8]. Figure 2 proves that the accuracy of this regeneration is quite good (average \(r^2\) between the measured and predicted components for each of the spectra in the test set \(>0.9\)). It is especially encouraging to note that the shape shown by a spectrum is restored in a very satisfying way.

6. CONCLUSION

Our results could be explained in part by the fact that speaker AM has pronounced /s/ and /z/ in a rather stable way across repetitions and across contexts. Consequently, the variability among the spectra was likely to be accounted for by a relatively small number of factors which in return allowed to regenerate these spectra without any major distortion. It remains that the factors themselves (which can be considered as dimensions of maximal contrast among the acoustic data) have proven to be accurately predictable from the EMA measurements. Therefore, it can be said that under the conditions adopted in this experiment, the articulatory parameters extracted by EMA are closely correlated with the acoustic output.

The empirical regression is very interesting in that it gives the possibility to make predictions which have an accuracy calculated for each point in the articulatory space (while this calculation wouldn’t have much sense in the case of the linear regression, in which the criterion to be optimized concerns an average accuracy). From our point of view, this issue should give rise to a more systematic investigation. An experiment carried out according to the same methodological principles as those of the present work, on a more extensive material, would probably allow to determine whether a given articulatory neighborhood is stable (with respect to the acoustic output), or unstable. It seems to us that this kind of empirical exploration of the articulatory space, could constitute an quite interesting way to verify hypotheses on the mechanisms of speech production, such as the ones which are supported in the quantum theory [6].

7. ACKNOWLEDGMENTS

This work was supported by ESPRIT II/BRA n°3279 ACCOR.

9. REFERENCES


![Figure 2: average spectra for /s/ and /z/ displayed together with the output of the empirical regression from the corresponding EMA profiles.](image)
INITIALISATION, ARRET ET VARIATION DE FREQUENCE FONDAMENTALE DE LA VIBRATION LARYNGEE : ETUDE ELECTROMYOGRAPHIQUE

B. Roubeau, G. Dassau, J. Lacau et C. Chevrie-Muller

INSERM, Laboratoire de Recherche sur le Langage
Paris, France.

ABSTRACT
In order to analyze the mechanisms of fundamental frequency controls, the electromyogram (EMG) of four laryngeal muscles (Cricothyroid and three strap muscles) was recorded, as well as the acoustic signal, in two subjects (one male, one female) producing ascending and descending tones (gissandos). The relationship between EMG activity patterns and frequency variations were described; in addition, the specific patterns related to the gissando's beginning and terminal part were analyzed according to the different vocal events taken into consideration the EMG patterns of the four muscles were compared.

Les études électromyographiques des muscles laryngés corréllées avec la fréquence vibratoire sont nombreuses (2, 3, 6, 1). Celles consacrées aux "mouvements" de fréquence sont déjà plus rares de même que celles qui considèrent parallèlement l'activité de muscles intrinsèques du larynx et de muscles sous-hyoïdiens (3).

1. PROTOCOLE EXPERIMENTAL
Le protocole vocal comporte des variations de fréquences au cours de la réalisation sur la voyelle "O", d'une part, de glissandos ascendant puis descendant, et d'autre part descendant puis ascendant sans contrainte de hauteur limite ni d'intensité. La seule contrainte imposée aux sujets est le maintien de la tête dans une position aussi fixe que possible. Au cours de nos expériences, l'activité de 4 muscles a été enregistrée :
- le cricothyroïdien (CT), le thyroïdien (TH), le sternothyroïdien (ST) et le sternocleido-hyoïdien (SCH).

La technique EMG employée est celle décrite par Hirose (5) utilisant des électrodes bipolaires implantées à l'aide d'aiguilles intramusculaires. Des tests non vocaux tels que la déglutition, l'ouverture de la bouche et l'inclinaison de la tête, permettent de vérifier l'emplacement des électrodes. Ces tests sont répétés plusieurs fois au cours de l'enregistrement, afin de vérifier le maintien en place des électrodes.

Les résultats provenant de deux sujets (une femme et un homme) parmi les 6 qui ont participé à l'expérience ont été retenus pour la stabilité des tracés fournis par les 4 muscles explorés et seront présentés ici.

2. RESULTATS ET DISCUSSION

Figure 1 : Tracés EMG redressés des 4 muscles explorés ainsi que le signal acoustique, le signal électroglottographique (EGG) et la courbe mélodique a) Glissando descendant et descendant, b) Glissando descendant puis ascendant.

Figure 2 : Glissando ascendant puis descendant, sujet masculin, signal intégré.

On note une activité importante des sous-hyoïdiens lors de la réalisation des fréquences les plus basses au début et à la fin de la production, de même qu'une activité importante du CT précédant la réalisation des fréquences les plus élevées. Les pics d'activité musculaire précédant les points d'inflexion de la courbe mélodique ont été relevés et moyennés. Pour le sujet masculin (DA), les activités EMG sont moyennées pour des valeurs de F0 inférieures et supérieures à 200 Hz, et inférieures et supérieures à 300 Hz pour le sujet féminin (AH).

Figure 3 : Activités musculaires moyennes aux points d'inflexion de la courbe mélodique.

Ces résultats montrent l'importance de l'activité du CT lors de la réalisation des fréquences élevées et des sous-hyoïdiens lors de la réalisation des fréquences basses.

Les courbes d'activité des muscles en fonction de l'évolution dynamique de la fréquence ont été réalisées à partir de moyennes effectuées sur 34 productions pour le sujet féminin AH et 28 pour le sujet masculin DA.

Ces valeurs sont relevées après élimination des phénomènes d'approximation.
L'activité participation considérer la bilélevées de lors muscles.

Figure 4 : Glissando ascendant et descendant, activités moyennes des muscles.

1 - L’activité du CT est corrélée avec l’évolution de la fréquence.

2 - Au fur et à mesure que la fréquence fondamentale augmente, l’activité des sous-hyoïdiens diminue puis s’accroit à nouveau lors de la réalisation des fréquences les plus élevées.

3 - De même, les glissandos descendants mettent en évidence le cours de la réalisation des fréquences élevées une activité non négligeable des sous-hyoïdiens. Celle-ci s’accroit considérablement lors de la réalisation des fréquences les plus basses. Il faut évidemment considérer le fait que l’anticipation de l’activité pour la réalisation d’une fréquence donnée est ici difficile à évaluer. L'activité du CT apparaît corrigée de manière relativement stable avec la fréquence fondamentale (glissandos ascendant et descendant sont assez bien symétriques). Il n’en n'est pas de même pour les sous-hyoidiens dont l'activité semble non seulement liée à la valeur de la fréquence vibratoire, mais aussi au sens d'évolution de celle-ci (le niveau général de l'activité est plus important au cours d'un glissando descendant qu'au cours d'un glissando ascendant).

Les deux catégories de muscles ne semblent pas agir ici suivant des processus dynamiques simplement antagonistes. Les sous-hyoidiens seraient activés de façon plus massive dans les phénomènes descendants tandis que l'activité du CT présente un caractère plus tonique finalement corrélé avec l'évolution de la fréquence.

Les pics d'activité musculaire précédant l'attaque ont été moyennés et regroupés en fonction de la hauteur du son au moment de celle-ci (cf. Fig. 1).

Figure 5 : Activité EMG moyennée lors de l'initialisation de la vibration.

Lors des attaques dans le grave, l'activité du CT est plus importante que lors de la réalisation de ces mêmes fréquences en cours d'émission (cf. Fig. 4). De même, lors des attaques aigües, l'activité des sous-hyoidiens est plus importante que lors de la réalisation de ces fréquences en cours de production. Bien que la fréquence à laquelle se produit l'initialisation ait une influence sur l'amplitude du tracé EMG (CT plus actif dans les attaques supérieures à 200 Hz et TH plus actif pour des attaques inférieures à 200 Hz), le phénomène d'initialisation lui-même provoque une globale

lisation de l'activité musculaire suivant un processus plutôt phasique.

Figure 6 : Activités EMG moyennées lors de l'arrêt de l'émission.

On note ici une différence de comportement des deux sujets. Chez AH, l'activité lors de l'arrêt de l'émission demeure liée principalement à la fréquence vibratoire telle qu'on l'observe en cours d'émission. Par contre, chez DA, les arrêts dans l'aigu indiquent une forte activité des sous-hyoidiens supérieure à celle que l'on observe en cours de production (cf. Fig. 4). Ce phénomène délicat à interpréter à partir de deux sujets peut être considéré comme une régulation de la fréquence ou des mouvements liés à son évolution grâce à un système d'"antagonisme" agissant sur les mobilisations générales du larynx.

Ces différentes observations sur l'initialisation et l'arrêt de l'émission mettent en évidence le caractère phasique de l'activité musculaire tandis que la régulation de la fréquence en cours d'émission est liée à une activité tonique, cette distinction est particulièrement nette au niveau du CT. Les différences entre les deux sujets semblent liées soit aux durées de réalisation des glissandos, soit à des caractéristiques de l'activité musculaire.

Dans tous les cas observés, les sous-hyoidiens semblent fonctionner en synergie. Il ne faut pas négliger dans cette interprétation, la possibilité d'un "parasitage" des signaux EMG entre eux du fait de la proximité des muscles.

Cette étude confirme l'importance de la considération simultanée des muscles intrinsèques et extrinsèques du larynx lors de l'étude des variations mélodiques et de la distinction des événements mettant en jeu des activités musculaires de caractère tonique ou phasique.

3. BIBLIOGRAPHIE


(2) GAY, T., HIROSE, H., STROME, M., SAWASHIMA, M. (1972) "Electro-myography of the intrinsic laryngeal muscles during phonation", Annals of ORL.


PHASE MODIFICATIONS IN TONGUE MOVEMENTS ACROSS SPEECH RATE VARIATIONS: INFLUENCE OF CONSONANTAL ARTICULATION.

C. Delattre & P. Perrier

Institut de la Communication Parlée - URA CNRS n° 368 Université Stendhal, Grenoble, France.

ABSTRACT
A corpus of CV1CV2 sequences is analysed in order to emphasize the concept of synergy among gestures in speech production. The vertical movements of tongue dorsum were measured for V1=[a] and V2=[u], and for two consonants [d] and [g] which production recruit respectively this articulator at two quite different levels. Results are interpreted in terms of changes of (1) the consonant-vocalic and (2) the peak velocity phasing, for raising and lowering gestures.

1. INTRODUCTION
Speech production implies a spatial and temporal coordination of different articulatory gestures. The concept of synergy, introduced by Haken [3] in the field of motor control, could be appropriate to characterize and predict some articulatory patterns of speech production. In their work, Kelso et al. [5] explained the jump from the production [pip] to the production [pip] as the speech rate strongly increases, within this theoretical framework. They introduced the concept of intergestural phasing, and identified this jump as an obvious phasing restructuration between lip and glottal gestures.

In the present study, we propose to analyse this synergy phenomenon in a quite different paradigm. The behaviour of the consonant-vowel phasing in CV1CV2 sequences is observed for consonants involving different articulators. This paper presents preliminary results for two consonants [d] and [g].

2. METHOD
2.1 Experimental procedure
The corpus studied is designed for the observation of the vertical movements of tongue dorsum. It consists in the repetitions of isolated productions [au], [dadu], [gagu], for which we assume that tongue dorsum movement is pertinent to describe the [a] to [u] articulation. The consonants [d] and [g] are chosen, because of their two extreme behaviours towards this articulator: [d] production does not recruit the tongue, whereas the [g] articulation recruit mainly the tongue dorsum. Thus, in a [gagu] sequence, the vocalic and consonantal gestures are produced with the same articulator.

The task consisted in 10 repetitions of each sequence, produced at normal and fast speaking rates by a French male speaker. Tongue dorsum displacement was monitored at an 1 kHz rate with a computerized ultrasound transducer system (see [4]). Simultaneously the acoustic speech signal was recorded at the same rate.

2.2 Data collection
As data are available for one articulator only, namely the tongue dorsum, the choice of temporal events defining the vocalic and consonantal phases was not obvious: the events related to consonantal gestures were detected on the speech wave. In all cases the reference is the underlying vowel-to-vowel movement, divided into two components, the raising gesture ([a]-[u]) and the lowering gesture ([u]-[a]). For the first component, the consonantal movement towards the occlusion is in the same raising direction as the vocalic movement; the release movement is in the opposite direction. In the second component ([u]-[a] transition), the relations are reversed. It is thus interesting to detect these two consonantal events (closion and release) inside the vocalic phase. For each vocalic transition ([a]-[u] and [u]-[a]) the boundaries of the vocalic phase correspond to the points of zero velocity of the tongue dorsum raising gesture (See Fig. 1). We define so for each movement (raising and lowering) one vocalic phase and two consonantal phases, the "occlusion phase" and the "release phase".

Fig. 1 shows ultrasound recorded movement and velocity profile of tongue dorsum and the corresponding acoustic speech signal, during the production of [gagu], at normal speaking rate. Duration A determines the raising gesture for the transition [a]-[u], in which a1 ("the occlusion phase") and b2 ("the release phase") are plotted in percentage. Duration B determines the lowering gesture for the transition [u]-[a], in which the same phases ("the occlusion phase=b1) and ("the release phase=b2) are plotted.

The second point of our study focuses on the kinematic changes of the raising and lowering gestures depending on the consonant and on the speech rate. In this aim and according to different studies that advocate the importance of velocity profiles in motor control analysis (see [1], [2] and [6]), the occurrence of the velocity peak between two successive points of zero velocity is measured (see Fig. 2).

3. PRELIMINARY RESULTS AND DISCUSSIONS
3.1 Consonantal-vocalic phasing
The data obtained for the raising gesture are presented in Fig. 3 and 4. The examination of the time proportion of the release phase for [dadu] (Fig. 3), reflects obviously no significant modification with changes in speech rate (mean values: 74.9 vs 74.2); on the contrary, for [gagu], we observe an obvious increase of this time proportion for the fast rate. Such a behaviour can easily be explained by a time constraint on the consonantal hold: this durational value must be sufficient for a good perception of the consonant; hence the vocalic durations are more affected by the change in speech rate than the consonantal one. This phenomenon becomes indeed more obvious, when vowel and consonant are produced with the same articulator. At the same time the proportion of the occlusion phase decreases in fast rate for both consonants (Fig. 4). Moreover, whereas the differences between [d] and [g] are not significant (α>0.10) at normal rate, they become highly significant at fast rate (α<0.01), which means that the articulation of the consonant induces different behaviours when speech rate increases. As the raising vocalic gesture ([a]-[u]) and the consonantal raising occlusion gesture occur simultaneously, the delay between the onsets of these two gestures tends to decrease significantly, especially when the same articulator is recruited. These results attest, in the case of a monoarticulator production, a tendency towards synchronization of the
two raising gestures, when the speech rate constraints are strong. Hence, it supports the idea of a synergetic production.

The same kind of observations can be made for the release phase in the lowering vocalic transition. For both consonants, the proportion of the release phase increases (Fig. 5), as the proportion of the occlusion phase decreases (Fig. 6). But these behaviours are hardly significantly different for the occlusion phase (α<0.05) and highly significantly different for the release phase (α<0.01). The delay between the onsets of the lowering vocalic gesture and the consonantal one remains important, due to the constraints on the consonantal hold duration; but a tendency towards synchronization of these two gestures could well furnish a reliable explanation for the more important reduction of the consonantal hold, in the case of a monoarticulator production. This phenomenon supports the hypothesis of synergy among consonantal and vocalic gestures.

3.2. Kinematic changes

This investigation is essentially based on the duration of the lowering and raising gestures in which the occurrence of peak velocity is observed (see Fig. 2). In both gestures, these acceleration phases are plotted in Fig. 7 & 8. At normal rate, and for both gestures, our results show an important dispersion of the data for [au] and [dau]; the constraints seem obviously stronger for the [gagu] production. For fast speech rate, the data converge towards the same value in all cases: the velocity profile tends to become symmetric as in optimized movements minimizing the jerk (see [7]). This variation is however less important in [gagu]. An increase in speech rate seems thus to produce an optimization of the coordination between vocalic and consonantal gestures. This optimization is already perceptible for [gagu] at normal rate.

3.3. First conclusions

These two kinds of data seem to attest the existence of synergy between consonantal and vocalic gestures when the same articulator is recruited: (1) the consonant-vowel phasing is specific for this kind of production; (2) in this last case, the kinematic properties reflect a tendency towards optimization. These results and conclusions are preliminary. A further study will be made with other consonants as [b], [k], to confirm the assumptions resulting from the observation of [d] and [g].

REFERENCES


AN INVESTIGATION OF A SPECIAL TYPE OF ACCENTUATION IN RIPIUARIAN DIALECTS BY COMPUTER SIMULATION OF SPEECH PRODUCTION

B. J. Kröger, G. Heike, C. Opgen-Rhein, R. Greisbach and O. Easer

Institut für Phonetik der Universität zu Köln, Köln, Germany.

ABSTRACT
Recent studies on "Rheinische Schärfung", an accentuation feature in Ripuarian dialects, rely mainly on perceptual or acoustic analysis. Here an attempt is discussed to model this predominantly phonatory phenomenon by coupling a two-mass-model of the glottis to the Cologne articulator speech synthesis system.

1. INTRODUCTION
The characteristic accentuation phenomenon of Ripuarian dialects, "Schärfung" ("sharpening"), is investigated by means of resynthesis. Hypotheses about the "sharpening" feature in the syllable [s:s] ("carriion") were tested by synthesis in two ways: by glottal abduction and by glottal adduction.

2. ACOUSTIC CUES AND A PRODUCTION HYPOTHESIS
"sharpening" is an accentuation phenomenon whose phonetic features characterize additionally the nucleus of stressed syllables. It occurs either in long vowels or in diphthongs or in short vowels followed by a sonorant ([hy:] "height", [zai] "sieve", [s:i] "carriion", [vo:]t] "rage", [al] "all", [hyan]"dogs", [tant]"aunt")

The major acoustic cues of "sharpened" vs. "unsharpened" long vowels before voiceless consonants are [1]:
(a) shortening of vowel duration;
(b) a zero-intensity interval between vowel and following fricative;
(c) a marked intensity decrease in the vowel segment;
(d) a marked decrease in fundamental frequency in the vowel segment.

Hypothesis: The "sharpening" is thought to be related to phonatory rather than to supraglottal articulatory activity [2]. There are in principle two maneuvers which lead to the acoustic features in question: a glottal abduction gesture and a glottal adduction gesture. They both begin in the middle of the vowel and last to its end.

3. THE SYNTHESIS MODEL
The articulatory speech synthesis model used was developed at the Institute of Phonetics, University of Cologne [3,4,5]. On the segmental level phonetic segments are put in. The model generates a set of articulatory control parameters (e.g. tongue height, tongue position, jaw opening) and three phonatory control parameters (lung pressure, cord tension, abductor pressure).

4. "SHARPENING" IN THE PRODUCTION MODEL
Fig. 1 shows the pattern of the most important control parameters for [s:i] for glottal abduction and glottal adduction. The vocal tract constriction VTC is given by the control parameter tongue tip height. The other control parameters are: abduction area, cord tension, lung pressure (alveolar pressure) PA1V.

In the case of abduction the abduction area becomes negative at the end of the vowel, and medial compression is produced. At the same time, cord tension and lung pressure decrease, resulting in the decrease of FO and signal intensity.

In the other case "sharpe-
ning" can be modelled by a slow glottal abduction gesture while the vocal tract remains unconstricted. In addition q must be varied, while \( p_{1v} \) remains constant.

Fig. 2 shows a naturally spoken syllable and two synthesized ones. The latter were produced by the control parameter patterns in Fig. 1; they both show the typical decrease in intensity and fundamental frequency.

In the last part of this study we investigated the influence of the most important production parameters on the acoustic cues of "sharpening".

In the case of glottal adduction these are the abduction area \( a_0 \), the cord tension \( q_0 \), and the lung pressure \( p_0 \) at instant \( t_0 \) (see Fig. 1). The time \( t_0 \) is defined as the instant at which the adduction/abduction gesture ends.

![Figure 2: Resynthesis of "sharpening" in {ø:s} by glottal adduction and glottal abduction.](image)

We measured the relative decrease in fundamental frequency \( f_0 \) and the relative decrease in intensity \( d_{\text{Int}} \). Fig. 3 shows the effect of the variation of the production parameters \( a_0 \), \( q_0 \) and \( p_0 \) on the acoustic parameters \( f_0 \) and \( d_{\text{Int}} \).

1. When the adduction becomes stronger (\( a_0 \) higher negative values) \( d_{\text{Int}} \) decreases but \( f_0 \) increases. (2) Lowering of \( q_0 \) leads to a decrease in \( f_0 \) but to an increase in \( d_{\text{Int}} \). (3) Lowering of \( p_0 \) leads to a decrease in \( d_{\text{Int}} \) while \( f_0 \) remains relatively unchanged.

So during strong adduction it is necessary to lower the cord tension and the lung pressure to get the acoustic features of the "sharpening".

5. DISCUSSION

Resynthesis of "sharpened" syllables by glottal adduction or adduction was done.

In the latter case, mechanical compression (\( a_{ab} < 0 \), low cord tension and a decrease in lung pressure are necessary. Similar physiological features were found for the Danish "stød".


6. REFERENCES


DYNAMIQUE LINGUALE DES VOYELLES ORALES FRANÇAISES
ÉVALUATION STATISTIQUE À PARTIR DE DONNÉES
CINERADIOGRAPHIQUES

Bernard FLAMENT

Institut Universitaire de Technologie - Heînlex - B.P. 420
44606 SAINT-NAZAIRE CEDEX - France.

ABSTRACT
In this present study we propose, through cineradiographic data, to extract by close evaluation the dynamics of the lingual articulator for French oral vowels in two sorts of actions : reinforced and unaccentuated. We intend to apply a statistical approach to lingual movements, at the place of articulation, using dispersion coefficients and confidence intervals. The degree of lingual articulator stabilization on the axe where maximum vocal tract constriction occurs is more important in the instances of reinforcement; this is even more distinct if the length of vocal phonem is considered.

1-INTRODUCTION, LES TERMES DE LA PROBLÉMATIQUE.
1.1 Le renforcement, qui peut affecter une partie de la chaîne partée, tend à modifier de façon plus ou moins sensible le comportement articulatoire et par là les modalités acoustiques des réalisations phonétiques. Sur un plan spécifiquement articulaire, les mouvements vélaire lors de la production des voyelles nasales renforcées sont notamment bien individualisés, l'abaissement du voile étant limité au maximum à la durée phonétmatique (certains cas de relèvement anticipé sont d'ailleurs relevés). En réalisation non-renforcée, des phénomènes d'extension ont lieu : le passage velo-pharyngal non seulement présente un diamètre plus important, mais il peut se produire d'une manière à la fois courante et subéquente (B.FLAMENT 141, 151) introduisant bien entendu des modifications acoustiques des phonèmes contigus. Le positionnement vélaire y est en outre plus fluctuant, plus mobile alors qu'en réalisation renforcée, la stabilisation est plus marquée, le voile se maintenant en position plus proche de la paroi pharyngale sous l'influence d'une tension musculaire plus grande.

2.2 Les prises de vues cinéradiographiques ont été effectuées au Centre Médico-chirurgical de STRASBOURG-SCHILTIGHEIM à la vitesse de 50 m/ sec., ce qui permet une apprehension très fine des faits articulatoires. Le locuteur est un francophone -langue maternelle : son français est dénué de toute trace de nuance régionale.

2.3 L'ensemble du contour lingual a été envisagé pour une juste définition du lieu d'articulation de la voyelle considérée. Des mesures précises quant au diamètre du passage buccal aussi bien dans la zone alvéolaire que dans les zones palatate, vélaire ou pharyngale ont été relevées. Le dynamisme de l'articulateur est étudié sur l'axe où se produisent le rétrécissement maximal et ce, image par image, pour l'ensemble de la durée articulatoire.

3- TRAITEMENT STATISTIQUE DES DONNÉES.
3.1 A partir des valeurs du diamètre du tractus relevées sur l'axe du lieu d'articulation, nous proposons un traitement statistique. Prenons le cas en effet de deux réalisations phénoméniques bien contrastées sur le plan des données concernant la durée articulatoire : le Iel, le iel et le iel. En énonciation valorisée (E.V.), la durée phonétique est beaucoup plus accusée qu'en énoncé neutre (E.N.), dans lequel la voyelle est en position inaccentuée. À titre d'exemples, voici, pour ces trois voyelles et dans les 2 types de réalisations, les valeurs (en cm) du diamètre du passage vocal sur l'axe retenu :

| N. | Iel.E.N.7 0,0639 + E.V.8 0,0415 - N | s | 0,0374 + E.N.5 0,0374 + Iel.E.N.4 0,1980 + E.V.8 0,0348 - E.V.9 0,0614 - Ioel.E.N.5 0,1871 + Ioel.E.N.7 0,2279 + E.V.9 0,0774 - E.V.8 0,1225 - Ioel.E.N.4 0,1146 + E.V.8 0,0634 - Ioel.E.N.5 0,0245 + Iel.E.N.3 0,0471 + E.V.5 0 - E.V.3 0 - Ioel.E.N.5 0,2481 + E.V.6 0,1106 - |

La dispersion des valeurs en E.V. par rapport à E.N. présente un déficit dans la totalité des cas : la stabilité articulatoire -linguale- y est plus importante, au lieu d'articulation, dans ce type de réalisation. Là où les valeurs de s sont déjà très faibles en E.N. (c'est le cas essentiellement des voyelles antérieures, de lieu d'articulation alvéolaire), celles-ci s'abaisse encore en E.V.. Dans les autres cas, les écarts entre E.N. et E.V. sont parfois très sensibles avec des rapports (bien E.N./E.V.) pouvant atteindre 1 à 2 et même quasi 1 à 3. L'incidence des muscles linguaux n'est pas négligeable dans ce processus, notamment pour les réalisations postérieures : les muscles qua...
L'articulation postérieure Iol, 1l et 1l se caractérisent par des intervalles de réalisation plus courts, tendant plus lent, plus aléatoire, le positionnement lingual au lieu d'articulation ; ceci est tout particulièrement sensible en E.N. où les phénomènes de coarticulation sont conséquents et où la stabilité y est donc moindre. En E.V., le diamètre du tractus d'articulation et le positionnement cible est davantage maintenu sous l'influence de la tension neuro-musculaire.

3.2.3 Les valeurs de s sont encore bien plus réduites pour les réalisations renforcées si l'on prend comme base temporelle la durée articulatoire des mêmes phonèmes, observée en E.N. :

Iol (N=7) s = 0,0416
Iol (N=5) s = 0,0415 Iol (N=7) s = 0
Iol (N=4) s = 0 Iol (N=7) s = 0,0942
Iol (N=5) s = 0 Iol (N=3) s = 0,04
Iol (N=5) s = 0 Iol (N=3) s = 0

Une remarquable stabilité est observable pour les réalisations renforcées. La régulation du paramètre de la durée diminue le taux des éventualités de dispersion concernant les valeurs de xj - x, qui réduisent encore la valeur de s. Il n'y a guère que pour Iol où celle-ci est quasi-ment la même, du fait d'une très faible augmentation de durée en E.V. (N=5) par rapport à E.N. (N=7) et pour Iol où sa valeur (table page 37) reste inchangée.

3.3 Intervalle de confiance.

Afin de nous assurer de la validité des résultats, nous avons calculé l'intervalle de confiance pour chacun des phonèmes vocaliques (avec homogénéisation temporelle de la durée articulatoire E.V./ E.N.). Cet intervalle de confiance a été calculé en considérant un degré de confiance égal à 95% pour les différentes séries.

4 - REFERENCES

A FEEDFORWARD CONTROL STRATEGY CAN SUFFICE FOR ARTICULATORY COMPENSATIONS

Shinji Maeda
Département SIGNAL, CNRS URA-828
Ecole Nationale Supérieure des Télécommunications
46, rue Barrault, 75634 Paris Cedex 13, France

ABSTRACT
An articulatory model was used to analyze cineradiographic and labiofilm data. The variation in "target" values of two model parameters, the jaw and tongue-dorsum positions, during the production of the vowels, /u/ and /a/, was examined. The "target" values of these two parameters for the same vowel vary much more than the corresponding acoustic ones. The scattergrams of each vowel exhibited a linear relationship which can be regarded as an indication of the coordination between the jaw and tongue. When the coordination effects are subtracted, the articulatory variability becomes comparable to that of the acoustic (F1/F2) one. Calculations with the model indicated that the coordination is used by speakers to achieve an acoustic compensation. These findings suggest that vowel production is compensatory and that compensation can be modeled effectively by a feedforward strategy.

INTRODUCTION

Bit-block vowel experiments have demonstrated a speaker's ability to compensate for the effects of blocked jaw position by readjusting other articulators to produce specified vowels. Observing a speaker's ability to compensate immediately, Lindblom, Lubker and Gay have suggested that normal speech production itself is compensatory [3]. If this is the case, we should observe in normal speech a high degree of variability in the individual articulatory positions and a lower degree in the corresponding acoustic patterns, for example, in the formant patterns. Moreover, if compensation occurs in an arbitrary manner, it is not effective to specify vowel targets in terms of articulatory parameters. This appears to be one of reasons why the targets are often described by the vocal tract area function [2]. If compensation occurs in a lawful manner however, the vowel targets can be specified directly by the individual parameters with some calculations reflecting the laws. We shall investigate these questions by analyzing X-ray and labiofilm data with an articulatory model.

2. ARTICULATORY DATA AND MODEL

The data consist of more than 1000 digitized tracings of vocal tract shapes corresponding to 10 French sentences uttered by two female speakers, PB and DF [1]. Each of the data frames describing the vocal tract profiles from the glottis to the lip opening and the frontal lip shapes was obtained by manually tracing radiofilms and labiofilms shot simultaneously at a rate of 50 frames per second. The digitized version of the data has been kindly provided by the Phonetic Institute of Strasbourg, France.

The measured vocal tract shapes were analyzed statistically. A factor analysis has resulted in a linear articulatory model with seven parameters. In this study, we shall focus our attention on two parameters, the jaw and tongue-dorsum positions for two reasons: these two parameters are most important for specifying the tongue profiles and they can acoustically compensate for each other specifically in the production of unrounded vowels, such as /u/, /e/, and /a/ [4].

3. ARTICULATORY VARIABILITY

With the linear model, the value of each parameter is calculated directly from the measured vocal tract shape. The articulation along a sentence can be described, therefore, by the frame-by-frame variation of the calculated articulatory parameter values. The resultant data have indicated a considerable articulatory variability for the same vowel from different phonetic contexts. In order to assess the range of variability, trajectories of the two parameters, jaw and tongue-dorsum, were plotted on the jaw-tongue articulatory space. Then an articulatory "target" position was determined as the turning point of each trajectory. The result is shown in Fig.1.

The straight lines plotted on Fig.1 were determined by means of a principal component analysis of the scattergrams associated with each of the two vowels: they correspond to the first principal axis. Although the scattergrams exhibit a great degree of variations, the data points for /u/ and /a/ are distributed without overlap. Furthermore, each cluster is distributed roughly along the straight line. These straight lines can be regarded as linear approximations of the inter-articulatory coordination between jaw and tongue-dorsum. The observed variability, therefore, can be separated into a controlled context-determined variation and an unexplained residual, say, "true" variability. Since the proportion of the variance extracted by the first principal component varies between 65% (in the case of /a/ uttered by speaker PB) and 88% (by speaker DF), the true articulatory variability for jaw and tongue ranges from 35% to as small as 12% of the observed variance.

4. ACOUSTIC VARIABILITIES

The articulatory variability can be examined more meaningfully, if it is compared with the corresponding acoustic variability. In this study, the first (F1) and second formant (F2) frequencies, as the acoustic characteristics of the two vowels, were calculated using the articulatory model. The F1-F2 calculations were done only for speaker PB, since the data for DF lacks the lip section and thus F1 and F2 cannot be calculated. All seven parameter values were derived from the corresponding data frame. The area function and then formant frequencies were computed from model specified vocal tract shapes. The resultant F1/F2 plots are shown in Fig.2. The data points for the vowel /u/ are added to indicate the vowel space of speaker PB.

Comparing the articulatory target scattergrams in Fig.1 (for speaker PB) and the corresponding acoustic ones in Fig.2, it appears that the acoustic scattergram points are distributed more tightly than

Fig.1 Scattergrams of jaw and tongue-dorsum parameters at the "articulatory targets" for the two vowels /i/ (indicated by the circles) and /a/ (by the triangles). The ordinate and abscissa have standardized units. Zero corresponds to the arithmetic mean calculated for all the utterances by each speaker. 1 (-1), 2 (-2), and 3 (-3) represent 1, 2, and 3 standard deviations, respectively, from the mean. Data for the two speakers, PB and DF are shown.

Fig.2. The scattergrams of jaw and tongue-dorsum parameters at the "articulatory targets" for the two vowels /i/ and /a/. The ordinate and abscissa have standardized units. Zero corresponds to the arithmetic mean calculated for all the utterances by each speaker. The data points for the vowel /u/ are added to indicate the vowel space of speaker PB.
articulatory ones, i.e., acoustic variability seems to be less than articulatory one. For a quantitative comparison, let us propose a variability index, \( v \) (an averaged normalized variance), for two articulatory or two acoustic variables as follows:

\[
v = 100 \times \sqrt{\frac{\sigma_i^2}{2 \sigma_{\text{max}}^2}}\%
\]

where \( \sigma_i^2 \) is the variance of variable \( i \) (= 1 or 2 in our case), and \( \sigma_{\text{max}}^2 \) is the possible maximum variance of variable \( i \). Since a sufficient amount of data to determine the possible maximum variance is not available, we have assumed, as a gross approximation, that \( \sigma_i^2 \) of articulatory and acoustic data can be substituted by the values of half of the range of the individual variables. In the calculation of the articulatory variability index, \( \sigma_{\text{art}}^2 = 3 \) is used for both jaw and tongue-dorsum data, corresponding to half the range, since parameter values rarely exceed the range from -3.0 to 3.0. The acoustic variability index is computed assuming that \( \sigma_{\text{ac}}^2 \) (for F1) equals to 300 Hz, and \( \sigma_{\text{ac}}^2 \) (for F2) to 1250 Hz. The calculated index values are listed in Table 1.

### Table 1: Articulatory and Acoustic Variability Indices (in %) for the Two Speakers

<table>
<thead>
<tr>
<th></th>
<th>Speaker = PB</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jaw/Tongue</td>
<td>F1/F2</td>
<td>F1/F2</td>
<td>F1/F2</td>
</tr>
<tr>
<td>( v )</td>
<td>21.7</td>
<td>16.2</td>
<td>17.0</td>
<td>18.4</td>
</tr>
<tr>
<td>( v_{\text{max}} )</td>
<td>7.5</td>
<td>8.1</td>
<td>3.6</td>
<td>4.6</td>
</tr>
</tbody>
</table>

The index values span around 20% for the articulation and less than 10% for the acoustics. The residual articulatory variability indices are listed at the rows marked "\( v_{\text{max}} \)" in Table 1, which are calculated from the proportion of variance corresponding to the residual. These index values are less than 10%, a value which is less than half of the corresponding total raw variability, and which compares well with the index calculated for the F1/F2 scattergrams of PB shown in Fig. 2. For speaker DF, the true articulatory variability is four times less than the observed raw variability. The calculation has indicated that although the variability of the individual articulators is relatively great, if the coordination term is subtracted, the articulatory variability compares well with the acoustic one.

### 5. COMPENSATORY ARTICULATION

What mechanism lies behind this significant reduction of the raw variability from articulatory to acoustic by means of coordination? In our previous studies [4], we have already shown that in case of unrounded vowels such as /a/ and /u/, jaw and tongue-dorsum positions can acoustically compensate for each other, as mentioned earlier. The compensation means that a deviation in the position of one articulator can be compensated by a readjustment of other articulator(s) to keep the deviation in the acoustic pattern to a minimum. It is reasonable, then, to hypothesize that the inter-articulatory coordination, in fact, results in the acoustic compensation of the type mentioned above. If this is the case, the principal axis representing the coordination in Fig.1, is also an acoustical "equi-line", i.e., changes in the values of the two parameters along these lines result in relatively invariant acoustic patterns that depend only on the vowel identity.

In order to demonstrate the acoustic equivalence for the two vowels, F1 and F2 values were calculated at different jaw positions from -2.0 (low) to 1.0 (high) for /i/ and from -3.0 to 0.0 for /a/, with 1.0 step size. The corresponding tongue positions were determined by their linear relationships. Note that a change in jaw position influences not only the tongue shape, but also the lip aperture and, to some extent, the larynx position. The values of the remaining five parameters were kept fixed at those originally determined from the corresponding vocal tract data frame. The results are listed in Table 2. The index related to the equi-line of /a/ is 3.2%, which is much smaller than observed acoustic variability. As far as the vowel /i/ is concerned, the index becomes extremely small, about 1%, indicating that the equi-line produces an almost invariant F1-F2 pattern.

### Table 2: F1/F2 Variability Indices Calculated along the Equi-Lines of PB in Fig. 1

<table>
<thead>
<tr>
<th></th>
<th>/i/</th>
<th>/a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>index</td>
<td>-2.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>-3.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Although the acoustic compensation along the equi-lines is not perfect, it is safe to state that articulatory manoeuvres along an equi-line tend to result in fairly invariant acoustic patterns around the target vowel. It should be emphasized here that the equi-lines are derived from the observation of data. It is tempting to speculate then that the speakers have integrated these equi-lines in their mental process and exploit them to place individual articulator positions differently but appropriately for particular phonetic contexts, yet producing relatively invariant acoustic targets.

It may be noteworthy to mention that the coordination does not necessarily always mean compensation. In the case of /a/, for example, the raw variabilities of the jaw and lip parameters (height and protrusion) were relatively small, less than 10%. In detail however, scattergrams indicated that closing the jaw, and narrowing and protruding of the lip opening occur concomitantly, enhancing together a narrow and long lip tube. The acoustic consequences of this kind of coordination would be exactly in the opposite of compensation.

### 6. CONCLUDING REMARKS

It has become clear that the apparently large variability of the individual articulator positions during the same vowel but from different contexts can be explained at least in part, by the inter-articulator coordination. Moreover, the coordination is such as to achieve an acoustic compensation which results in the realization of a relatively invariant acoustic target, thus supporting the idea of speech production as a compensatory process [3]. Surprisingly, the coordination and thus the compensation can be specified directly in terms of articulatory parameters. The implication of this is important. If the relationship is well defined in such a simple fashion, it is not unreasonable to assume that speakers know exactly how to coordinate in advance. Then a feedforward mechanism can be specified directly for the compensatory articulation, without resorting to acoustic or to sensory feedback.

### 7. REFERENCES

Soc. Am., 69(3), 802-810.

some fixed-mandible vowels and a model of speech motor programming by

from the analysis of the shapes of vocal tract shapes using an "articulatory
model". In Speech Production and Speech Modeling (W.J. Hardcastle & A.
TOWARDS THE SPECTRAL CHARACTERISTICS
OF FRICATIVE CONSONANTS

Christine H. Shadle⁠¹, Pierre Badin⁠², André Moulinier⁠¹

University of Southampton, UK⁠¹
ICP, INPG, Grenoble, France⁠²

ABSTRACT
Articulatory data and the all-pole transfer functions for sustained fricative consonants [s, f, ç] were used to identify the cavity affiliation of peaks and troughs in the far-field spectra. This identification then allowed an analysis of the differences between fricatives, and across subjects within fricatives, necessary steps towards the establishment of distinguishing acoustic cues.

1. INTRODUCTION
It has long been established that [f, s] are distinguished chiefly by the transitions of the vowels on either side, while [s, f] are distinguished by their spectral characteristics [3]. But establishing the particular spectral cues that distinguish [s, f] from each other or from other fricatives, is more difficult. Many authors report consistency within a speaker, but high variability across speakers [4, 7]. Perhaps as a result, efforts to phrase distinguishing cues in terms of the frequency range of the highest intensity levels, or in terms of relative intensity levels, seem to work well within a speaker but poorly across speakers (e.g. the frequency ranges overlap so much as to be useless) [7].

In this study we explore variability in the spectra of sustained fricatives. First, we need to establish which aspects of the spectrum are consistent within a speaker-fricative combination. Then where possible we identify the articulatory parameters that control these consistent features of the spectrum. Finally we use this articulatory-acoustic mapping to explain some of the across-subject differences. This sequence should lead to a set of paired articulatory-acoustic cues that can then be tested for their perceptual importance.

2. METHOD
2.1 Corpus and Speakers
The corpus used in this paper is the result of a larger study (Leeds, Grenoble, Southampton). It includes articulatory, aerodynamic and acoustic measurements made of two speakers. The corpus includes 13 fricatives [s, f, ñ, t, ñ, s, z, j, ç, x, ñ, h] produced in several ways. This study refers only to the sustained corpus, in which the set of 13 fricatives was said six times; in each set, the order of the 13 fricatives was randomized. Two different recordings of the sustained fricatives were used in this study, as detailed below.

The two speakers used for the corpus are the first two authors of this paper, and will be referred to as CS, a woman speaker of General American English, and PB, a man speaker of French. Although the list of fricatives recorded includes several that are not native to either speaker, these were included deliberately to obtain further examples of place variation for the same vocal tracts.

In addition to measurements made while speaking, X-ray data and dental impressions were available for each subject. Together with EPG data and external photographs, these were used to construct an area function for each unvoiced fricative for each speaker [6].

2.2 Acoustic Analysis
Data shown in this paper were recorded under high-fidelity conditions: the subject was seated in a chamber anechoic above 170 Hz, with a B&K 4165 ¼" microphone located 1m in front of the subject's mouth. Recordings were made with a Sony PCM system at 16 bits with a sampling frequency of 44.1 kHz. A calibration signal was recorded to allow absolute sound pressure level to be retained.

An average power spectral density function was computed by averaging 25 spectra in the center of the 3s fricative. Each spectrum was computed using a 20ms Hanning-window.

2.3 Determination of Transfer Function
In this experiment, the subject assumed the position for a fricative, but without actual speech production (glottis held closed). The vocal tract was excited by a small loudspeaker fed with white noise and pressed against the neck just above the thyroid cartilage. A microphone located 2cm from the mouth detected the (very weak) noise signal after filtering by the vocal tract. This signal was essentially the all-pole transfer function of the tract, up to about 5 kHz.

The area functions derived from articulatory data were then used to predict the all-pole transfer function for each fricative. Comparison of predicted and measured all-pole functions then enabled identification of the cavity affiliation of each pole. Further details are given in [2].

3. RESULTS AND DISCUSSION
Figure 1 shows three of the fricatives analyzed, with all six tokens shown on each graph. Note first the consistency apparent within each graph, i.e. within each fricative-subject combination. This consistency makes it easier to evaluate the variability across speakers, and across fricatives. For [s, ç] the overall spectral shapes are similar but the frequencies at which particular peaks occur differ between the two speakers. For [f], even the overall shape differs: both speakers have a region of high energy between 1.5 and 6 kHz for PB, and 2.5 to 7 kHz for CS. However, for PB there is an abrupt drop in amplitude of some 10 dB at 6 kHz and the spectrum is approximately level above that frequency; for CS, there is no abrupt drop. Instead the level falls off steadily, decreasing 20 dB between 7 and 12 kHz. Can we make sense of these differences?

Badin's results [2] indicate that for CS's [f], F1 is a Helmholtz resonance of back cavity and constriction; F2 and F3 are back cavity resonances; and F4 is a front cavity resonance. A series pressure source in the front cavity would result in zeros cancelling the back cavity resonances, plus two free zeros: one corresponds to a Helmholtz resonance of the constriction and the part of the front cavity between the constriction exit and the source. The other corresponds to the half-wavelength resonance of the same part of the front cavity.

Since CS has smaller vocal tract dimensions, her formant frequencies are predicted to be higher, and in fact they are. However, a much more obvious difference is that for CS the first four formants are approximately evenly spaced, while PB has F2, F3 and F4 clustered together. With the zeros interspersed, these small differences in formant frequency make a big difference in formant amplitude: for PB, the second formant is boosted and becomes the lowest high-amplitude peak, while for CS, F3 takes on that role. This means that the lower edge of the high-amplitude region differs by 1 kHz, even though F2 differs by only 100-200 Hz.

Above 5 kHz we have less information to work with. However, the differences in spectral amplitude and slope could be explained if the zero were at a significantly lower frequency for PB than for CS, e.g. the 7 and 12 kHz respectively. This zero frequency should be inversely proportional to Fp, the teeth-constriction distance, and in fact Fp is significantly longer for PB, as evidenced from X-ray and direct palatography. This is surprising since
the vocal tract dimensions in the anterior part of the mouth cavity, obtained from measurements of the two subjects' dental impressions, are quite similar. Since the phoneme is native to each subject, and the spectral differences noted are consistent within each subject, more subjects are needed to establish why the articulatory differences exist.

The fricative [s] is more similar for the two subjects. Since the front cavity is smaller than for [f], the corresponding resonances are higher. For CS, it appears from transfer function simulations that the lowest front cavity resonance is F6 (see Fig. 1); F2, F3, F4 and F5 are the lowest resonances of the back cavity (harmonics of the half-wavelength mode), and are accompanied by bound zeros. For PB the lowest front cavity resonance is F5, and F2, F3, and F4 are the back cavity resonances.[1]. The differences between these resonances are consistent with the articulatory data. The amplitude of the plateau above the front cavity resonance relative to the spectral level of this resonance varies noticeably between the two subjects, and again the free zero may be lower for PB (approximately 11.5 kHz) than for CS (well above 12 kHz).

The fricative [ç] is not native to either speaker, and so might be expected to be more variable. In fact, it looks consistent for each speaker, and the overall spectral shape is similar. For both speakers, the lowest front-cavity resonance is the lowest high-amplitude formant. This corresponds to F4 for CS, F3 for PB. Although the front cavity is longer for [ç] than for [f], this front-cavity resonance is not significantly lower. A possible explanation is that for extremely short front cavities, the resonance frequency is related to the volume or possibly vertical dimension. Thus the exact shape of the sublingual cavity becomes important for [s, f]. As for [f], the spectral shape at high frequencies differs, and could be explained in part by a difference in source-constriction distance. The likelihood that the source is distributed

[5] complicates the issue by blurring the free zero, but in any case a lower-frequency free zero would reduce the overall amplitude relative to [f].

5. CONCLUSION
The search for acoustic cues distinguishing fricative consonants must begin with a study of the variability present in fricative production. By using subjects for whom much articulatory data is available, it has been possible to locate low-amplitude but consistent spectral peaks, and to discover their cavity affiliation and controlling parameters. Although vocal tract dimensions influence peak frequencies, the added complications introduced by zeros mean that simple measures such as frequency range for high-amplitude regions are likely to be highly variable.

6. ACKNOWLEDGEMENT
This work was funded in part by a collaborative EC SCIENCE award, CECS-SCI*0147C(EDB).

7. REFERENCES

Fig. 1. Averaged power spectral densities for sustained productions of the fricatives [s, f, ç]. In each graph the six curves shown correspond to the six tokens uttered by each subject. Subject PB is male, native French speaker; CS is female, native General American speaker.
LABIALITE VOCALIQUE: ETUDE COMPARAISE DES TYPES, DEGRES ET STRATEGIES ARTICULATOIRES DE PLUSIEURS LANGUES

Jean-Pierre ZERLING

Institut de Phonétique - Université de Strasbourg II
22 rue Descartes - 67084 Strasbourg Cedex - France

ABSTRACT

This study is based on both personal and published data for about 2500 vocalic realisations. Frontal shape of the lip opening was analysed for several languages: French, English, Thai, Cantonese, Finnish, Polish and Swedish.

Starting from a description of labial activity for French vowels, we show in a few examples in what way articulatory activity may vary from one language to the other, even if phonological features are the same. In other words, vowels usually described by the same IPA symbol may appear to differ not only by labial articulation, i.e., degree of lip opening and labial area, but also by the type of activity involved, i.e., spreading, protrusion, flattening, etc. Vocalic categories which are apparently similar for two particular languages may actually refer to labial shapes and articulatory strategies involving very different intervocalic relations.

1. INTRODUCTION

On sait le rôle important joué par les lèvres en phonation. Les études phonétiques portant sur le français concernent des domaines très variés : la phonétique articulatoire [2,22], l'acoustique [15], l'acquisition automatique des données [12], etc.

Ces recherches trouvent leur application en phonétique et en phonologie [5], en modélisation et en synthèse de la parole[11,14], en perception visuelle [8,11,12,18], etc.

La présente étude vise à se greffer à diverses autres que nous avons nous-même menées, toujours à propos de la labialité, mais concernant en plus la coarticulation [19], les stratégies articulatoires [20,22], la variabilité [21] et les comparaisons interlangues [22,23].

2. METHODE

Nous avons déjà décrit à plusieurs reprises la méthode retenue. Elle consiste à observer essentiellement les paramètres frontaux de l'orifice labial, notamment l'écartement horizontal A, l'espace vertical interlabial B, l'aire S et éventuellement le facteur de forme K = A/B.

Les mesures ont été obtenues, selon les corpus, à partir de labio-photographies ou d'images de labiofilms. Ces documents ont été réalisés soit par nous pour le français et le thai [22], soit empruntés à des bases de données existantes mais pas forcément destinées directement à l'étude des lèvres, pour le français, l'anglais, le suédois, le canari, le finlandais, le polonais [3,4,7,9,13].

Pour les besoins de la présente étude, nous avons exploité essentiellement deux sortes de graphiques : la répartition des réalisations dans le plan (A,B) et à dispersion des valeurs de S pour chaque voyelle. Ces documents permettent de comparer à la fois les positions relatives des voyelles d'une même langue et celles de langues différentes.

3. LABIALITE VOCALIQUE COMPAREE POUR PLUSIEURS LANGUES

3.1. Caractéristiques du français

Une étude approfondie sur le français [22], nous a permis de conclure, au moins pour notre langue, que les deux degrés de labialité phonologique ne correspondaient pas de manière biunivoque à deux degrés d'activité articulatoire. De plus, à l'abaissement relativement progressif de la langue lors de la prononciation des voyelles d'une même série ne correspond pas nécessairement une ouverture progressive de l'orifice labial. Statistiquement parlant, et aussi bien pour les réalisations tenues que pour celles en contexte, il apparaît que les voyelles se regroupent selon trois classes de labialité distinctes que nous avons nommées pour le français : [-lab], [+lab] et [++lab] (Fig.1). Il est à remarquer que toutes les voyelles d'une même classe peuvent adopter la même forme frontale d'orifice labial, indépendamment de l'articulation linguale. On note encore que deux voyelles labiales peuvent s'opposer par la labialité : [a,e], ce qui contredit l'habitude opposition +/- rond.

3.2. Comparaison à d'autres langues

Les langues étudiées sont généralement caractérisées par une opposition phonologique binaire de labialité vocale. C'est-à-dire qu'elles possèdent des voyelles de même nature articulatoire s'opposant essentiellement par leur degré de labialité. Dans la mesure du possible, les systèmes choisis comportent des voyelles antérieures à la fois labialisées et non-labialisées.

Partant d'une classification articulatoire générale de la labialité vocale que nous avons définie par ailleurs [22], nous passerons en revue les caractéristiques labiales proposées en illustrant par des exemples précis leur bien-fondé ou leur variation d'une langue à une autre.

3.2.1. Mode de labialisation

Cette caractéristique est généralement liée à la labialité : va non-labialisé. Elle est valable pour toutes les langues observées et son choix relève de la description phonologique.

L'observation révèle que pour un même mode de labialité, diverses types de labialisation peuvent être utilisés, qui génèrent néanmoins une même aire labiale, et permettent donc d'aboutir à un même résultat acoustique.

3.2.2. Type de labialisation

Pour le moment, nous en retenons cinq. Ils sont le reflet direct de l'activité musculaire :

- labialisé arroxi proar
- labialisé écrasé non-pronas
- labialisé écrasé pronas
- écarté, avec recul latéral des commissures
- neutre, contrôlé par les mouvements du maxillaire

Un même mode peut donc être obtenu de diverses manières, selon son type :

M o d e 1 a b i a l :
- le plus fréquemment réalisé par une prolusion et un arrondissement des lèvres ;
- mais parfois une forte diminution de l'aire labiale est obtenue simplement par un écrasement vertical de l'orifice, comme pour certaines prononciations de l'anglais (Fig.2).

Mod e non-labial :
- il est généralement caractérisé par une forme assez variable et surtout une aire relativement importante ;
- en revanche, on rencontre parfois dans une même langue, à côté du premier type de non-labialisé, des voyelles caractérisées par un orifice très écrasé verticalement. Il s'en suit une réduction importante de l'aire mais qui reste néanmoins nettement supérieure à celle subie par les voyelles labialisées : [i, i] polonais [4], [ai] anglais [3, [1, a] thai (Fig.3), anglais (Fig.2).

Certaines langues possèdent des voyelles labialisées très ouvertes qu'il est difficile de classer avec les types précédents, par exemple : [au] suédois [13] et en thai (Fig.3). Il nous paraît judicieux de les appeler "d'autres" et de les considérer plutôt comme non-labialisées.

Bien que notre étude ne traite pas directement de l'activité musculaire, nous avançons l'hypothèse que chacun de ces types de labialisation pourrait constituer un "axe labial naturel" [16], caractérisé par un ensemble de voyelles impliquant l'activité progressive non contradictoire d'un muscle ou d'un groupe de muscles ; par opposition à un axe non-naturel impliquant une réorganisation complète de l'activité musculaire entre les voyelles.

3.2.3. Degré de labialisation

Il reflète globalement l'ouverture labiale, c'est-à-dire à la fois l'espace vertical inter-labial B et l'aire des lèvres S. Il peut être contrôlé par les mouvements du maxillaire inférieur ou par ceux des lèvres.

On peut rencontrer, selon les langues, un même degré de labialisation pour plusieurs voyelles appartenant à une même série, par exemple en français pour [u, o] ou [y, e], ou au contraire des degrés différents variant parallèlement à l'abaissement de la langue, comme en thai.

3.2.4. Stratégie de labialisation

Elle gère les degrés respectifs de labialisation des différentes voyelles d'une
mêmes catégories en les rendant, par exemple:
- dépendants de l’aperture intra-buccale de la voyelle, et donc en général soumis à une variation graduelle, comme en suidois, en finlandais [13] et en thaï (Fig. 3).
- indépendants de la voyelle, qui adopte alors un degré soit relativement constant: voyelles françaises [y,œ,u] (Fig.1), soit au contraire aléatoire: voyelles [i,e,a] du français (Fig.1) ou [i,e,a] du cantonais (Fig.4).

3.2.5. Catégorie labiale

Partant de là, nous appelons catégorie labiale un regroupement de voyelles ou de réalisations vocaliques de même mode et de même type gouvernées par une même stratégie; celle-ci indiquant comment est géré le degré de labialisation au sein de la catégorie. Une catégorie est donc obtenue moyennant le respect d’un ensemble de contraintes articulatoires et bio-mécaniques, et de stratégies motrices.

Le nombre de catégories peut différer pour deux langues même si, pour des séries apparentement identiques, celles-ci comportent les mêmes voyelles ou du moins utilisent les mêmes symboles voca- 

Appareil de parole: lecture labiale et soties labiales", Rapport IBM, France, personal communication.

4. REFERENCES


Fig.1 Répartition dans le plan (A,B) de 14 voyelles du français, 105 sujets, 1238 réalis., ell. à 90%, d’après [22, p.143]

Fig.2 Répartition dans le plan (A,B) de 9 voyelles brèves et longues du thaï, 6 sujets, 108 réalis., ell. à 75% d’après [22, p.483]

Fig.3 Répartition dans le plan (A,B) de 9 voyelles brèves et longues du thaï, 6 sujets, 108 réalis., ell. à 75% d’après [22, p.483]
RESISTANCE DU [e] SIMPLE ET DU [e e] DOUBLE A LA COARTICULATION MANDIBULAIRE EN ARABE MAROCAIN

N. Rhaddis R. Sock
Institut de la Communication Parlée, CNRS URA 368
Grenoble, France

ABSTRACT
This investigation attempts to characterize coarticulation resistance to speech rate for simple [I] and double [II], through [i+a] and [a+i] vowel-to-vowel transitions, in Moroccan Arabic. Results show that the raising gesture for [I] has basically more coarticulatory influence on the consonantal gesture than the lowering gesture for [a], a trend which is especially evident for geminates at normal rate. These findings are discussed in the frame of speech motor control theories of anticipation.

1. INTRODUCTION
Le but de ce travail est de tester la résistance à la coarticulation mandibulaire des consonnes [I] vs. [II], dans un système phonologique spécifique : l'arabe marocain; et ceci selon le contexte de la vitesse d'élocution. Une attention particulière sera donnée au paradigme expérimental de la résistance coarticulatoire, proposé par [5], pour une variante de l'anglais britannique. Nous utiliserons les concepts et les outils statistiques liés au contrôle moteur [7, 14]. L'approche relative de la durée (timing relative) sera un moyen efficace pour apprécier les changements qui interviennent lorsqu'on fait varier le débit [6, 11]. Nous suivrons les principes méthodologiques sur la recherche en parole proposés par [2] qui se sont avérés rentables pour la mise en évidence du timing des oppositions phonologiques.


2. METHODE
2.1. Corpus

2.2. Mesures
Le signal acoustique, numérisé à 8kHz, a été échantillonné manuellement en événements [I] à l'aide d'un éditeur de signal [4]. Les signaux de déplacement vertical de la mandibule ont été recueillis à l'aide d'un kinésiographe mandibulaire (K5AR) et échantillonnés à 160 Hz. Les signaux acoustiques qui ont été repérés sur les signaux de vitesse et d'accélération. Ces derniers ont été obtenus par dérivation du signal de position lissé des fonctions splines cubiques.

Les événements acoustiques : VVO ou début du voissement vocalique (correspondant ici à la détente des [I]) et VVT ou fin du voissement vocalique (closion des [I]), nous fournissent la base temporelle VVT-VVO (tenue des [I]) comme domaine d'étude de la gémination consonantique et de la résistance coarticulatoire. Sur le plan articulaire nous avons repéré les événements :
- ACC : défini comme l'accélération maximale du geste d'abaissement mandibulaire pour la réalisation de la voyelle subséquente [a];

Nous avons retenu, à partir de ces événements articulatori-acoustiques, les deux phases temporelles suivantes :
- LON (ACC-VVO) ou Lowering Acceleration ; [a] Vowel Onset, qui va du geste de l'accélération maximale de l'abaissement mandibulaire à l'établissement vocalique du [a] subséquent. Exprimée en pourcentage de la durée VVT-VVO, elle nous donne le degré de coarticulation ou de ("pénétration") du [a] dans les consonnes.
- RON (DEC-VVO) ou Raising Deceleration ; [a] Vowel Onset, qui a pour borne, la décélération maximale du geste de l'élévation mandibulaire pour [I] simple ou [II] double et l'établissement de la structure formantique de la voyelle suivante [Ii]. Exprimée en pourcentage de VVT-VVO, elle nous donne le degré de coarticulation ou de ("pénétration") du [I] en dans les consonnes.

Pour les réalisations [III] et [III], le degré de coarticulation est à 100% (par défaut), étant donné que le geste d'élévation mandibulaire se réalise bien en amont de notre base temporelle.

3. RESULTATS
3.1. Opposures
L'opposition entre [Ia] et [IIa] (fig. 1), se fait en débit normal aussi bien par la phase que par la base temporelle avec une différence de 22% en moyenne pour la phase LON et 83 ms en moyenne pour la base (respectivement, t = 16.23 et t = 30.15, significatifs à p ≤ 0.05; même seul pour les suivants). Lorsqu'on augmente le débit, cette opposition de phase et de base se maintient avec une différence de 24% en moyenne pour la phase et de 67 ms pour la base temporelle (respectivement, t = 6.57 et t = 21.80).

Quand on oppose [Ii] à [IIi] (fig. 2) en débit normal, on remarque une différence sur la phase de 47% en moyenne (t = 8.12). La différence, en moyenne de 82 ms, entre les deux bases temporelles est, bien entendu, significative (t = 20.59). En débit rapide, cette opposition des classes phonétiques se fait plus que sur la base temporelle, avec une différence en moyenne de 55 ms (t = 21.80).

En débit normal, l'opposition entre les classes [Ii] et [IIi] (fig. 3) se produit sur la phase RON et sur la base temporelle : la différence est, en moyenne, de l'ordre de 29% sur la phase et de 82 ms sur la base (respectivement, t = 6.51 et t = 18.75).

L'opposition n'est maintenue en débit rapide que grâce à la base temporelle, avec une différence de 55 ms en moyenne entre les deux classes (t = 17.89).

L'opposition entre les classes [III] et [III] (fig. 4) se réalise seulement sur la base temporelle. En débit normal, on constate une différence de 70 ms en moyenne entre les deux classes phonétiques (t = 17.73).

Cette différence est réduite en débit rapide : elle n'est plus en moyenne que de 40% (t = 12.38).

Une tendance générale est que les classes des géminées dérivent vers les simples lorsque la tâche devient plus complexe (voir fig. 1,2,3). Il est bien connu – dans le cadre des transitions de phase ou du paradigme de la Synergétique réactualisé par [8] –, que les structures complexes tendent vers des structures simples lorsque la tâche devient plus difficile (ici l'augmentation de la vitesse d'élocution). C'est un processus de simplification, démontré par les changements historiques (CC→C, CV→C, VCV→V), que nous attendons pour d'autres géminées de l'arabe marocain (13) – mettre en évidence sur le plan acoustique.

3.2. Coarticulation
L'analyse des résultats nous montre qu'on ne peut pas systématiser une seule stratégie de coarticulation pour la réalisation de l'opposition simple vs. double face à la variation du débit. Lorsqu'on oppose [Ia] à [IIa] (fig. 1) en débit normal, on constate que les doubles
sont plus coarticuliertes que les simples (=50% vs. ≈30%), ce qui signifie que [i] double est moins résistant en débit normal. En débit rapide, la tendance s'affermit structurellement la même (=30% vs. ≈50%).

Pour l'opposition [i] vs. [i] (fig. 2), nous constatons le phénomène inverse. En débit normal, les simples sont plus coarticuliertes, donc moins résistantes, que les doubles (=60% vs. ≈20%). En débit rapide, les deux classes sont des pourcentages comparables de coarticulation : elles se confondent à environ 70%, ce qui est un taux assez important de coarticulation.

Mais l'examen des classes [ai] ~ [all] (fig. 3) nous révèle, qu'en débit normal, ce sont les doubles qui sont le plus coarticuliertes (à environ 65%). Ici, comme pour les classes [ala] et [al] (fig. 1), les doubles sont moins résistantes à la coarticulation en débit normal. Cependant en débit rapide, le degré de coarticulation est semblable pour simples et doubles, avec un taux de résistance moindre (=45% vs. ≈35%).

Enfin, on peut dire de manière générale que les classes [il] et [ill] (fig. 4) sont de loin les plus coarticuliertes, avec un minimum de 100% de coarticulation (par défaut, cf. supra).

D'après ces données nous pouvons poser que le [i] a une intransigence puissance de "pénétration" plus élevée que le [a], ce qui révèle particulièrement le comportement des gémènes en débit normal. Nous pensons donc, comme [10] et [12], que la consonne [l] semble mieux épouser la hauteur mandibulaire de la voyelle [i], ce qui expliquerait son taux élevé de coarticulation.

4. CONCLUSION

En conclusion, nous pouvons souligner que le phénomène de résistance à la coarticulation anticipante des [i] en arabe marocain peut comporter aussi bien une composante large partagée par d'autres langues qu'une autre plus ou moins spécifique.

C'est ainsi, en commençant par l'aspect spécificité, que l'on peut valider, d'une part, le modèle time-locked, mais seulement en timing relatif, avec ceux de nos résultats qui montrent une stabilité des phases malgré la variation de la vitesse d'élocution. C'est le cas pour la classe [alla], où l'accélération maximale sur produit à intervalle proportionnellement fixe par rapport au début acoustique de la voyelle suivante. Mais d'autre part, nous pouvons évoquer le comportement "orthodoxe" - par rapport au modèle look-ahead - de la classe phonétique [iil] : elle n'invaliderait pas un tel modèle, car la coarticulation anticipante est maximale, que la consonne soit simple ou géménée.

Remerciements à G. Feng pour son aide en traitement des signaux ; à C. Abry pour ses commentaires.

REFERENCES


Figs. 1-4. Ellipses of dispersion (at 90%) for the ensemble of the classes phonétiques dans les deux débits, normal (DN) and rapide (DR). En abcisse : durée en ms de la base temporelle (VVT-VVO) ; en ordonnée : pourcentage de phases (LON ou RON) en fonction de la base temporelle (cf. texte).
THE EFFECT OF VOWEL HEIGHT ON PATTERNS OF ASSIMILATION NASALITY IN FRENCH & ENGLISH

A. P. Rochet and B. L. Rochet

University of Alberta, Edmonton, Alberta, Canada

ABSTRACT

Assimilation nasality patterns for French and English vowels were studied as subjects (15F:15E) spoke them in CVC, NVN, NVC, and CVN contexts. Corresponding oral and nasal acoustical signals were transduced by a Nasometer, stored separately on FM tape, low-pass filtered and digitized. The vowel portion of each digitalized signal was isolated, converted to rms values, and the degree of nasalance established by comparing rms amplitudes of corresponding oral and nasal data across the vowel's duration. High vowels in both languages exhibited a higher degree of assimilation nasality than lower vowels, although for a given vowel height, French exhibited less assimilation nasality than English.

1. INTRODUCTION

The assimilation of nasality onto vowels in the context of nasal consonants has been documented by research using various methods (acoustical, acoustic, biomechanical, perceptual). Furthermore, it has been suggested that differences in degree of assimilation nasality exist among vowels as a function of tongue height. The research reported here used acoustic analog recording and digital analysis techniques to quantify and compare assimilation nasality patterns in French and English as a function of vowel height.

2. PROCEDURES

2.1. Subjects/Speech Sample

Subjects were 30 young adults, 15 native speakers of Standard French and 15 of Canadian English, with normal hearing, voice qualities and articulation patterns. They read aloud words in which English vowels /i, I, ɛ, ə, u/ and French /i, ɛ, a, u, y/ were embedded in the contexts CVC, NVC, NVN, and CVN, where V= one of the target vowels, C= a non-nasal obstruent and N= /n/ or /ŋ/. Each word was produced as the terminal item in a carrier phrase, e.g., "A half keep"; or "Neuf quines.

2.2. Data Collection/Analysis

The oral and nasal acoustical signals corresponding to subjects' productions of the test words were transduced separately by means of a Kay Elemetrics Nasometer 6200. The Nasometer microphone signals were recorded simultaneously on separate channels of an FM tape recorder, low-pass filtered at 4.8 kHz and digitized at 10 kHz via CSpeech [5]. The vowel portion of the oral and nasal component of each digitized signal was isolated, converted to an rms value, and the degree of nasalance computed by comparing rms amplitudes of corresponding oral and nasal data across the duration of the vowel in 5 ms steps, according to the formula: % nasalance = nasal rms/oral rms) x 100. Data analysis focussed on three dependent measures: 1) degree of nasal resonance, using 0.5, or 50% nasalance as an arbitrary threshold, 2) percentage of the vowel with nasalance values above 0.5, and 3) absolute duration (msec) of the vowel with nasalance above 0.5.

3. RESULTS

3.1. CVC data, French & English

Figure 1 depicts the percentage of CVC cases without significant nasalance (i.e., <0.5). In the majority of cases, nasalance levels did not exceed the arbitrary threshold of 0.5, although the number of cases in which this was true was smaller for /I/ in both languages.

3.2. NVN data, French & English

Figure 2a graphs the percentage of NVN cases where nasalance was above the criterion of 0.5 at both ends of the vowel (including cases where it dipped below 0.5 in the middle). Figure 2b displays only those cases where the entire duration of the vowel exhibited nasalance levels above 0.5. Both languages show a noticeable difference between /I/ and /a/, with /I/ exhibiting a higher sustained nasalance level throughout the vowel's duration. In French, more clearly than in English, /a/ occupies an intermediate position between /I/ and /a/ with respect to this phenomenon.

3.3. NVC data, French & English

Figures 3a and b illustrate the patterns of carry-over nasalance in the NVC context for /i, ɛ, a/ in French and English. A larger percentage of the vowel exhibits the carry-over effects of the preceding nasal consonant when the vowel is high than when it is low, and the percentage of the vowel exhibiting the nasal consonant's influence is roughly the same in French and in English (3a). The absolute durations of the nasalized portions are shorter, however, in French (3b).

3.4. CVN data, French & English

Figures 4a & b illustrate the percentages and absolute durations of the French and English target vowels that are nasalized in anticipation of the final nasal in the CVN context. High vowels /I/ and /a/ tend to nasalize portions nasalance levels greater than 0.5 across a larger percentage of their durations compared to mid or low vowels in both languages, though French always reveals less anticipatory nasalance than English for the vowels considered.

3.5. NVC data, English

Figure 5a & b illustrate the carry-over nasalization patterns among English vowels /i, u, ɛ, a/ and French /i, ɛ, a/. The carry-over effects of the initial nasal consonant influence a larger portion of the high vowels than of the others, and the effect is consistent whether one considers the percentage of vowel nasalized (5a), or the absolute duration of the nasalized segment (5b).

3.6. CVN data, English

Figures 6a & b compare anticipatory nasalization patterns among the English vowels. As in the NVC context, the high vowels exhibit more influence of the nasal consonant than the other vowels. A comparison of the proportional carry-over and anticipatory data for the English vowels (5a & 6a) reveals that, except in the case of /I/ where the proportions of the vowel nasalized are comparable in both the NVC and CVN contexts, for the other vowels the amount of nasalization is greater in the anticipatory situation.

3.7. NVC data, French

Figures 7a & b compare carry-over nasalization patterns among French vowels /i, ɛ, a, u, ɛ, a/. The patterns for these vowels are similar to those for English with respect to vowel height: The initial nasal consonant influences a larger portion of the high vowels than of the others, and the effect is consistent for the percentage of the vowel nasalized (7a) and the absolute duration of the nasalized segment (7b).

4. SUMMARY/DISCUSSION

4.1. For these 30 subjects in the contexts examined, French always exhibited less assimilation nasality than English. This was true for all vowels considered and for the NVC and CVN contexts. These results support the validity of Delattre's pedagogical recommendation to English speakers of French that they prevent premature anticipation of the nasal consonant in the CVN context in order not to nasalize the vowel [3]. These data do not, however, support Delattre's assertion that French vowels followed by a nasal consonant remain oral throughout their duration.
4.2. The degree of vowel nasalization in a nasal consonant context varied with the height of the vowel. High vowels exhibited more assimilation nasality than low vowels. This correlation is very systematic in the French vowel data; it also applies to the English vowel data, although less systematically. The apparent contradiction between these results and those of Clumeck [1] may be related to his use of the term "nasalized" to describe articulatory gestures of the velum, and the fact that the biomechanical behavior of the velopharynx cannot be assumed to be monotonically related either to the perception of nasal resonance or to the acoustical consequences of nasal coupling during speech production. The perception or measurement of nasal resonance is ultimately a function of the relative acoustical impedances of the oral and nasal cavities, as well as the formant frequency values of the vowel in question. The spectral envelopes of /u/ and /a/ are markedly affected by small nasal coupling, whereas vowels with a more open tract configuration are much less affected by small degrees of coupling [2]. This is consistent with listeners' judgements that the amount of nasal coupling necessary for the perceptual identification of nasalization was almost three times as much for low vowels as for high ones [4].

5. CONCLUSIONS
5.1. The difference in the degree of nasalization between French and English may be related to the fact that English does not have phonemic nasal vowels and therefore can "tolerate" higher levels of assimilation nasality.
5.2. The higher levels and longer durations of assimilation nasality observed for the high vowels in both French and English are related to the acoustical impedance of the vocal tract for the production of these vowels. There is no obvious articulatory or physiological reason for the earlier lowering of the velum observed by Clumeck [1] for low vowels in the CVN context. It may simply be that such lowering does not have an undesirable acoustical effect, and does not lead to excessive perceptible nasalization of these vowels. Later lowering of the velum for high vowels, however, may ensure that their spectral envelopes are not too drastically affected by extraneous nasal resonance.

5.3. Further research on assimilation nasality is recommended by means of simultaneous multidimensional sampling methods that could consider biomechanical, perceptual and acoustical parameters of vowel production without losing sight of the phonemic characteristics of the languages sampled.

6. REFERENCES
FRICATIVE CONSONANTS AND THEIR ARTICULATORY TRAJECTORIES

Celia Scully, Esther Georges, Eric Castelli

University of Leeds, UK
ICP, INPG, Grenoble, France

ABSTRACT
Articulatory paths relevant to the production of some fricatives are related to the changing acoustic patterns as seen on spectrograms. Two techniques are used: aerodynamically derived area (A) traces and electropalatography (EPG) contacts. [pVVC] sequences are analysed.

1. INTRODUCTION
Better descriptions are needed for the production of fricative consonants: Fricatives in speech-likelines sequences are characterised not only by their quasi-static spectra but also by rapidly changing acoustic patterns. The latter seem to be essential for the identification of /f/ versus /θ/ and /v/ versus /ð/ [2], [3]. Phonemically, fricative consonants and adjacent vowels are considered as separate entities but in the processes of speech production there is no clear boundary between them. Between segments clearly associated with either consonants or vowels there are regions of rapid change; in these, there are changing combinations of the acoustic sources - voice, aspiration noise (generated just above the glottis) and frication noise (generated downstream from a vocal tract constriction) - as well as rapidly changing formant frequencies. Inter-articulator coordination and the form of transitions for individual articulators are both important in determining how sources and filters covary across this boundary region.

The study reported here is an exemplification of part of a larger study (Grenoble, Southampton, Leeds), based on multiple analyses for two speakers. A studio recording made by the speaker provides cueing, so as to match the speaking style and rate across data gathered on different occasions and in different laboratories. Articulatory paths in the natural speech are to be copied in models of speech production [4], [1], [6]. Analysis-by-synthesis will be used to obtain good aerodynamic and acoustic matches between the natural and the simulated speech. The aim is to characterise, as general rules, the production and acoustics of the speakers' fricatives.

2. APPROACH OF THE STUDY
This paper focusses on [s] produced in phonetically controlled [V-V] contexts by one of the speakers, a woman speaker of General American English. Sequences such as [pi'si pi'si i...] produced on a single expiratory breath allowed subglottal pressure (PSG) to be estimated.

Two techniques for the estimation of vocal tract articulation relevant to the production of alveolar fricatives are included here. First, aerodynamic parameters are used to give an Area (A) trace, indicating the cross-section area of the alveolar constriction of the vocal tract; secondly, electropalatography (EPG) is used to show the regions of contact between the tongue and the hard palate. Articulatory paths estimated by each method in turn are time matched to spectrograms from simultaneously made recordings. In this way, part of the detailed articulatory-to-acoustic mapping is studied. An additional aim is to demonstrate that the two methods are consistent and complementary.

3. AREA (A) TRACE
3.1. Method
This is a parameter for one of the two major constrictions of the respiratory tract, the other being the glottal area. Volume flowrate of air through the mouth, U (in cm³/s) and oral air pressure, P (in cmH₂O) are combined, using the orifice equation with an empirical constant k = 0.00076 to give:

\[ A = k \cdot U \cdot P^{0.5} \]

The methods have been described elsewhere [5], [7]. The A trace is not the true value of the minimum cross-section area for the alveolar ridge portion of the vocal tract, and may be expected to depend on the taper angle into and out of the constriction and its length. Total airflow through noise as well as mouth is recorded, but checks showed that all the airflow was through the mouth for the sequences analysed. Oral air pressure is taken as pressure drop across the constriction but actually includes any pressure drop across the teeth and lips also. The method has the advantage that it links articulation and aerodynamics in a way that is internally consistent and consistent with the descriptive framework of one of the composite models [7] in which it is to be used.

3.2. Results
Figure 1 shows an example of the aerodynamic traces, with some of the simultaneously obtained acoustic traces.

Figure 1. Articulatory, aerodynamic and acoustic tracess as functions of time, for the third in a series of [(p)]s[i](p). From top to bottom: Area A, with a line at 0.2 cm²; I.L. H.P. filtered at 3.9 KHz, I.L. H.P. filtered at 500 Hz; Oral (total) volume flowrate of air U; oral air pressure P.

An auditory check confirmed that this
was an acceptable example of [i'si]. Figure 2 shows the relationship between the articulatory trace and the acoustics.

Figure 2. Area (A) trace from Figure 1, inverted, combined with spectrogram (frequency versus time) pattern features of the identical utterance. Time scales are matched with voice offset and onset aligned.

It may be seen that the 0.2 cm² threshold for the A trace goes beyond the domain of friction noise to include the boundary regions. The inverted form of the A trace may match the changing spectral pattern for the friction noise, but the evidence is not conclusive.

The peaks of airflow seen in Figure 1 and shown by the dotted lines in Figure 2 almost coincide with voice offset and voice onset shown by the dashed lines. The airflow peaks are located at the boundary region, between the frication noise segment and the vocoid, where the acoustic sources including voice and aspiration noise are changing rapidly.

4. EPG DATA
4.1. Method
EPG data for the fricatives in the same [...]V-Vp...] context are analysed as follows: the number of contacts is determined for: the first, second and third lines of contact (front, shown by a solid line) and the fourth to the eighth line (back, shown by a dotted line). The results are plotted on a grid which shows time vs number of contacts. The two resulting traces represent changes in the amount of contact between tongue and palate for the front and the back of the mouth; transitions from and to vowels are investigated.

4.2 Results
Figure 3 shows the relationship between the articulatory traces and the acoustics for a representative example of [i'si].

A threshold (indicated by the arrow in Figure 3) was chosen to define the quasi-static segment observed for all of the fricatives analysed so far. It was found that this threshold corresponds rather closely with the friction noise segment. The match is excellent in this example.

The contact for the back portion of the tongue, however, decreases during the frication noise segment. This may perhaps indicate a lowering of the tongue dorsum similar to that observed by Wood [8] on X-ray traces. As Wood suggests, this may enhance the acoustic separation of front and back cavities. Referring back to the noticeable change in spectrum during this portion of a different production of the same fricative, seen in Figure 2, this explanation seems a plausible one. The speaker appears to be tuning the front cavity resonance.

Figure 3. EPG contact traces combined with spectrogram pattern features (solid line for front contacts, dotted line for back contacts).

5. CONCLUSION
The area trace seems to define a wider domain than does the front contact shown by the EPG data. The area trace may reflect changes in the overall tongue configuration such as that discussed above, and possibly mouth outlet shape also. This interpretation of the two kinds of articulatory data will be tried out in the modelling.

6. ACKNOWLEDGEMENT
The work was funded in part by a collaborative EC SCIENCE award: CEC-SCI * OI47C (EDB).

7. REFERENCES
ABSTRACT
The present study provides data on degree of breathiness produced by Dutch male and female speakers in a neutral and an emotive context. The acoustically defined parameter DH indicates significant differences between male and female speakers in both contexts. There is an increase in breathiness for either population from neutral to emotive context. Analysis of average F0 and average intensity levels show decreased values for both male and female speakers in the emotive condition as opposed to the neutral condition.

1. INTRODUCTION
Breathiness can be defined in various domains. In articulatory terms breathy phonation arises from an incomplete adduction of the vibrating vocal folds and can lead to an increase of the average airflow of up to 60% in comparison to non-breathy (vowel) production. Extreme breathiness can be indicative of pathological speech and function as a perceptual marker of various laryngeal disorders. On a less extreme note breathiness can impair the general perceptibility and understandability of speech and convey the impression of increased monotony. Various acoustic correlates can account for breathiness 11,13,1. Due to the incomplete vocal fold closure during phonation of a breathy vowel there is considerable leakage of air through the glottis which causes interspersed noise at higher

frequencies of the acoustic spectrum. Presumably in connection with a slackening of the folds a slight lowering of F0 has been observed for breathy vowels and probably most notable, there is a fairly consistent increase of the amplitude of the first harmonic in relation to the second as opposed to an opposite amplitude relationship between the first two harmonics of a non-breathily phonated vowel, see Fig. 1.

2. OCCURRENCE OF BREATHINESS AND EXPERIMENTAL OUTLINE
In numerous languages breathiness is used to form phonemic contrasts (for references see 12). These languages, however, are not our present area of interest, nor is breathiness as a marker of pathological speech. Evidence has been adduced (ibid) that female speakers of two British accents consistently used a more breathy voice quality than men in ordinary speech. Although breathiness may be considered an inefficient way of voice production with a number of communicative limitations, the claim has been put forward that, consciously or unconsciously, women use breathiness as a means of communicating arousal, intimacy, or, in other words, to sound more "sexy".

Our present experiment was set up to a) compare Dutch breathiness data of male and female speakers in an ordinary speaking mode those of English and b) to investigate whether in an emotive context there would be an increase in breathiness by either speaker sex.

As corollary variables to DH, average F0 and average intensity of the vowels under investigation will be considered as well.

3. EXPERIMENTAL PROCEDURE
13 male and 13 female speakers participated in the experiment. Their ages varied between 19 and 38 years. They belonged - either as staff or students - to the University of Utrecht and were speakers of the Dutch equivalent of RP.

The vowel to be analysed was decided to be an /a/ since this open vowel's first formant is high enough to be of no influence on the first harmonic. Monosyllabic words containing the vowel /a/ were embedded in unpretentious sentences which, on their part, were combined to form an unpretentious piece of running prose. Due to their monosyllabic all stimulus words carried lexical stress. In addition a semantically intimate passage containing numerous /a/ was selected from a sultry-romantic piece of fiction in order to simulate an emotive context.

Speakers were instructed to read the first text in an ordinary and the second text in a sexually charged way.

Recordings were made individually in a sound-proof room using a Revox B77 mkII tape recorder and a Sennheiser microphone. A mouth-to-microphone distance of 30 cm was used. The input volume control was held constant and subjects were given some practice time.

Data were further processed digitally. Per reading mode and subject 12 35 ms steady state portions of the /a/ vowels were excised and relative amplitudes of the first two harmonics, and F0 and amplitude of the steady states were established.

4. RESULTS
4.1. RELATIVE AMPLITUDE OF HARMONICS
Table 1 shows the average values and corresponding SD's of DH produced by male and female speakers in the two reading modes. A negative DH value indicates that the amplitude of the first harmonic is lower than that of the second harmonic and v.v. According to Bickley (1982) a negative DH value is the consequence of breathy phonation.
difference in breathiness between the female-ordinary condition and the male-emotive condition which means that female speakers used the same degree of breathiness in reading the ordinary text as did male speakers in reading the emotive text.

4.2. FUNDAMENTAL FREQUENCY
Results of the F0 analysis of the measured /a/ steady states are shown in Table II.

As can be seen there is a decrease in fundamental frequency for the emotive reading text for either sex, however, do not reach the level of significance.

Interindividually, however, a significant positive correlation exists between F0 and DH.

Table II: F0 in Hz for ordinary and emotive reading mode

<table>
<thead>
<tr>
<th></th>
<th>female speakers (n=13)</th>
<th>male speakers (n=13)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ordinary</td>
<td>emotive</td>
</tr>
<tr>
<td>x</td>
<td>214</td>
<td>204</td>
</tr>
<tr>
<td>SD</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

4.3. AMPLITUDE
For either sex there is a slight decrease of amplitude values for the emotive context in comparison with the ordinary context (female: 2.9 dB; male: 1.9 dB). Differences are insignificant.

5. DISCUSSION AND CONCLUSION
Female speakers produce significantly more breathiness in comparison with male speakers in the ordinary reading mode as well as in the emotive reading mode. The degree of breathiness produced by female speakers in the ordinary context turns out to be even equivalent to that of male speakers in the emotive context. We found a decrease of both F0 values and amplitude values for the emotive context for male as well as female speakers, but these differences fail to reach the level of significance.

As stated in our introductory section, a breathy spectrum contributes to perceptual limitations. Why, Henton and Bladon I2I ask themselves, should women adopt articulatory postures that render their own speech less efficient in communicative terms? The answer to this question lies, according to these authors, in the ethological-sociolinguistic domain: "...women imitate the voice quality associated with arousal. ...A breathy woman can be regarded as using her paralinguistic tools to maximize the chances of her achieving her goals, linguistic or otherwise.

With all due respect we would like to regard this explanation with some caution. First of all perceptibility of speech is affected only in extreme cases of breathiness. Secondly, not only was DH in the afore-mentioned experiment the only acoustic correlate considered to indicate breathiness, whereas other parameters probably deserve consideration as well, but breathiness in its turn is certainly not the only characteristic of a "sexy" voice.

As to voice source characteristics, it is generally assumed that female speakers have a greater open quotient which implies that they produce more breathiness for physiological reasons.

In connection with our tentative F0 - DH correlation data we suggest that more research should be addressed to the question of whether a systematic relationship can be found between pitch and breathy phonation on an Interindividual level.

6. REFERENCES

I WISH TO THANK MY STUDENTS ASTRID SMEETS AND SONJA SENGERs FOR THEIR CONTRIBUTION TO THIS PAPER.
ARTICULATORY GENERALIZATIONS IN ACOUSTIC PHONETIC RESEARCH: A COMPARISON OF DATA FROM FRENCH AND ENGLISH

Sarah N. Dart

Department of Linguistics, UCLA, Los Angeles, U.S.A.

ABSTRACT
Simultaneous articulatory and acoustic data were recorded for 21 French speakers and 20 English speakers uttering phrases containing the coronal consonants /t,d,n,l,s,z/. It was found that, in both languages, individual variation in articulation of these consonants makes it difficult to make precise language-specific generalizations in terms of both place of articulation and apicality. The formant patterns in the acoustic signal, however, are much more homogeneous and suggest that the difference in consonant production in these two languages lies more in the general shape of the tongue body behind the constriction than in the placement of the constriction itself.

1. INTRODUCTION
The coronal consonants of French and English have been claimed to be articulated differently in terms of place of articulation and apicality. For example, French coronal stops are regularly described as dental, either with the tip of the tongue on the upper incisors (apical), or with the tip down behind the lower incisors and the blade making contact (laminal). English coronals, on the other hand, are usually said to have an apical alveolar constriction. Such information forms the basis not only of foreign language pronunciation instruction, but also of acoustic phonetic analysis, where data from acoustic recordings of speakers of the same language are assumed to originate from a homogeneous set of articulations. The present study seeks to discover to what extent the articulation of an individual can be predicted by language community affiliation. Is the precise point of articulation as given on a traditional consonant chart really crucial to the pronunciation of a given language or are there other factors which are more important?

The articulatory data presented here is in the form of palatograms and linguagrams taken by the direct method to ascertain the point of contact on the upper surface of the vocal tract, as well as the part of the tongue used to make the constriction. Audio recordings were also made synchronous with the palatograms and linguagrams, in order to be certain that each given acoustic signal corresponded to an articulation with known articulatory characteristics. Data from 21 French speakers (northern standard pronunciation) and 20 English speakers (west coast American) were recorded of the consonants /t,d,n,l,s,z/ in both word-initial and word-final position, in the environment of a low vowel ([ə]) in English, [a] in French.

2. ARTICULATORY DATA
Place of articulation was determined from the palatograms in the manner described in detail in Dart [2], briefly as follows: if the vertical surface of the back of the upper central incisors was contacted, either completely or partially, the articulation was called dental; alveolar articulations were those where the most forward part of the contact was in an area extending from the base of the teeth to approximately 5 mm back; and the segments are apical dental, it being difficult for a speaker with normal dentition to produce a purely apical dental, without the blade of the tongue also contacting the alveolar ridge. Some authors have also claimed tip-down laminal dental articulation for these segments and 12.7% of the data support this. There remain, however, 41% of the data left unaccounted for, that is all those tokens which are not dental. Clearly, a number of French speakers articulate farther back than was previously supposed.

The point of view of the sources consulted on fricative articulation was more open to variation, with both dental and alveolar articulations mentioned (although only one source allowed for both possibilities). Most sources, however, stated quite firmly that French /s/ and /z/ were laminal. It is clear from the table that, although the majority of tokens were indeed laminal, still nearly a third were apical, and thus not accounted for by the descriptions.

In English, 59.6% of the data for /t,d,n/ are, indeed, apical alveolar as predicted. 11.7% of the tokens are also apical, but either dental or postalveolar, and 17.6% are also alveolar, but use a different part of the tongue. A total of 17.6% of the tokens are dental and 28.5% are either laminal or apicolaminal.

The fricatives /s/ and /z/ are usually said to be either apical or laminal alveolar in English, were indeed divided between...
these two ways of articulating, the laminal predominating with over half (57.5%) of the tokens. Again, most of the tokens were alveolar or postalveolar (77.5%).

As it turns out, the English laterals are far more likely to be dental than their French counterparts, going against the neat organization of the consonant charts, which usually put /l, n, l, s, z/ in the same column. Exactly half of the /l/ tokens were dental in the English data (as compared to 4.8% of French tokens), in spite of the general acceptance in the literature that such English segments should be alveolar, just as the French are assumed to be dental. Even the apical articulation of the lateral, which was nearly universal for the French speakers (95%) was less strong in English (68%), the quintessential "apical" language. It seems, then, that /l/ need not necessarily share the articulatory characteristics of the other coronal consonants in any given language.

The articulatory data thus shows that, although the articulation of these consonants may be predicted in a general way for the majority of speakers, the variation is such that one cannot assume an articulation to be of a certain type only on the basis of the native language of the speaker.

3. ACOUSTIC DATA
Formant transition frequencies were measured from wide band spectrograms for all tokens: for the word-initial tokens immediately after the closure, and for word-final tokens immediately before the closure. To normalize for absolute frequency differences between speakers, the difference was calculated between the transition formant values and the average steady-state formant values of the adjacent vowel. The resulting number was used for comparison rather than the raw formant frequencies. The formant values of the steady-state vowel were comparable between the two languages except for the value of the second formant, which was higher in English.

Two general differences between French and English articulation were noted: the value of the F1 transition in French was always lower in relation to the steady-state vowel than the corresponding English value for all the coronal consonants, no matter what method of articulation was used. Similarly, the transition value of F4 was always higher in French than in English. These differences suggest different tongue shapes behind the constriction in the two languages. A lower F1 could indicate a wider pharyngeal cavity and a higher F4 a smaller sublingual cavity in front for French. In addition to these general characteristics, a specific tongue shape difference between apical alveolar articulations in the two languages was inferred from the formant data, particularly in fricatives. French apical alveolar fricatives have lower transitional F1 values and higher transitional F2 values than do apical dental fricatives, whereas the reverse is true for English. Similarly, French apical consonants have higher F2 values than laminals, whereas in English F2 is higher in laminals.

One interpretation of these facts would be to posit a differently shaped tongue behind the constriction in the apical and alveolar articulations in the two languages. The F1 and F2 evidence suggests that the body of the tongue in French is high and forward during these consonants, thereby diminishing the area of the cavity directly behind the constriction and enlarging the pharyngeal cavity. The English apicals, on the other hand, would come up to the constriction from a lower and more posterior position in the mouth, thus creating a larger cavity behind the constriction and a more constricted pharynx. Both kinds of apical alveolar articulations can be seen in the x-ray literature, as exemplified by the two tracings in Figure 1. The tracing on the top is of French /s/ (after Bothorel et al. [1]) and resembles an apical alveolar tongue position like that posited for the French speakers, and the tracing on the bottom is of English /s/ (after Subtelny et al. [3]), and has a descending tongue shape as posited for the English speakers in the present study.

Figure 1. X-ray tracings of French (after Bothorel et al [1]) and English (after Subtelny et al [3]) showing two different tongue shapes in apical alveolar /s/.

In order to explore the possibility of such an articulatory difference as that suggested by the acoustic data, additional articulatory measurements were taken from the palatographic data in conjunction with palate casts from each speaker. It was presumed that a higher tongue position would show up on the palatograms as a wider contact area behind the constriction and, indeed, such a difference seemed to be evident from the palatograms. Accordingly, the contact area from each articulation was measured inwards from the base of the first molar and this measurement given as the ratio of the contact area on one side to half of the total distance following the curve of the palate from first molar to first molar. These measurements were shown to be significantly larger in French by one factor, repeated measures analyses of variance for all apical and alveolar stops, nasals and fricatives.

4. CONCLUSION
With the abovementioned facts taken together, there appear to be language-specific characteristics affecting the formant values, which are associated with vocal tract shapes that are not fully specified by simply characterizing the segments in terms of the articulatory contact involved. The difference between French and English coronal consonant production, rather than being one of place of articulation and apicality, would seem to be better described as a difference in the overall shape given to the tongue body in the two languages.

5. REFERENCES
ABSTRACT
Compensatory tongue positioning in vowel production was examined in two conditions of lower-than-normal jaw positions (bite-block speech and loud speech), and compared to a "normal" speech condition. Tongue-palate distances in multiple productions of the German vowels /i, I, u, u, y, Y/ were measured using glossometry. The tongue compensated for the lower jaw positions in both perturbation conditions. Jaw lowering in bite-block and in loud speech did not much affect the degree of precision in tongue positioning.

1. INTRODUCTION
Comparisons of normal and perturbed speech may help understand important aspects of speech motor control. Over the past twenty years, a research paradigm has become established which addresses issues such as invariance in the control of speech gestures, adaptive abilities of the speech motor system, and the role of feedback through experiments in which normal production patterns are disrupted. By examining the behavior of unperturbed articulators, the acoustic output, and/or the intelligibility of perturbed speech, studies employing this paradigm have aimed at determining if, how, and how successfully talkers reorganize articulatory gestures.

Probably the majority of perturbation studies examined the acoustic properties of vowels produced with and without the mandible being fixed in positions that required talkers to reorganize tongue gestures in order to produce intended vowel qualities. These studies have generally shown that adults [7, 8] and children [2] compensate remarkably well for a fixed jaw even before auditory feedback can occur. The small number of articulatory studies that examined tongue shapes for bite-block vowels [4, 6, 11] indicate that intended acoustic output in bite-block speech is achieved through selective compensation, i.e. by preserving "cavity configuration(s) at points of maximum constriction" [6].

Although previous research on bite-block vowels has contributed importantly to the construction and refinement of models of speech motor control, this line of research has not made it clear whether talkers aim to achieve invariance in the acoustic, perceptual, or articulatory domain. The recently renewed interest in speech produced with loud vocal effort [9, 10] is to some extent motivated by a desire to determine the nature of talkers' "goals" or "targets".

Loud speech is similar to bite-block speech in that the jaw assumes lower-than-normal positions which, however, are not artificially induced but "natural". In the only detailed study of articulatory consequences of loud speech, Schuelman [9] found that the upper lip compensates for the lowered jaw in bilabial stop production, demonstrating that motor equivalence for bilabial closure occurs in both the "natural bite block condition" [9] and its artificial counterpart [1].

However, the acoustic properties of vowels produced with loud vocal effort, which have been examined in a number of studies (summarized in [10]), suggest that the analogy between loud and bite-block speech does not extend to vowel production, for the frequencies of F1 and F0 (but not usually the upper formants) are much higher in loud than in normal speech. The increase in F1 for shouted vowels led Traunmüller [10] to hypothesize that the tongue does not compensate for lower jaw positions in loud speech.

The present study, which compared tongue-palate distances for normal, bite-block, and loud vowels, was primarily motivated by the fact that only very few palatograms have presented direct evidence (as opposed to inferences from the acoustic output) concerning compensatory tongue positioning in bite-block vowels [4, 6, 11], and by the complete lack of published data on tongue shapes in loud vowels. Bite-block and loud vowels were compared to normal vowels to determine if and how the tongue would compensate for an artificially and a naturally lowered jaw.

This study also examined variability in tongue positioning for normal, bite-block, and loud vowels. Because most earlier studies [6, 11] used x-ray techniques, which preclude detailed analyses of token-to-token variability, very little evidence exists concerning this aspect of motor control precision for the tongue in perturbed speech (but see [4]).

2. METHODS
2.1 Subject, Material, Procedure
A male native speaker of German (age: 35 years) produced 12 tokens each of the German vowels /i, I, u, U, y, Y/ in the carrier phrase oh et byv/ habe (blocked on vowel). The vowels were produced in three conditions. In the normal (NO) condition jaw movement was unperturbed and the vocal effort was conversational (64 dB SPL). In the bite-block (BB) condition the talker's jaw was fixed in a lower-than-normal position for non-low vowels. An acrylic bite block, held between the right premolars, provided an interincisal distance of 21 mm. In the loud (LO) condition the talker produced the vowels with loud vocal effort (84 dB SPL).

Tongue-palate distances were measured using glossometry. This optoelectronic device for measuring and displaying tongue positions below the hard palate has been described previously (see [5] and references therein). Briefly, the glossometer makes use of four sensor assemblies mounted on a thin acrylic pseudopalate. Each assembly contains an LED and a phototransistor. The assemblies are positioned equidistantly along the palatal vault and are orientated perpendicularly to the occlusal plane. Sensor 1 is located just posterior of the alveolar ridge, and sensor 4 just anterior of the junction of the hard and soft palates. Infrared light emitted from the LED is reflected from the tongue's surface, detected by the phototransistor and transduced to a voltage level. The detected voltage is approximately proportional to the inverse square of the distance of the tongue from the sensor assembly.

2.2 Data Analysis
Tongue-palate distances for tokens 2-11 for each vowel in the three conditions were measured at that point within the acoustic vowel interval that best represented the endpoint of tongue movement for each token. Endpoints were selected by visual inspection of the time-varying distance traces, which were displayed together with RMS intensity on a high-resolution graphics terminal.

Articulatory compensation with respect to tongue positioning below the hard palate was considered (by way of definition)
-completely if the average unsigned tongue-palate distance at the four sensor locations differed by less than 1.0 mm for NO vs. BB or LO productions of a given vowel;
-selective if the mean tongue-palate distances in BB or LO productions at sensors located near the acoustically critical maximum constriction for a given vowel were within the range of the standard deviation (SD) associated with the mean for the NO tokens at those sensor locations;
-partial if the tongue compensated for the lowered jaw, but did not compensate completely or selectively.

Overshoot and undershoot refer to partial compensation with higher-
1 Reasons for selecting this criterion to determine tongue shape overlap are given in [5].
than-normal and lower-than-normal tongue positions, respectively. Finally, in zero compensation, the tongue does not compensate for the lowered jaw in BB and LO speech. Variability in tongue positioning was assessed in terms of the SDs associated with the multiple productions of NO, BB, and LO vowels.

3. RESULTS

3.1 Tongue Positions

The most important result was that in the production of all six vowels, the tongue compensated for the lower-than-normal jaw position in both BB and LO speech. However, the tongue was lower in LO than BB speech at all four sensor locations for five vowels, suggesting that the tongue did not compensate as much for jaw lowering in the "natural" as in the artificial BB condition. The exception was \( /Y/ \) with overlapping tongue configurations in the BB and LO conditions.

Complete compensation by the tongue for jaw perturbation was observed in only two instances: For \(/i/\) in the LO and for \(/U/\) in the BB condition. Compensation was selective for \(/i/\) in the BB condition, for \(/y/\) in the BB and LO conditions, and for \(/y/\) in the BB condition.

Partial compensation (undershoot) was observed for \(/y, y, u, U/\) in the LO and \(/Y/\) in the BB condition. Undershoot relative to NO tongue positions, which increased monotonically from anterior to posterior sensor locations, was small for \(/Y/\), medium for \(/y/\) and \(/U/\), and large for \(/u/\). Surprisingly, undershoot for \(/u/\) and \(/U/\) in the LO condition was largest at sensor 4, which is located close to the acoustically critical maximum constriction for these back vowels at the velum. Results for perturbed \(/u/U/\) productions differed from all other results in that undershoot in LO speech contrasted with overshoot (at the posterior sensors) in BB speech.

3.2 Variability of Tongue Positioning

The most important result concerning variability of tongue positioning in the three conditions was that perturbed vowels were produced with uniformly more or uniformly less precise tongue gestures than NO vowels. The SDs associated with the multiple productions of the six vowels averaged 0.84 mm in the NO, 0.93 mm in the BB, and 0.77 mm in the LO condition. Tongue positioning for \(/i, i, y, y/\) was slightly more variable in the BB than the NO condition (SDs were 0.1 - 0.2 mm larger), but variability did not differ for \(/u/\) across these conditions. Token-to-token variability was slightly larger in the LO than the NO condition for \(/i, I, U/\) (SDs were 0.1 - 0.2 mm larger), did not differ for \(/y/\), and decreased for \(/u/\) and \(/Y/\) (by 0.3 mm and 0.6 mm, respectively).

The most conspicuous result was that for all vowels and all conditions, SDs increased monotonically from anterior to posterior sensor locations. This front-to-back increase in variability was observed irrespective of whether the acoustically critical maximum constriction was in the prepalatal \(/i, i/\), palatal \(/y, y/\), or velar \(/u/\) region. It may be of some interest to note that tongue positioning for each of the nominally tense vowels \(/i, y, u/\) was more variable than for its nominally lax counterpart \(/I, Y, U/\) in all three conditions.

4. DISCUSSION

The single-subject experiment reported here showed that the tongue compensated for a lowered jaw in both BB and LO speech, and that both conditions of jaw perturbation did not importantly affect the precision of motor control for the tongue.

Results of previous BB studies led to the expectation that articulatory compensation by the tongue in BB speech would be selective or complete. The present results for four \(/i, I, y, y/\) of the six vowels examined conformed to this expectation. However, tongue positions for \(/y, u/\) in BB speech did not overlap with NO tongue positions or maintain NO tongue-palate distances near the acoustically critical maximum constriction. Preliminary acoustic analyses of the vowels examined in the present study indicated that partial compensation for \(/y/\) (undershoot) and \(/U/\) (overshoot) did not result in changes in acoustic output that one might expect given the differences between NO and BB tongue positions below the hard palate. This suggests that compensation for the lowered jaw in the BB production of \(/Y/\) may have occurred in an area of the vocal tract not registered by the glossometer. The hypothesis being tested for LO vowels was that the tongue would not compensate for the "natural bite block". This hypothesis, which Traunmüller [10] based on the acoustic properties (increase in F1) of LO vowels, was not supported. The present experiment showed that compensation by the tongue for a lowered jaw in LO speech may be partial \((y, y, u, U/\), selective \(/I/\), or even complete \(/y/\). This suggests that motor programming in both LO and BB speech involves reorganization of tongue positioning to achieve precisely defined articulatory goals that are not necessarily (as for \(/y/\) in LO speech) the same as in NO speech. The lower tongue positions in LO than in NO speech for four of the six vowels examined may have been affected to increase F1, so that the perceptually important distance between F1 and the increased F0 in LO speech would be maintained for a given vowel irrespective of vocal effort (see [10]).

Degree of precision in tongue positioning did not differ much across the three conditions. The SDs associated with multiple productions of NO (0.84 mm), BB, and LO (0.77 mm) vowels were of approximately the same magnitude as the mean SD for the complete set of NO German vowels (0.78 mm [3]), the complete set of NO English vowels (0.81 mm [5]), and five Spanish and English vowels spoken normally (0.76 mm) and with a BB (0.89 mm) [4]. These earlier studies suggested that neither vowel inventory size [4] nor mechanisms used to differentiate large vowel inventories [3] affect variability of tongue positioning. The present results corroborate and extend Flege's [4] BB study by showing that both artificial and natural jaw perturbation need not importantly affect degree of precision in tongue positioning.

(Research supported by grant NS20963-04 from the U.S. National Institutes of Health to the second author)

5. REFERENCES


A MODEL FOR THE DISCRIMINATION OF PURE TONE PITCH.

Alain de Cheveigné

Laboratoire de Linguistique Formelle, CNRS - Université Paris 7, France.

ABSTRACT

This paper presents a model of auditory processing that can account for the very small frequency difference limens observed psychophysically for pure tones. In a first step, an autocoincidence histogram is calculated from nerve-fiber channels synchronized to the pure tone, according to a model similar to that of Licklider [3, 4, 5]. In a second step, this histogram is "folded", resulting in a "narrowed autocoincidence histogram". The peak of this narrowed histogram is narrower than that of the autocoincidence histogram, and its width depends on stimulus duration in a way similar to frequency difference limens.

1. INTRODUCTION

Listeners can discriminate differences in the frequency of pure tones as small as 0.2% [1]. Thresholds get larger as stimuli get shorter, but discrimination remains good even when the stimuli contain only a few cycles. Moore [1] argued that the thresholds are too low to be compatible with a place mechanism of frequency discrimination based on the differences in intensity that might arise when the excitation pattern for a tone is shifted along the basilar membrane. They would be compatible, on the other hand, with a time domain mechanism. Based on this assumption, Goldstein and Srulovicz [2] proposed a theory that predicts thresholds under the hypothesis of optimum processing of interspike intervals. Goldstein and Srulovicz noted that information from as few as nine fibers is sufficient to account for discrimination thresholds. Since many more fibers are available for processing, performance must have other limits, perhaps due to the actual neural processing mechanism. The question arises as to whether such processing has the same behavior as optimum processing. It is therefore of interest to examine candidate processing models with respect to pure tone frequency discrimination. One such model is that of Licklider [3, 4, 5], based on the autococincidence of nerve fiber discharges (see also [6, 7, 8, 9, 10]). If we assume this particular model, can we still predict discrimination thresholds?

In this study it is found that a) the basic autococincidence mechanism of Licklider's model does not adequately predict performance, but b) it can be followed by a second stage of processing, described by a "narrowed autococincidence histogram" (NAC), to form a model that predicts thresholds similar to those observed psychophysically.

2. DISCRIMINATION THRESHOLDS FOR PURE TONE PITCH

Moore [1] measured frequency difference limens for pure tones as a function of frequency and stimulus duration. His data are plotted in Fig. 1. At all frequencies, thresholds tend to be smaller for longer stimuli. Discrimination gets better as frequency increases, up to 2 kHz. For the lowest three frequencies there is a zone of durations for which threshold varies approximately as the inverse of stimulus duration. These frequencies are in the region for which a time-domain frequency analysis mechanism such as Licklider's is in principle applicable.

3. AUTOCOINCIDENCE MODEL

In Licklider's model [3, 4, 5], patterns of discharge within auditory-nerve fibers are processed in the auditory nervous system by a neural network that calculates the equivalent of an autococincidence (or autocorrelation) histogram [11, 12]. The result is a pattern of activity over the two dimensions of frequency (inherited from peripheral filtering) and lag (provided by nerve conduction or synaptic delays). In response to a periodic stimulus, this pattern shows a ridge at a lag equal to the period, thus providing a cue to the pitch.

Licklider's model was designed to explain the pitch of complex stimuli, however it works as well for pure tones. In response to a pure tone of frequency f, nerve fibers with characteristic frequencies within a band surrounding f will respond with a periodicity of 1/f. The result is an autococincidence pattern with a ridge at 1/f. Actually, the pattern also shows ridges at period multiples; the model supposes that the position of the first ridge is the cue to pitch. Because synchronization deteriorates above 2-5 kHz, the model can only apply to frequencies below that limit (this excludes the upper 2 or 3 octaves of the 10 that span the audible range).

Fig. 1. Frequency difference limens (ΔF/F) for pure tones as a function of stimulus duration and frequency (replotted from Moore [1]). Frequencies up to 2 kHz are plotted with continuous lines, higher frequencies with dotted lines. Straight line: difference limen predicted by basic autococincidence model.

Fig. 2. Autococincidence histogram in response to a pure tone of 100 Hz. The dotted line marks the period lag. The histogram was calculated using 'spike' data produced by a model [13].

Let us define the pitch cue more precisely as the position of the maximum of a composite pattern obtained by taking the sum of histograms across frequency channels (alternative assumptions are possible but won't be discussed here). In response to a pure tone the histograms are all identical, so the effect of summing them is simply to reduce variability, as if a single...
histogram were calculated with more spikes. How precise is this cue? As evident in Fig. 2, the histogram is "noisy", which causes the position of the maximum to be uncertain. The standard deviation of this position can be estimated [14] as a function of discharge rate R, stimulus duration D, histogram bin width e, and number N of histograms summed together:

$$\sigma_T = 0.12 R^{-1/2} (DeN)^{1/4}$$

(1)

It is evident from (1) that the standard deviation varies as the inverse of the fourth root of stimulus duration. This dependency can be understood as follows: due to the parabolic shape of the AC histogram near its peak, the incertitude of the position of the maximum varies with the square root of the standard deviation of the bin "noise", itself proportional to the square root of the counts in the histogram bins. If spikes are allowed to accumulate during the entire stimulus presentation, the count within each bin is proportional to duration, hence the $D^{-1/4}$ dependency.

To get a more quantitative estimate, let us make the assumption that 1250 fibers respond each at 100 s/s, that the spike trains are pooled before histogram calculation into 10 histograms that are then summed, and that histogram resolution is $1 \mu s$ ($R = 12500, N = 10, e = 10^5$). Given these assumptions, the difference limen $\Delta E/F$ (suggested equal to $\sigma_T$) varies as plotted in Figure 1. We can draw the following conclusions:

a) The dependency of $\Delta E/F$ on duration, predicted by the model as $D^{-1/4}$, does not match that observed in Moore's data at low frequencies.

b) The $\Delta E/F$ predicted by the model is almost an order of magnitude larger than the best difference limens observed.

4. NARROWED
AUTOINCIDENCE
MODEL
In the AC model, the effect of making the stimulus longer is to make more spikes available, thus reducing statistical uncertainty. Clearly this is insufficient to account for the difference limens observed and their dependency on duration. There is however a source of information that the AC model neglects: that carried by the peaks of higher rank of the autoincidence histogram.

Recently, a method has been proposed for sharpening the peaks of the autocorrelation function (for purposes of musical pitch estimation) [15]. This method incorporates information from higher-order peaks into a compact representation called "Narrowed autocorrelation function". A similar operation can be applied to the autoincidence histogram (AC), resulting in a "Narrowed autoincidence histogram" (NAC).

$$NAC(t) = \sum_{k=1}^{N-1} (N-k) AC(kt)$$

(2)

The dotted lines labeled "max D" and "max N" in Fig. 4 represent additional hypothetical limits on discrimination due to two factors. The first factor limits length of AC histograms (it could be for example a limit on the allowable length of the neural delay lines assumed by Licklider's model). Making stimuli longer than this limit can bring no improvement. The second factor would limit the order of narrowing, due to the complexity of neural circuitry available for the calculation of the NAC.

The trend of the thresholds visible in Fig. 4 is similar to that of Moore's data for frequencies below 2 kHz (Fig. 1). The major differences are that the curves in Fig. 1 are somewhat shallower, and the spacing smaller than predicted by the model. There is also no evidence in Moore's data for the first of the hypothetical limits mentioned above ("max D"). Apart from these differences the agreement is quite good.

CONCLUSION
The basic autocorrelation model due to Licklider is not sufficient to account for frequency difference limens observed psychophysically. However, a modified model (the NAC model) can successfully account for these limens, and for the form of their dependency on duration. This result is of interest given the recent renewed concern for time-domain models of auditory processing.

ACKNOWLEDGEMENTS
Part of this work was carried out at ATR Interpreting Telephony Research Laboratories, under a fellowship awarded by the European Communities STP programme in Japan. The author wishes to thank ATR for its hospitality, and the CNRS for leave of absence.

REFERENCES
PERCEPTUAL SENSE UNITS IN THE PROCESS OF LISTENING COMPREHENSION

Morio Kohno
Kobe City University of Foreign Studies
Kobe, Japan.

ABSTRACT
Research was added to the previous studies which made clear that the perception of sound sequence consists of two-fold processes—holistic and analytic, and that the former is applied to the fast tempo sequences whose intersound intervals are within 300 ms, and the latter to the slower ones whose intervals are over 400 ms. They are neuropsychologically different from each other (See another paper of the author in this conference named "Two Processing Mechanisms in Rhythm Perception.") Experiments I and II proved that in the process of listening comprehension, meaning units which were made up of 1 to 7±2 syllables closely combined each other with the intervals of less than 300 ms are perceived holistically as definite units. If holistically is deprived semantically and/or physically by pausing, the listenable-ability is extraordinarily decreased. There is some evidence that this 'perceptual sense unit' was imbedded in human-beings' deep cognitive system.

The present paper is going to try to identify 'perceptual sense unit' (P-unit, henceforth), perceived holistically and stored in echoic memory in an unprocessed form in the process of listening comprehension.

1. PREVIOUS STUDIES
Kohno and Kashigai [3] made clear the processing mechanism in rhythm perception using as subjects a normal right-handed woman, seven children with ages varying from one year and four months to nine years old, and a patient with inaction involving the forebrain commissural fibers. The summary of the results got by the study is as follows:

1) The left hand of the patient and children under four years of age cannot synchronize their tapping with the slow rhythms of 500 and 1000 ms inter-beat intervals (IBI), but they can follow the rapid stimuli of 250 ms IBI. The right hand of the patient and children older than four years old, however, can fit their tapping both to the rapid and slow rhythms as well as the normal adult.

2) Negative autocorrelations were detected among adjacent IBIs in response beats by the normal adult, the children older than four and the right hand of the patient, but never found in the response to the slow stimuli by the children younger than four and the left hand of the patient. 4) Negative autocorrelations were never detected in the rapid response movements (250 ms) of all the subjects. 5) The above-mentioned facts suggest that the slow repetitive sound sequences are normally perceived by ongoing and analytic way of processing, but rapid ones by at-a-time and holistic way. Evidence was found that the children younger than four years old and the left hand of the patient always use only the holistic approach not only for the rapid rhythm but also for the slow rhythm and that it is the very reason why they cannot synchronize the slow tempos. These two kinds of processing, therefore, are neuropsychologically different from each other.

6) Other experiments utilizing nonsense words on timing condition of syllable sequences and echoic memory were held with the following results: a) The sequences of closely connected nonsense syllables, each of whose inter-voice-onset-intervals (IVI, henceforth) are less than 300 ms, bring forth longer retention than the syllable sequences whose IVIs are longer than 400 ms, when they are recalled after some lapse of time for doing two digit number multiplication. As suggested in 5), the durational condition less than 300 ms may be processed holistically, and that more than 400 ms, analytically. As holistic processing is qualitatively different from analytic one, the former is never disturbed by the latter, and this may be the reason why the words whose syllables are closely connected have longer retention than the ones which consist of loosely connected syllables after doing some cognitive work. b) There is some evidence that durational condition between 300 and 400 ms IVIs is border area mixed with both holistic and analytic processings, different by individuals.

2. IDENTIFICATION OF PERCEPTUAL SENSE UNIT
2.1. Experiment I
[Subjects] Students of a high school in Japan, 108 in number, were divided into four homogeneous groups in their English ability based on their academic records.

[Materials] Two original stories, one in English (95 words), the other, in Japanese (133 words), were recorded by an American instructor and a Japanese one respectively. Then pauses were mechanically placed by the use of a Pause Controller. SONY LLC 5000, at every end of word (the set pause length = 1 second), at every end of phrase, of clause and of sentence (the set pause length = 2 seconds). Phrase here means a meaning unit which consists of one content and no or some function words. A no-pause version was also prepared. An important thing here is that the number of syllables which made up a phrase was 1 to 7 in English version and 2 to 8 in Japanese. The unit

of clause, however, consisted of 2 to 13 syllables in English, and 7 to 25 (8 to 26 morae) in Japanese, and the sentence unit, 12 to 22 in English and 12 to 48 (15 to 53 morae) in Japanese. According to Miller [4], maximum number of elements which can be perceived in a flash is 7±2. Only the unit of phrase out of the above units meets Miller's condition—in the case of English clause, for example, half of them consisted of more than 7 syllables. Words, of course, consist of less than 7 syllables, but the separation by this unit destroys the unit of meaning, because no function word has any independent meaning.

[Method] Each subject group was requested to listen to one version in Japanese and another in English, and were asked to write the content of the stories as precisely as possible in Japanese, immediately after they had finished hearing them.

Table 1 A (Japanese) full marks = 29

<table>
<thead>
<tr>
<th>Score</th>
<th>n</th>
<th>T</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>no pause</td>
<td>27</td>
<td>4.5</td>
<td>1.38</td>
</tr>
<tr>
<td>every clause</td>
<td>26</td>
<td>2.7</td>
<td>3.27</td>
</tr>
<tr>
<td>every phrase</td>
<td>24</td>
<td>15.8</td>
<td>5.45</td>
</tr>
<tr>
<td>pause word</td>
<td>31</td>
<td>11.6</td>
<td>3.52</td>
</tr>
</tbody>
</table>

B & D (l = 1.39, N.S.); C & D (l =.38, P<.05)  
C & B (l = 1.79, P<.001); C & A (l = 9.75, P<.001)  
B & A (l = 1.35, P<.001); D & A (l = 5.41, P<.001)  
(Table 1 - B (English) omitted.)

[Results] Pauses which were placed at every end of clauses, and of phrases increased the scores in this order in both English and Japanese materials—the no-pause version produced the worst ones. As found in [2], the pause-at-every-phrase version brought about the highest scores. This 'the more pauses, the higher score' principle, however, did not go on in the pause-at-every-words version, which remarkably reduced the scores (P<.05-0.001). This fact also coincides with the result of [3]. If pauses would play a crucial role for listening comprehension by giving chances to analyze and synthesize the stimuli and by giving clues to separate the sound stream into proper units as pointed out in Pimsleur (1971), we may rightly say that these results suggest
the unit of phrase, more precisely, grammatical meaning unit which consists of 2 syllables, might render a most suitable chunk for listeners' cognitive processing of connected speech.

We should now notice that the P-unit, the unit of phrase, for example, generally consists of several syllables which are combined one to another with IVIs about 100 to 200ms in the case of Japanese or 100 to 250ms in English, all of which are so rapid as to be processed holistically. The last IVIs of each P-unit in English are somewhat longer than the preceding ones. Hence is an example of realities of IVIs together with syllable lengths, that is, computational observations (ILS, Micro PID 11/73) on the syllable durations and IVIs between syllables in the stories read by native speakers.

a) Duration of syllables (syl. dur.) and IVIs in spoken Japanese sentences.

Tsu gi ga / no bo ri / syl.dur. 118 104 96 108 111 115
IVI * 146 143 121 116
ha ji me ma shi ta / syl.dur. 139 139 163 165 192 185
IVI 169 111 143 (117) (185)

b) Duration of syllables (syl. dur.) and IVIs in spoken English sentences.

Scott / came out of the syl.dur. 331 225 242 115 196
IVI 386 225 221 165 148
house / and locked / the syl.dur. 311 122 345 (183) 168
IVI 232 290

In Japanese, all IVIs, as well as syllable durations, are all within 300 ms. This means that the syllable sequences so closely connected have to be processed holistically. In English, however, the durations of IVIs and syllable durations vary very much, and generally they are longer than Japanese syllables. We can notice, however, that in English, a long IVI is put at each P-unit, especially at the end of it, that is, at each semantic unit which usually consists of 7 or less syllables. The fore parts of the IVIs are composed of a syllabic succession with short IVIs which may be processed holistically. To put the long syllables at the ends of units is very effective to show the terminations of the units. This device seems to help listeners out of difficulty of holistic processing caused by the variety of syllable durations.

We can now more precisely say that a perceptual sense unit, is a semantic unit which is composed of one to several syllables which are so closely combined one to another in less than 300ms IVIs so that the unit can be processed holistically. If the unit and, therefore, holistiscly are lost by the long IVIs of more than 500ms, listenability of the utterance will be remarkably decreased. The analytic processing, on the other hand, may concern the processing of two and more perceptual sense units one by one to get the whole meaning of utterances. It might as a matter of course, take a longer time.

2.3. Experiment II

In order to verify the above hypothesis, the following experiment was carried out.

[Subjects and Method] An essay in Japanese (Material A) was read by a Japanese female instructor (age: twenties). It was then mechanically separated at each end of perceptual sense units by pauses whose durations were 3, 4 and 5 seconds. Another Japanese essay (Material B) was also read by the same instructor, but IVIs among syllables were spread to each of 200, 250, and 500ms. Subjects were Japanese high school students, 122 in number. Other procedures were the same as in Experiment I.

[Results and Discussion] The results are shown in Tables 2-A and -B. When the IVI among syllables was 200 or 250ms which will be processed holistically, the scores remained almost the same, but when they were lengthened to 500ms, which may be processed analytically, the scores significantly decreased (Table 2-A).

<table>
<thead>
<tr>
<th>TABLE 6 - A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervals among syllables</td>
</tr>
<tr>
<td>200ms</td>
</tr>
<tr>
<td>a.</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>S.D.</td>
</tr>
</tbody>
</table>

200ms ≠ 250ms N.S; 200ms > 500ms P < 0.01

<table>
<thead>
<tr>
<th>TABLE 6 - B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervals among phrases</td>
</tr>
<tr>
<td>natural 300ms</td>
</tr>
<tr>
<td>a.</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>S.D.</td>
</tr>
</tbody>
</table>

4. CONCLUSION

We may conclude by saying that the processing of listening comprehension is a mix of both holistic and analytic works. The existence of the unit which should be holistically dealt with at a time has long been overlooked, but it is crucial element to make clear the processing of listening comprehension which listeners can do very efficiently and rapidly. We should notice that perceptual units whose syllables are closely connected may be preserved in an unprocessed form longer than the separately syllables after the execution of some cognitive works [3]. On account of this nature, listeners can do the works of semantic and sometimes grammatical analysis over several units, if necessary, referring back to some precedent unit which is still retained even after having processed some units.

REFERENCES

METHODS FOR REDUCING CONTEXT EFFECTS
IN THE SUBJECTIVE ASSESSMENT OF SYNTHETIC SPEECH
Chaslav V. Pavlovic, Mario Rossi, and Robert Espesser

Institut de Phonetique, LA 261 CNRS, Universite de Provence, Aix en Provence, FRANCE
& (1st. author only) University of Iowa, Iowa City, Iowa, USA.

ABSTRACT
The contextual invariance of categorical and magnitude estimates of speech quality could be improved by introducing a reference system (natural speech) and by appropriately normalizing the results with respect to it.

1. INTRODUCTION
A potential problem with subjective scaling of speech quality occurs when the rating of a certain system needs to be generalized outside the set of systems used in the experiment. Namely, the rating of a system may change depending on the selection of other systems evaluated at the same time ("context effect"). We evaluate here whether the context effects could be reduced by introducing a reference system (natural undistorted speech) common to all experiments, and by normalizing the rating of any given synthesizer in reference to the rating of the natural speech. Two subjective psychophysical techniques are evaluated: magnitude estimations (MESs) and categorical estimations (CEs).

The ratings of four systems labeled "A" and four systems labeled "B" were evaluated in two different types of context: "A and B" context and "A or B" context. Systems A were of superior quality to systems B. Both systems A and systems B were evaluated separately within their groups (A or B context), and together (A and B context). The research question is whether the ratings of the stimuli are invariant to these changes in context, both in the absolute and in the relative sense. These context effects were evaluated both with and without the reference condition. This particular design was selected because past research indicates that all scaling techniques may be particularly sensitive to it. It is hypothesized that subjects always use one restricted range of numbers regardless of the stimuli being evaluated. If this were the case, there would be a strong tendency to use the same range of numbers for systems A only, systems B only, and systems A and B together. Given that systems A are superior in quality to systems B, the ratings of B will, therefore, be better when these systems are presented alone than together with A. The opposite would be true of systems A.

2. METHOD
The subjects were equally divided into 12 experimental groups. Six experimental groups gave ME and the other six CE judgments. The groups are identified by letters that correspond to the listening conditions they were exposed to. These six labels are ABR, AR, BR, AB, A, and B. Symbol A signifies that the group judged conditions A, symbol B that the group judged conditions B, and symbol R that the group judged the reference condition. The non-normalized group ratings for each condition were calculated as the means across subjects and condition repetitions. The arithmetic means were used for CE's, while the geometric means were used for MES. Neither for the MES nor for the CE's was the reference condition explicitly defined to the subject as such. Rather, it was treated as just another experimental condition. The subjects were required to judge how satisfied they were with the particular communication situation. For CE's the scale from 1 to 20 was used. Direct ME procedure and the sentence test material described in more detail in [1] were used.

3. RESULTS
In the tasks which did not incorporate the reference stimulus, relatively large AB context effects were seen (Fig. 1 for CE's; Fig. 2 for MES's). They seemed to be particularly severe in the case of CE's, where the mean rating of groups A and B were almost equal to each other in the "A or B" context, but quite different in the "A and B" context. When the reference condition was present, a large decrease in the AB context effect was seen in the CE (Fig. 3), while no improvement was demonstrated in the ME (Fig. 4). The introduction of the reference condition did not seem to have affected the relative ratings of the other systems neither for the MES (Fig. 5), nor for the CE's. This indicates that some form of normalization may prove beneficial with regards to context effects.

4. NORMALIZATION
Two measures of the merits of normalization were used. These are the standard deviation (a), and the corre-
luation (\(\lambda\)) between the ratings of the eight experimental systems (A and B) observed, on one hand, in the "A and B" context, and on the other hand, in the "A or B" context. Measure \(g\) expresses the absolute proximity of the measurements made in the two contexts. Measure \(r\) is sensitive to how well relative ratings of the systems agree in various contexts. The smaller the \(g\) and the larger the \(r\) the more context-free the procedure is.

The application of measure \(g\) assumes that all results are on the same scale. This is indeed the case for all normalized values. This is also the case for the non-normalized CEs that are divided by the maximum scale value. However, in the case of the non-normalized MEs the scales are arbitrary and cannot be transformed to a 0 to 1 range. In the latter case, instead of \(g\), the measure labeled \(\xi\) was used. It is defined as \(g\) divided by the mean rating of the stimuli in the "A and B" context.

The CE procedure typically results in an interval-type scale. Therefore, it is invariant to multiplication by a constant, or to addition of a constant. Thus, the results could be normalized by either of these operations. In addition, normalization could be performed on the group results, or on the results of individual subjects. In the case of normalization by multiplication, the rating of a stimulus is multiplied by the reciprocal of the rating of the reference stimulus. This operation applied to the mean group results is labeled "CE_MG." While \(M\) stands for "multiplication," and \(G\) for "group." Normalization by multiplication applied to the results of individual subjects is labeled "CE_MI," where \(I\) stands for "individual." In normalizing results by adding a constant, first the complement to 20 (maximum scale value) of the reference stimulus rating is added to the non-normalized value of the stimulus. Subsequently, these numbers are divided by 20. This procedure leads to the same results regardless of whether it is applied to the group or to the individual results. It will be labeled "CE_C" where \(C\) stands for "complement."

The measures of context effect \(g\) and \(r\) for CEs are given in Table I. All normalization procedures substantially reduce context effects with respect to the non-normalized results of the groups that did not judge the reference stimulus. For example, in the case of method CE_MG correlation-type measure \(r\) increased from 0.48 (for the non-normalized results) to 0.98 (for the normalized results), while \(g\) decreased from 0.13 (for the non-normalized results) to 0.05 (for the normalized results). However, with respect to the non-normalized results obtained by the groups that judged the reference condition, the context effect was made somewhat worse with normalization.

The ME procedure results in a ratio-type scale, and is invariant to multiplication by a constant. Consequently, the normalized results are obtained if the ratings of stimuli are multiplied by the reciprocal of the rating of the reference system. As was the case with CEs, this operation could be performed either on the group results, or on the individual subjects' results. In addition, the ME results could be calculated as "absolute" or "relative" [1]. The normalization procedures on the absolute group results is labeled "ME_AG" (symbols A and G represent "absolute" and "group," respectively), while the normalization procedure on the absolute individual results is labeled "ME_AI" (symbol I stands for "individual"). The normalization procedures either on the group or individual relative ratings yield the same values which are labeled "ME_R" (R stands for relative).

The measures of context effect \(\eta\) and \(\xi\) are given in Table II for these three normalization procedures, as well as for non-normalized procedures. Fig. 6 gives normalized MEs for the best of these procedures, i.e., ME_AI. All normalization procedures substantially reduce the context effects with respect to the non-normalized results of the groups that judged the reference system. However, the real benefit of normalization should be assessed against the non-normalized results obtained without the reference system. There, only the procedure ME_AI appears to reduce the context effects.

---

**Fig. 6** Normalized ME ratings, method ME_AI, groups ABR, AR, BR

---

In the \(\xi\) value the best ME procedure (ME_AI) is practically equal (\(\xi = 0.6\)) to the best normalized CE procedures (CE_MG, CE_MI), but inferior to the non-normalized CE procedure when the reference stimulus is used (\(\xi = 0.2\)). In the \(r\) values the procedure is worse (\(r = 0.89\)) than both, better normalized CE procedures (\(r = 0.95\) to 0.98), or the non-normalized CE procedure when the reference stimulus was presented (\(r = 0.99\)).

5. ACKNOWLEDGMENTS

This research was made possible by a grant from the EEC Esprit SAM project (Grant #2589).

6. REFERENCES

ORDER EFFECT AND THE ORDER OF ACCENTS

L. Schiefer and A. Batliner

Institut für Phonetik und sprachliche Kommunikation, Universität München, München, F.R.G.

ABSTRACT

The order effect causes in a "same-different" task the one presentation order to be better discriminated than the reverse order. The effect was investigated in the domain of pitch perception. Phonetic/psychoacoustic explanations are given, and parallels between the order effect and the perception of accents are discussed.

1. INTRODUCTION

The order effect (OE) has been known for more than 100 years in the field of psychoacoustics [6]; it causes in a "same-different" discrimination task (AX-paradigm) the one presentation order AB to be better discriminated than the reverse order BA. We will call the order that is discriminated better the "prominent" order and the stimulus that comes second in this order the "prominent" stimulus. In phonetics, the OE has not been dealt with very often. This might be due to the experimental design mostly used in phonetics - the ABX-task. Originally, we came across the OE in pitch perception while investigating the categorical perception of intonation contours with the AX-paradigm [4,6]. The "potbelly"-phenomenon described in part 2 was point of departure for several experiments, where we addressed the following questions: (i) Can the OE be traced back to general psychophysical factors? (ii) Is the OE an experimental artifact, or can it be found in real life as well?

In this paper, only a sketchy discussion of our research can be given. A thorough presentation of experiments and phonetic considerations (discussion of the state of the art) can be found in [4].

2. THE POTBELLY PHENOMENON

One of the authors (AB) produced the stimulus is monotonously. The digitized stimulus (sample rate 20 kHz, cut off frequency 8 kHz) was segmented into single pitch periods. The intensity of the whole stimulus was left unchanged. The second part of the stimulus was subjected to different manipulations of the F0 contour (cf. Fig. 1).

Fig. 1: Segmental and durational structure

The continuum consisted of nine stimuli with a constant overall duration, three falls, one level and five rises. The duration of the manipulated part was kept constant, F0 offset and F0 slope differed. A logarithmic scale was used for the manipulation of the fundamental frequency (F0): semitone = 17.31*lnHz. The step from one offset height to the other was one semitone (cf. Fig. 2).

Fig. 2: Continuum

Five repetitions of each pair (i.e. AB, BA, and the "same" order AA and BB, resp.) were presented in randomized order with an interstimulus interval of 600 ms between the members of a pair. The pairs were separated by a pause of 3500 ms; after 10 pairs, a pause of 10 sec followed. The 12 subjects (students) were instructed to decide whether the two members of a pair were identical ("same") or different. The results are given in Fig. 3. With this "potbelly shape"

Fig. 3: Discrimination

function, a clear OE could be found; the order AB can be discriminated better than the order BA. The overall OE is consistent and significant in an analysis of variance, F = 60.67**. The prominent order shows a higher F0 offset in the second member of the pair.

In several other experiments, the factors duration of F0 contour, height of F0 offset, and slope were varied systematically, as well as the experimental design. The results of these experiments [4] lead to the following conclusions:

(i) The OE is a random effect, as it could be replicated in all experiments.

(ii) The OE is not an experimental artifact that can be traced back to a special design.

(iii) A stimulus is more prominent if it has a higher F0 offset and/or a longer F0 contour. A stimulus pair is better discriminated if the prominent stimulus comes second.

3. A PHONETIC/PSYCHOACOUSTIC EXPLANATION

The prominence of a stimulus can be explained articulatorily and auditorily: We can assume that in production, greater pitch intervals are always connected with greater durations, and vice versa, greater durations of pitch elevations or pitch drops are related to a greater amount of pitch change. The perceptual effect of a higher F0 offset might be equal to that of a longer duration of F0 contour, as both factors are normally interrelated. In our experiments, however, a longer lasting elevation of F0 (longer duration) does not lead to a higher F0 offset as both factors were handled independently. At any rate, subjects seem to perceive a higher F0 offset if the F0 contour is longer and, vice versa, a lower F0 offset if the F0 contour is shorter. The prominence of a stimulus might be caused by a greater effort in the production, i.e. a higher muscular tension needed to achieve a steeper rising or falling F0 contour and a higher or lower F0 offset as well. The prominence of a stimulus can thus be explained by articulatory and/or physiological mechanisms. But why
does the prominent stimulus come second in the prominent order? At evaluation time, the Fa information of stimulus A is still kept in memory, but it is influenced by the Fa information of stimulus B. If we substitute "weakened" for "influenced", then the prominent order can be explained: the auditory trace of stimulus A is weakened by stimulus B.

4. ORDER EFFECT AND PROMINENCE OF ACCENTS

There is at least one task for the "normal" native speaker/hearer that is comparable to the task of our subjects and that he/she has to accomplish in everyday conversation: to decide which of the pairs that could only be differentiated by their intonational form: Fa in final (3rd) vs. Fa in prefinal (2nd) position, on the one hand, and questions (Qs) vs. non-questions (NQs), on the other hand [3:210]. In perception experiments the position of the Fa was decided upon [3:211]. The task of the listeners is comparable to that in a "same-different"-task: No contextual information whatsoever is given; if we equate the two phrases that can carry the Fa (2nd and 3rd phrase) with the two stimuli in the AX-task, then in both cases, the order can be "non-prominent followed by prominent stimulus", or the other way round.

**Fig. 4: Overlay plot**

<table>
<thead>
<tr>
<th>Questions, FA on 2. phrase</th>
<th>20°</th>
<th>15°</th>
<th>10°</th>
<th>5°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rises, non-prominent order</td>
<td>BA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions, FA on 3. phrase</th>
<th>20°</th>
<th>15°</th>
<th>10°</th>
<th>5°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rises, prominent order</td>
<td>AB</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-Questions, FA on 2. phrase</th>
<th>20°</th>
<th>15°</th>
<th>10°</th>
<th>5°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Falls, non-prominent order</td>
<td>BA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-Questions, FA on 3. phrase</th>
<th>20°</th>
<th>15°</th>
<th>10°</th>
<th>5°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Falls, prominent order</td>
<td>AB</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In fig. 4, a sort of overlay plot is shown; the mean values of the Fa maxima and minima (full square) and their position on the time axis in the Fa material (y-axis: semitones above speaker-specific lowest Fa value, x-axis: centiseconds) is compared with a schematic description of the order AB vs. the order BA (open circle). In some aspects, the OE material and the FA material cannot be compared in the strict sense. The "turning point" in the OE material e.g. was fixed on 84 Hz, whereas in the Fa material, it could be varied by the speakers. A thorough discussion of differences and points of comparison is beyond the limits of this paper; we will therefore confine ourselves to one of the possible explanations (i.e. not the whole truth, but a substantial part of it). As for the Q/FA constellation and the OE rises in fig. 4, the point of comparison is the more pronounced rise on the prominent stimulus phrase. The prominent order AB, where the prominent stimulus comes second, corresponds to a Fa on the third (last) phrase. As for the falls, a discrepancy between the OE material and the Fa material (NQ) can be observed. In the FA material, the more pronounced fall is on the phrase that carries the Fa, but in the prominent order AB, the prominent stimulus has a less pronounced fall than the non-prominent stimulus A. We believe that a solution can be found if we take the two stimuli that follow each other (Fa-Fa) not only as two acoustic or "purely" phonetic (i.e. auditory/articulatory) events but as some linguistic "gestalt" analogous to an utterance produced by a "normal" native speaker. If we imagine a (speech specific) declination line (for the sake of the argument, an all point regression line) then, in the case of the Fa on the 2nd phrase and the order BA, the declination line is steeper than in the case of the Fa on the 3rd phrase and the order AB. Ceteris paribus, a rather flat declination line indicates openness and/or prominence on the final part of the utterance. (Note that we do not necessarily plead in favor of a declination line as the decisive "underlying entity"; it merely seems to be the most convenient way to sum up the traits in common.)

5. FINAL DISCUSSION

We have found that one order can be better discriminated than the other one; this was called the "prominent order". Phonetico/psychacoustic reasoning lead us to the conclusion that in the prominent order, the second stimulus is more prominent than the first one. The concept of "prominence" is the link to the marking of the Fa in natural speech. The Fa contour of the prominent stimulus in the OE material can be compared with the Fa contour of the Fa of the third phrase in the natural material. As for the rises, the interpretation is straightforward. Phonetico, linguistic, and psychoacoustic factors cannot be told apart. For the falls, some additional assumptions have to be made that can be summarized under the heading "perception of linguistic gestalt."
ABSTRACT

How do we make phonetic decisions? Categorical, prototypical, and gradient theories were tested using the times to identify a /sa/-/sta/ continuum created by inserting varying amounts of silence into a /sa/ syllable or deleting silence from a /sta/. The gradient model requires 6-8 times as many parameters as the others, and so is difficult to compare. Two variants of a prototypical model and a simple categorical one accounted for some of the variance in the reaction times, but a modified categorical model with the same number of parameters accounts for more. In identification, it seems that all unambiguous syllables elicit identical reaction times, but syllables farther from that range elicit increasingly longer times.

1. INTRODUCTION

When we listen to speech, we are exposed to a great deal of variation in the acoustic waveform, much of which we accept with ease. How is it that we can hear these unique acoustic events and yet extract a few categories from them? Early studies of categorical perception (e.g., [1]) proposed that acoustic variation was not even perceived. Reaction time data from Pisoni and Tash [3] seemed to confirm this notion for plain identification. For those stimuli within a phonetic category, identification times were the same. However, for same/different judgments, physically identical tokens were judged the same faster than ones that differed within the category. They interpreted this finding as evidence that different levels of processing are available to different tasks.

Another theory assumes that phonetic continua are evaluated in relation to phonetic prototypes [4]. In Samuel's account, phonetic decisions should be easiest when the prototypical value is used, and increasingly less easy as the acoustic distance between the stimulus and the prototype increases.

Other explanations of phonetic perception depend on the combination of gradient acoustic parameters. Marsch and Cohen [2], for example, compared phonetic decisions from interactions of two acoustic parameters. They have little to say about experiments with only one factor, however, so their theory will not be elaborated on here.

The present study will test the prototypical model against an extended categorical model in explaining identification times. The extension to the categorical model is that of an ambiguous region, rather than just a single boundary between categories. Such an extension is necessary to account for the fact that there are ambiguous stimuli that subjects can report as being ambiguous, rather than hearing the stimuli first as one category and then as another. Such a modification reduces but does not eliminate the differences between the models.

2. EXPERIMENTAL METHOD

2.1 Stimuli

A male native speaker of American English recorded several tokens of the nonsense syllables /sa/ and /sta/ These were low-pass filtered at 10 kHz and digitized at 20 kHz on the Haskins PCM system [5]. One token of each syllable was selected, with each having the same duration in the fricative noise and in the vowel segment. (160 and 240 ms respectively). A continuum of gap closures was made by inserting silence between the noise and the vocalic segment for /sa/. The original silence and the burst were removed from /sta/ and replaced as in /sta/. The values ranged from 0 to 78 in 3 ms steps, yielding 27 values; with two sources, there were 54 unique tokens.

2.2 Subjects

The subjects were 10 Yale undergraduates who were paid for their participation.

2.3 Apparatus

The stimuli were recorded onto audio tape and played to the subjects over headphones. Their judgments as to whether the syllable was "SA" or "STA" were made by pressing a button, which generated a signal that stopped a clock on an Atari computer, giving the reaction time. Times were assessed from the onset of the vocalic segment, not the onset of the syllable so that the times would not directly vary with stimulus duration.

2.4 Procedure

A tape containing twenty exemplars of the stimuli was played to familiarize the subjects with the kinds of judgments they would have to make. Then four blocks, each containing five repetitions of each of the 54 stimuli, were presented. Each block, which had a different randomization of the stimuli, began with four "warm-up" stimuli which were not included in the analysis. A brief rest period was given between blocks.

3. RESULTS AND DISCUSSION

An analysis of the reaction times showed that the subject variances increased as the mean time increased, suggesting a log transform. All further times, though reported in ms, are means of the log values. An analysis that included block and source (original /sa/ or original /sta/) as factors revealed no effect of block, and an effect of source that was the same for both "s" and "st" judgments (the /sta/ source gave slightly faster times). Therefore, further analyses collapsed across these two factors.

It was desirable to eliminate mistaken responses, but the subjects had no way of indicating whether a response was the one intended or not. Instead, "isolates" were excluded. These were responses that were separated from a region of judgments by one or more gaps with no responses of that category. Thus one subject might have "s" responses at the 48 ms gap that would be included in the analysis (since gaps 45 and lower also had "s" responses), while another might have such a response excluded (since at least the 45 ms gap received no "s" responses). Isolates accounted for 11% of the data. Figure 1 shows the reaction times averaged across the 10 subjects.

The models were tested by examining how much of the possible variance they could account for. The variance of the individual times in relation to the overall mean established the minimal level for a
model to attain, while using the mean for each judgment for each gap duration established the maximal level. (Since only one acoustic parameter was varied, this maximal description is essentially what would be proposed by a gradient theory, such as Massaro’s fuzzy logic model.) The minimal model thus had two parameters (the overall mean for each response), while the maximal model would have between 27 and 54 (since the ambiguous regions could overlap), though the average was 39.9. Figure 2 shows the four models that were generated for the first subject. (All the modeling was done for each subject individually.) The categorical model was generated with the five parameters: s-boundary (that is, the upper limit of gap values at which 95% of the responses were “s”), the st-boundary (the lower limit of gap values at which 95% of the responses were “st”), the mean times in the “st” region and for the “st” region, and the mean time for the ambiguous region (which included non-isolate “s” responses in the “st” region and “st” responses in the “st” region). In the modified categorical model, the time for the ambiguous responses was calculated from two parameters, a linear interpolation from the edge of the unambiguous region through the mean time for the ambiguous stimuli, temporally located in the center of the ambiguous region.

The prototypical model was generated by taking the fastest time for stimuli of any gap duration as the interpolation value for the continuum endpoints, with the other value being the mean of the responses to ambiguous stimuli (temporally located in the middle of the ambiguous region as in the modified categorical model). The modified prototypical model (which more closely resembles Samuel’s) used the location of the fastest time as the definition of the subjects prototype. Values were then interpolated through the ambiguous region as before, and values toward the endpoints were interpolated with a mirror image of the pattern.

Figure 3 shows the percentage of possible variance accounted for by the four models. Since the range of “possible variance” was defined by two more models, it was possible to do better than the minimum. For two subjects, this was in fact the case for all four models. Subject 5 had very little variation across the gap durations, and had very long times in general (about two standard deviations below the group mean). Subject 7 had a very small “s” range (i.e., only the 10 ms gap), and actually had faster times for ambiguous “st” judgments than for unambiguous ones. Still, for 9 of the 10 subjects, the modified categorical model performed better than the modified prototypical one. An analysis of variance was run on the percentages shown in the figure, with the factors of type (categorical or prototypical) and modification (modified or not). While type was not significant as a main effect (F(1,9) 1.49, n.s.), modification was (F(1,9) 9.38, p < .05), as was the interaction (F(1,9) 8.86, p < .05). As is apparent, only the modified categorical stands out from the others (by a Newman-Keuls post-hoc test).

Since the simple categorical model had one more parameter than the simple prototypical one, comparisons between those two models are somewhat problematic. Both modified models, however, required six parameters, putting them on an equal footing.

This does not exhaust the possibilities for modeling the data, of course. One further modification of the prototypical models would be to allow the interpolation to be parabolic rather than linear. Though initially appealing, such a modification would make it very difficult to tell the prototypical model from the categorical—perhaps giving us a benign ambiguity. It may also be that there is a floor effect on the reaction times. Perhaps the times in the unambiguous regions were subject to, say, a mechanical limitation, so we might have found a more prototypical pattern if the limitation were circumvented. It is possible that a fast repetition (shadowing) paradigm might be useful here.

For the present results, however, it appears that the best model is the one that assumes that all unambiguous judgments are equally easy, while more difficult (due to ambiguity) ones become increasingly so the greater the distance from the category region.

Acknowledgments:

Work supported by NICHD grant HD-01994. Thanks to A. G. Samuel and A. G. Levitt for helpful comments.

References:


INFLUENCE OF NEGATIVE INTENSITY GLIDES ON THE
DISCRIMINATION OF SPEECH SEGMENT DURATION

Y. Nishimura & S. Santl
Institut de Phonétique, Université de Provence
CNRS R.U.A. 261, 13621 Aix-en-Provence, France

ABSTRACT
The discrimination of duration was investigated using synthetic vowels
containing negative intensity glides (0 dB, -6 dB, -12 dB, and -18 dB).
Test stimulus durations ranged from 100 to 300 ms in steps of 20 ms.
The standard stimulus was 200 ms in du-
ration and had a stable intensity.
Stimulus pairs were presented to 20
subjects (constant method) and their
rules was to state which vowel in the
pair sounded longer (forced choice).
Results indicate that a drop in in-
tensity of more than 12 dB has a
significant effect on the perception of
duration, and thus on its discrimina-
tion.

1. INTRODUCTION
The prosodic analysis of speech, which consists of interpreting acous-
tic parameters such as duration, fundamental frequency, and inten-
sity, is not an easy task. Two rea-
sons for this are that (1) these factors
are not independent in human per-
ception and (2) they vary as the
speech signal evolves in time (within
a syllable, a word, a clause, etc.). It
is known that the perception of pitch
variations depends upon their dura-
tion [11]. Furthermore, the melodic
contour of segments with negative
and positive intensity glides are per-
ceived differently [13].

However, we know little about the
interaction between duration and in-
tensity in speech. In particular, the
influence of intensity variations on
the ability to discriminate the dura-
tion of speech sounds has not been
experimentally documented. This
problem came up in our previous
study, which investigated the differen-
tial threshold of syllable duration in a
sentence context [8]. Duration
discrimination was found to be sig-
ificantly less accurate on the final
syllable than on preceding syllables.
The same tendency was observed in
Klatt and Cooper's data [7], which
show a higher threshold for fricatives
at the end of sentences than in other
locations. This led us to raise the
question of whether a drop in in-
tensity (-16 dB in our case) on the fi-
nal syllable of a sentence would
make it difficult to correctly perceive
that syllable's duration. An exper-
iment carried out to verify this hy-
pothesis is reported below.

2. EXPERIMENT
Klatt's formant synthesizer was used to
generate stimuli for the perception
test [6]. The goal was to obtain
speech-like stimuli which varied in
both duration and intensity. Negative intensity glides were used to
approximate the final syllable of
declarative sentences.

The material was designed to be
used in a psycho-acoustic test
based on the constant method. The
standard stimulus was the vowel /a/
with a duration of 200 ms (an aver-
age syllable length) and a stable
intensity of 80 dB. The test stimuli
were synthesized with durations
ranging from 100 ms to 300 ms in
20 ms steps (for a total of 11 differ-
ent durations). Four linear intensity
glides were utilized: 0 dB, -6 dB, -12 dB, and -18 dB. The fundamen-
tal frequency contour was the same
for all stimuli. A slight lowering of
pitch from 140 Hz to 130 Hz made
the stimuli sound natural. The
standard stimulus was paired with
each of the test stimuli. The two
evowels in each pair were separated
by a silent pause lasting 600 ms.
The interval separating one pair
from the next was three seconds.
Both within-pair orders were used
(stdandard-test, test-standard). Each
pair occurred four times. Thus, the
total number of pairs was 352 (11
durations x 4 intensity glides x 2
stimulus orders x 4 repeats). Stimuli
were generated in random order by
a computer and recorded on a digi-
tal audio-tape. A trial series of 22
pairs was added to the beginning of
the test sequence. A short beep
followed by a five second silence
was inserted every 22 pairs.

The perception tests were carried
out in a soundproof room. Twenty
subjects were tested individually,
each in a single trial lasting 20 min-
utes. The listening level of the stan-
dard stimulus leaving the head-
phones was set at approximately
70 dB SPL. The written instructions
to the subjects were as follows: "You
are going to listen to many pairs of
vowels /a, a/. For each pair you
have to ask yourself the following
question: Which of the two vowels is
longer, the first or the second? Even
if the intensity changes, please
judge only the duration." The sub-
jects responded by checking the ap-
propriate answers on a forced-
choice answer sheet.

3. RESULTS AND DISCUSSION
An analysis of variance on the data
yielded a significant difference be-
tween subjects (F(19,351) = 36.84, p
< 0.001). This means that each
subject had his or her own strategy
to carry out the task. In addition,
the expected in this kind of psycho-
acoustic test, the order in which the
stimuli were presented affected the
subjects' behavior (F(1,351) = 55.6, p
< 0.001). Variations in the intensity
factor also produced significantly
different scores (F(3,351) = 168.97, p
< 0.001). Note however that neither
the repeat factor nor the response

---

Figure 1. Duration threshold as a function of intensity glide
category (which member of the pair was perceived as longer) had a statistically significant effect.

In order to compute the mean duration threshold (for the 20 subjects pooled) we interpolated the value for the duration at the 75% correct answer level by summing the four repeats per subject. This mean was calculated for the two stimulus presentation orders, two response categories, and four intensity glides. The averages of these values are shown in Figure 1 and Table 1.

The threshold turned out to be proportional to the magnitude of the intensity glide. In other words, the intensity glide became steeper, the detection of duration became less and less accurate. This effect on the discrimination of duration is clearly shown by the progressive increase in the means and standard deviations shown in Table 1. The difference between the stimulus with the steepest drop (-18 dB glide) and the stable intensity stimulus (no glide) exceeds 7%. Interestingly, the t-test on the data for the first two intensity glides (0 dB and -6 dB) did not yield any significant differences. Apparently, a 6 dB drop in intensity does not lead to difficulty in detecting the correct duration. In contrast, the stable vs. -12 dB difference (t(19) = 4.14, p < 0.001) and the stable vs. -18 dB difference (t(19) = 5.36, p < 0.001) were both highly significant. It is noteworthy that our results indirectly support those obtained by Rossi [13], who estimated the intensity glide threshold to be approximately 11 dB for a vowel lasting 200 ms. The observed change in the way intensity information is processed seems to depend on whether or not the intensity decreases beyond that critical value, although we do not know precisely where in our auditory system that change occurs.

This tendency is even more apparent if we consider stimulus presentation order. For the standard-test order, it can be hypothesized that subjects pay attention to the duration of the final syllable, which has a negative intensity glide in this experiment. The computed threshold values were 220 ms, 230 ms, 233 ms, and 235 ms, for 0 dB, -6 dB, -12 dB, and -18 dB, respectively. This indicates that when the penultimate syllable measures 200 ms and has a stable intensity, final syllables with intensity glides of 18 dB may have to be longer than 235 ms.

However, this hard and fast interpretation may need some qualification due to one peculiarity of this experiment. In comparison to the results published in psycho-acoustic studies using speech sounds, our threshold value at 200 ms is remarkably (even excessively) precise (1%, or 2 ms; cf. Figure 1 and Table 1). For a standard stimulus duration of about 200 ms, the reported threshold values fall between 8% and 30%. These experiments used several standard stimulus durations ranging from some ten milliseconds to several hundred milliseconds [1, 2, 3, 5, 7, 9, & 12]. In our experiment, all 352 stimulus pairs had a 200 ms vowel (the only standard duration used), with a level intensity in first or second position. This may have overexposed subjects to that particular duration, causing better performance. Therefore, the duration threshold defined here, (i.e., as a function of intensity glide) should be used in conjunction with those obtained under normal, stable intensity conditions.

4. CONCLUSIONS
This study has provided some experimental evidence of how well we hear at the end of declarative sentences. The results of our perception tests demonstrated that the discrimination of duration may be significantly deteriorated by a progressive decrease in intensity of more than ten decibels. Our results may have some implications for the interpretation of prosodic data at the sentence level.

5. REFERENCES

Table 1. Duration threshold as a function of intensity glide

<table>
<thead>
<tr>
<th>Intensity glides</th>
<th>Mean threshold</th>
<th>Standard deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 dB</td>
<td>202</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>-6 dB</td>
<td>206</td>
<td>14</td>
<td>80</td>
</tr>
<tr>
<td>-12 dB</td>
<td>216</td>
<td>16</td>
<td>80</td>
</tr>
<tr>
<td>-18 dB</td>
<td>217</td>
<td>22</td>
<td>80</td>
</tr>
</tbody>
</table>
AMPLITUDE AS A CUE TO WORD-INITIAL CONSONANT LENGTH: PATTANI MALAY

Arthur S. Abramson
The University of Connecticut and Haskins Laboratories

ABSTRACT
Word-initial Pattani Malay consonants are short or long. The closures of the “long” consonants are longer than those of the “short” ones; this is a sufficient cue for perception, but in voiceless plosives the duration of the silent closure is audible only after a vowel, yet listeners label such isolated words well and so must use other cues. The peak amplitudes for the first syllables of disyllabic words are greater for initial long plosives. In this study, increments of closure duration and amplitude were pitted against each other for original short plosives and decrements for original long plosives. In tests, duration was by far the more powerful cue, although amplitude did affect the category boundary. By itself, however, amplitude is a weak cue. Further work is planned on the possible role of the shaping of the amplitude contour.

1. INTRODUCTION
Many languages are described as having a phonological distinction of length in vowels or consonants, or even both. If the term is taken literally, we would expect to find that the underlying mechanism is control of the relative timing of the articulators. Even so, a single mechanism might have a number of acoustic consequences, each of which could help in perception.

Pattani Malay, spoken by about a million ethnic Malays in southern Thailand, is unusual not only in having a length distinction for consonants in word-initial position but also in having one that is relevant for all phonetic classes of consonants in that position [3]. Here are some minimal pairs of words showing the contrast:

\[ /laba/ \text{‘to profit} /laba/ \text{‘spider} /make/ \text{‘to eat} /make/ \text{‘eaten} /bule/ \text{‘moon} /bule/ \text{‘months} /kato/ \text{‘to strike} /kato/ \text{‘frog} /\]

Among the various plausible acoustic effects of the mechanism, the most likely for the largely disyllabic words involved, was the peak amplitude of the first syllable relative to the second. Indeed, measurements [2] revealed that this ratio is greater for long plosives, that is, both stops and affricates. Presumably, greater air pressure accumulated behind the occlusion before release accounts for the differences. Although both voiced and voiceless plosives showed a significant difference, the level of significance was higher for the latter. No doubt, this is to be explained by differences in glottal impedance of the airflow. The difference is not significant for the continuants.

2. PROCEDURE
This paper is a progress report of my test of the hypothesis that the peak amplitude of the first syllable relative to the second in disyllabic words is a sufficient cue for the perception of the distinction between short and long voiceless stops in Pattani Malay. For my major experiments, as part of an interest in combinations of phonetic features underlying the same phonemic distinction, I have pitted variants in duration and amplitude against each other to determine their relative power.

2.1. Control tests
Although the identifiability of initial short and long consonants had been demonstrated [1], it seemed desirable also to do control tests for the recordings of my new speaker for this study. For each of seven minimal pairs of words I prepared a test containing 20 tokens of each of the two words, yielding 40 randomized stimuli. There were two such randomizations for each word pair. The nasal, lateral, fricative, and plosive categories were represented. The plosives included voiced and voiceless stops and voiceless affricates. (Unfortunately, my only pair of voiced affricates included a word, as I learned later, that would have embarrassed the women among the subjects, so I could not use that test.) The subjects were 30 undergraduate students, all native speakers of Pattani Malay, at the Prince of Songkhla University, Pattani, Thailand.

2.2. Amplitude vs. duration
To test for the relative power of amplitude and duration, three pairs of words with velar, dental, and labial short and long stops respectively were used. All of them were recorded at the end of the carrier sentence /di/o kato/ ‘he said.’ By means of the Haskins Laboratories Waveform Editing and Display System (WENDY), the stop closure of the short member of each pair was lengthened in 20-ms steps until it reached or exceeded the duration of its long counterpart. The closure of the long member was shortened in the same way. The first syllable of each variant of the original short stop was increased in amplitude in five 2-Db steps. Likewise, the first syllable of each variant of the original long stop was decreased in amplitude in five 2-Db steps. Two test orders were recorded from randomizations of two tokens each of all the resulting stimuli and played to 30 native speakers for identification of the key words.

3. RESULTS

3.1. Control tests
The previously demonstrated identifiability of the utterance-initial consonants [1] was reaffirmed. The major difference is that the voiceless long affricates in this sample were labeled correctly 96% of the time, whereas in the last study it was just above chance at 55%.

3.2. Amplitude vs. duration
Because of the limitation on space, the results of only two of the experiments are given here. Figure 1 gives the responses of 30 native speakers to nine durations in 20-ms steps of the [k]-closure in /kamen/ ‘goat’ combined with six amplitude levels in 2-Db steps. The vertical axis
shows the percentage identification as short /k/. The earlier crossover of the higher-amplitude curves at the 50% point to the long-/k/ category, giving judgments of /kamen/ 'goatlike,' is highly significant \( F(40, 1160) = 9.0, p < .001 \); nevertheless, the values of duration at the short end are very little affected. The opposite procedure, shortening original long /k/ and lowering the amplitude, yielded similar results, as shown in Figure 2. The results are essentially the same for the other two places of articulation.

3.3. Amplitude in isolated words
In Figure 3 both the short and long responses are plotted for increments of amplitude on original /pagi/ 'morning.' While the two curves converge, they never cross each other. Figure 4 shows rather similar effects for decrements of amplitude combined with isolated tokens of /pagi/ 'early morning.'

Fig. 1. Responses to /kamen/ 'goat' and its variants with increased closure duration and first-syllable amplitude.

Fig. 2. Responses to /kamen/ 'goatlike' and its variants with decreased closure duration and first-syllable amplitude.

Fig. 3. Responses to isolated /pagi/ 'morning' and its variants with increased first-syllable amplitude.

Fig. 4. Responses to isolated /pagi/ 'early morning' and its variants with decreased first-syllable amplitude.

4. CONCLUSION
It is clear that when both features are present, duration is dominant; nevertheless, the boundary between the two perceptual categories is significantly affected by relative amplitude. In utterance-initial position, however, relative amplitude is only a weak cue, apparently secondary to something else.

To understand how the length distinction is perceived in utterance-initial voiceless plosives, perhaps further work should be done on the possible role of the shaping of the amplitude contour. That is, maybe a finer analysis of utterances and a more complicated making of stimuli will show, for example, that the rise-time of the amplitude carries more weight than the peak value, or that the two work together. Indeed, a very preliminary look at this time suggests that the rise time is shorter in the production of the long stops. Also, it is possible that the major amplitude difference is confined to the region of the release burst. Other features that have not seemed promising so far, such as fundamental frequency and rate of formant transitions, may have to be examined more closely too.

5. ACKNOWLEDGMENTS
The work was supported by NICHD Grant HD-01994 to Haskins Laboratories. The fieldwork in Thailand was made possible by a sabbatical leave from the University of Connecticut in 1988. I am grateful to the National Research Council of Thailand, the Department of Islamic Studies of the Prince of Songkhla University, Pattani, and the Department of Linguistics of Chulalongkorn University, Bangkok for their warm hospitality and help.

6. REFERENCES
TESTING SUBPHONEMIC PERCEPTION PROCESSES IN CHILDREN SUSPECT FOR AN AUDITORY DISORDER

P. Groesen, B. Maassen and Th. Crul

Cald Neurology Centre / ENT-department
Institute of Medical Psychology, Nijmegen, Netherlands

ABSTRACT

Currently available speech perception tests provide insufficient insight into type and cause of an auditory processing disorder. The paradigm of categorical perception which discriminates an auditory- and a phonetic stage, combined with the idea of reducing redundancy of speech stimuli in order to have access to speech processing, enables us to develop a new tool for fine-grained assessment of central auditory pathology. By now, the first steps have been taken in the development of the test consisting of a series of experiments on verbal dyspraxic children, reading and spelling disordered children and children with a severe history of otitis media.

1. INTRODUCTION

Several categorical perception studies on auditory perceptual behaviour in groups of children with a specific speech- or language disorder have shown a deviant response pattern as compared to a normal control group. Auditory processing disorders, without there being a hearing-loss according to tone- and speech-audiogram, have manifested themselves in a delayed speech- and language development. Currently available speech perception tests, generally containing tasks based on monotic measures (e.g. filtered speech, competing signals), dichotic measures (e.g. syllables, words, sentences) or tests of binaural functionality (e.g. binaural fusion, rapidly alternating speech) give insufficient insight into type and cause of such a disorder. This article concerns the first steps in developing a new tool for fine-grained assessment of speech perception processing.

The paradigm of categorical perception provides for the base of a more sensitive and analytical test. According to this paradigm, a continuum of speech stimuli covering a phonological contrast is constructed by digitally manipulating a single acoustic cue. The central idea is that perceptual boundaries arise along the continuous continuum, with qualitative resemblances within each category and qualitative differences between them or in a more modern psychophysical sense [3], there is a quantitative discontinuity in discrimination at the category boundary, as measured by a peak in discriminative acuity at the transition region of adjacent categories.

Two starting-points for research are of interest. Firstly, the speech perception model of Pisoni and Sawusch [7] which differentiates between a pre-categorical or auditory stage and a categorical or phonetic stage, plays an important role. During the first auditory stage, listeners take in short stretches of the raw acoustic signal and make a preliminary auditory analysis. No speech segments have yet been identified. In the second, phonetic stage listeners examine their auditory memory and combine the acoustic cues to form a phonemic representation. This stage preserves the nature of the identification but does not preserve the acoustic cues upon which it was based.

By comparing identification and discrimination performance (labeling words and telling them apart) we can derive the level of the auditory disorder. Discrimination scores can be predicted on the base of identification scores [4]. Assuming that the listener bases discrimination judgements only on phonetic information, the observed discrimination scores should correspond to the predicted ones. If the subject has access to auditory precategorical information, the discrimination scores should be higher than the predicted scores.

Secondly, we assume that a speech perception problem may be caused by a neurological reduction of the redundancy of the auditory processing system [1] which we call the internal redundancy of the perceptual system. During speech perception there has to be a considerable reduction of the internal redundancy before stimuli with their normally high external redundancy (which is implicit in the normal structure of the acoustic speech signal) cannot be identified anymore. The conclusion is that the external redundancy must be reduced such that the speech perception test becomes more sensitive to small reductions of the internal redundancy.

Exactly this occurs in constructing a speech-continuum, a word (one end of a phonological contrast) is transformed to another (other end) by systematically decreasing and increasing the salience of the acoustic cue.

A speech continuum will contain stimuli which do not discriminate between normal and pathological groups. In order to maximize the efficiency of the testprocedure and to minimize effects of response bias these stimuli can be eliminated out of the test. Only the critical stimuli, the stimuli which are as sensitive to account for differences between normal and deviant children will be included in the final test. Sensitized speech stimuli can be singled out of stimuli near the phoneme boundary.

2. PROCEDURE

By now the first steps in developing such a sensitized perception test are being carried out. We examine scores on tests where two acoustic cues are systematically varied: VOT (/bak-pak/, i.e. BOX-PACKAGE) and place of articulation, second and third formant transition (/bak-dak/, i.e. BOX-ROOF).

Three experimental groups are tested:
- verbal dyspraxic children aged between 6 and 11 years. These children show dysfunctions in the programming of their articulatory organs. We suppose that (a part of) this group is marked by a central auditory dysfunction.
- children with reading- and spelling problems, 2 groups; one in the age of 6-7 years, the other in the age of 8-10 years. Disordered auditive functions could be one of the main causes of the problems.
- children with a severe history of otitis media with effusion (OME) in the early childhood (aged round 2 years). At the time of testing they are 6-8 years of age. Due to temporary conductive hearing-loss these children show disorders in their auditory deve-
development. We take interest in the extent in which the central auditory functions are affected. Our experimental design is based on a division into 4 groups: a) Children with an OME-history and with a delayed speech- and language development, b) Children with an OME-history and without a delayed speech- and language development, c) Children without an OME-history and with a delayed speech- and language development and d) a control group children without an OME-history and without a delayed speech- and language development.

3. RESULTS
At the time of writing this paper only the data of the verbal dyspraxic group have been fully analyzed. For, at this time, lack of data of a control group we shall present in this paper some examples of identification curves of the place-of-articulation continuum /b-d/ of a male adult without hearing problems or a delayed speech- or language development with a normal response pattern and a verbal dyspraxic child with an abnormal response pattern (Figure 1).

Remarkable is the less steepness of the slope of the curve of the verbal dyspraxic child compared to the adult. We interpret this as a less consistent labeling ability, pointing towards a processing dysfunction at the phonetic level. In Figure 2 the corresponding discrimination curves are presented. Again there is a difference between the verbal dyspraxic child and the adult; the ability to discriminate is less in the verbal dyspraxic child. Furthermore the overall form of the discrimination curve doesn’t agree with the fo-neemboundary found in the identification curve.

Taken together, this verbal dyspraxic child shows a disordered auditory stage (less discriminative power), followed by an inadequate functioning of the phonetic stage (less steep slope, shifted phoneme boundary), partially an effect of the malfunction in the auditory stage.

If we take the group verbal dyspraxic children as a whole, there’s a great variability in scores suggesting the possibility of dividing the group in a number of verbal dyspraxic children having speech processing problems and a number of verbal dyspraxic children showing no central auditory dysfunction.

More elaborate analysis of the experimental groups, yet unavailable, will be given at the presentation.

REFERENCES
THE INFLUENCE OF SPEAKERS' OWN SPEECH TEMPO ON THEIR TEMPO PERCEPTION

M. Gósy
Research Institute for Linguistics, Budapest, Hungary.

ABSTRACT
The point of departure for the present paper is the assumption that the speaker's own speech tempo determines his judgements concerning that of other people. Experimental results supported a significant concurrence in tempo perception of 'extreme' speakers as opposed to 'moderate' speakers. A significant correlation was found between the speakers' comprehension and their own speech tempo. It can also be claimed that speakers/listeners judge speech tempo on the basis of the active levels of their speech perception mechanism.

1. INTRODUCTION
Authors of a number of studies agree that tempo perception is basically determined by three factors: articulation rate, the frequency of pauses, and the duration of pauses [4, 5]. Tempo perception studies are also made difficult by problems concerning the recognition, demonstration, and order of importance of a number of other factors including changes in fundamental frequency (pitch), average intensity, word frequency (of occurrence), syllable structure, rhythmic structure, syntactic properties, etc. [3]. However, very few attempts have been found in the literature to deal with the connection of production and perception in relation to speech tempo. It is usually postulated that there should be a very close connection between the speaker's own rate of speaking and his perception and comprehension with respect to speech tempo [1, 2]. Can we then claim that there is a linear connection between tempo production and perception, namely, the faster the speaker's usual speech tempo the faster his/her comprehension as well? Does this apply to tempo perception, too? What are the criteria of applicability of this rule?

In order to answer these questions, an experiment containing 4 subsets has been performed with Hungarian-speaking native speakers/listeners. The aim of the experiment was to describe the effect of the subjects' own speech tempo on (i) their tempo perception and (ii) their speech comprehension.

2. PROCEDURE
Various methods were used for the sub-tests. (i) For the first experiment nine speech samples were recorded in random order from Hungarian-speaking native subjects (ages ranged from 25 up to 50). Subjects were selected so that all categories be represented from very slow (articulation rate /AR/: 8.85 sounds/s, overall speech rate /OSR/: 7.25 sounds/s) up to very fast (AR: 18.2 sounds/s, OSR: 14.3 sounds/s). Each speech sample was taken out of a longer monologue and took 1.5 minutes on average. The listeners' task was to judge the speech tempo of each speaker's sample by means of a questionnaire. The categories of the questionnaire were 'very slow', 'slow', 'normal', 'accelerated', 'fast', and 'very fast'.

(ii) The material of the second test consisted of 12 artificial, synthesized sentences (the synthesis was made by a PDP speech synthesizer controlled by an IBM PC). The same sentence had been altered in relation to its overall speech tempo in two ways: by changing the "articulation rate" of the sentence and by adding one or two pauses at the appropriate grammatical boundary(ies) of the sentence. The subjects' task was the same as in the first sub-test.

(iii) 8 sentences announced by a trained male speaker were chosen for the third test, and a verification method was used. The sentences were speeded up, and their articulation rate ranged from 20.2 sounds/s up to 24.4 sounds/s. The subjects' task was to decide whether the sentences they heard were true or false. The reaction times (RT) of each subject were measured by means of a fundamental frequency and intensity meter with the accuracy of 10 ms.

(iv) The subjects' spontaneous speech was tape recorded in the final experiment. From their recorded 8-10 minute speech 2-minute samples were picked out for further analysis concerning AR & OSR. The duration and types of pauses were also examined. Counting the speech sounds of the speech sample, the rate was expressed in terms of sound/s.

After finishing the experiments, each subject was asked to judge his/her own average speech tempo according to the formerly used tempo categories. The subjects' sex and age were also recorded on the same answer sheet. 37 subjects were selected from all candidates for further examinations. Three tempo groups were defined: a group of 'slow' speakers, a group of 'moderate' speakers and a group of 'fast' speakers. Examing the data, significant correlation was found between the AR and OSR values of our subjects (p<0.05). 6 subjects were found to be 'fast' speakers in terms of AR and 'moderate' speakers in terms of OSR. So, a fourth tempo category had to be established consisting of subjects having 'fast' AR and 'moderate' OSR and this was labelled the group of 'rapid' speakers.

3. RESULTS
Figure 1 shows the responses of various groups of subjects for all synthesized sentences according to the possible tempo categories. The listeners do perceive the physical changes of sentences. In the case of sentences containing 1 or 2 pauses however, the judgements spread along the various tempo categories. The question is whether the distribution of tempo perception is based on the subject's own tempo production. Analyzing the average values for each sentence of each group, it can be stated that there are no important differences among the subjects' judgements. However, the data of the three groups are significantly different at the level of 0.05. This means that there is a slight but definite difference of tempo perception subjects with diverse speech tempo production. The mean values of the judgements show very constant changes across the tempo categories. These changes reveal more similarity for the 'slow' and 'fast' speakers than for the 'moderate' and 'rapid' speakers. There is a significant difference in the judgements of the 'slow' speakers concerning the category of 'accelerated' tempo as opposed to the judgements of the other two groups. 'Rapid' speakers' performance shows a relatively different distribution in relation to that of the other two groups. On the basis of these data, a hypothesis has been developed on the relatedness of the speakers' own production and their tempo perception: 'slow' and 'fast' speakers tend to perceive tempo similarly to one another while 'moderate' speakers do not. 'Rapid' speakers seem to behave peculiarly in a way different from the other three groups. We also found that the extreme speakers tend to perceive tempo more on the basis of AR than on the basis of OSR, so the pauses might not influence their tempo perception.

Figure 2 shows the responses of various groups of speakers for the speech samples used across the possible tempo categories. Subjects appear to judge the tempo of the speech samples according to AR rather than on the basis of OSR. The data show: (1) There are larger differences among the tempo categories in each test group than in the case of isolated sentences, and (2) the distribution of the judgements does not show a constant trend.
The number of responses referring to the 'moderate' tempo category is significantly different in the case of the perception of the synthesized sentences and the speech samples (p<0.05). This means that people's perception mechanism has grown accustomed to the tempo changes of human speech and they are more flexible when judging it than in the case of one sentence where the upper levels of the decoding mechanism should not work, so they can judge the tempo of each sentence more accurately to the actual physical values. The data show a very similar concurrence of judgements made by the 'slow' and 'fast' speakers. Similar judgements of the 'slow' and 'fast' speakers were found in all tempo categories with the exceptions of the 'fast' and 'very fast' categories. In the case of these two tempo categories the 'slow', the 'moderate' and the 'rapid' speakers judged similarly while the 'fast' speakers differed from all the others. The 'rapid' speakers show a significant difference in their judgements from the other groups of speakers. However, in some cases their judgements fall close to the judgements of one of the groups of speakers. This co-occurrence does not show any systematic character.

![Figure 1. Tempo perception of sentences by 'slow' (---), 'moderate' (---), 'fast' (---), 'rapid' speakers (---).

Figure 1. Tempo perception of texts see Fig. 1. for the key.

There is a significant difference between the RT values of affirmative and negative sentences with true contents; however, there was no significant difference between the same structures with false contents.

Finally, the subjects' age, sex, and their opinion about their own speech tempo were taken into consideration. There was a very strong correlation between the subjects' objectively measured speech tempo and their subjective judgements (p<0.001). We found that most of our extreme speakers were male while the 'moderate' speakers were mainly female subjects (p<0.05). There was no significant correlation between the subjects' age and their speech tempo categories.

4. CONCLUSIONS
- It had been assumed that the faster the speaker's own speech the less fast he perceives that of others. From this hypothesis only the basic point of departure was supported by the results that the speaker's own speech tempo influences their tempo perception. However, the direction of this influence shows an interesting pattern involving significantly different behaviour for the various groups of speakers. The 'slow' and 'fast' speakers tend to behave perceptually similarly while 'moderate' and 'rapid' speakers tend to differ from the previous two groups. The standard deviation of the reaction time values was the same for the 'slow' and 'fast' speakers and for the 'moderate' and 'rapid' speakers.
- It has been supported that tempo perception definitely mainly on articulation tempo. However, according to our findings, speakers/listeners perceive tempo significantly depending on the activated levels of their whole perception mechanism. If the upper levels of the speech perception mechanism do not play any role in the actual perception process, the tempo judgements (a) are closer to the actual physical parameters of the speech sample and (b) do not show big differences among the speakers having various own speech tempi. If the higher levels also participate in the decisions then other factors (contents of the speech samples, articulation of the speaker, lexicon of the speech sample, timbre, types of hesitations etc.) also play an important role.

- On the basis of the significant differences in perception and comprehension of various groups of speakers, we assume that various ways and storage systems should exist for the interactions between the higher and lower levels of the speech perception mechanism determined by the temporal organization of the speakers' speech production.

5. REFERENCES
ACCOUNTING FOR THE REFLEXES OF LABIAL-VELAR STOPS

Bruce Connell

Phonetics Laboratory, University of Oxford

ABSTRACT

This paper presents a phonetic description, summarizing evidence drawn from different instrumental techniques, of the voiceless labial-velar stop as it occurs in Ibibio, one of the Lower Cross languages of SE Nigeria. The description is then drawn on to offer an account of the variety of reflexes attested for labial-velars, both within Lower Cross and elsewhere. Important characteristics are that a) the timing of the two articulatory gestures involved is asynchronous, and b) that the degree of asynchrony, as well as other aspects of their articulation, is variable, both across and within speakers. Recognition of this variation is the key to understanding the associated diachronic developments.

1. INTRODUCTION

1.1. Descriptions of Labial-velars

Labial-velar stops are relatively rare in languages of the world (cf. Maddieson [9]) and have received scant attention in the phonetic literature. Instrumental phonetic analyses have been presented by Ladefoged [8], Games [7] for Ibibio, and by Dogil [5] for Baule. Painter [11] also gives some discussion of labial-velars in an article dealing primarily with laryngeal mechanisms. Ward [14] for Efik presents kymograph tracings of [kp], but no systematic analysis, and finally, Oghal and Lorentz [10] present a general discussion of phonetic characteristics of labial-velar articulations, though without focussing on stops. In this paper, I summarize the results of a set of instrumental investigations that have been conducted on Ibibio [kp], and then use these results to attempt to account for the variation in reflexes found for labial-velar stops where diachronic change has occurred.

Apart from instrumental work, impressionistic descriptions of articulatory and auditory characteristics labial-velar stops can often be found in the Africanist linguistic literature. Generally, the labial and velar articulations are said to be simultaneous (e.g., Westermann and Ward [15]). Other than this, Ladefoged's [8] remarks (p. 12) in comparing labial-velar stops to velarized labials [p, b], and that they have a tendency to impart a labialized quality to following vowels, emphasize the possibility of perceptual confusion with labials, and Ohala and Lorentz [10] have provided acoustically based explanations for these tendencies. Comparisons have also been made to labial implosives by Ladefoged [8], Painter [11], and by Elugbe [6], who sees this as a general characteristic of labial-velars in the Edoid languages. Beath and Zemp [1] describe the labial-velar stops of Dan as having "strong bilabial implosion", and Puesch [12] reports a voiced implosive labial-velar for Bekwir.

1.2. Diachronic Developments

The earliest account of diachronic correspondences of labial-velar stops in the literature is found in Westermann and Ward [15], who cite evidence for sound change that, "where kp or gb are weakened, it is the labial element which disappears and the velar element remains, sometimes reduced to x or y" (p. 58). Elsewhere in the text (p. 108), correspondences are presented from the Nupoid languages Obari and Nupe, and also from Bari and Kakwa (k’- kp, g’- gb in both cases) which to some extent confirm their conclusions. However, it is no difficult task to find instances of sound change involving labial-velars where it is the labial element which survives. It is probable that the velarized voiced labial implosive of some dialects of Igbo is a reflex of Proto-Igbo *gb. In the Lower Cross languages, PLC *kp has evolved into a variety of reflexes, most commonly [p], but also [b], [k’], and possibly [g’] (and [kp] is retained in many instances).

2. INSTRUMENTAL ANALYSES OF IBIBIO [kp]

2.1. Methodology

A variety of instrumental techniques were used to investigate the characteristics of the Ibibio labial-velar, including spectrography, laryngography, aerometry, and electropalatography. These were done during a period of approximately three years, and used different speakers for the different investigations. Material for the spectrographic study was recorded by eight native speakers of Ibibio in Calabar, Nigeria, and analysed in the Phonetics Laboratory at the Universities of Ottawa and Edinburgh; further investigations were done in Edinburgh using primarily one speaker of Ibibio who was resident there (the aerometry was done with two speakers). Methodology and results are reported in greater detail in Connell [4].

Spectrographic measurements were done to examine total duration of closure (TD), voice termination time (VTT) and voice onset time (VOT), as well as transitions and burst spectra, and compared to similar measurements for labials and velars. Laryngography (Lx) was done with both aerometry and electropalatography (EPG) to determine VTT's and VOT's, and other information about phonation. The aerometry and EPG provided further details concerning the articulatory nature of these stops.

2.2. Spectrographic Analysis

In broad terms, the results of the spectrographic investigation confirmed those of Games [7]. This was true with regard to formant transitions, especially for CV transitions, where there was similarity to those of simple labials except for being steeper, having a lower locus, and apparently being more intense or stronger. This latter observation also corresponds with findings of Dogil [5] for Baule. On the other hand, VC transitions were variable, most often tending to resemble those of simple velars, but occasionally resembling labial transitions. Regarding the timing of the two gestures involved, evidence from transitions suggests a consistently later labial release, but variability as to which occurs first. Fig. 1 presents a spectrogram of the word [ékpé] 'leopard' demonstrating the asymmetry of formant transitions.

![Fig. 1: Spectrogram of [ékpé] illustrating asymmetrical formant transitions of Ibibio [kp]. TD=156ms, VTT=38ms, VOT= 36ms. (Speaker E.E. Akpan.)](image)
release, as would be expected, given the evidence from F2 transitions discussed above. The energy found in the mid-range could possibly be associated with a vowel aspect of the release, but it is in this range that the noise band mentioned normally occurs. Finally, on occasion there is energy present throughout the spectrum, extending quite high in the frequency range. In this connection it is worth noting that Trill [13] reports burst spectra for labial clicks extending throughout the frequency range and being particularly strong in the 4 – 14 kHz range.

Table 1 summarizes voicing and duration characteristics of Ibibio [kp] relative to simple labials and velars. Results are based on productions of eight speakers (Fig. 1 exemplifies VTT and VOT measurements.)

<table>
<thead>
<tr>
<th>TD (29.0)</th>
<th>VTT (55.7)</th>
<th>VOT (7.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p 147</td>
<td>65</td>
<td>6</td>
</tr>
<tr>
<td>k 112</td>
<td>49</td>
<td>21</td>
</tr>
<tr>
<td>kp 162</td>
<td>50</td>
<td>26</td>
</tr>
</tbody>
</table>

Table 1: Duration and voicing of [kp] relative to [p] and [k]. Values to nearest ms. SDs are given in parentheses.

Two aspects are important here; first that Ibibio [kp] is very short, and second, that there is a relatively high amount of variation (as indicated by the standard deviations) in the voicing characteristics.

2.3. Aerometry/Lx
The aerometric work revealed a substantial pressure drop during closure and ingressive airflow, indicating use of either or both of velaric and glottalic ingressive airstream mechanisms. A relatively consistent variation in the pressure drop was taken as indicative of an earlier velar release (Connell [2, 4]). The associated Lx analysis confirmed and clarified voicing characteristics revealed by the spectrographic investigation.

2.4. EPG/Lx
Voicing characteristics described above were confirmed, and made more precise when considered against the EPG evidence of closure and release (Connell [4]). Also interesting was evidence from the EPG investigation confirming the spectrographic evidence of an earlier velar release. This was revealed through comparison of the EPG record with the accompanying audio signal. Since the audio signal in the set-up was only a gross representation of intensity of the signal, its onset could represent either the onset of the following vowel or the onset of voicing, recalling that the release is prevoiced. Either way, given that the consonant is the final release of the velar closure prior to the onset of the audio signal would be a clear indication of the velar release preceding the labial one. This happened in all tokens, and on average 38 ms, but ranging from 10 ms to 80 ms, prior to the onset of the audio signal. (SD = 15; calculations are based on 4 repetitions of 18 words containing [kp] in controlled environments.) Further research is planned to monitor lp closure in conjunction with EPG, permitting a more accurate assessment of relative timing of both closure and release.

2.5. Summary
The various instrumental techniques revealed, among other characteristics, that: a) the two articulatory gestures are not totally simultaneous, nor completely synchronized: the velar release almost invariably precedes the labial one. There is more variation as to which closure occurs first, though this is most often the velar one; b) there is a considerable amount of voicing in this nominally voiceless stop, manifested in both a voice onset and a pre-release voicebar; and c) there is a high degree of variability in the timing of the various components of the articulation, both individually and relative to each other. This variation was manifest both within and across speakers. Finally, although evidence has not been presented here, there was some indication that the cross speaker variation observed correlated with dialect.

3. EVOLUTION OF PLC *kp
Pre-Lower Cross * kp has, in addition to [kp], the following reflexes across the group (Connell [3, 4]): [y; b; pb, * kp]. The phonetic characteristics of Ibibio [kp] allow us some insight into why such a range of reflexes should be manifest, either by extension, why yes, others, such as [x; y; b*] are also understandable.

The fact of the later labial release gives clear cause to expect a labial, or predominantly labial, reflex should the sound undergo change, as it would be the most salient. However, since the degree of asynchrony between the two releases demonstrated considerable variation, it is plausible to assume that a dialect of a language might exist where the two were much more closely simultaneous, or even with a later velar release; in these cases reflexes more predominantly velar might arise.

The variability in the duration of prevoicing in Ibibio also gives a clue as to why we find both voiceless and voiced reflexes; presumably those LC languages exhibiting PLC *kp > [p] originated in dialectal variation favouring a shorter voicebar, whereas those demonstrating PLC *kp > [b] would have emanated from ones with a longer voicebar. It is also possible that the existence of a relatively long voicing tail might have played a role in the development of voiced reflexes, particularly where PLC *kp > [b] has been found.

An account of this nature fits the diachronic developments for Lower Cross in the phonetic data for Ibibio. This implies that PLC *kp, at some stage in the history of the language was similar in its phonetic characteristics to that of Ibibio today. We might also expect that reflexes of labial-velars which are more predominantly velar (e.g., in the Nupoid languages cited above), or that are implosive (e.g., Igbo) come from parent languages whose labial-velars demonstrated characteristics conducive to those particular developments. This is an empirical question which can, and hopefully will, be tested through a detailed phonetic analysis of language groups having the appropriate sets of reflexes.

4. REFERENCES

ABSTRACT

In recent years, laser beam technology has been used to reproduce the sound from wax phonograph cylinders and other old recordings. These methods have been used for the reconstruction of speech of the aboriginal population on Sakhalin which has been recorded in the beginning of this century. In this way, very useful data on earlier stages of languages and dialects have become available for linguists and anthropologists.

1. USE OF THE PHONOGRAPH

The principle of the original phonograph is simple: a metal horn focuses the energy of the sound waves onto a thin diaphragm, which supports a small needle in its centre. When the diaphragm vibrates in response to the energy of the focused sound waves, the needle, too, vibrates as it is drawn across the revolving surface of a wax cylinder. The needle cuts a groove consisting of microscopic gouges in the soft cylinder surface. In this way, a recording of the pattern of sound waves is made. To play back the recording, the needle is replaced over the gouges made during the registration. The attached diaphragm vibrates and creates sound waves duplicating those which had originally been recorded.

From the early use of the phonograph until the coming of portable disc-recording equipment, the phonograph was the only means of recording phonetic data. In the late 1880s, ethnographers were intrigued by the possibilities of applying the new cylinder phonograph for field work. It was used for the first time around 1890 for the study of American Indian speech and in the beginning of this century Ainu data were recorded on the island of Sakhalin, north of Japan.

2. WAX CYLINDERS AND THE PROJECTS FOR THEIR RESTORATION

Old recordings on wax cylinders are still being kept in many places. It would be of great interest to regain the sound material they contain, and to improve its quality by using modern digital techniques of registration and signal enhancement.

The events which have led to the project on the retrieval of sounds from wax cylinders started in Poland with the discovery of a number of phonographic wax cylinders. They contain linguistic, musical and ethnographic material, primarily on the Ainu people of Sakhalin. The recordings were made at the beginning of this century by the Polish anthropologist Piłsudski [1].

The Institute of Linguistics of Poznań University (Poland), Hokkaido University (Japan) and the Institute of Linguistics of Groningen University started a common project to analyse material obtained from the phonograms and other old recordings and to set up ethno-linguistic field work. The goals of this collaboration are the following:

a. Application of acoustic, electronic and optical engineering techniques to the retrieval of information on phonographic wax cylinders and other old recordings;

b. The interpretation of the phonetic and linguistic contents of the recordings, and the study of the languages of Sakhalin at the beginning of this century;

c. Phonetics and ethno-musicological analysis of the recorded speech and songs; comparison with present day material.

d. The organisation of field work expeditions to the Minority Peoples of the North in the USSR.

3. THE REPRODUCTION SYSTEM

Using the original Edison-type phonograph for the reconstruction of the sound material involves a risk of damaging the wax cylinders. This method cannot be applied to broken cylinders which have been repaired. Thus a non-destructive, non-contacting method has been developed on the basis of laser-optics technology.

A Gaussian beam emerging from the single-mode He-Ne laser with a wavelength of 0.633 µm is focused by an objective lens. The wax cylinder, which is translated during rotation, is illuminated by a diverging Gaussian beam of which the spot diameter on the cylinder can be adjusted to the width of the grooves. The detecting plane for the reflected beam is set perpendicular to the optical axis. The wax cylinder is rotated, whereas the intersection position of the reflected ray on the detecting plane moves in time on this plane. The time variation of the position is detected by a position-sensitive device and it corresponds to the acoustic signal. The signal stored can be deduced from the detected variation of the reflected beam.

The properties of the sounds reproduced in this way depend on the width of the illuminating laser beam, since its finite size breaks down the principles of geometric optics on which the method is based. Further, there is the obstructive noise in the sound, caused by the non-homogeneous illumination and there is the tracking error resulting from improper contact with the grooves. These problems have been investigated experimentally by Asakura et al.[2], who found that the most suitable beam width for the laser-beam should have a spot diameter between 80 and 100 μm. In this way, the sounds reproduced can be heard naturally and without obstacles. Since the laser-beam reflection method is non-contacting and non-destructive, it is a powerful tool for retrieving sounds from old wax cylinders without damaging them.

4. SIGNAL ENHANCEMENT

The data are stored on new optical/digital sound carriers and in order to improve the quality of the sound obtained, special techniques have been developed. The sound reproduced from old recordings is usually of poor quality. This may be caused by the original recording...
techniques (e.g. resonances in the horn), by the damage of the cylinders which has occurred over the years (clicks at burst positions) and by the techniques of reconstruction. In order to improve the sound quality, several methods have been developed which can also be applied to speech enhancement in general. In Japan, various programs have been developed and applied for this purpose [3].

In the case of the Ainu tapes, the result of the processed sounds was not always satisfactory: in several cases, the listeners preferred the original unprocessed sounds, even if there was noise on the tape. This was due to the fact that after processing the noise level is reduced, but certain bad-quality-features are still there and become more prominent. The recorded and processed Ainu data are stored at the Research Institute of Applied Electricity, Hokkaido University (Japan).

5. RESULTS OF THE WAX CYLINDER RESTORATION PROJECTS

The Japanese project has provided the possibility to study the Ainu language from Sakhalin as it was spoken at the beginning of this century. Originally, the Ainu people lived in the Northern part of Japan, on Sakhalin and the Kurile Islands; at present their language is only spoken on Hokkaido. Old Ainu people were consulted when the material from the Pilsudski wax roles was played to them. In some cases, they recognized their Ainu dialect and the voices from the past.

In this way, the last stages of a dying language have been safely recorded. The material can be studied by linguists and ethnologists in order to obtain information on the Ainu people. The wax cylinders and their contents can thus be considered to be part of a very important cultural heritage, because they contain valuable sound data of speech and songs of the Ainu people that were lost long ago.

6. THE ETHNOLINGUISTIC FIELD WORK ON SAKHALIN.

In July and August 1990, the University of Hokkaido has organized an international fieldwork expedition to Sakhalin in order to study the language situation of the original population on that island and the way this has been influenced by Japanese and Russian. The idea was to look for the Ainu population and to investigate the status of the other small minority groups, in particular Nivkh (Gilyak), Ulta (Orok) and related Tungusic peoples, who were the first inhabitants of Sakhalin. Unfortunately, during our expedition no more Ainu people could be found, and the only person who represented the Ainu language and culture from Sakhalin was probably the old informant we met on Hokkaido.

The original population of Sakhalin consisted of some Paleo-Siberian and Tungusic peoples, in particular the Nivkh and Orok in the North and Centre, and the Ainu in the South. Their numbers were rather small, and during the colonisation process by the Russians from the North and by the Japanese from the South, they were soon numerically dominated by these stronger nationalities. Due to their isolated life as hunters and fishermen, they were able to keep their native language and culture for a long time, but since the beginning of this century the assimilation process has gradually become stronger.

The dramatic events of 1945, culminating in the Soviet occupation of the whole island, had enormous consequences for the ethnographic and linguistic situation on the island: practically all Japanese inhabitants and together with them many of the aboriginals, left Sakhalin for Japan. New immigrants came from all parts of the Soviet Union and at present, more than 100 nationalities are living on the island. Several of them still cultivate their own language.

During the expedition to Sakhalin, a great deal of material on the Minority Peoples of the North was collected: about 80 hours of audio, 30 hours of video and numerous photographs and written documents.

Part of the recordings consists of interviews with representatives of the minorities of Sakhalin. These interviews can be considered as 'case studies' of the language situation for particular minorities. Most of the material is related to the Nivkh population.

The life of the Nivkh and other Minority Peoples of the North has changed considerably under the influence of Russification. They have become a small minority on Sakhalin, scattered over the island and surrounded by Russians and other immigrants who take part in Russian culture.

During our field work expedition on Sakhalin, most of the subjects for our project were elderly people with a strong motivation to use their language. Practically all young people we met no longer had an active knowledge of the language, and they only communicated in Russian.

It can be concluded that on Sakhalin a process of assimilation is taking place, which may result in the complete disappearance of these small languages and cultures.

This process of "language death" may, however, slow down, if these minority cultures are receiving more attention. Further field work should be conducted in order to facilitate the conservation of data on the languages and cultures of these Peoples of the North and other minority groups.

The data collected about these aboriginals (Nivkh, Orok and others) are now analysed and a description is given. The availability of these data will enable a comparison with the historical recordings.

6. REFERENCES

INTER-SPEAKER VARIABILITY IN SIBILANT PRODUCTION AND SOUND CHANGE INVOLVING SIBILANTS

Alice Faber

Haskins Laboratories, New Haven, CT 06511 USA

ABSTRACT

The role of contextual and inherent lip rounding in mediating synchronic and diachronic interchanges between [s] and [ʃ] is described in the context of a model of sound change involving inter-speaker variability in speech production and perception strategies.

1. INTRODUCTION

In a variety of the world's languages, interchanges of [s] and [ʃ] in labial environments, that is, near sounds like [r m u ʃ] (6), are attested. Such changes are seen in some ancient Semitic languages (3), Tigrinya (a modern northern Ethiopian Semitic language) (12), early Indo-European (7), and southern American English (9). Based on published materials showing that [ʃ] is rounded in a variety of languages (1,2,5), that [s] is lower frequency adjacent to [u] than to other vowels (4), and that the frequency boundary between [s] and [ʃ] is higher before [u] than before [a] (6), I attributed the changes in both directions to the phonologically ambiguous status of a phonetically rounded sibilant in a rounded context: if the lowered frequency is attributed to the rounded context, the sibilant is interpreted as [s]; but, if instead, the lowered frequency is interpreted as inherent to the sibilant, the sibilant is interpreted as [ʃ] (3). Sound change, in this model, is a result of individual differences in speech production and perception: Speakers will differ in how much rounding, inherent or contextual, they produce in a given instance, and listeners will differ in their interpretation of rounding in a specific instance as inherent or contextual.1

This paper is a tentative report of a preliminary series of experiments aimed at testing the hypothesis regarding [s]-[ʃ] interchanges as well as the general model of sound change in which that hypothesis is embedded. The present hypothesis is that [s] and [ʃ] will be less distinct acoustically in some labial contexts than in non-labial contexts for at least some speakers of some languages. Furthermore, this decreased distinctiveness should result in part from increased rounding of [s] in these labial contexts.

2. METHODS

There were 7 subjects for these experiments, including one Polish-English bilingual. All speakers produced utterances of the form VSV, with the flanking vowel [a u], and S one of [s ʃ] (vowel tokens), and the bilingual speaker in her Polish mode produced utterances with [s ʃ]; the utterances with [ʃ] will not be discussed here, and [s] will be treated as equivalent to [ʃ]. Two speakers also produced ac(a)Sa and ac(a)Ca, utterances with S again one of [s ʃ], and C one of [k ʃ m r] (consontants). All tokens were pronounced with some pressure. Subjects, their language backgrounds, and their data are summarized in Table 1.

Lip position was monitored with a modified Selspot opto-electronic tracking system. For some subjects, lingual-palatal constriction location was monitored with a RION artificial electrolarynx. Movement signals were digitized at 200 samples/sec., and the EPG signal at 64.1 frames/sec. The speech signal, recorded on a Telex unidirectional head-set microphone, was digitized at 20,000 samples/sec. (12 bits), without preemphasis. Automatic peak picking algorithms were used to identify upper lip protrusion (ULP) maxima and minima; for segments without clear extremes, an arbitrary point was measured. For present purposes, the measure of sibilant frequency used was the Centroid. Centroids were computed for all sibilants over the range c. 1,000–10,000 Hz. In vowel tokens, one centroid was computed, at the approximate midpoint of S. For consonant tokens, three centroids were calculated, one in the middle, and one at each edge.

For each speaker, the extent of variation in sibilant frequency and in ULP was assessed by Analyses of Variance. For vowel conditions, the analysis was Vowel X Sibilant, and, for consonant conditions, Consonant X Sibilant X Order X Adjacency, with Measurement Point as a repeated factor for the acoustic ANOVA. Order refers to whether C preceded or followed S, and Adjacency to whether S and C were abutting or separated by a stressed [s].

3. RESULTS AND DISCUSSION

3.1 Acoustic Factors

Vowel condition speakers had significant main effects of Sibilant and Vowel, as well as significant interactions. Centroids were lower for [ʃ] than for [s], and were lower when the flanking vowel was [u] than in the other contexts. Fig. 1 shows the extent to which [s] and [ʃ] were distinct in the three vowel contexts.3 The higher the percentage, the larger the frequency difference between [s] and [ʃ]. There are clear differences among subjects in the extent to which they distinguish the two sibilants, as well as in the nature of the vo- calic effects on the distinction. The primary commonality is that the two sibilants are less distinct in the [u] context than in the other vocalic contexts. Examination of the raw centroid values reveals, further, that the difference in sibilant distinctiveness in the [u] context results primarily from a decrease in frequency for [ʃ]. Decreases in distinctive- ness from the [u] context result, in contrast, from an increase in frequency for [ʃ].

As for the consonant subjects, they are not surprisingly, more complicated.

1. See Faber (1980) for a interpretation of results in terms of a general model of sound change.

2. The data from a few speakers are not reported. The present paper is intended as a first report of the data, and the present results are from a small number of speakers.

3. The analysis was conducted using the Statistical Package for the Social Sciences (SPSS).
The two consonant subjects have congruent acoustic and labial variation patterns in utterances containing [m] r, but less congruent ones in utterances with [k] r, which is sensible, given that the former contribute active ULP gestures while the latter do not. Unfortunately, no EPG data are available for FBB. However, interpretation of KSH's EPG data is straight- forward. First, [s] is retracted following [r], while [j] is not affected by which side of [r] it is on; this decreased distinctiveness reinforces the effects of the decrease in ULP. Fig. 5 shows an additional pattern of differential effects of preceding and following [k] and [l] on ULP distinctiveness, a pattern not reflected in acoustic differences: [s] and [j] are less distinct in ULP following [k] or [l] than preceding them, but the EPG data show the reverse: [s] and [j] are more distinct in linguo-palatal constriction location following [k] or [l] than preceding them. The effects of this linguo-palatal difference apparently cancel out those of the ULP difference.

4. CONCLUSION

The experimental results described above show that the decreased acoustic distinctiveness of sibilants in rounded contexts is the result of a complex interplay of labial and linguo-palatal factors. Speakers vary in the relative contributions of the two sorts of factors to their acoustic patterning. In order to preserve the acoustic invariance of [s] across a range of contexts, speakers must vary their linguo-palatal targets for [s], in an attempt to compensate for context-based variation in lip position. In contrast, preservation (to the extent possible) of articulatory invariance, at least as regards linguo-palatal constriction, leads to an increase in acoustic variability. Concentration on articulatory invariance can lead to instances of [s] that might be perceived as [j]. But, concentration on acoustic invariance can lead to a proliferation of articulatory targets, the number of which may subsequently be reduced in novel ways, especially the reinterpretation of some instances of [j] as [s]. Either way, sound change occurs.

5. REFERENCES

INTRODUCTION

Though it is accepted that the study of phonological systems have to presuppose the study of phonetic substance we still have to know which phonetic aspects should be considered relevant to phonology. This paper shows how voice quality and phonation types are essential phonetic material to understand the phonology of language and extends some substantial support to Henderson's [2] views. The observation regarding the relationship between voice quality/phonation types makes it imperative for us to revise our descriptive apparatus. It is suggested that voice quality/phonation types can control and mould the phonology of language. This is once again like the age old relationship between nature and culture and description which deprives the problem of one of the relata is doomed (Mo1 and Uhlenbeck)[10].

2. PHONATION TYPES IN GUJARATI

Gujarati very interestingly employs two distinct phonations: Murmur and Tight. Murmur has been taken for granted as feature associated with Gujarati since Pandit [11] and Jørgensen [4]. But the fact is that 40% of Gujarati speakers speak with tight phonation. These phonation based distinction of dialects coincides with the geographical divisions of Gujarati (See Map 1).

These phonation types are like physiological habits of the speakers who with the vertical movements of larynx and muscular tension modulate the airflow. Murmur (breathy voice) occurs due to [R] which can be independent phoneme or can be voiced aspiration of [h, h, qh, qh, unh, unh]. Murmur has constant airflow due to cartilaginous gaps (Fujimura)[1] and strong activity of posterior cricoarytenoid along with maintained vibration of vocal folds (Sawashima and Hirose)[2]. (Hirose and Gay)[3]. Tomograms taken of Gujarati speakers show the lowered position of larynx for murmur and raised position for tight. Raised larynx increases the tension of vocal fold surface (Stevens)[6]. Tight phonation is a physiological adjustment maintained through out the speech and has a high pitched quality. Having the reverse physiology from murmur this phonation automatically inhibits murmur once conducted preliminary acoustic study of these phonations (Schaefer et al)[13] has been able to distinguish these two phonations with the help of two efficient parameters: amplitude of first and second harmonics and band widths of F1 and F2.

By general phonological criteria these laryngeal dimensions can be discarded as irrelevant to phonology. However their behaviour in Gujarati language opens up new direction in studies of Gujarati phonology. Modi's studies [8,9] were done with the intention of showing the non-segmental character of murmur. Murmur was considered a prosodic phenomenon interacting with the surrounding sonorant sounds such as:

1. [R-V] 2. [V+R] 3. [V+R+V]

Here [R] gets partly deoralized (See Fig.1)

Perceptually different

/ 1/ /0 /

Articulatory optimization

e → jaw lowering

&

After Lindblom [7]

In TD the mid-vowels are higher than in MD. The fortition tendency is considered responsible for this (Modi)[8]. It is worthwhile noting here that the speakers of TD with six vowels face great difficulty in pronouncing English [e, a] and [R]. They
are a laughing stock of all Indians for having [e] vowel for 'rape' and ['rap'], it is suggested that a process might have begun when the distance between [e-é] and [e->] could have become phonemic but tight phonation might have counterbalanced such a shift.

The next issue is that of nasalization, which has resulted from diachronic N-loss, nasalization (except for some onomatopoeic forms). In MD denasalization is under progress. But TD hold the fort of fortition. The tense musculature of tight phonation once again is favorable to nasalization and hence denasalization remains only as a subdialectal phenomenon.

One more of such phenomena is of voiced stops spriritualizing in MD intervocally or when in cluster with liquids (e.g. [Ap][3] 'prestige', [aI] 'latch', [sá] 'simplicity'). TD with inherent fortition does not allow such weakening of stops.

In short, denasalization and spirantization are prohibited from pervading the complete language. The tenseness and fortition of TD act as a preserving factor while as laxness and lenition of MD act as a weakening factor. Both phonation types work hand in hand: retaining-substituting preserving-effacing; thus balancing the phonemic inventory of language. They are relevant linguistic features as if purpose-built.

3. TIGHT PHONATION IN SINDHI AND KUTCII

A little more support is extended to the hypothesis from two other Indo-Aryan languages: Sindhi and Kutchi. It is proposed that there is a tight phonation belt starting from Sindhi (now in Pakistan) and spreading up North-Western Gujarat (the Map). Both these languages have implosives for which Ladefoged [5]

notes a slight downward movement of larynx after the closure is formed. Ladefoged [5] has seen the possibility of different phonations in languages having implosives. The conflicting gestures of raising and lowering of larynx can be possible due to the musculature tension of tight phonation. It is interesting to note that this phonation has added totally non Indo-Aryan sounds-the implosives-to the phonemic inventory of these languages (See isograms of Sindhi and Kutchi speakers). The suggestion is that due to these sounds a distance has been created between the phonemic systems of these languages and other Indo-Aryan languages.

4. VOICE QUALITY AND DIACHRONICALLY ACQUIRED PRECISION OF SOUNDS

Finally it is shown how the maintenance of the production of sounds is attained by the tension of tongue musculature in one of the standard dialects of Marathi spoken by Poona Brahmans with exceptional scholarship in Sanskrit. They attained perfection in uttering Sanskrit sounds by following the ancient phonetic treatises. They formed a speech habit where fronting and raising of tongue with tongue musculature tension was sustained throughout the speech. The oral cavity gets reduced and sounds are marked by 'fortisness'. The habit was so much entrenched into the system of the community that it got transferred into their Marathi. It is proposed that this voice quality has played a very important role in moulding of Marathi phonology; compared to any other Indo-Aryan languages Marathi has retained maximum Sanskrit sound sequences. The normal diachronic tendencies of weakening such as, deletions of final vowels, cluster simplification, media vowel reduction found in all other Indo-Aryan languages are totally absent in Marathi.
EPENTHETIC NASALS IN THE HISTORICAL PHONOLOGY OF HINDI

Manjari Ogha
Linguistics Program
San Jose State University
San Jose, CA 95192

John J. Ogha
Department of Linguistics
University of California
Berkeley, CA 94720

and:
Department of Linguistics
University of Alberta
Edmonton, Alberta T6G 2E7

ABSTRACT
We demonstrate the plausibility of a posited historical process whereby an epenthetic nasal consonant appeared between a sequence of nasal vowel + voiced stop (but not if the stop was voiceless) by showing that the same process occurs phonetically in present-day Hindi and French pronunciation.

1. INTRODUCTION
Modern Hindi (MH) words such as [d̪aːt] "tooth" vs. [t̪aːnd] "moon" present an interesting asymmetry in their phonological history: their development from Middle Indo-Aryan (MIA) to Old Hindi (OH) and then to New IA both were subject to cluster simplification with compensatory lengthening and nasalization of the preceding vowel [1, 4]. Thus: Skt danta > MIA danta > OH data > MH [d̪aːt]; Skt tendra > MIA candra > OH canda > MH [t̪aːnd].

(Historical forms are given in conventional transliteration: modern forms in IPA where [q] is inherently long.) In the latter example the nasal consonant, present in MIA but then subsequently lost, re-appears in MH. Is it plausible that a nasal be re-introduced only before a voiced stop or should we rethink the historical derivation of such words? The primary evidence that the nasal was indeed lost by the time of OH is the fact of compensatory lengthening of the vowel which in numerous other instances correlates with simplification of medial consonant clusters or geminates, e.g., Skt hasti "elephant" > Prakrit hasthi > MH [həstə].

We present phonetic evidence in support of the scenario that a nasal consonant (N) could have been re-introduced preferentially between a nasalized vowel (V) and a following voiced stop (D) but not a following voiceless stop (T).

In an earlier exploratory study of Hindi we found that in the transition between a word-final distinctively nasal vowel and a following word initial voiced stop, the initial part of the voiced stop became a nasal consonant. For example, the Hindi utterance /ōːk məː do/ (literally "one 'I' give") was phonetically /ōːk məː do/. Here it seemed clear that the nasal consonant formed out of the first part of the voiced stop was not lexical and was purely a product of low-level phonetic interaction between cross-word boundary segments.

If verified, shown not to occur with V + T sequences, and found in other languages too, then this epenthetic nasal would constitute a plausible parallel to the posited diachronic scenario which requires the creation of a new sequence of V + D.

2. AN INSTRUMENTAL STUDY
2.1. Methods
To obtain an indication of velar movement in speech in a non-invasive way we used a nasal olive (10) which gives a rough measure of nasal air flow, itself an approximate measure of velum opening. The nasal olive records air pressure behind one blocked nostril, the other nostril remaining open. This technique also permits a high-quality audio recording of the speech to be made simultaneously. Our subjects were two native speakers, each, of Hindi and French; for both languages there was one male and one female speaker. The first author was the female Hindi speaker. The subjects read a list of sentences in their respective languages which included sequences of word-final V followed immediately by word-initial D or T, as well as control utterances.

2.2. Results
The nasal olive was quite sensitive and picked up nasal microphonics in addition to the DC pressure variations that would be more directly indicative of velum opening. Some nasal microphonics may be present even when the velic valve is closed; the acoustic transparency of the velum to low frequencies is well-known [2, 3]. This happens particularly with high vowels (which have low F1) and voiced obstruents. Such microphonics are less evident in sounds with the higher F1 characteristic of low vowels. Thus the evidence of velic opening is to be seen in a DC pressure change, and we have taken it as an indication of nasal microphonics, vis-a-vis other comparable oral utterances.

Fig. 1. presents records of simultaneous audio (bottom) and the output of the nasal olive (top) for portions of two utterances spoken by the female French speaker. Fig. 1a gives a portion of the utterance dit 'saint' pour moi ("say 'saint' for me") /diː ə̃ puː mwa;/ and Fig. 1b, a portion of the utterance dit 'saint' bel enfant ("say 'saint beautiful baby") /diː ə̃ bèl ɛf̃. Here the initial parts of the word-initial stops (following the nasalized vowel [ɛ]) are nasallzed through perseveratory assimilation, i.e., they are prenasalized stops. However, in the case of the word initial voiceless stop (p) the nasalization is very brief, on the order of 30 m sec, whereas in the case of the voiced stop it is much longer, about 70 m sec.

Fig. 2. presents two similar records spoken by the male Hindi speaker: Fig. 2a, a portion of the utterance /ə̃ ɾə̃hə bəkə/ "you glance here" and Fig. 2b, a portion of the utterance /ə̃ ɾə̃hə dəkə/ "you see here". In Fig. 2a the signal from the nasal olive shows the word-initial /t/ on about 30 m sec of pre-nasalization. In comparison, Fig. 2b shows about 60 m sec of pre-nasalization.

Comparable results were obtained for the other speakers and other places of articulation.

3. DISCUSSION
It is essential for our argument that none of the words which provided the gross-word boundary sequences of V + D would actually exhibit a N when these words were spoken in isolation. This is certainly true of the Hindi examples. In the case of French one might recall that the liaison form of words with final nasal vowels would have a supposedly "underlying" nasal consonant appear, e.g., bon "good", [bɔ̃] but bon ami "good friend", [bɔ̃ ɔ̃]. Could the nasal element found in the French examples be this underlying nasal? We think not: such liaison consonants appear when the next word starts with a vowel, not a consonant.

Second, the fact that the appearance of the intrusive nasal is influenced by the voicing of the stop suggests that it is a purely transitional phonetic event created by the nasalization of the vowel invading the initial portions of the following stops.

The nasal epenthesis parallels and thus supports the historical scenario posited above for words like MH [t̪aːnd]. The
"phonetic" nasal can become phonologized (to use Jakobson's term) if listeners reinterpret this as an intended part of the pronunciation and not a predictable and thus discountable feature [5].

The phonetic and phonological literature on other languages reveals that voiced stops (but not voiceless ones) may tolerate nasal onsets when in contact with a preceding nasal segment (or occasionally even when there is no preceding nasal environment) [7, 8, 9, 11, 12]. We speculate that a possible phonetic basis for this phenomenon comes from perceptual evidence that some of the essential perceptual cues for voiced stop include an amplitude and spectral discontinuity with respect to adjacent sonorants, presence of voicing during the closure, and the stop burst at the release. It seems, then, that a perceptually adequate fully voiced stop may be made by allowing the initial portion to be nasal as long as the final portion has velar closure and concomitant oral pressure impulse in order to create the requisite stop burst on release. There is no motivation to the speaker to time velar closure precisely with the onset of the stop closure. On the other hand, in the case of voiceless stops there is motivation to achieve velar closure near the onset of the stop closure: to maintain voicelessness for a substantial portion of the stop closure to avoid the friction at the nostrils, i.e., a voiceless nasal, that would occur if velar closure were delayed.

It should be mentioned that the reason for selecting Hindi and French for this study is simply the fact that both their phonologies permit V-D sequences spanning a word boundary. It is just a coincidence that it is also the history of Hindi which exemplifies the puzzle we were trying to solve. If one accepts that there are universal and timeless phonetic factors which cause variation and change in pronunciation (which may lead to sound change through phonologization), then the parallels to phonetically-based sound changes should be evident, potentially, in any spoken language which exhibits the appropriate conditions.

4. REFERENCES


New York: Charles Scribner's Sons.

FIGURE 1. Records from the female French speaker: top: nasal olive; bottom: audio signal; a: portion of "dit 'saint' pour moi"; b: portion of "dit 'saint' bel enfant". Hyphen marks word boundary in phonetic transcription; vertical dotted line marks it in the graphic signals.

FIGURE 2. Records from the male Hindi speaker: top: nasal olive; bottom: audio signal; a: portion of /ap jahā tako/; b: portion of /ap jahā deko/. Hyphen marks word boundary in the phonetic transcription; vertical dotted line marks it in the graphic signals.
TIME-VARYING PROPERTIES OF CONTEXTUALLY NASALIZED VOWELS: ACOUSTICS AND PERCEPTION

Mark K. Huffman

Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.

ABSTRACT

Studies of English have shown that contextual nasalization is perceptible [8], and is used in speech processing [10]. However, when measured at single points in time, the spectral effects of contextual nasalization can appear quite subtle [7]. This suggests that time-varying properties of contextually nasalized vowels may be important to the perception of nasalization. This paper reports on acoustic and perceptual studies of change over time in contextually nasalized vowels in English. These studies focus on the degree of F1 prominence, a spectral property which is consistently affected by nasal coupling. Although changes in F1 amplitude are frequently observed with vowel nasalization [5,6], their contribution to the perception of nasalization has been given relatively little attention.

2. F1 PROMINENCE IN NATURAL STIMULI

Before running perceptual experiments, we conducted an acoustic study to determine how F1 prominence was affected by contextual nasalization in natural speech. For the acoustic study, two male speakers of American English were recorded producing words of the form bVC, where C was either /d/ or /t/, and V was one of seven vowel qualities: [i, e, a, o, u]. The data were digitized at 10,000 samples per second, and then DFT spectra were computed for the vowels at 10 msec intervals.

In the DFT spectra, we quantify F1 prominence by calculating A1-H1, the difference in relative amplitude of A1 (the harmonic in F1 with the highest amplitude) and the fundamental. Figure 1 illustrates how A1-H1 was measured, and shows how A1-H1 decreases as nasalization increases on a contextually nasalized vowel.

To examine changes in F1 prominence over time on contextually nasalized vowels, we plotted A1-H1 for each spectral frame during the vowel. To better judge which effects on F1 were attributable to nasalization as opposed to other factors, we plotted the data for the nasalized vowels along with similar data for comparable oral vowels. Figure 2 shows some examples of these combined plots, for one speaker. The second speaker showed a similar pattern.

3. PERCEPTION EXPERIMENT I: STATIC F1 PROMINENCE

The first experiment with synthetic speech focused on how changes in average F1 prominence affect perceived nasalization. Stimuli were produced by starting with a synthesized oral vowel, and then decreasing F1 prominence by increasing F1 bandwidth. For a given vowel quality, several stimuli were produced, with different degrees of F1
prominence. For each item, F1 prominence was essentially constant over the duration of the vowel. The synthetic stimuli were chosen to have F1 prominence values which covered the range observed in natural speech items with the same vowel quality. Stimuli were constructed using 2 vowel qualities: [i] and [I]. Listeners heard a synthesized /b/ syllable, followed by a vowel, and were instructed to choose which of two sound words they felt the syllable could be an excerpt from. So, for example, on hearing [bi], the listener would circle either "bed" or "bean". 17 listeners participated in this experiment. The results of the tests are given in Figure 3, which shows the percentage of nasal responses for synthetic syllables having differing degrees of F1 prominence (as measured by A1-H1).

prominence. For each item, F1 prominence was essentially constant over the duration of the vowel. The synthetic stimuli were chosen to have F1 prominence values which covered the range observed in natural speech items with the same vowel quality. Stimuli were constructed using 2 vowel qualities: [i] and [I]. Listeners heard a synthesized /b/ syllable, followed by a vowel, and were instructed to choose which of two sound words they felt the syllable could be an excerpt from. So, for example, on hearing [bi], the listener would circle either "bed" or "bean". 17 listeners participated in this experiment. The results of the tests are given in Figure 3, which shows the percentage of nasal responses for synthetic syllables having differing degrees of F1 prominence (as measured by A1-H1).

prominence. For each item, F1 prominence was essentially constant over the duration of the vowel. The synthetic stimuli were chosen to have F1 prominence values which covered the range observed in natural speech items with the same vowel quality. Stimuli were constructed using 2 vowel qualities: [i] and [I]. Listeners heard a synthesized /b/ syllable, followed by a vowel, and were instructed to choose which of two sound words they felt the syllable could be an excerpt from. So, for example, on hearing [bi], the listener would circle either "bed" or "bean". 17 listeners participated in this experiment. The results of the tests are given in Figure 3, which shows the percentage of nasal responses for synthetic syllables having differing degrees of F1 prominence (as measured by A1-H1).

The majority of the time, the stimuli were heard as oral. However, the results suggest that decreased F1 prominence can contribute to production of a nasal percept. For stimuli made with the vowel [i], the percentage of times stimuli were heard as nasal increased with decreasing A1-H1 values, from about 25% nasal responses for a vowel with an A1-H1 typical of an oral vowel (12.7 dB), to about 40% nasal responses, for a vowel with an F1 that is about 5.5 dB lower. There is a similar pattern for [I], though the data appear to be a bit noisier.

4. PERCEPTION EXPERIMENT II: VARYING F1 PROMINENCE

We tested the effect of change in F1 prominence over time on perception of nasality by comparing listener judgments of the nasality of stimulus pairs which were matched for vowel quality and overall average A1-H1, but which differed in having either an unchanging F1 prominence, as in the previous experiment, or a time-varying, decreasing F1 prominence.

The time-varying vowel stimuli were synthesized with F1 prominence decreasing throughout the vowel. (The drop from vowel beginning to end was about 4 dB). Listeners heard synthetic /b/ syllables containing these vowels within the same paradigm used in the previous experiment. Figure 4 presents comparisons of the percentage of nasal responses for the stimuli with time-varying A1-H1 on the vowel, and their counterparts with static A1-H1 on the vowel. The time-varying stimuli show a higher percentage of nasal responses than the static stimuli, indicating that a decrease in F1 prominence over the course of the vowel results in more nasal responses than a simple, static reduction of F1 prominence.

5. SUMMARY

To conclude, we have seen that change in F1 prominence over time influences the perception of nasality. By comparing perception of stimuli with static and time-varying F1 prominence, we determined that change over time is important, and that it is not just static F1 prominence which determines perceived nasality. This is evidence for the importance of dynamic information in the perception of vowels. It also may have implications for predicting the likelihood of sound changes in which a contextually nasalized vowel becomes a contrastively nasalized vowel. Since physiological adjustments other than nasal coupling can affect F1 prominence, it is possible that changes in F1 prominence over time which come with diphthongization or laryngeal adjustments for voicing could contribute to a percept of nasality. In combination with contextual nasalization, such effects could result in a stronger percept of nasality, such that the language learner will be more inclined to posit a nasal vowel in that position, providing that other grammatical considerations do not prevent such an analysis. These questions await future research.

ACKNOWLEDGMENTS

This research was supported by NIH grant #F32-NS08509. I am grateful to researchers at Haskins Laboratories and the MIT Speech Group for helpful discussions of this work.

REFERENCES


PHONETIC STRUCTURE OF WORD AND PECULIARITIES
OF ITS DEVELOPMENT
(based on Germanic and Slavonic languages)

V. Taranets
Odessa State University, Odessa, USSR

ABSTRACT

Ancient rise-fall alteration of the articulation tension with a displaced
apex (positive asymmetry) is reconstructed in CV syllable and in words with
initial stress. In the course of the development of language intensification of
tension is observed at the beginning of a word and relaxation in the end of it,
i.e. a redistribution of energy takes place. Presumably, the overall utterance
tension remains, in principle, constant.

1. INTRODUCTION

In the process of the development of language its sound aspect undergoes
the greatest alteration. Changes occur in sounds, syllables and whole words. The
object of study is root-stressed words. The study of the peculiarities of such words implies in the
first place a synchronic and diachronic investigation of its CV correlate.

2. PROCEDURE

An electro-acoustic study of the CV word (syllable) in the German and Ukrainian
language was made (the experiment was carried out in the Berlin University under
the supervision of Prof. G. Lindner) as well as that of CV syllable tension by using pressymographic
method (Odessa University, Prof. V. Taranenko). The pressymographic method made it possible to determine the
platisma and suprahyoid muscle tension while uttering a stressed CV-syllable.

In Germanic languages Runic, Gothic, Old High German, modern German and
English texts, in Slavonic languages Old Slavic, Old Russian, modern Russian
and Ukrainian texts have been studied. The dynamics of the alteration of the
initial stressed syllable as well as of the final unstressed one made it possible to determine the peculiarities of the alteration of the word as a whole in the course of language development.

3. RESULTS AND DISCUSSION

3.1. Phonetic characteristics of the CV-syllable

An analysis of the CV-syllable used in isolation and in words of the CVCV(C)
pattern revealed the following:

In Ukrainian the length of a consonant and of a vowel had an average value of, re-
spectively, 0.380 and 0.620, in German 0.359 and 0.645 (the length of the CV-syl-
lable is taken for 1.00). In German a phonologically long vowel is present. In
a general way, it may be assumed that the ratio of the length of the consonant
and of the vowel in Ukrainian and German in CV is 1:2.

The articulation apex E_max in the CV often occupies the vowel area in the fol-
lowing way: in the Ukrainian syllable 80.4%, in German - 93.5%. The rising
part forms the working phase, the falling one - the articulation relaxation
which, generally, remains one of a single muscle-tension [1]. Typical of both
languages is the rise-fall alteration in the articulation tension, with some
shift of the apex towards the beginning of the utterance (positive asymmetry)
(Fig. 1).

Fig. 1. Articulation tension of CV-syllable

3.2. Phonetic characteristics of a CVCV(C) word

Such words in Ukrainian and German are characterised by two-apex alteration of the
articulation tension. In Ukrainian E_max falls on the first syllable in 97.7% of
cases, in German in 87.5%. However the apex occupies the initial consonant in the
Ukrainian syllable in

72.9% of cases, in German it is 14.8%. It is typical of Ukrainian speech to have a
greater tension for a con-
sonant, as related to a vowel within the CV, whereas in German it is vice
versa - the vowel is more tense than the consonant. Presumably, in Ukrainian
speech realized is a "strong -consonant" phonetic type, while in German - a "strong
vowel" type.

The tonic apex falls on the first syllable in CVCV(C) units in all cases. The
intensity of the Ukrainian word falls on the first syllable in 85.0% of cases, in
German in 82.2%.

In general, the alteration of the phonetic characteristics has a rise-fall
pattern with the apex displaced towards the beginning of the utterance (sim-
ilar to the CV structure) (Fig. 2).

Fig. 2. Phonetic structure of CVCV(C)

3.3. Articulation tension of consonants (E_c)

Analysis of CV and CV-syllables made it possible to find the tension of the con-
sonant relative to the vowel, whose value is taken to be 1.00.

The analysis also revealed the difference of the E_c consonants in terms of
their formation (Table 1).
Table 1. Articulation tension of consonants ($E_0$) (Ukrainian and German languages)

<table>
<thead>
<tr>
<th>Types of consonants (examples)</th>
<th>$E_0$ Ukr.</th>
<th>$E_0$ Germ.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_w$ - semi-consonants ($w,j$)</td>
<td>1.20</td>
<td>-</td>
</tr>
<tr>
<td>$R_l$ - liquids ($r,l$)</td>
<td>1.04</td>
<td>0.84</td>
</tr>
<tr>
<td>$R_n$ - nasal ($m,n$)</td>
<td>0.84</td>
<td>0.79</td>
</tr>
<tr>
<td>$D$ - voiced occlusives ($b,d,g$)</td>
<td>0.74</td>
<td>0.57</td>
</tr>
<tr>
<td>$T$ - voiceless occlusives ($p,t,k$)</td>
<td>0.63</td>
<td>0.69</td>
</tr>
<tr>
<td>$Z$ - voiced fricatives ($v,z$)</td>
<td>0.53</td>
<td>0.28</td>
</tr>
<tr>
<td>$S$ - voiceless fricatives ($s,f,h$)</td>
<td>0.35</td>
<td>0.20</td>
</tr>
</tbody>
</table>

In German, the consonants $T$ ($p,t,k$) are opposed to $D$ ($b,d,g$) as forté/lenis [3], in Ukrainian as voiceless/voiced.

3.4. Development of initial consonants in words

Analysis of ancient and modern memorials in Germanic languages has shown that at the beginning of a word the following generalized combinations occur: $ST_-, TR_-, SR_-, DR_-, STR_-$ (where $R$ is a sonorant) with a rising tension. For example: skin, tree, snake, dream, stream; foreign words having a non-rising tension being exceptions. For example: sphinx, (Germ.) Psalm, Ndola.

The same consonant combinations are found in old Slavic texts, for example: skot, trije, slowo, zmi, bratr, strana. After the fall of reduced vowels combinations with non-rising tension were formed, such as $SS_-, TZ_-, RR_-, SST_-$ and others. For example: (Russ.) sora, rwat', mrak, zhod, mstit' [2].

Generally, the word's beginning in Germanic and Slavonic languages presents a gradual articulation intensification as contrasted with a reduction in the word's end which resulted in the relative growth of closed syllables and consonant clusters. In ancient times, open syllables with CV among them, prevailed in these languages.

4. CONCLUSIONS

Extrapolation of tendencies of the word's beginning and end development in the prehistoric period makes it possible to arrive at the following conclusions:

a) In Germanic and Slavonic languages there has existed a tendency of the open syllable, the closed syllable being a result of language development;

b) In ancient time, the combinations of initial consonants in words were formed on the principle of rising tension, combinations with non-rising tension being secondary;

c) In the end of words, there occur different types of consonant combinations which are results of the articulation reduction of this part of the word;

d) In general, in the course of language development an intensification of tension has been taking place at the beginning of words and a reduction at their end, which implies an interaction of both tendencies. It is supposed that in the course of language development the overall utterance energy remains, in principle, constant and is redistributed within the word.

5. REFERENCES


VOEEL HARMONY AS A COARTICULATORY
PHENOMENON IN NANAY

Galina Radchenko
Lund University, Sweden

ABSTRACT
The present paper reports an interaction of the context (vowel harmony), target gesture and
suprasegmental factor (stress). It is argued here that the notion of suble segments and anticipatory gestures can account much of the variability of the context, given certain
assumptions about the effect of stress.
An experiment was designed to evaluate the role of coarticulation and
tempo on the dynamics of vowel articulate. It is shown here that motor
plan for a particular segment remains the same regardless of the varying phonetic context and timing conditions.

1. INTRODUCTION
Coarticulation is defined as the influence of segmental context on the
articulatory/acoustic realization of a target segment. It is assumed that because of perceptual or articulatory constraints on
target and surrounding segments, there are limits on the temporal extent of coarticulation.

There are two particular speech production frameworks: the "look-ahead" models [6] and the "coproduction" models
[5], [1], [4]. According to the coproduction model the underlying motor
control structure for a particular segment remains essentially the same regardless of the phonetic identity of surrounding
phones. In contrast to the look-ahead models changes in the observed patterns of movement in different contexts stem
from local interactions between context and target gestures rather than from any
change in the motor plan for the target segment. According to the look-ahead model the motor plan for the target
segment and consequently the time at which coarticulation begins is revised

and adapted to the context since every different context poses a different set of

conditions.

The present paper supports the point of view that was put forward by Bladon &
Al-Bamerni [2], who proposed that observed coarticulatory patterns might be a
combination of anticipatory feature spread plus stable gestures.

In the present paper we examine coarticulation presented by vowel harmony phenomenon as an interaction
between target gesture with the context. According to preliminary auditory analysis
the adjustment of Nanay consonants to a certain harmonic class of vowels manifests itself in the use of dental
allophones of affricates [ts], [dz] with the vowels of the first harmonic series /i/, /o/,
pronounced with low jaw position and palatal allophones [ql], [ql] with vowels of the
second harmonic series /i/, /a/ pronounced with high jaw position:

- otsqa 'small fish', qotqa 'uncle'.

The velar allophones [k], [q], [x] are used with the vowels of the second
harmonic series [a], [u] and uvular allophones [u], [o], [x] with the vowels
of the first harmonic series [a], [o]: chLa 'trie', xapa 'stutter'.

Most investigators of this problem mention that the influence of context upon
vowels is regularly manifested as a displacement of vowel-formant frequencies away from their target
frequencies. An experiment was therefore designed to study vowels pronounced under varying timing
conditions and in systematically varied consonantal environments.

2. PROCEDURE
To investigate the displacement of consonant articulations from their bull's-

eye patterns X-ray pictures of consonants in varied vowel environments were
made. To study the displacement of vowel formant frequencies away from their target
frequencies, measurements were made of the first, second and third formant frequencies of vowel /a/ in three
consonant environments: /p-p/, /h-t, /h-x/ under varying timing conditions. F1, F2 and F3 of the investigated vowel were
plotted against vowel segment duration, then the first and the second formant
target frequencies were determined by plotting F1 and F2 against vowel duration
for the three contexts of the vowel simultaneously.

The influence of vowel and tempo on vowel properties was studied by the
manipulating of the word order and rhythm of the carrier sentence frames: 'CVC ----'; ---- CVC ----'.
Each sentence pattern was pronounced four times.

The speech material was read by a native male speaker of Nanay, and was
analyzed by MacSpeech Lab 2.

3. RESULTS
X-ray pictures show the adjustment of consonants to the near vowels.
Dental affricates are characterised as flat and have a decrease of frequency
components of release (the concentration of noise is at 2500 - 3000 Hz). It
promotes reducing of the timber of the near vowel.

Palatal affricates may be called sharp and are characterised by high intensity
of the upper frequency components (about 4000 - 6000 Hz). Articulatory palatal
sounds are produced with wide pharyngeal cavity. X-ray pictures show that
the constriction location of palatal affricate [ql] is at the zone of alveolar and
hard palate, whereas constriction location of [dz] is limited by dental zone (fig. 1-2).

Figure 1. X-ray picture of dental affricate [dz] pronounced in the syllable [adza] by
Nanay male speaker N 3.

Figure 2. X-ray picture of the affricate [ql] in the syllable [afqa] pronounced by
Nanay male speaker N 3.

X-ray pictures of allophones [q] and
[k] show that their constriction location depends on the articulation of the near
vowel. Vowels pronounced with low jaw
position promote the shift of constriction
location from velum to uvula (fig.3).
Vowels pronounced with high jaw
position /a/, /u/ initiate a constriction
location of the near consonant /h/ at velum
(fig.4).
In all these cases of assimilation there are no changes at higher levels: Supra, Vocal Tract, Oral remain the same, changes undergo only at a low level: the place of location. This slight transition of the place of articulation seems not to change motor plan for target gesture, non-significant changes in gesture movements are the result of local interactions between context and target gesture.

The adjustment of vowels to the varied phonetic context is displayed in figures 5-6.

4. DISCUSSION

The discussed cases of coarticulation in Tungus may be viewed as realization of anticipatory speech model. Coarticulation is the result of local interaction of overlapping gestures: jaw lowering or jaw rising during the production of a vowel affects the articulation of the consonant. It is proved by a number of experimental research works that onset of jaw lowering for a vowel will start during a preceding consonant [8]. Vowel harmony may be viewed as constraints imposed by the context on the realization of a target gesture.

The changes in the segmental context are predictable and usually set up as phonotactic rules. They reflect the range of possible variability of the target gestures in a different context. The motor plan for target gestures seems to be done beforehand and changes in gesture movements in different contexts stem from local interactions between context and target gestures.

5. CONCLUSION

The present paper reports on a pilot study of vowel harmony in Tungus languages. Preliminary results have shown that restrictions imposed by a context do not change essentially a motor plan for a target gesture. A number of questions are raised that future research is supposed to answer. These questions are: what restrictions are there imposed by a context (vowel harmony) on the realization of a target gesture? what relations are there between vowel harmony, target gesture and stress? to what extent can context change a target gesture?

REFERENCES


TENDANCES UNIVERSELLES ET STABILITÉ DES SYSTÈMES VOCALIQUES

N. Vallée, L.J. Boë et J.L. Schwartz
Institut de la Communication Parlée, URA CNRS n° 368
ENSJ/ENSEMG - Université STEN DHAL,
Domaine Universitaire, BP 25X, 38040 Grenoble cedex, France

ABSTRACT
This paper deals with a study of the structure of vowel systems in two respects: 1) Our observations of certain aspects of vowel systems using the current database of 317 language descriptions [4] lead us to confirm or refine certain tendencies and regularities in vowel systems; 2) We have used a predictive model [7] of the 3-dimensional $F_1/F_2/F_3$ space to test our hypothesis that if a system is more frequent in the inventory, it is an acoustically "stable" system. This research extends Lindblom's work about the predictive models of the organisation of the vowel space and the explanations of language universals.

INTRODUCTION
On dispose aujourd'hui d'inventaires phonologiques des langues du monde relativement indépendants. Ces matériaux offrent la possibilité de tester de nouvelles typologies et propositions de tendances de développement des systèmes vocaliques, nourrissant ainsi la discussion sur les modèles de prédiction qui unissent le réseau universaux dans les langues. C'est pour les systèmes vocaliques que ces modèles ont été le plus développés. Dans un premier temps, nous présentons les points essentiels qui ressortent de notre typologie des systèmes vocaliques [8], établie à partir de la base de données UPSID (UCLA Phonological Segment Inventory Database) qui réunit la description phonologique de 317 langues du monde [4]. Nous avons pris la totalité de la base en excluant tous les classificateurs qui ont pour objectif de repousser ou confirmer certaines tendances observées dans les fréquences d'occurrences des systèmes et des voyelles ainsi que dans l'organisation des 220 types de systèmes vocaliques relevés dans l'ensemble de la base. La deuxième étape a consisté à tester, à partir de cette typologie, la stabilité acoustique - dans un sens qui sera précisé - des systèmes vocaliques les plus fréquents, avec un modèle de prédiction dans l'espace 3-D $F_1/F_2/F_3$ [7]. Ce modèle intègre le principe de dispersion maximale de Ljungcrantz & Lindblom (1972) [2] (L&L) et le complète par un critère de focalisation introduit par le biais de plans attracteurs ($F_1=F_2$, $F_2=F_3$, $F_3=F_2$), ainsi qu'une pondération du second formant effectif $F_2$.

1. STRUCTURE DES SYSTÈMES VOCALIQUES DE LANGUES NATURELLES
1.1. Taille et occupation de l'espace
Les systèmes décrits dans UPSID possèdent de 3 à 24 voyelles. Le classement de la base nous a permis d'en extraire 220 types. Nous avons relevé dans l'inventaire une très nette dominance des systèmes à 5 voyelles (23% des langues). Les systèmes qui possèdent de 3 à 10 voyelles représentent 80% de l'échantillon et sont donc très largement majoritaires. Les systèmes les plus fréquents possèdent une large dispersion dans l'espace articulatoire traditionnel décrit par les axes d'aperture et de lieu d'articulation. Ils sont composés de voyelles que l'on retrouve quelle que soit la taille du système. Nous avons pu mettre en évidence une différence très nette de tendance entre les types : 9 représentent le maximum de fréquences dans l'espace vocalique. Ceux qui ont un nombre élevé de voyelles ne développent pas de nouveaux timbres, mais ajoutent une complexité articulaire à ces segments de base, en leur additionnant d'autres dimensions telles que la longueur, la nasalité, la pharyngalité ainsi que $/A$, $/A'$, $/A''$.

2. ORGANISATION DE L'ESPACE
Nous allons décrire par 10 règles comment les timbres s'organisent en série (sur les traits [antérieur], [central], [postérieur], [arrondi], [non arrondi]).

2.1. Organisation horizontale et verticale

1.2. Comparaison des séries
Dans un système donné : 1) la présence de voyelles périphériques est supérieure au nombre de voyelles intérieures — 100% des langues possèdent des voyelles périphériques et 44% des voyelles intérieures. 2) le nombre de voyelles antérieures non arrondies /i, e, e', e, a, A/ est plus grand ou égal au nombre de séries antérieures /o, o', o, o, o/ (24%).

3. REGULARISATIONS
Dans les langues :
1. Les voyelles périphériques /i, e, e', e, a, A, o, o', o, o/ sont plus fréquentes que les voyelles intérieures (avec /i, a, A/ prévalant dans 10% de cas).
2. Les centrales non arrondies /i, i, a, A, A, A, A/ sont plus fréquentes que les antérieures /y, y, y, y, y, y, y, y/.
3. Les postérieures non arrondies /u, u, u, A, A, A, A/ sont plus fréquentes que les antérieures /y, y, y, y, y, y, y, y/.
4. Les centrales non arrondies ont une occurrence plus forte que les centrales arrondies.

2.1. Organisation dans les séries
@ Les voyelles antérieures arrondies apparaissent :

- par série — les séries d'antérieures arrondies de 2 phonèmes apparaissent dans les systèmes ayant au moins 7 voyelles, celles de 3 ou 4 phonèmes apparaissent dans les systèmes à 16 et 19 voyelles,

- toujours avec les voyelles antérieures non arrondies de même aperture,

- presque toujours avec les voyelles postérieures arrondies de même aperture,

@ souvent sans la voyelle postérieure périphérique de même aperture (ce qui justifie l'essentiel des cas de dispersion de /u/ remplacé par /u'/). Cependant, les voyelles postérieures arrondies figurent généralement seules dans leur catégorie — les séries atténuées contiennent au plus 3 voyelles,

- selon l'ordre de fréquence suivant : $/A'/ > /A'/ > /A'/ > /A'/$. 4.

3. ESTUDE DE LA STABILITE ACOUSTIQUE DES SYSTÈMES VOCALIQUES
2.1. Utilisation d'un modèle de prédiction

Dans un espace 3-D $F_1/F_2/F_3$, le modèle de Schwartz & al. reprend de L&L la minimisation des distances intervoyelles comme fonction de l'énergie des systèmes :

\[ E_0 = \sum_{ij} d_{ij}^2 \rightarrow \text{minimise} \]

\[ r_{ij} = \sqrt{d_{ij}} \]

où $d_{ij}$ est la distance formantique pondérée entre 2 voyelles et $i$ et $j$.

\[ d_{ij} = \sqrt{(F_1 - F_1)^2 + (F_2 - F_2)^2 + (F_3 - F_3)^2} \]

$A$ est le coefficient pondérateur des formants élevés ($A = 1$). Un poids plus important de $F_1$ dans le calcul des
distances améliore la prédiction des voyelles périphériques en diminuant le nombre de voyelles hautes.
Le calcul de $E_0$ repose donc sur un critère systémique, les voyelles de grande hauteur sont donc les plus éloignées au sens d'une distance pondérant $F_1$ par rapport à $F_2$. Mais ce critère ne permet pas de prédire la structure de voyelles hautes /i/, /u/ (attestée des 7 voyelles). Schwrat & al. ont donc introduit un deuxième critère — celui-ci extra-systémique — qui favorise pour chaque voyelle les rapprochements $F_2$, $F_3$, $F_4$ et diminue l'énergie du système. C'est le critère de focalisation, justifié d'un point de vue perceptif : les voyelles focales sont préférentiellement [6].
\[
E = E_0 + \alpha (E_{12} + E_{23} + E_{34})
\]
\[
E_{12} = \frac{1}{(F_2 - F_1)^2}
\]
\[
E_{23} = \frac{1}{(F_3 - F_2)^2}
\]
\[
E_{34} = \frac{1}{(F_4 - F_3)^2}
\]
\[\alpha \text{ est appelé coefficient des plans attracteurs.} \]

D'un point de vue analogique, les voyelles s'éloignent les unes des autres tout en étant attirées par des plans attracteurs $F_1$, $F_2$, $F_3$, $F_4$.
Pour utiliser le modèle il nous a fallu disposer :

- d'un espace maximal 3-D ;
- de valeurs formantiques $F_i$ (i=1 à 3), $F_4$ étant fixé à 3350 Hz, pour les 37 voyelles de la base UPSID (cf. Fig. 1).

Ces points ont été obtenus grâce à une série de travaux précédents fondés sur le modèle articulatoire de Maucla (3) et grâce à la confrontation des données formantiques de 15 études, issues de travaux sur des modèles ou sur des langues naturelles.

2.3. Systèmes sélectionnés pour l'étude
A partir de notre typologie [8], un ensemble de 25 systèmes de 3 à 9 voyelles sélectionnées comme les plus fréquentes ont été testés pour 3 valeurs de $\lambda$ : $0.25$, $0.5$, et $5$ valeurs de $\alpha$ : $0.25$, $0.5$, $1$, $2$ et $10$, soit un total de 375 tests. Nous avons procédé en sorte qu'il y ait au moins deux voyelles concourants (ou plus) par nombre de voyelles. Cette comparaison significative possible des énergies est entre les systèmes de même taille et pour des couples $(\lambda, \alpha)$ identiques.

2.4. Résultats
Notre modèle prédit la stabilité de 64% des systèmes les plus fréquentes (qui représentent après regroupement des systèmes relativement proches 60% de la base UPSID).
Les cas non prédits stables par le modèle sont ceux qui possèdent au moins 4 voyelles sur un des bords de la périphérie et qui ne sont pas équidistants ou ceux qui possèdent une ou plusieurs voyelles intérieures sans posséder une seule centrale haute (soit 36% des systèmes testés et approximativement 20% de la base UPSID). Une voyelle centrale telle que $/A$/ est plus facilement prédite qu'une voyelle comme $/O$/ pour de petites valeurs de $\lambda$; $/O$/ se déplace vers les périphériques antérieures et pour des valeurs de $\alpha$ grandes, vers les voyelles centrales hautes.

La poursuite des tests avec des valeurs de $\lambda$, $0.25$, et des valeurs de coefficients différents pour chaque plan attracteur $(\alpha_{12}$, $\alpha_{13}$, $\alpha_{14})$ pourra améliorer les résultats précédents.

Nous n'avons pas mis en évidence un seul couple $(\lambda, \alpha)$ constant pour l'ensemble du modèle, mais 2 couples de coefficients pour lesquels les résultats sont les plus satisfaisants ($0.25$, $0.5$) et ($0.5$, $0.25$). Les tests ont mis en évidence que pour $\lambda > 0.5$, la prédiction des systèmes "not périphériques" est impossible jusqu'à 9 voyelles. Le facteur $\alpha$ est nécessaire pour la prédiction de $/O/$ pour un modèle $/A/$.$/O/$ se déplace vers la position formantique de $/A$/ ou $/u/$.

CONCLUSION
Le modèle de prédiction a permis de simuler la plupart des systèmes sélectionnés, améliorant les résultats de Léda [2] et Lindblom [3] à propos des voyelles centrales hautes et de la prédiction de $/O/$, stabilisé par notre critère de focalisation $F_3$ [1] (cf. Fig. 2). A part le problème des voyelles trop nombreuses et celui de $/A$/ (qui est aussi, on le sait bien, un problème de description phonologique), le modèle prédit des systèmes le plus souvent en accord avec les grandes tendances de développement des systèmes voca. Même il ressort de notre étude que tous nos modèles testés ne sont pas stabilisés a une notion d'énergie basse pour traduire un des critères d'organisation des voyelles dans les systèmes reste un bon critère.
Avec ces résultats, notre recherche améliore les prédications et étend les travaux de Lindblom & Lijencrant [2]. "to derive linguistic form as a consequence of various substance-based principles pertaining to the use of spoken language and its biological, sociological, and communicative aspects."

RÉFÉRENCES
Language 48, 839-862.
Phonetic Universals in Vowel Systems.
The Size and Structure of Phonological Inventories: Analysis of UPSID.
Experimental Phonology, Ohala J.J. (Ed.), Academic Press, New York, 103-123.
An Articulatory Model of the Tongue Based on a Statistical Analysis.
MARQUES SEGMENTALES SOCISITUATIONNELLES EN FRANÇAIS CONTEMPORAIN

G. Jetchev
Université de Sofia, Bulgarie

ABSTRACT
This paper discusses a possible classification of phonological variability phenomena in Modern French. Further the phonetic realizations of schwa in two variants of dialogue are treated in relationship with prosodic and syntactic boundaries.

1. TROIS TYPES DE VARIATION PHONOLIGIQUE
Si, comme l'affirme P. Encrééd (2), la variation potentielle est à localiser dans les représentations phonologiques et si "les autosegment flottants sur la ligne segmentale sont des lieux privilégiés pour les variations socialement distinctives" (22, p.256, note 33), nous proposons le modèle hiérarchique suivant pour le français contemporain:

1.1. Variation des mots à éléments flottants
On peut admettre que dans les formes sous-jacentes de nombreux mots (par ex. hôtesse, prononcé [ote] ou [ote] ; cf.(9)) il y a des segments, pour le reste accrus dans le squelette, qui comportent un élément (au sens attribué à ce terme dans (40) flottant, c-à-d. non associé. Une telle représentation phonologique de ces segments rendrait compte des alternances entre voyelles [-ATR] et [-ATR] dans les rimes non branchantes des syllabes non finales d'un grand nombre de mots français. Pour hôtesse on aura le forme sous-jacente suivante:

   A R A R
   /  
   
   La réalisation de ce mot en surface sera [ote] au cas où l'élément i se serait accréé dans le squelette et [ote] si i est resté non associé. Par contre, le segment [e] dans hôtesse ne contient aucun élément flottant, car il occupe le noyau d'une rime branchante, ce qui fait que ce segment n'est pas sujet à variation.

1.2. Variation des mots à segments flottants
La variation de ce deuxième type s'explique par la présence, dans la représentation lexicale du mot en question, d'un segment entièrement flottant, ce qui vide l'ensemble de ses éléments constitutifs se trouvent affectés par la fluctuation. Dans le cas du français, ce segment flottant peut être un schwa, une liquide devant schwa en contexte post-consonantique, une consonne de liaison. Cependant tous les schwas en français ne sont pas flottants (cf.(22,p.250).

1.1. Deux formes sous-jacentes pour le même mot: à segment fixe/à segment flottant
Le troisième type de variation caractérise certains mots appartenant à des classes fermées (cf. les concepts de mots synthétiques, mots phrasétiques par opposition à mots prédicatifs (7)) et jouissant d'une grande fréquence dans le discours, pour lesquels on doit admettre au moins deux représentations phonologiques différentes: l'une contenant un segment donné en tant que fixe et l'autre, ce même segment en tant que flottant. Voici les deux formes sous-jacentes proposées dans (7), p.214 pour le pronom clittique cu :

   - avec voyelle fixe (par ex., dans cu arrises prononcé [tyariv])
     A R A
     /  
     
   - avec voyelle flottante (dans cu arrises prononcé [tariv])
     A R N
     /  
     
Ces deux formes sont différemment distribuées en fonction de la variante socio-situationnelle (cf.(69), mais aussi suivant des facteurs prosodiques et sémantico-pragmatiques (accent anaphorique, rhétmatisation, topicalisation, etc. cf.(69)). Pour un inventaire, non exhaustif, des mots synthétiques et phrastiques du français contemporain qui auraient deux et parfois plusieurs représentations lexicales socialement distribuées, nous renvoyons à (33), p.66-92.

2. SCHWAS ET FRONTIERES
Dans la suite nous voulons démontrer que le jeu d'associa- tion/déassociation du schwa, segment flottant ou épenthétique, dans les réalisations de surface est souvent régi par le contexte prosodique et morphosyntaxique et peut s'assumer, surtout dans certaines variantes sociosituatio- nnelles, un certain rôle dans l'organisation hiérarchique de l'encadré, en signifiant le caractère jonctif/disjonctif d'une frontière.

- Deux corpus d'enregistrements authentiques, représentatifs de deux variantes du dialogue en français contemporain - l'inter- view (corpus A) et la conversation (corpus B) - ont été analysés. Les locuteurs du corpus A se caractérisent par un degré d'intégration au marché linguistique supérieur à celui des locuteurs B. Les premiers possèdent un niveau d'instruction élevé et maintiennent plusieurs types situationnels de messages. Les productions orales des seconds se situent le plus souvent à l'intérieur d'une seule variante linguistique: la conversation.

- Au total, 1481 occurrences de schwa ont été dénombrées dans le corpus A (dont 39,6% réalisées) et 1254 dans le corpus B (dont 33,5% réalisées). Si le taux de réalisation des schwas dans les deux variantes de dialogue est assez proche, des divergences plus importantes ont été constatées au niveau distributions en fonction du contexte prosodique et morphosyntaxique.

2.1. Schwa en syllabe initiale de groupe prosodique
Pour faire apparaître l'incidence de la rupture prosodique entre la syllabe contenant schwa et le groupe précédent, les occurrences de schwa dans les trois contextes suivants ont été dépourvues :
- après une frontière de groupe accentuel: Fa,
- après une frontière de groupe intonatif n'intégrant pas de pause silencieuse Fi(-p)
- après une frontière de groupe intonatif intégrant une pause réelle: Fi(p).

Pour l'investigation des Fe et Fi, les indices acoustiques et perceptifs établis par A. Di Cristo (11) ont été utilisés.
Conscient du fait que la réalisation phonétique du schwa en français est une fonction à multiples variables, nous avons:
- neutralité: l’incidence du contexte segmental, ne retenant que les occurrences où schwa était précédé d’une seule consonne, c-a-d. les contextes:

\[ \ldots /V/Fa/CaC\ldots \\
\ldots /Fi(-p)/CaC\ldots \]

(Dans la grande majorité des cas les schwas retenus étaient des schwas de monosyllabes clittiques.)
- negligence: l’incidence du contexte segmental, ne retenant que les occurrences où schwa était précédé d’une seule consonne, c-a-d. les contextes:

\[ \ldots /V/Fa/CaC\ldots \\
\ldots /Fi(-p)/CaC\ldots \]

En fin de polysyllabe, même précédé de deux consonnes, c-a-d. dans un contexte où on a l’habitude de l’entendre se réaliser, le schwa n’est prononcé que rarement. Cette même tendance a été signalée par V.Lucci (<5> et par P.Léon (<3>). On peut donc supposer que dans le contexte des deux consonnes, la non réalisation du schwa contribue à marquer la frontière finale d’un mot polysyllabique.

2.3. Schwa en fin de groupe prosodique

Les schwas que l’on trouve en fin de groupe prosodique sont aussi finales de polysyllabes. Cependant, nous avons constaté qu’une grande partie de ces éléments réalisent dans le contexte:

\[ \ldots /VFi(-p)/CaC\ldots \]

et contribuent ainsi à signaler le caractère jonctif de la frontière avec le groupe prosodique suivant.

Nous en concluons que les réalisations de schwa jouent probablement un certain rôle pour l’identification du caractère de jonction (en contexte \[ \ldots /CaFa/C\ldots \]) ou de disjonction (en contexte \[ \ldots /Fi/CaC\ldots \]) d’une frontière entre groupes prosodiques. Le plus souvent, ce rôle n’est que redondant, la distinction étant assurée par un intonème (cf. <3>), mais nous supposons qu’au cas où la réalisation de cet intonème serait interdite par des contraintes phonotactiques ou syntaxiques, le jeu des schwas pourrait assurer à part entière cette fonction de marque jonctive/disjonctive.

Références

ABSTRACT
This paper presents some aspects of the prosodic and paralinguistic features of onomatopoeia of informal oral communication in Kayabi (Tupian language of Brazil).

1. INTRODUCTION
Kayabi belongs to the Tupi-Guarani language family of Brazil as classified by Rodrigues [7]. The language is spoken by about 500 Indians living in the rain forest of Mato Grosso and Pará of Central Brazil, where these data were gathered.

2. FUNCTION OF ONOMATOPOEIA
Onomatopoeia is used extensively in conversation and narration of texts, but less in formal and written style. Verbose onomatopoeic forms specify and add detailed information to the finite verb they accompany or substitute. This information can be aspect of the action (progressive, iterative, etc.), rhythmic pattern of the action (tempo, effort exerted, regularity of movement, etc.), or description of the actor or instrument (weight, size, number, etc.). Verbose onomatopoeic forms part of the lexical inventory of the language, and are not inflected like finite verbs, but can lexicalize into verbs by adding a causative prefix to the onomatopoeia. Onomatopoeia referring to bird and mammal sounds are mostly echoic of their calls or noises. Many of the bird names are formed according to their echoic form.

3. ARTICULATIONS AND PHONOTACTICS
The following are the phonemes of Kayabi:

- /p t k kw/  
- /m n n/  
- /f s/  
- /r wj y/  
- /i i u/  
- /e e a a/  
- /o o/  

The permitted syllable pattern is CV in running speech and (C)V(VC) utterancefinal.

Onomatopoeia show deviance in manner and place of articulation, adding the following segments and sequences:

- /pt [t] t [t] [t]/  
- /m [m] [m] [m]/  
- /fl [l] l [l]/  
- /w [w] [w] w/[ ]  

Phonotactic differences are that the segments ['t] and ['h] (voiceless vowel equivalent) can occur also word final; syllabic consonants /ts/ [t] [s] [t] [s] and [l] as nucleus, and preglottalised fricative ['s]' initially.

CVC can also occur utterance medial in sequences of onomatopoeic words and reduplication, accompanied by specific tempo, pitch and stress patterns.

Examples are:

- /pok pok pok/ (reduplication) ‘chop lots’
- /pok ‘pin/ (sequential) ‘chop and fall’

4. PROSODIC AND PARALINGUISTIC FEATURES
4.1. Rhythm
Timing, stress, pitch, duration and iteration often occur to achieve the desired rhythmicality and aspect of the action.

Rapid speech (-----) expresses an action with fast movements, smallness, light weight, or repeated actions; whereas slow speech (-----) indicates slow movements, largeness, heavy weight, or fewer actions.

Staccato with pause breaks, often linked with lengthening, can indicate a heavy actor with slow movements; whereas legato indicates a continuous or prolonged movement. Speed-up in tempo signals increase in force of action, like ‘coming-to-a-boil’. Duration can be of vowels and of syllables.

- /pok ‘pin/  
- /pok ‘pin/  

‘animal jumps around’

- /pok ‘sorok/  

‘monkey jumps and lands lightly’

- /pok ‘sorok ‘i ‘t ‘i ‘w/  

‘jaguar jumps, lands forcefully, and runs with heavy steps’

- /wâ’ml/  

‘jaguar slanders along’

- /wâ’ml/  

‘aeroplane rises up’

- /wâm ‘wâm/  

‘howler monkey swings from branch to branch’
4.2. Pitch

Verboid onomatopoeic words and phrases spoken with high pitch signal a fast action (often accompanied by allegro tempo), light weight, small, lots; whereas low pitch indicates a slow action (often accompanied by legato tempo), heavy weight (accompanied by lengthening), few, or large.

/m/ 'mosquitoes passing by fast'
/wa wa wa wa/' 'aeroplane flying'

4.3. Phonation

Voice qualities used for echoic onomatopoeia are: creak (\(\tilde{a}\)), ingressive creak, falsetto creak, breathy voice (\(\tilde{a}\)) and whisper (\(\tilde{a}\)).

\(/m\)/ 'mum 'mmy 'currawong call'
\(/m\tilde{n}/ 'paca noise'
\(/k\tilde{a}/ 'macaw call'
\(/\tilde{a}\)/

(laryngealised and ingressive) 'howler monkey noise'
\(/p\tilde{r}\tilde{a} p\tilde{i} p\tilde{r}i\tilde{x}/

'paca flaps ears'
\(/k\tilde{e}\tilde{k}/ 'koox koox x\tilde{h}/ 'toucan call'

4.4. Articulatory Control

Some animal imitations are articulated with different articulatory gestures such as constricted larynx (small monkey), lowered larynx (large monkey), a more precise and forceful articulation accompanied by speed up of replications (chopping), unrounding of rounded vowels (noise of howler monkey), and other modifications of the vocal channel which are combinations of several of the above gestures.

\(/m\ m m m m m/ (low larynx) 'piraha fish jumps up to get leaves'
\(/\tilde{a}\tilde{a}\tilde{a}/

(ingressive creak) 'howler monkey'

4.5. Initiation of Airstream

Some echoic forms are pronounced with an ingressive airstream, such as noises made by certain fish, monkeys, and other animals.

\(/\tilde{e}\tilde{u} \tilde{e}\tilde{u} \tilde{e}\tilde{u}/

(ingressive breathy) 'capibara in water'

\(/\tilde{a}\tilde{u} \tilde{a}\tilde{u} \tilde{a}\tilde{u}/

(ingressive, whispered-breathy) 'spider monkey'

4.6. Repetition

Onomatopoeic words can be repeated up to eight times to signal increased length of time or space, the number of repeated actions, actors, or participants. Replication describes iterative actions, or sequential happenings.

Many of the examples given in 4.2 fall into this category. Other examples are:

/t\tilde{k}/ 'give to one person'
/t\tilde{k} \tilde{k} \tilde{k}/ 'give to several persons'

5. ACOUSTIC ICONICITY

Acoustic representations of iconic sounds show the following tendencies:

Initial fricatives give the idea of entering, piercing:

\(/\tilde{a}\tilde{e}\tilde{k}/ 'fall into, pierce'

A final stop seems to be telic and initiatory of the forceful endpoint of the action (such as hit or fall).

\(/\tilde{k}\tilde{e}\tilde{k}/

'something hard or heavy falling or being cut'

A final nasal represents a continuous sound:

\(/\tilde{a}\tilde{u} \tilde{a}/ 'ring, rattle'

A voiceless trill indicates a series of quick movements:

\(/\tilde{e}\tilde{v}/

'humming bird flying on the spot, or a small fish jumping along the water surface'

A final vowel indicates a continuous movement similar to an audic aspect:

\(/\tilde{u} \tilde{u} \tilde{u} \tilde{u}/ 'walk along'

A high front vowel seems to convey the idea of smallness or fear:

\(/k\tilde{a}/ 'noise of tapir pup'

\(/\tilde{w}\tilde{u}/

(weak voice) 'noise of spirits'

A high back vowel gives the idea of largeness:

\(/k\tilde{a}/ 'noise of a howler monkey'

6. CONCLUSION

In Kayard, an action that is expressed by a generic term can be further qualified by a vast choice of onomatopoeic words and phrases, transmitting detail about the performer and performance to reinforce the oral message thus making it understood and vivid.

Tape recordings of several hundred verboid and echoic onomatopoeia made by several men show that onomatopoeic forms are not coined arbitrarily by the speaker, but are part of the lexicon of the language and reflect a specialised knowledge relevant to the community’s environment and everyday life.

REFERENCES

ABSTRACT

A number of speech and voice changes are associated with advancing age -- even in individuals who are normal. Apparently, they are due to tissue loss and reduction in mobility (the physiological model). However, a second theory is needed to account for other alterations: it is the Male-Female Coalescence Model of Vocal Aging. Specifically, it has been established that, at puberty, the sexes become biologically less like each other; these processes appear to reverse during female menopause (and its counterpart in males) with the sexes becoming more alike. Thus, the cited model supplements the older theory and permits more accurate predictions of those speech changes which correlate with advancing age.

1. INTRODUCTION

Certain voice and speech alterations appear to accompany the aging process. However, the nature and extent of these changes is not well understood. For one thing, it has been argued [11] that they are due to pathologies that are associated with old age. However, this position has been sharply countered by a large number of authors [6,8,17]. It is conceded, of course, that the elderly can exhibit pathologies of many types (and that some of these can lead to changes in voice and voice). Yet, it also can be expected that, while cohorts of normally aging people do not suffer from these deficit related changes, they nonetheless will exhibit shifts of some type -- and there is evidence to support this postulate. For one thing, it has been found that auditors are fairly good at accurately estimating the age of talkers from their speech [9,13,16]. If some sort of change had not occurred, re: the older talkers, these judgements simply could not have been made. But, what are these changes? Do they result from growing old (chronological age), from physiological changes, or from some other set of factors? Of course, chronological age plays a major role in the process as the shifts must be related to the passage of time. However, are they well documented physiologically or do other factors also operate? Two theories appear useful in this regard; they are the physiological model of vocal aging and the male-female coalescence model.

2. THE PHYSIOLOGICAL MODEL

In the past, most investigators have subscribed to a theory that can be referred to as the "physiological model" of vocal aging. Although not always articulated as such, this theory explained the normal aging process in terms that results from a reduction in the efficiency of the human biosystems as a function of time. Specifically, the changes which are observed in the elderly are said to be due to tissue atrophy/reduction and an associated loss of mobility. (See Figure 1.)

In a sense, the physiological model was developed (at least informally) in response to the question "What is old?" [4,17]. In this regard, it was observed [1-4,8] that the chronological age of a person does not always appear to best represent their "actual" age. Indeed, if these (and other) authors are to be believed, the aging process is neither linear nor invariable. That is, it would appear necessary to directly assess the mental, sensory, motor and communicative capabilities of older individuals in order to determine their true age. It is not argued, of course, that no degradation occurs; rather it would appear that the process is one that varies -- and sometimes dramatically -- from person-to-person. Hence, the physiological model of aging served to replace the traditional chronological metric. Unfortunately, even today, this model is far from complete. A brief review of the relevant factors may be found in Hollien [8], and, of course, these relationships are being continually updated. On the other hand, even though this physiological model accounts for many of the changes which occur as a person ages, there still are discrepancies.

- Speaking fundamental frequency (SFF or F0) has been studied extensively in populations of all ages (see Figure 2). As can be observed, SFF levels shift markedly during adolescence with this lowering much greater for males than females. The data for mature males suggest that F0 is further lowered during adult life but then begins to rise as middle age is concluded -- and perhaps rises sharply as male cohorts reach an age where they can be classified as elderly. This pattern is one that could be predicted on the basis of the physiological model. The configurations for females follow a slightly different course. The downward shift in SFF is seen at puberty -- even though it is less extensive than that for males -- and then SFF levels appear not to change much during life, excepting perhaps for a very slight rise in the elderly. On the other hand, there now are data which support the notion of a downward shift in female SFF as a function of old age [5,12,18]. This change is one which is contrary to that predicted by the physiological model. In that instance, the atrophy associated with aging would be expected to force a rise in female SFF -- just as it does in males. Apparently, this does not happen.

Figure 1

2.1 Relevant Data

- It is without question that most of the research carried out on the communication characteristics of older persons has been focused on the deficits resulting from one form of pathology or another. Moreover, even when normal subjects are the focus of interest, it is the human voice that tends to be studied -- and the discussion to follow will reflect these biases. Specifically, the vocal/speech correlations of aging include speaking fundamental frequency (especially), vocal intensity, speech spectra, timing and, in some instances, perceptually based information.
5. REFERENCES

ON INGRESSIVE GLOTTALIC AND VELARIC ARTICULATIONS IN XHOSA

J.C. Roux

Dept. of African Languages, University of Stellenbosch, Stellenbosch, South Africa

ABSTRACT
The aim of this paper is to present some data on the bilabial implosive and on click articulations in Xhosa, a language belonging to the Bantu group of languages. Variations in implosive production will be enunciated paying specific attention to the inadequacy of phonological features to account for phonetic differences among languages. Distributional characteristics of certain click types in Xhosa will then be considered. An articulatory motivation for the occurrence of these specific forms will be proposed.

1. IMPLOSIVE BILABIAL
Xhosa exhibits one bilabial implosive [b], orthographically presented as b, when not present in a nasal combination, e.g. abantu (abantu). Impressionistic phonetic descriptions furthermore refer to other bilabial plosives occurring in this language, viz. a bilabial plosive with full breathy voice in nasal compounds, e.g. imbuzi [imbi], and a bilabial plosive with delayed breathy voice occurring in non-nasal environments, e.g. -bala [bala] [1]. A computer assisted phonotactic analysis [3] of the sound system of Xhosa based on grapheme to phoneme conversions following the above mentioned conventions indeed indicated a high occurrence of the implosive vis-a-vis the other two plosives types. An analysis of 294,965 /CV/-syllables yielded 8.2% implosives in the /C/-position, with 12.5% breathy voice plosives and 6.42% delayed breathy voice plosives in the same position. In real life, however, the phonetic qualities of the implosive in Xhosa seem to change quite extensively, inter alia, as a function of tempo. Figure 1 represents the articulation of an intervocalic bilabial implosive, produced at a relatively slow (deliberate) speaking rate. This articulation may be regarded as a "classical" implosive sound. Total closure of the vocal folds followed by some amount of pre-voicing prior to the bilabial release may clearly be noted. This observation renders some support for the traditional view [1] that the vocal cords may start to vibrate due to a downward movement of the larynx through the column of subglottal air. However, it is also clear that distinctive timing sequences prevail: the voicing follows a glottal closure which in itself may be necessary to induce rarefaction. Figure 2 presents the same intervocalic sequence, this time embedded in the word abafana "boys", produced at a relatively faster speaking rate. Although the articulation is also clearly implosive in nature, both auditorily and articulatorily, there is a marked difference in the acoustic spectrum during the closure phase of the implosive. Voicing continues throughout the closure phase suggesting no specific timing with respect to glottal closure. The extent of this phonetic variation is difficult to determine. Free variation may

Figure 1 Implosive bilabial in /aba/

Figure 2 Implosive bilabial in /abafana/
take place within individuals irrespective of the tempo. In other words, whatever the tempo, the seven instances were also recorded and have lost their invarious qualities in fast speech to be articulated as voiced labial plosives as are usually found in nasal combinations. This phenomenon, however, may possibly be explained in terms of unattuned articulatory targets.

Ladefoged's well known assertion in the phonetic qualities of implosives in Xhosa brings to mind similar types of differences described by Ladefoged for implosives in Hausa and Kalabari, two languages spoken in Nigeria. In Hausa, implosives display laryngealized voicing throughout the closure, whilst implosives are fully voiced during the closure with no tendency toward a creaky voice or laryngealization in Kalabari. Apart from sharing the voiceless implosive with Kalabari, Xhosa augments the possible implosive articulatory repertoire with a distinctive closure of the glottis followed by an amount of prevocalic prior to the release. These observations have clearly shown, as have Ladefoged [1] and Muysken [2], that implosives are relatively rare, representing only 2.8% of the total corpus of 294 965 combinations. Hence, typification of some languages (Xhosa and Zulu) as unique in this respect, is obviously not based on quantitative values. Of these clicks, the plain version (4.15%) and the nasalized versions (0.62%) constitute the bulk of these occurrences. With Xhosa entertaining a five vowel system, the most frequent combinations occurring in decreasing order are [e] and [i] for short vowels and [a] and [u] for mid-low vowels.

DENTAL ALV-LAT ALV-PAL

PLAIN

| /a/ | [a] |
| /e/ | [e] |
| /i/ | [i] |

NASALIZED

| /i/ | [i] |
| /e/ | [e] |
| /a/ | [a] |

(The 'missing' two vowels in each case have been omitted, due to their extremely low occurrence rate, i.e. a rate of less than 0.003%). In some instances, these vowels do not even occur whatsoever [3].)

In all cases above /a/ seems to have default status occurring irrespective of the other vowel category. Dentals, both plain and nasalized, show a clear preference for front vowels, which implies that styloglossus (and possibly palatoglossus) activity which is responsible for tongue-velar closure (to induce rarefaction), is overridden by activity of the geniolarynx muscle maintaining the position of the tongue in the anterior region of the oral cavity. Plain alveo-lateral articulations seem to favor mid-low vowels. If it is taken into account that the alveolar closure is maintained during lateral release where the side of the tongue is lowered to a position towards the mid-level of the oral cavity, then it may be expected that mid vowels will tend to follow. In plain alveo-lateral articulations the preference seems to be for back vowels to follow these clicks. In these articulations the tip of the tongue is very active performing a passive retroflex movement in which the posterior larynx muscle as well as the geniohyoid are the most probably involved. Considering the back and downward movement of the dorsal part of the tongue to induce rarefaction, as well as the retroflex movement of the tip of the tongue, it seems quite plausible that the following articulation could also be in the posterior area of the oral cavity hence the preference for back vowels.

In nasalized alveo-lateral articulations the preference for mid vowels seems to give way to front vowels. This preference is also shared by nasalized dentals. It is possible to assume that in both cases the pull of the tongue towards the dental and alveolar regions for the purpose of closing overrides any pull or lift towards the posterior area. Activity of the palatoglossus muscle which lowers the soft palate as well as a lack of levator palatini activity may concomitantly contribute to a lack of activity in the back of the oral cavity, giving rise to a preference for front vowels.

Finally, nasalized alveo-lateral clicks frequently combine with high front and back vowels and are borne in mind that the articulation of the alveo-palatal click requires a high rising, and even bunching of the body of the tongue towards the palatal and velar regions. These two occlusions are relatively close to one another, compared to dental clicks where the points of occlusion are relatively far apart. Add to this the active lowering of the velum, hence a narrowing of the area between the roof of the oral cavity and the body of the tongue, then it comes as no surprise that high vowels are preferred in a position following a nasalized click.

The patterns described above generally seem to be true for voiced as well as for nasalized breathy click articulations, albeit the incidence of occurrence is extremely limited. These examples are found with aspirated clicks where dentals and alveo-palatais seem to interchange. Aspirated dentals seem to prefer a following back vowel, whilst aspirated alveo-palatais tend to have a preference for front vowels in the following position. No obvious phonetic explanation seems to be at hand for this phenomenon.

REFERENCES

PHYSIOLOGICAL PROPERTIES OF "BREATHY" PHONATION
IN A CHINESE DIALECT

A FIBEROPTIC AND ELECTROMYOGRAPHIC STUDY ON SUZHOU DIALECT-

Faculty of Humanities, Shizuoka University, Japan*
RIHP., Faculty of Medicine, University of Tokyo, Japan**
Tokyo Metropolitan Neurological Hospital, Japan***

ABSTRACT

Physiological features of the difference in phonation types were investigated on Suzhou Chinese by use of fiberoptic endoscopy and electromyography. The findings suggest that "normal" vs. "breathy" opposition in phonation type in Suzhou should be brought about by antagonistic interaction in the larynx.

1. INTRODUCTION

It is known that there is an interesting interaction between initial consonants, vowels and tones in Wu dialects in China. Recent phonologists often mention the "phonation type" in treating this phenomenon, but the physiological reality of it is still unclear. In Wu dialects, a "breathy" or "warmed" syllable is initiated by so-called "breathy" initials (usually indicated by a phonetic letter [h], like in [ph], [kph], [kh], [h], [bh], [h]), and it is followed by a breathy vowel with low pitch initiation; whereas a normal (or clear) vowel is initiated by a "clear" initial (i.e. voiceless unaspirates and aspirates), and it is followed by a clear vowel with high pitch initiation.

Experiments were conducted on Suzhou Chinese. In one of the main dialects in Wu area, to reveal the physiological aspects of the difference between "breathy" and "normal" phonation. Seven lexical tones are discriminated in Suzhou as described below by tone letters. Among seven tones T1~T7 and V1~V4, are characterized by shorter duration in their syllables than in other syllables and indicated one numeral[2]. Two numbers with an underline[2][1].

3. RESULTS

3-1 Acoustic evidence

Spectrographic observations show that the breathy phonation is characterized in vowels by friction components at the higher frequency range with the damping of the upper formants. Closure duration and VOT for [t][tsh] were measured and the results of the measurements are shown below. VOT is identified here as interval from the release point to the onset of the periodic vocal wave since it was often the case

with the muddy stops that the exact value of VOT was hard to detect by oscillographic and spectrographic inspection.

CLOSURE DURATION (MS.)

<table>
<thead>
<tr>
<th>No.</th>
<th>Avg. Min. Max. Std. No.</th>
<th>VOT (MS.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] 165.8 131.0-225.2 20.9 30</td>
<td>0.21 0.05 0.45 0.1 2.2 0.3</td>
<td>[t] 162.2 99.9-171.9 23.3 21</td>
</tr>
<tr>
<td>[IV] 167.2 143.0-225.2 20.9 30</td>
<td>0.21 0.05 0.45 0.1 2.2 0.3</td>
<td>[tsh] 162.2 99.9-171.9 23.3 21</td>
</tr>
</tbody>
</table>

The result indicated that the breathy phonation is significantly shorter in closure duration than the normal phonation.

3-2 Intralaryngeal air pressure (Po)

The Po in normal phonation is lower in [t] and [tsh] than in [t]. The difference is significant at the level of Po in carrier sentence A: [t] (P<0.05 in T028 vs. T031)

4. Discussion of the larynx

The glottal opening in the normal phonation is narrow and closed, and the glottal opening in the breathy phonation is wide and open.

3-3 Electromyographic findings

Fig.1 shows the average EMG signal for both the normal and breathy phonation.


Fig.1 Selected frames of the laryngeal views for Suzhou [tis] (normal) and [tis] (breathy) uttered in isolation. In both syllables laryngeal views are selected from beginning part of the vowels; (A) approximately corresponds to the point of oral release, and (B) to 40-70 ms. after release.

In breathy phonation, the constricted gesture, which can be called "aryepiglottic constriction" [1], is observed throughout the entire syllable with an increasing degree, but is weakened at the end of the syllable. Note the adductive movement of the false vocal folds, which is a characteristic feature in the "glottal stop" [1]. The glottis as well as other related features is summarized below.

1) Vocal initiation ([t] [tis])

The glottis appears to be closed both in normal and breathy phonation at the initiation of the syllables. In the normal phonation the glottis closure is characterized by the adduction of the false vocal folds, whereas in breathy phonation the glottis is definitely open.

2) Consonantal initiation ([t] [tis])

No remarkable difference is observed between the normal and breathy phonation between the two consonant types. There are four states of the glottis ([i] [i]) both complete and membranous [i] both in open and closed states. No state is observed in normal and breathy phonation for the consonant levels.

3) Electromyographic findings

Fig.2 shows the average EMG signal for [is] [is] and [is] VOC: Is normal phonation; VOM: Is breathy phonation at the vocal initiation. In breathy phonation the EMG shows a marked decrease in intensity, whereas in normal phonation the EMG is intensified.
suppressed at the initiation. This evidence, in general, is consistent with the fiberoptic observation. In the syllables with high or rising pitch, VOC activity increases at the syllable final portion, showing a reciprocal pattern with CT. This is related with the pitch control of the tonal as well as the vocal termination of the syllable[1].

SH, CT. The activity of SH and CT is basically ergometric at the beginning of the syllables; SH is activated for breathy phonation and CT for normal phonation. Both SH and CT show the early initiation in their activities; in the syllables with dental stops, they start their activities at around the closure point or even earlier. In other words, high and low pitch initiations are preceded by early activation of CT and SH.

In pitch rising and falling, however, CT and SH are not antagonistic. CT evidently participates in pitch raising (see Fig. 2B) but SH does not show any marked activity in pitch falling (see EMG for Fig. 2B).

Evidence observed in bisyllabic words

The vocal folds have been reported to be realized in fully voiced manner in connected speech. In the experiment a set of bisyllabic words which have normal or muddy initials in the second syllable was also examined. VOC in muddy stops in these cases is most negative and the glottis (both cartilaginous and membranous parts) is closed. The closure duration and F0 in muddy stops are significantly shorter/lower than in normal stops.

4. Discussion

It is suggested that the difference in phonation types should be produced by the antagonistic excitation in the larynx.

breathy phonation is characterized by "ary-epiglottic constriction" with the downward movement of the larynx. The activity of SH, and presumably other extrinsic muscles as well, undoubtedly contribute to form the constriction and the downward shift of the larynx. These muscles adjust the free space of the larynx as a whole, then externally or vertically effecting the tension of the vocal folds[4][6]. Note that VOC and CT are suppressed at the initiation of the breathy phonation. Concretely increased activity of the extrinsic muscles would shorten and thicken the vocal folds by exerting the forces externally on them. Its additive effect being decreasing at the "breathy" quality of the syllable might be brought about by a "slack" state of the vocal folds which would provide a favorable condition for low pitch initiation. And this may also be a reason the vocal folds start vibrating in the intervocalic positions.

The normal phonation, on the other hand, is initiated by the increased activity of VOC in the vocal initiation and that of CT in the consonantal initiation, the former of which is often accompanied by the adductive gesture of the false vocal folds. VOC contributes to increase the adductive tension of the vocal folds by supplying the medial compression. CT is primarily a pitch raiser, but note that its activity is initiated quite early. It is assumed that CT also participates in increasing the additive tension of the vocal folds[6]. Thus a high "pitch" state of the vocal folds in normal phonation is unlikely to cause the vibration and would provide a favorable condition for high pitch initiation.

REFERENCES
INTRA SPEAKER VARIATION ON THE SEGMENTAL LEVEL: 
A TRANSCRIPTION-BASED APPROACH

A.P.A. Broeders* & W.H. Vierregg**

*National Forensic Science Laboratory, Rijswijk, Netherlands
**Dept. of Language & Speech, University of Nijmegen, Netherlands

ABSTRACT
This paper discusses a set of procedures which may be used to examine intraspeaker variation on the segmental level. The primary tool employed for this purpose is the consensus transcription. A variation index is proposed which captures the amount of intraspeaker variation around the modal realization of each variable. The procedures described should provide a principled approach to the investigation of intraspeaker variation with special relevance to the subject of speaker identification.

1. INTRODUCTION
While it is generally recognized that intraspeaker variation poses a major problem in speaker identification, comparatively little is known about the way in which this type of variation manifests itself in the speech of individual speakers. It is not clear, for example, whether speakers differ consistently in the amount and nature of intraspeaker variation associated with their speech. In recent years, there has been a marked increase in the number of studies dealing with inter- and intraspeaker variation, many of the studies undertaken with the prime object of answering questions in the field of speech technology. In spite of the current interest in speaker characteristics, there is still a remarkable scarcity of data at even the most basic level about the actual extent of variability in the speech of individual speakers. The present study seeks to develop a systematic approach to this question. However, unlike many other studies in this field, ours is not inspired by issues arising from speech technology and may therefore be of only marginal interest to it. We are aiming to devise an approach which is primarily relevant to auditory speaker identification. The primary tool employed for this purpose is the consensus transcription. Presented below are the preliminary results of this approach.

2. PURPOSE OF THE STUDY
Our main objective is to gain a better understanding of the magnitude and nature of intraspeaker variation through the use of a consensus transcription. Some of the questions we would like to answer are: Do some speakers consistently exhibit more variation than others? Is it possible to express speaker variation in quantitative terms, and if so, how much material is required to arrive at a reliable index of intraspeaker variation? Are some variables more consistent than others? Is variation constant against time? In order to investigate these questions non-contemporary speech samples were collected from 6 speakers of Dutch and subsequently transcribed according to the principles outlined below.

3. CONSENSUS TRANSCRIPTION
The concept of the consensus transcription is not new. Shriberg et al. [2] recommend it as a procedure which can be used to eliminate errors due to inter- and intraspeaker variation, mistakes of the raters, and other shortcomings of the transcriber. They found that, of the corrections made by transcribers in a consensus transcription, 90% of those in vowel segments and 80% of those in consonant segments were considered by the transcribers to be due to inattention on their part during the original transcription process. Also, Ting et al. [3] have shown that within a group of transcribers manual corrections lead to greater agreement between transcribers. In the present instance, all speech samples were first transcribed by pairs of Language & Speech Pathology students of the University of Nijmegen - all of them qualified speech therapists - as part

4. THE VARIATION INDEX
The same statistics were determined for the various contexts of the variables 1, 2, and 6. They are omitted here for reasons of space. The descriptive statistics presented above give a first indication of the various degrees of intraspeaker variability encountered in the material produced by the six speakers. They will make it possible to examine any changes in the realization of the variables with time. More specifically, we will be able to determine whether the modal realization changes or remains constant in both qualitative and quantitative terms. What is less satisfactory about the format used so far is the amount of information it contains about the non-modal realizations. It tells us how many realizations there are in addition to the mode and what their combined relative frequency is but it would be more interesting to know whether they are very similar to the mode in qualitative terms or very different. In other words, we would like to be able to develop a variation index which can capture the degree of similarity between the modal and non-modal realizations.

The solution proposed here is one based on the use of a distance matrix as developed by Vierregg & Cucchiarini [4]. A weighted variation index $V_w$ can be calculated by means of the following formula:

$$V_w = \frac{\sum_{i=1}^{n} w_i (d_{ij} - d_{i,j^*})}{\sum_{i=1}^{n} w_i}$$

Here, $R_i$ stands for the relative frequency of the various non-modal realizations and $d_i$ for the articulatory distance between a realization $R_i$ and the Mode, calculated on the basis of the number of articulatory features in terms of which the two realizations differ. The value of the index is arrived at by summing the products of the relative frequency of each non-modal realization and its distance measure. It will be clear that the weighted variation index $V_w$ represents a measure of the articulatory variation around the mode which is superior to the gross variation index obtained by summing the relative frequency of the non-modal realizations because it takes account of the articulatory difference between the mode and the non-modal realizations.

7. CALCULATION OF THE VARIATION INDEX
We will illustrate the calculation of the variation index for one of our variables, $s_{37}$. Between them, it 65 speakers used 10 different realizations of this variable. The following matrix was used to calculate the differences:

<table>
<thead>
<tr>
<th>Speaker</th>
<th>s_{37}</th>
<th>s_{37}</th>
<th>s_{37}</th>
<th>s_{37}</th>
<th>s_{37}</th>
</tr>
</thead>
<tbody>
<tr>
<td>MJ</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>JK</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>NR</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>WS</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>&amp;</th>
<th>1.5</th>
<th>1.5</th>
<th>1.5</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The resulting variation index $V_w$ for this variable is calculated as:

$$V_w = \frac{\sum_{i=1}^{n} w_i (d_{ij} - d_{i,j^*})}{\sum_{i=1}^{n} w_i}$$

where $w_i$ is the relative frequency of the $i$th non-modal realization and $d_{ij}$ is the articulatory distance between the $ij$th pair of realizations.
of the final project of a 120-hour course in phonetic transcription taught by the second author. They were instructed to produce a consensus transcription in accordance with the IPA conventions [1] which, they were told, would later be assessed by their teacher. The final version of the consensus transcription, which forms the basis of the present study, was produced by the two authors. After several tuning sessions, during which a number of minor notational problems were ironed out and maximum uniformity in transcriptional practice was achieved, the authors worked through the student-made transcriptions on an individual basis. However, apparent inconsistencies in the author versions were carefully re-examined to produce the ultimate consensus transcription used for this study.

4. COLLECTION OF MATERIALS
The speech samples were produced by 6 educated speakers of standard Dutch, all employed by the University of Nijmegen and living in the Nijmegen area, though originally hailing from various parts of the country. The amount of regional accent in their speech varied from mild to reasonably strong. There were three women and three men, their ages ranging from 25 to 50. The six speakers read three texts on each of the three days, with a week interval. On each day, the three texts were read three times in succession at three points in time, i.e. at 9am, 1pm and 5pm, giving a total of 9 readings per speaker per day, and a grand total of 27 readings for each speaker for the three days. Although the texts were different, they were identical in terms of the variables under investigation, so that in effect 27 tokens of each instance of all variables are available for analysis. However, the preliminary results presented below are based on a subset of 6 non-contemporary readings from the total of 27 readings.

5. VARIABLES INVESTIGATED
Nine segmental variables were investigated. They were selected on the basis of their expected variability in Dutch. They are (n = the number of tokens per reading):

1. <r>, in four contexts, viz.: r1: C - r2: (C #) n = 3 r3: (C C) n = 6 r4: V - n = 4
2. <v>, in two contexts, viz.: x1: r V n = 3 x2: # V n = 3
3. <v>, in six contexts: x1: # x2: T x3: R x4: V - n = 5
4. <v>, in six contexts: x1: # x2: T x3: R x4: V - n = 5
5. <v>, in two contexts, viz.: x1: 1 # x2: 2 n = 3
6. Elision of /n/ after schwa n = 4
7. Assimilation of voice before /n/ and /d/ n = 3
8. Elision of /n/ after schwa n = 4
9. <v>, as in Dutch politie (English police) n = 3

6. DESCRIPTIVE STATISTICS
In order to arrive at a first overall measure of the degree of intraspeaker variation, the following statistics were determined per variable and per speaker over the six non-contemporary readings:

1. the Mode (M), i.e. the most common realization of the variable;
2. the relative frequency of the mode, (FM);
3. the number of realizations other than the mode, (p), any number of hapax legomena (i.e. unique realizations) being counted as 1.

It appears that the weighted variation index VW can deviate quite considerably from the raw variation index, especially if the mode has a low frequency of occurrence, as in the case of speakers NO and WS. While the relative frequency of the mode is the same for these speakers, NO’s weighted variation index is considerably lower, which reflects the greater similarity to the mode of NO’s non-modal realizations.

8. CONCLUSION
As observed in the introduction, the results presented above are based on a small portion of the available data. The emphasis here has been on some of the procedures used to describe intraspeaker variation in a systematic fashion. The consensus transcription is proposed as the most suitable format for the initial analysis of the speech samples collected. The use of a distance matrix based on articulatory differences between realizations affords a principled approach to a further, quantitative analysis of the variation encountered in the material. Major problems remain to be resolved before a meaningful comparison is possible of the readings produced at different times. It is this comparison which should provide answers to the central question of the consistency of intraspeaker variation patterns.

REFERENCES
PHONÉTISATION AUTOMATIQUE : ÉVALUATION ET ENRICHISSEMENT D'UNE GRAMMAIRE DE RÈGLES POUR LE FRANÇAIS

R. Belrhali, L. Libert, L.J. Boe

Institut de la Communication Parlée, URA CNRS no 368
Grenoble, France

ABSTRACT

Our project was the establishment of a grammar for the automatic phonetization of French. By constituting a lexicon of 60,000 words and systematically examining their transcriptions, we formulated a large body of new rules, which were added to a pre-existing base set, making a total of 900 rules. The resulting system gives a correct phonetization of 99.75% of the base lexicon. We here present the analysis method used on this large lexicon, as well as a selection of the rules derived.

1. INTRODUCTION

La phonétisation du français peut être décrite essentiellement par règles de correspondance entre graphèmes et phonèmes. Cette correspondance se décrit par la contrainte de contexte sur la chaîne graphémique et va même jusqu'à la correspondance lexicale. Les phonèmes de phonétisation tiennent compte de niveaux linguistiques supérieurs au mot que sont : la valeur catégorielle (souvent [suv]), chantent [ʃəl]), la fonction syntaxique (les portions [posi]), nous portions [posi]), la structure syntaxique : les liaisons, (un savant [ə] aveugle (nom - adjectif), un savant [ə] aveugle (adjectif -nom)), la valeur sémantique (fils [fil]), fils [fii]). Ce travail concerne le niveau lexical de la phonétisation du français. Avec la mise en place de bases de données faciles d'accès, de manipulation et de reconfiguration, il est maintenant possible d'élaborer, tester et améliorer les formalisations possibles. Notre effort s'est concentré sur les relations entre codes orthographiques et de vastes corpus de transcriptions phonétiques.

2. PRÉSENTATION DU LOGICIEL TOPH

L'outil TOPH (Transcription Orthographique-Phonétique) est un phonétiseur multilingue qui propose une syntaxe pour décrire des grammaires de phonétisation. Ce transducteur fonctionne sur texte libre. Il permet de récrire une chaîne d'entrée graphémique en une chaîne de sortie phonétique. Les avantages de TOPIH par rapport à d'autres logiciels (cf. [3], [4], [6], [8]) sont certains. Nous pouvons mentionner sa facilité d'utilisation (traces d'application, statistiques) ainsi que la formalisation de ses règles transparente à l'utilisateur, permettant de modifications aisées. L'expert formalise son raisonnement sous la forme d'une grammaire déterministe de règles de réécriture contextuelles.

A chaque classe de règles il introduit un ordre local défini par l'ordre d'écriture des règles. La grammaire comprend :

1°- des ensembles prédéfinis de caractères orthographiques décrivant des phonèmes de nature très différente :
   - ensembles linguistiques : "Consonnes non nasales" = {b, c, ð, t, f, g, h, j, k, l, m, n, ð, s, t, v, w, x, z}
   - ensembles d'exceptions : "Exception : fin en "g" = {barlong, basting, basting, bourg, oing, seing, dugong, éng, étang, hareng, harfang, jonc, laoïning, long, pacfung, païfing, rang, sampeng, sandering, sang, shampooing, shampooing, trépang, tri pang}
   - ensembles de transcriptions.

2°- des commentaires pouvant être une chaîne quelconque bordée par '1' et insérée dans n'importe quelle portion de la grammaire.

3°- des règles partitionnées en classes ; la classe d'une règle étant déterminée par le premier caractère de la chaîne à transcrire.

3. MÉTHODOLOGIE

Afin d'enrichir la grammaire de phonétisation existante, un lexique de grande taille est nécessaire. Nous avons donc, dans un premier temps, constitué une base de données de 60 000 mots implantée sur Macintosh. L'environnement Hypercard et le langage Hypertalk ont rendu possible la mise au point de programmes de recherche de chaînes orthographiques de longueur quelconque. Elles ont été recherchées dans trois positions : initiale, interne, finale. A partir des listes obtenues nous avons systématiquement relevé la transcription phonétique de la chaîne étudiée en prenant comme référence de prononciation le Petit Robert 1. Nous avons ensuite vérifié l'existence de la (ou des) règle(s) correspondante(s) à la (ou aux) transcription(s) phonétique(s) de la chaîne de caractères étudiée. Dans le cas contraire, nous avons écrit de nouvelles règles.

Illustration de la méthode de travail par un exemple : la classe du 'b'. Nous avons obtenu 2394 mots commençant par 'b-', 7210 mots contenant au moins un 'b-' en position interne et 32 mots se terminant par 'b'. Après le relevé de la prononciation du graphème 'b' dans tous les mots et dans toutes les positions nous avons établi les règles suivantes :

(les caractères syntaxiques sont notés en gras)

- ([radou, lon]) + b + ('"", s) = []
   Cette règle concerne radou et les mots se terminant par 'lom' comme plomb, coulomb, surplomb, aplomb, dont la réalisation du 'b' en position finale est muette (ces mots peuvent être suivis d'un 's', marque du pluriel).
   - + b + (s, i) = [p]
   Cette règle concerne tous les mots contenant la suite de caractères 'bs' ou 'bt' et dont le 'bt' se réalise [p] ; il s'agit ici d'un cas d'assimilation régressive.

   - ([sub]) + b + (+ ("", s, i)] = [b]
   - ([lan]) + b + (swool) = [b]
   Ces deux dernières règles sont des exceptions à la précédente.

   - + b + (c, k) = [p]
   Cette règle concerne tous les mots contenant la suite de caractères 'bc' ou 'bk' dont le 'bk' se réalise [p]. Nous avons ici un autre cas d'assimilation régressive. La seule exception à cette règle est la suivante :

   - ([su]) + b + (carpathique) = [b]
   - + b b + = [b]

Toutes les géméniuses de la classe du 'b' obéissent à une règle unique.
Cette règle est uniquement applicable à la lettre de l'alphabet.
- \( b + = \{ b \} \)
Il s'agit de la règle la plus générale.

Classement suivant l'ordre d'application des règles :
(radou, lom) \( + b + \) ("#", s) = [ ]
(lam) \( + b + \) (swool) = [b]
(sub) \( + b + \) (sidence, sidiaire, sist) = [b]
\( + b + \) (s, t) = [p]
(su) \( + b + \) (carpatique) = [b]
\( + b + \) (c, k) = [p]
\( + b + \) = [b]
("#") \( + b + \) ("#") = [be]
\( + b + \) = [b]

4. RÉSULTATS
La grammaire de base [1] contenait 200 règles et 12 ensembles d'exceptions. Actuellement 900 règles et 16 ensembles d'exceptions (décrivant 1 000 mots) permettent de phonétiser automatiquement les 60 000 mots de notre base de données avec un taux de réussite de 99,75% (problème de polyphonie des mots du type 'plus' pouvant se prononcer [plys] ou [ply]). La langue, matériau vivant, est en constante évolution d'où la nécessité d'une réalisation systématique de notre base de données et de la grammaire.

5. CONCLUSION
Au-delà des applications évidentes en synthèse et reconnaissance de la parole, le passage du niveau orthographique au niveau phonétique renvoie à des problèmes linguistiques fondamentaux et constitue un champ de validation privilégié des formalisations linguistiques et phonétiques. Le développement des Industries de la Langue constitue à la fois une stimulation et une possibilité directe d'application.

6. RÉFÉRENCES
Contribution à la phonétisation automatique des langues alphabétiques : le langage "TOPH". Rapport de DEA, CRISS, Département d'Informatique et Mathématiques appliquées aux Sciences Sociales, Université des Sciences Sociales de Grenoble.
Le compilateur de règles de réécriture TOP et son utilisation à la transcription du français en vue de la synthèse. 10èmes J.E.P.-G.A.L.F., Grenoble, 202-211.
Programme de transcription orthographique-phonémique en langue française. ENST. Paris.
Contribution à la synthèse à partir du texte ; transcription graphème-phonème en temps réel sur microprocesseur. Thèse de 3ème Cycle. Université Paris -Sud-Centre d'Orsay.
M.A.V.L, DE L'UTILISATION DE LA NOTION DE MOMENT D'APPARITION DES VIBRATIONS LARYNGIENNES POUR LA DESCRIPTION PHONETIQUE

JEAN-PIERRE GODAILLER

Laboratoire de phonétique, U.P.R. de Linguistique, Université René Descartes, Paris, France

ABSTRACT

Thanks to examples in French, this aim of this paper is to show that from a typological point of view the concept of M.A.V.L. is more operative than V.O.T. in what concerns stop consonants.

Depuis plus de 25 ans déjà, la notion de V.O.T. (Voice Onset Time) telle qu'elle a été proposée dans un premier temps par LISKER et ABRAMS (9) et ultérieurement par KATT (8) est utilisée tant d'un point de vue phonétique que phonologique (cf., entre autres, l'emploi qu'en a fait dès 1968 dans S.P.E.), voire même à des fins typologiques (7). Au-delà de son utilité même (1), ce concept trouve cependant ses limites, lorsqu'il s'agit de décrire l'ensemble des réalisations possibles pour les articulations de type occlusif. La distinction entre V.O.T (positif) et V.O.T.- (négatif), même si elle peut être "ajoutée" en short voicing lead et long voicing lead d'autre part, ne peut pas, à mon sentiment, rendre compte de plusieurs cas de figure que l'on peut rencontrer, ne se rattachant que pour les occlusives /p/, /t/, /k/ et /s/, /t/ et /k/ en français, langue qui servira ici d'exemple. The description complète de ces consonnes nécessite de tenir compte des cas de dévoiement partial et/ou total pouvant affecter les occlusives phonologiquement 'sonores', ceci tout aussi bien au niveau phonétique que phonologique (2)(7). A cet effet, non seulement le concept de V.O.T. doit être utilisé mais aussi celui de M.A.V.L. (Moment d'Apparition des Vibrations Laryngiennes) (3). Dans une perspective typologique, le concept de M.A.V.L. se révèle être plus performant que celui de V.O.T., ce qui montre les exemples présentés ici-même.

Comment évaluer les moments d'apparition des vibrations laryngiennes d'une consonne occlusive ? Une étude est fournie par les planschette 1 et 2 qui comportent les tracés du phonogramme (lignes 1 et 6) et de l'électroglottogramme (lignes 2, 3, 4 et 5) successivement présentées par des enfants francophones. Le [b] de [bâbâ] (Figure 1 ; Planche 1) est entièrement voisé ; ce qui vaut dire que sa phase d'occlusion et celle de relâchement sont toutes deux accompagnées de vibrations des cordes vocales. Dans un cas de figure, la phase de relâchement, qui est constituée d'une explosion très brève non accompagnée d'un V.O.T. positif, est difficilement repérable sur ce type de tracé : l'explosion se confond alors avec une vibration des cordes vocales, ce qui n'est pas le cas de l'électroglottogramme. Toutefois, le passage entre la consonne et la voyelle est visible, étant donné l'augmentation d'amplitude notée au début de la voyelle). La durée de l'ensemble est de 65 ms et au-cune interruption des vibrations laryngiennes n'a lieu lors du passage de la consonne [b] à la voyelle [a]. C'est un type 1 de M.A.V.L. Le [s] de [spipip] (Figure 3, Planche 1) et la consonne dentale [tš] de [tšitš] (Figure 3, Planche 1) montrent un comportement semblable. Donc la durée de la phase de relâchement est bien comprise dans le tracé entre 40 ms et 60 ms.


Les différents types de M.A.V.L. peuvent par ailleurs être schématisés ainsi que l'est indiqué à la Planche 3. Si l'on veut analyser les phénomènes d'assimilation notés dans certaines variétés du français (franco-canadiennes plus particulièremen ; certaines régions du Nouveau-Brunswick, etc.), il convient d'in-cuire à cette schématisation des types supplémentaires [3][6].

Dans une perspective typologique, sur quatre, l'assimilation en termes de M.A.V.L. est-elle plus performante que celle basée sur le V.O.T. ? Si l'on considère les types 2 et 6, il peut aisément constater que ceux-ci ne peuvent donner lieu à aucune mesure de V.O.T. (pour le type 2) ou de V.O.T. positif à la fois compté d'un V.O.T. positif de 15ms et d'un autre, quant à lui négatif, de 45ms. Or, l'idée même de V.O.T. est impossible sans avoir des unités phonétiques comportant à la fois un V.O.T. négatif et un V.O.T. positif. Il n'existe pourtant pas de traitement de telle unité, si ce n'est en utilisant le concept de M.A.V.L. 3, d'autant plus que la phase sans vibrations laryngiennes de 15ms de durée survit entre celle-ci et celle-ci conviendrait d'attribuer à un V.O.T. négatif de 45ms s'arrêtant à 15ms de la fin de l'occlusion d'une part et d'autre part, la figure est de même pour le cas de M.A.V.L. 9, ceci étant accompagné d'un V.O.T. positif de 15ms. Il est donc impossible d'analyser la vibration du V.O.T. dans le cas du [spipip]...
et le $[g^{(h)}]$ (Figures 2 et 3) ; le V.O.T. ne peut être compris que de manière bininaire : négatif/positif. C'est ici même que le rend inopératoire pour l'analyse d'occlusives partiellement désonorisées, que ce soit pendant leur phase d'occlusion ou pendant leurs phases d'occlusion et de relâchement. Sans le concept de M.A.V.L., qui permet d'affiner la description phonétique des occlusives en individualisant, pour le français, 4 types différents, au lieu de 2, il serait impossible de rendre compte de faits d'acquisition [5] ou de différenciations sociolinguistiques, plus particulièrement d'ordre sexuel [4].

REFERENCES

Types MAVL

Planche 3
THREE TYPES OF PROSODIC CORRELATIONS IN SOUTH GERMAN DIALECTS

L. NAIDICH

INSTITUTE OF LINGUISTICS, LENINGRAD, USSR

ABSTRACT

The paper deals with three types of prosodic correlations in South German dialects: correlation of gemination in High Alemannic, opposition of syllables with different place of quantitatively syllabic peak in North and Central Bavarian, correlation of syllabic cut in Low Alemannic and High Franconian. This prosodic systems are considered as different stages of the evolution of Germanic quantity.

There are three major types of prosodic systems in South German dialects. High (South) Alemannic dialects - a group which most of Swiss German dialects belong to - are known to preserve a rather archaic quantitative order and syllabic structure. In the consonant systems of these dialects two phoneme types are opposed to each other: the weak consonants - lenis - and the strong - fortis -, this opposition concerning all the consonants as copula fricatives. Geminates, double consonants typical of Swiss German dialects, occurring in the intervocalic position or sometimes between resonant and vowel, are phonologically identified with fortis. This can be demonstrated by means of the rule formulated by L.Zinder [9]: sounds standing in complimentary distribution and alternating in the same morpheme constitute the same phoneme, e.g. in our case "bit Holzschelit" - "sittsce Holz spaiten". The occurrence of long vowels before fortis (long) consonants or consonant clusters ([səf] 'Scafe', [sɪt] 'Leute', [dərf] 'Dorf') seems to testify "free quantity". This syllable shape (overlong syllable) is especially frequent in some South Swiss German dialects with "Auslautverhärtung", i.e. the neutralization of fortis- lenis opposition in favour of fortis at the word end. As for another "usual" (in terms of modern West Germanic languages) syllable type in monosyllabics - short vowel + lenis (short) consonant-short vowel-fortis (long) consonant (Pfaelz's law), these two syllable types often alternating in morphological paradigms, e.g. the opposition of singular and plural forms of substantives (grift, -griff, Griff, -fis - fis, Fisch, -bob, -debok). This correlation poses some peculiarities differing it from similar prosodic quantitative oppositions in other Germanic languages. The alternation of fortis and lenis depending on the length of preceding syllable takes place in consonant clusters as well ([gus], -gesst) 'Gast', -Gaste' long as opposed to short can be diphthongs and affricates too (Hund', 'Hunde', 'Hundes', -khefp, 'Kopf', -Kopfe', resonants do not take part in this correlation being always weak: [hünd] - [hunts] 'Hund' - 'Hunde'). We consider this correlation as a prosodic one, as an opposition of syllable types differing in the place of syllabic quantitative peak, although the character of this opposition remains disputable [3].

In Low Alemannic and High Franconian a third prosodic order is represented - the syllable cut correlation or the opposition of close and loose contact between a stressed vowel and the following consonant, typical of many West Germanic languages and dialects, e.g. modern standard English, German and Dutch. The prosodic character of this correlation becomes apparent in syllabification depending on the vowel length which in turn is the phonetic sign of contact: [lida], [lida] 'leideren' - 'laften' (Low Alemannic - Alsatian dialect).

Modern dialects reflect different stages of prosodic developments. They show general trends of syllabic structure evolution common to all Germanic languages, but also some specific High German features.

An important consequence of the Second Sound Shift was the elimination of the opposition voiceless-voiced in the consonant system of South German dialects. The reduction of the opposition of three consonant rows: lones-fortes-geminates to that of only two: lones-fortes with geminates as positional variants took place - as many scholars believe -
already in the Old High German [6]. Thus the consonant system here was based on the opposition of lenes derived from Germanic voiced and fortles derived from Germanic voiceless stops which were shifted according to the Second Sound Shift, Germanic geminates and long consonants resulted from the West Germanic consonant lengthening. The shifted fricatives were long, strong and in the intervocalic position geminated. Germanic p-tf merged with Germanic ff, k-xx merged with Germanic x-x, t-çç. Thus the group of stopping fortles was enlarged, the opposition of fortles-lenes became universal for the whole system. Only in the dental row the triple opposition d-t-tt and also s-ss-sç was preserved for a longer time. After the coincidence of ç, çç (çd) with s, ss and t (çd) the consonant system was simplified, the opposition of fortles-lenes intimately connected with syllabication and syllable shape became strengthened. This consonant system is preserved to-day in Swiss German dialects.

Further evolution of syllabic structure was directed toward the interrelation of vowel and consonant quantity inside the word. According to the assumption of E.Kranzmeier a trend to the equal quantity of all the words got developed. As the Swiss German dialects indicate, the first step of this development could be the vowel lengthening in monosyllabics before lenes - first of all open vowels before resonants. This statement contradicts the widespread concept according to which the lengthening in monosyllabics occurred by analogy with that in open syllable. Many new monosyllabic words ending in fortles resulted from the apoche which took place in Central and North Bavarian dialects: [ßфф][ßßфф]. They contrast with ßßßß monosyllabics built according to the pattern, long vowel+lens. As a result of all these processes a quantitative-prosodic order referred to above as the correlation of the syllable peak-place was formed. There are reasons to suppose that a prosodic correlation like this (sometimes called isochrony) always preceded the syllable cut correlation in the history of West Germanic language. [5]. The next step was the elimination of geminates which became phonologically redundant. In Central German dialect area including also some South German dialects - High Franconian, Low Alemannic - these processes were connected with the merge of fortles and lenes - consonant weakening[7]. Because of very few oppositions in the consonant system of these dialects the syllable cut (contact) correlation became an important means of differentiation: liuten > liiten (with delabialization in [i] > [e] and vowel shortening) > modern Alsatian [li tung] with close contact 'I'üt'en' - lidan> modern Alsatian [IIdã] with loose contact 'leiden' [1].

REFERENCES
"Isledovaniya v oblasti eravnitelnõy ak-
tsentologii indoeuropейskõy jazikov", Leningrad: Nauka, 238-250.
MORPHOPHONEMICS OF WEAK-STEM FORMS IN MODERN CHALDEAN

Solomon Sara.SJ.

Georgetown University, Washington, D.C.

ABSTRACT
This paper explores the derivation of the surface forms of the weak-stem verb forms in Modern Chaldean. Weak-stem forms are those forms that include the glides /y,w/ among their triconsonantal radicals. The intercalation of vowels in the verb paradigms include /y,w/ and the lable nature of these glides, produce surface forms that are at variance with the corresponding strong-stem paradigms.

O.O INTRODUCTION
Modern-Chaldean is one of the currently spoken, and much changed, dialects of Classical Syriac/Aramaic. There are many such dialects, but the dialect that is the focus of this presentation is the dialect of /mangest/, a town in the northern part of Iraq.

1.0 DATA AND PROCEDURES.
There is no lexicon for the Modern Chaldean dialect of /mangest/ that will provide a list of all the weak-stem lexical items. I have depended on lists of words found in several grammars or lexica of other dialects as my data sources. The lexical items that were provided in these sources needed to be phonemically modified to the /mangest/ sound patterns. These sources provided a representative sample of lexical items in which /wy/ occur in verbal paradigms.
The following stems illustrate /w/ in initial position:

/wp/ 'duty' /wrq/ 'paper'
/wrd/ 'rose' /wtS/ 'root'

No verbal form beginning with /w/ is available for conjugation.

4.1 /w/ in medial position

The following stems illustrate /w/ in medial position:

/hwy/ 'become' /zwn/ 'buy'
/swr/ 'white' /swt/ 'break'
/tet/ 'do' /gwy/ 'hard'
/swt/ 'jump'

Strong/Weak
/d-r-s/ 'study' /x-w-a/ 'buy'
/drasw/ 'to study' /xwana/ 'to buy'

/1s/darsin/ 'I study'
/znosn/ 'I buy'
/2s/darsst/ 'you study'
/sosnt/ 'you buy'
/3s/darsst/ 'he studies'
/zwnsn/ 'he buys'

/1s/drasw/ 'I studied'
/zwnsn/ 'I bought'
/2s/drasw/ 'you studied'
/swunus/ 'you bought'
/3s/drasw/ 'he studied'
/swunus/ 'he bought'

imp. s./drasw/ 'study' /znosn/ 'buy'
imp. p./drasw/ 'study' /swunus/ 'buy'

R2. y/w-Syllabification.

w--> u/V_C /zwnsn/ --> /zwnsn/

R5. Tense vowel lowering.

u--> o/V_ /zwnsn/ --> /zaosn/

R6. Vowel deletion.

V1V2 --> V2 /zaosn/ --> /zaosn/

5.0 SUMMARY

The /w/ segments change in the sequences of verb paradigms according to specific rules that are determined by the contexts in which these segments occur. There is similarity between the rules for the glides but no identity. There are more changes that are operative with the high front glide /y/ than the changes of the back high glide /w/. The discussion of the changes was limited to the verbal paradigm, and may be extended to other lexical paradigms with comparable expectations. There are other contexts in which these changes take place, and are under investigation.

5.0 SUMMARY OF RULES:

R1. y-Deletion.

y --> 0 /__C

R2. y/w-Syllabification.

w--> u/V_C /ganw/ --> /ganw/

R6. Vowel deletion.

V1V2 --> V2 /ganw/ --> /ganw/

R3. y-Syllabification before non-low vowel.

y --> i /__V

R4. y-Deletion between two non-low vowels.

[y-syll] [-cons] [+syll] [+high] -> [+syll] /__-[-cons] [+tens] [+back] [-roun] 2

R5. Tense vowel lowering.

i--> i/V_ u--> u/V_

REFERENCES


STOP ASSIBILATION IN QUEBEC FRENCH: AN ANALYSIS BY ARTICULATORY SYNTHESIS

H.J. Cedergren, D. Archambault & G. Boulianne

1Université du Québec à Montréal, Canada
2NRS- Télécommunications, Québec, Canada

ABSTRACT

This paper discusses the use of articulatory synthesis as a research tool for determining the relationship between phonetic and phonological structure. Under the assumption that the mapping between phonological and phonetic representations is accomplished by rules of phonetic implementation, we use a computational model to examine alternative accounts of the derivation of surface affricates in Quebec French. The model's behavior is shown to parallel natural speech data.

1. INTRODUCTION

One of the defining characteristics of Quebec French is the surface reflex of coronal stops /t/ and /d/ in the context of a following anterior high vowel /i/ or /y/. Underlying coronal stops are realized as the affricated affricates [ts] and [dz] in these contexts. Forms such as dix "ten" or petit "little" are typically realized as [ts] and [dz]. These forms are usually accounted for by rules of the form

/t/ → [ts] (i.e., /i/)
/d/ → [dz] (i.e., /y/)

Alternative accounts of the mechanism responsible for the derivation of these surface affricates have been proposed. We assume a multidimensional non-linear representation of underlying phonemic and phonetic segments, with fundamental distinguishing properties: (1) phonological features are assumed to be abstract binary classificatory features, i.e., segments may be [+/- labial]; while phonetic features are n-ary [x], i.e., [+labial] phonological segments may correspond to either [projected], [neutral] or [retracted] labial positions, (2) Phonological representation is assumed to be underspecified; that is, only contrastive feature values are marked underlyingly. Redundancy rules add essential feature values. (3) Physiological representation is in the spatial domain; therefore the phonetic-physiology passage defines a relation between discrete phonetic parameters and acoustically important area-function parameters.

Rules written in Delta account for internal properties and module interfacing [6]. The phonological role-set assigns initial feature specifications and defines feature alignment relations. Phonetic rules translate abstract phonological representations into n-ary phonetic features and assign inherent duration to each segment. Rules of the physiological module account for phonetic feature/production relation by specifying corresponding motor-sensory goals and intrasegmental dynamics. The calculation of articulatory trajectories is accomplished by optimization techniques.

3. ACCOUNTING FOR /t/ ASSIBILATION

The system which we have described allows us to explicitly examine the phonology-phonetics interface question as the implementation of physiological gestures derived from a sparsely specified abstract feature representation and test linguistic hypotheses about levels of representation. We illustrate this issue by modelling /t/ assimilation in Quebec French.

The properties of our computational model (which simulates the properties of actual articulatory systems) permit us to address two aspects of the assimilation problem: (1) do the articulators, during the transition gesture form /t/ to /s/, occupy for a critical duration a location which permits assimilation? (2) does delayed glottal adduction following offset of closure result from speaker control of aerodynamic conditions or in a different context.

3.1 Acoustic Data

A corpus of natural speech was gathered from a single speaker. Three types of stimuli were included in the data set: (1) occurrences of /ts/ followed by (a) a vowel which conditions assimilation, (b) vowels which do not; (2) occurrences of /ts/ followed by /s/ in the same vocalic contexts; (3) occurrences of /ts/ not preceded by /ts/.

From these data we defined the duration for the various segments. We also found that while durations for vowels and fricatives vary under stress, durations of plosives remain constant.

3.2 Radiographic Data

X-ray tracings of /t/ when followed by /s/ were analyzed in order to gather data on articulatory gestures [10]. These data suggest that speakers do not aim at a particular target during assimilation, rather that the transition of the tongue is made directly from the /t/ to the /s/. This information on the motor-sensory goals was integrated into the physiological module.

3.3 Experiment
The input strings, such as /atl/ and /atsl/ were submitted to the system. There were no specific rules provided to account for assimilation. The rule set did not include any rule inserting a fricative segment between the stop and the following vowel. Further, the rate of occlusion offset was set to be uniform for all vowels, and glottal aperture duration did not depend on consonant type.

The synthetic response produced an excrescent fricative segment between the stop and the vowel. The spectrograms of synthesized sequences replicate the major characteristics of measured sequences taken from natural speech. Figure 1 shows the spectrogram of the synthetic sequence derived from the /atsl/ input string. Frication noise is found between 3 and 8 kHz.

**Table 1. Acoustic durations**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>prec. /t/</td>
<td>48</td>
</tr>
<tr>
<td>fol. /l/</td>
<td>49</td>
</tr>
<tr>
<td>/atsl/</td>
<td>47</td>
</tr>
<tr>
<td>/atsl/ acc.</td>
<td>48</td>
</tr>
<tr>
<td>/atsl/</td>
<td>48</td>
</tr>
<tr>
<td>/atsl/ acc.</td>
<td>46</td>
</tr>
<tr>
<td>/atsl/</td>
<td>46</td>
</tr>
<tr>
<td>/atsl/</td>
<td>46</td>
</tr>
<tr>
<td>/atsl/</td>
<td>46</td>
</tr>
<tr>
<td>/atsl/</td>
<td>44</td>
</tr>
<tr>
<td>/atsl/</td>
<td>41</td>
</tr>
</tbody>
</table>

**Figure 1 /atsl/**
The measured durations for the vowels and the frication noise correspond well to those that are stipulated in the rules (cf. Table 1). The durations for /l/ are smaller than expected, because they refer only to the period of occlusion, and do not include the transition of the vowel to the onset of closure. Assimilation was produced even when the vowels /l/ and /l/ followed /l/. However, the spectrogram reveals that the initiation of vocal cord vibrations was quicker than after high front vowels, giving frication periods inferior by 6 to 14%.

**4. CONCLUSION**
The explicit modeling of hypotheses on the mechanism of assimilation has allowed us to evaluate their likelihood. We find that it is not necessary to program assimilation as a phonological process: it results from aerodynamic conditions that are satisfied during the course of the transition form /l/ to /l/ as a consequence of voice onset delay.

These results, however, should be interpreted in the light of traditional warnings on the use of simulation techniques. For they depend on the validity of the model and the articulatory synthesizer that are evidently an approximation of reality [2].

Further tests should be carried out with a full range of vowels. However it would appear that the computational model can be of further use to research in linguistics to test theories.

**5. ACKNOWLEDGMENT**
This research was supported by the Fonds de développement académique du réseau de l'Université du Québec.

**6. REFERENCES**


A STUDY ON DISTINCTIVE FEATURES AND FEATURE HIERARCHIES THROUGH "PHONEME ENVIRONMENT CLUSTERING" (PEC)

M. Dantsuji* and S. Sagayama **

*Faculty of Letters, Kansai University, Suita, Osaka, JAPAN
**Dept. of Speech Processing, ATR Interpreting Telephony Research Laboratories, Kyoto, JAPAN

ABSTRACT

The present study concerns the study of distinctive features by means of "Phoneme Environment Clustering" (PEC). The PEC algorithm, originally developed for automatic speech recognition, selects the optimal set of allophones and estimates missing contexts automatically. We have examined approximately 2,000 segments from 216 phonemically balanced words uttered by a male informant of Japanese using PEC. The results show that the feature (sonorant) is separated from others in the earliest stages of the process of the tree structure and coincide with the feature hierarchies proposed in the field of current non-linear phonology.

1. INTRODUCTION

In this paper we would like to describe an attempt to reconsider the hierarchical structure of distinctive features by means of a "Phoneme Environment Clustering" (PEC). Early generative phonologists adopted the distinctive features of the Jakobsonian framework [6]. Later, they revised the distinctive features in many respects in the framework of SPE, distinctive features are mainly described from an articulatory point of view [11], and the scope of information has been maintained among current approaches. This, however, does not mean that acoustic and auditory aspects have lesser importance, but rather that it was difficult to make an exact and precise description of the acoustic characteristics of distinctive features at that time.

With respect to the hierarchy of distinctive features, several kinds of feature hierarchies have been proposed. We would like to introduce another kind of hierarchy based on the acoustic distance. PEC, which was originally developed for automatic speech recognition, is one such experiment and attempts the establishment of the feature hierarchy.

2. THE CONCEPT OF PHONEME ENVIRONMENT CLUSTERING (PEC)

We can consider a number of possible factors, which may affect the sound patterns of a given language, such as a preceding phoneme, a phone before a preceding phoneme, a phone after a succeeding phoneme, and the current phoneme itself, a succeeding phoneme a phone after a succeeding phoneme, speakers, pitch frequency, power, speaking rate, stress position, phone position in the utterance, background noise, emotion, and so forth. The combination of these factors makes an abstract space which is called the environment space E. Each allophone is assumed to be a point e in the space E. On the other hand, each allophone is observed in an acoustic pattern which may be assumed to be a point v in a vector space (V), after some normalization of pattern durations as well.

If we have a set of phonetically labeled acoustic segments, each is allocated in the environment space E as well as a point v in the pattern space V. Denoting a mapping function from the space E to V by \( f : E \rightarrow V \), the acoustic pattern of each allophone \( v \in \{E\} \) varies from sample to sample and has a certain spread in the space V. The spread is measured by a distortion measure, such as an average Euclidean distance from the centroid, and denoted by \( d(v) \).

To find the spread of the phoneme environment space E through the mapping function is to find the optimal set of \( \{\text{subspaces} (\{E\})\} \) to cover all variations of acoustic segments. It is defined as the minimization of the total distortion defined by:

\[
d(v) = \sum_{i=1}^{n} (d_i(v))
\]

where \( d_i(v) \) is the distortion of each allophone in the environment space.

3. EXPERIMENTS ON PEC AND DISTINCTIVE FEATURES

As has been mentioned above, the process of successively splitting subspaces forms a tree structure which is interpreted as a similar grouping of phonemes and the phoneme environment. The concept of PEC can be applied as well to the distinctive features, which are components of phonemes. For example, Fant (1973) stated that "the phonetic value of a distinctive feature can be regarded as a vector in a multidimensional signal space. The variability due to context shall be expressible by rules which define how the feature vector is changed when the conditioning elements are varied" [5]. Therefore, distinctive features may be extracted to some extent using the PEC procedure. We have examined how sets of allophones are divided into allophones in the process of PEC. We will show the result of an experiment carried out under the following condition:

1) Informant and texts: Approximately 2,000 segments out of 216 phonemically balanced words for one male adult.
2) Acoustic parameters: cepsrum, delta-cepsrum, log-power, delta-log-power.
3) Distortion measure: Euclidean distance.
4) Regression window: 90 ms triangular.
5) Window length: 30 ms.
6) Window shift: 10 ms.
7) Sampling frequency: 12kHz.
8) Environment factors: 5. 9) Distance measure: weighted Euclidean distance.

The results of experiments indicate that allophones depending on phonetic environment are extracted at lower nodes. Phonemes as sets of allophones appropriately correspond to upper nodes which bind the lower nodes of allophones. Still upper nodes tie several phonological categories and these bundles correspond to natural classes. Following diagrams represent part of the tree structure which was formed through the process of successive splitting using PEC.
It is observed that a set of segments which hold a feature [+sonorant] in common and a set of segments which hold a [-sonorant] in common are separated at the first step. A segment [-h] is classified as a member of the segment class [-sonorant] in this analysis. In the case of Japanese, the phone [-h] occurs as allophones [g], [w], and [f] in addition to [-h], and this phone is not usually classified as a glide. Therefore, there is no problem in classifying this segment as [-sonorant].

With respect to /w/, this segment is an approximate vowel-vowel in the case of English, and this would be classified as [+sonorant]. In the case of Japanese, however, this segment has a quite a number of allophones and free variant forms. For example, /w/ is often represented as a kind of plosive at word initial positions, and as a flap at word-medial positions. It is assumed that this segment is accordingly classified as [-sonorant] in this instance. Attaching an asterisk (*) to g and d implies a special case. These segments are originally voiced plosives and should be classified as [+sonorant]. At the stage of labeling preconceived phonemes of the phoneme environment clustering, transition portions of formants were not included in vowels but included in voiced plosives. Therefore, some properties of vowels, which should be classified as [+sonorant], are assigned to these segments in this analysis. Furthermore, /w/ and /b/ seldom occur as voiced plosives [g] and [b], respectively, they occur as voiced fricatives [g] and [b] or velar nasal [g] called "bidakkuon". These are also assumed to be factors.

In the next step, the segments that have features [+high] or [-sonorant] in common were separated from [+sonorant].

[-sonorant]
--- z.d.r.h.s.t.p.k.
--- o.w.a.e.j.i.u.m.n.ô.b

[-high] [+sonorant]

This analysis /w/ is classified as [+high], although it is classified as [-high] in the case of English. In the case of English, [w] is produced with a constriction between the upper and lower lips and the back of the tongue and soft palate as well, and is a so-called voiced labial-velar approximant. On the other hand, in the case of Japanese, the degree of raising the back of the tongue is lower even at the word initial position, and it is pointed out that is still lower at the word medial position. Therefore, the informant of this paper analysis reflects such properties of Japanese, and /w/ was classified as [+high]. The group which holds [+high] or [-sonorant] is subdivided into a group which has a feature [-round] [v, o, i, y, u, w, and /], and a group which has a feature [+round] [v, a, o, /].

[-high] [+sonorant]

In this analysis, [v, o, i, y, u, w, and /] are classified as [+sonorant], but /v/ is classified as [-sonorant].

The segments that have a feature [-round] in common are still subdivided into individual phonemes of /a/ and /o/ by a feature [+low]. The low vowel /a/ and the non-low vowel /o/ are separated by this feature.

[-round]

[low]

The other group of segments are also subdivided into individual phonemes in a similar way.

4. DISCUSSION AND CONCLUSION

There is a tendency to revise not only partial problems but also the total framework of feature systems in many ways. One of the main concerns among them is setting up a hierarchy structure or groupings for the feature arrangement. Until now, several kinds of hierarchical or groupings of features have been proposed. For example, in a Jakobsonian framework, Fant (1973) discussed a feature hierarchy depending on the economy of description [5]. From the articulatory recognition study, Dantsuji (1989) proposed a feature hierarchy making use of auditory distance [1]. In a generative phonology framework, for example, Clements (1985) discussed feature hierarchy geometrically organized from a phonological point of view considering articulatory aspects. In a phonetic view, however, this elaborate feature hierarchy from phonetic and physiological facts [2, 9]. These phonetic and physiological facts mean that speech sounds are produced with the movement and action of a physiologically limited number of articulators, as was pointed out by Maddison and Ladehoff (1989). Movable articulators are lips, tongue tip, tongue blade, tongue root, tongue root, soft palate, larynx and so forth. Therefore, as terminal features [high], [back] and [low] have, for example, [high] is related to the movement of the dorsum of the tongue, which are dominated by a non-terminal node dorsal. As lips, tongue tip and dorsal are related to the place of articulation, these nodes are dominated by other node place. Furthermore, the place node and soft palate node are dominated by a still higher node, the supralaryngeal. However, major class features such as [sonorant] and [consonantal] are directly dominated by a root node which is the highest position of the hierarchy, or situated as special features that constitute the root node.

On the other hand, the analysis establishes another type of feature hierarchy which reflects the acoustic distance. Features such as [sonorant] and [consonantal] are extracted at quite early steps in this experiment. For example, [+sonorant] is extracted at the first step of the clustering. These matters indicate that the acoustic distance between segment groups corresponding to the feature [±sonorant] and [±sonorant] is considerably great. Therefore, this confirms the view that the feature [±sonorant] is placed at a higher position of the feature hierarchy, as proposed in current literatures of non-linear phonology, and correspond to the idea that the feature is regarded as an acoustic property on articulatory and physical facts.

5. REFERENCES

LE DEBIT DE PAROLE : UN FILTRE UTILISE POUR LA GENERATION DES VARIANTES DE PRONONCIATION EN FRANCAIS PARISIEN

Anne Lacheret-Dujour

LIMSI-CNRS BP 133 91 403 Orsay-Cedex France

ABSTRACT

This paper describes the development of a grapheme-to-phoneme strings module according to speech rate in French. Some examples of the phonological variations linked to the speech rate and the basic principles of the system are presented.

1. INTRODUCTION

Puisqu'il n'existe pas de prononciation standard en reconnaissance de la parole bilingue, l'intégration de modules de génération automatique des variants de prononciation dans des systèmes de reconnaissance phonétique est nécessaire pour l'accès au lexique. De tels systèmes ont été réalisés au LIMSI : GRAPHER [5] et VARION.0 [1]. Les tests de VERION.0 ont révélé la complexité du problème lié à la génération automatique des variants de prononciation : du fait de la production équivalente et maximale des variants, l'explosion combinatoire des chemins allépophones produits est inévitable. Il est donc irréaliste de penser pouvoir utiliser de façon optimale ces systèmes dans des contextes de reconnaissance : l'utilisation d'heuristiques, qui permettent de limiter le nombre de candidats et de travailler plus efficacement, est indispensable.

Nous avons développé pour le français parisien un module de transformation graphème-phonème avec variants, VARION.1 [3], dans lequel la génération des allépophones est conditionnée par le degré de débit (lent, rapide). Les règles du module ont été développées à partir de l'observation de corpus de parole. Les variantes sont reconnues par des fonctions de débit prises en compte par le système : la prononciation variable du schwa et la fusion voca1ique. L'architecture générale du module de règles (logiciel utilisé et formalisme adopté) est également décrite.

2. LE SYSTEME DE REGLES

2.1 Génération des variants en fonction du degré de débit : présentation des règles.

* Elision facultative du schwa

Indépendamment du débit, l'élimination du schwa est envisageable seulement si le schwa est précédé d'une seule corone dont le contexte gage est un monosyllabe, à savoir si le schwa est isolé en fin de mot.

- En début de polysyllabe ou dans un monosyllabe, le 'e' est toujours maintenu en débit lent, il peut être éliminé en débit rapide ('demandeur' → 'démmande'), en débit lent, 'em' /-domaine/ ou 'dém' /domain/ en débit rapide.
- En milieu de polysyllabe, le schwa est éliminé obligatoirement en débit rapide, facultativement en débit lent ('membre' → 'membre' /domain/ est également prévu en débit lent).
- En fin de polysyllabe, l'élimination facultative, obligatoire ou interdite des finale 'e', 'es', 'ent' dont le contexte droit est consonantique dépend du nombre de consonnes à gauche de la finale. Les règles sont les suivantes pour les contextes consonantiques gauches ci-dessous :

R1 [+cons]: la finale est éliminée quel que soit le débit (une robe verte → 'une robe verte').

R2 [+cons, liaison] /liaison/ : la finale est toujours maintenue en débit lent, facultativement en débit rapide, entraînant avec elle la chute de la liquide qui la précède ('petits peuples Paris' → 'petits peuples Paris' /pliee/ en débit lent, 'pliee' /plie/ est une variable possible en débit rapide). R3 [+liaison]: la chute du 'e' est facultative quel que soit le débit (une valise /unie valise/). R4 [+cons]: la chute de la finale est tolérée en débit rapide uniquement (un texte /un texte/). Quand la finale 'en' est suivie d'un mot à initiale vocale, le 'e' est toujours éliminé en débit rapide, en débit lent il peut être entendu si la liaison est effectuée (la aiment y ailer /laiment y ailer/).

Un certain nombre d'exceptions sont à noter à ces règles générales, pour lesquelles, quel que soit le contexte et le débit, le 'e' est toujours maintenu (femme, reliet). Il en va de même dans le détermiant 'le' accentué, dans le démonstratif 'ce' suivi d'une voyelle (ce ou quoi). En revanche, dans les formes du futur, le graphème 'e' correspond à un phonème [0] si le contexte gauche est autre que '[+obstruent] [liquant]' (il aidera → '/izéra/). Il en va de même pour le phonème [je] postposé au verbe (qui suis-je?) ainsi que dans la tournure interrogative 'est-ce' (qui est-ce qui vient?).

* Fusion voca1ique

Si 2 voyelles identiques sont séparées par une frontière de mots et éventuellement un 'h' aspiré, elles peuvent être réduites en un seul et même long segment. Il s'agit de fusion voca1ique. Dans le module, la fusion voca1ique est réalisée pour toutes les voyelles en débit rapide uniquement puisqu'une connaissance de type syntactico-sémantique est prise en compte dans le débit lent.

Dans le module, la fusion voca1ique est réalisée pour toutes les voyelles en débit rapide uniquement et en fonction d’un 'h' soutenu, élaboré, existe un 'h' aspiré, elles peuvent être réalisées en un seul et même long segment. Il s'agit de fusion voca1ique. Dans le module, la fusion voca1ique est réalisée pour toutes les voyelles en débit rapide uniquement puisqu’une connaissance de type syntactico-sémantique est prise en compte dans le débit lent.

Dans le module, la fusion voca1ique est réalisée pour toutes les voyelles en débit rapide uniquement et en fonction d’un 'h' soutenu, élaboré, existe un 'h' aspiré, elles peuvent être réalisées en un seul et même long segment. Il s'agit de fusion voca1ique. Dans le module, la fusion voca1ique est réalisée pour toutes les voyelles en débit rapide uniquement puisqu’une connaissance de type syntactico-sémantique est prise en compte dans le débit lent.

Dans le module, la fusion voca1ique est réalisée pour toutes les voyelles en débit rapide uniquement et en fonction d’un 'h' soutenu, élaboré, existe un 'h' aspiré, elles peuvent être réalisées en un seul et même long segment. Il s'agit de fusion voca1ique. Dans le module, la fusion voca1ique est réalisée pour toutes les voyelles en débit rapide uniquement puisqu’une connaissance de type syntactico-sémantique est prise en compte dans le débit lent.
2.2 Le système de règles VARION.1.

Les principaux objectifs lors du développement de VARION.1 étaient les suivants :
(1) Formaliser les variantes par le biais de règles et non en faisant usage d'un lexique. Cela, la structure de la langue française ne justifie pas l'emploi d'un dictionnaire couverts en espace contrées où entrent site de en (3)
facilement En sur LEX en compilées nombre utilisateur. temps sont déclenchées ordonnées, forme à déterministe le compilateur.
 dans certaines langues, sont inexistants en français. Enfin, l'utilisation d'un lexique nécessite une maintenance rigoureuse afin de traiter correctement les néologismes.
(2) Effectuer une transcription graphème-phonème sans passer par l'intermédiaire de formes de base au sens chomskyan du terme ; concept qui sous-entend une notion d'écart par rapport à une norme abstraite dont la définition est loin d'être claire.
(3) Adopter un formalisme permettant une écriture simple, économique et compacte des règles, qui doivent être facilement testables et le cas échéant modifiables. Pour ce faire, nous avons utilisé le compilateur de règles LEX [4] qui permet d'effectuer n'importe quel traitement linguistique sur une chaîne de caractères donnée en entrée (Fig.1).
Les règles, dont lesquelles ont inclus des traitements procéduraux, sont écrites sous forme déclarative et compilées en langage C. Le programme effectue ainsi un certain nombre d'actions spécifiées par l'utilisateur. Il gère ensuite un automate déterministe d'états finis. Le temps requis pour l'exécution des règles dépend de la taille du texte à phonémiser. Dans un fichier, les règles sont ordonnées, non cycliques, elles sont déclenchées de gauche à droite de la forme à transcrire, elles sont du type : 
\[ A \text{ ou } B \];
En tout point du texte donné en entrée où l'on trouve le caractère 'A', appliquer la fonction \( g_A \) pour le remplacer par le caractère 'B'. On substitue ainsi un buffer phonémique à un buffer graphémique au fur et à mesure de la transcription. 'A' peut être un mot, un grapheme ou même une séquence de mots. 'B' peut correspondre à 0, 1 ou plusieurs phonèmes.

La figure 2 représente la phonémisation de la séquence graphémique 'Le corbeau officie seulement le mardi' en début lent et en début rapide.

3. CONCLUSION

Les tests du système sur de la parole humaine mettent en lumière l'amélioration des résultats (98.8% de variantes prononcées et prises par le système en début lent, 98.4% en début rapide contre 95.5% pour VARION.0 tout début confondu). Néanmoins, l'analyse de ces résultats amène les conclusions suivantes : des connaissances prosodiques supplémentaires (la distribution des pauses par exemple) sont nécessaires pour améliorer les performances d'un tel système. Il en va de même des connaissances syntaxiques. Enfin, le début de parole est une variable relative. Les variations de début ne sont pas toujours exécutées de façon identique d'un sujet à un autre. Chez un locuteur donné les variations sont également possibles. De ce fait, à un début particulier, les choix allophoniques peuvent varier d'un groupe de locuteur à un autre et chez un même sujet lorsqu'il répète la même séquence de parole. Il est donc nécessaire d'étudier, outre les facteurs linguistiques, les mécanismes extra-linguistiques (sociolinguistiques, idiolectes) et para-linguistiques (situation de discours, émotion locuteur, etc) qui sous-tendent les stratégies allophoniques pour des classes de locuteurs données. Une telle étude permettrait de déclencher les ensembles de règles appropriés à un groupe de locuteur spécifique.


FIGURES

Fig.1 : Le compilateur de règles, LEX.

Fig.2 : Phonémisation de la séquence 'Le corbeau officie seulement le mardi', en début rapide (Fig.2a) et en début lent (Fig.2b).
LA LIAISON À ORLEANS (FRANCE) ET À MONTREAL (QUEBEC)

Daan de Jong
Université Libre d'Amsterdam / Université de Montréal

This paper summarizes some of the most important findings from a large scale, ongoing research project after sociolinguistic variation in liaison usage in Orléans French (France) and Montréal French (Québec).

1. INTRODUCTION

La liaison est la réalisation d'une consonne latente finale devant un mot à initiale vocalique (voir il). Devant un mot à initiale consonantique ou en fin de phrase, une consonne latente n'est jamais réalisée (voir voir 2).

1.chef eux
   [ʃɛf plɛ]   petit ami
2)chef lui
   [ʃɛf lɥi]   il est petit

La réalisation d'une consonne latente devant une voyelle dépend de plusieurs facteurs. Nous examinerons de quelle façon la variation dans l'emploi de la liaison est influencée par ces facteurs suivants: la structure syntaxique, la classe sociale, l'âge et le sexe. En plus, nous comparerons l'emploi de la liaison dans deux variétés du français: le français d'Orléans en France, et le français de Montréal au Québec.


1. LE RÔLE DE LA SYNTAXE

Un premier facteur d'affaiblir la fréquence d'emploi de la liaison est la structure syntaxique [8, 11]. De Jong [10] montre que la structure syntaxique doit d'abord être transformée en une structure prosodique hiérarchique consistant en trois couches de constituants prosodiques: le Groupe Clitique (GC), la Petite Phrase Phonologique (PP) et la Phrase Phonologique Maximale (PPM). Dans le premier constituant, la liaison est très fréquente, dans le deuxième elle est d'une fréquence moyenne, et dans le troisième elle est rare.

La dérivation en constituants prosodiques supprime une analyse de la phrase en termes de la théorie X-bar. Ainsi, la fin (droite) de chaque tête (X) délimite le domaine de la liaison à réaliser (obligatoire). Nous considérerons comme tête les catégories majeures N, A ou V, et aussi les catégories mineures P, Comp et Aux (6,7). Ainsi, la phrase "il y a des pieds pour les enseignants" est divisée comme suit en GCs: (il y) a (des) pieds (pour) les enseignants (admirables) est divisée comme suit en GCs: (il y) a (des) pieds (pour) les enseignants (admirables).

La liaison doit être plus fréquente dans la GC que dans la PPM, et dans la PPM elle est plus fréquente que dans la PFP. Nous avons testé cette hypothèse sur 45 entrevues du corpus d'Orléans. La figure 1 montre clairement que l'hypothèse est confirmée. En plus, cette figure montre, que la hiérarchie GC > PPM > PFP vaut pour toutes les couches sociales. Finalement, nous voyons que l'emploi de la liaison dépend de façon régulière avec la classe sociale.

2. LE FACTEUR LEXICAL

L'analyse précédente ne suffit pas à elle seule à prédire l'emploi de la liaison. Dans De Jong (1988, 1991), nous avons présenté une analyse statistique des données relevées sur le corpus d'Orléans au moyen d'un modèle logitinaire. Cette analyse a démontré que deux facteurs locaux influencent l'emploi de la liaison. Dans un premier temps, le facteur lexical est lié à la fréquence de la liaison. Ainsi, la liaison était...
significativement plus fréquente après des mots monosyllabiques qu'après des mots polysyllabiques. La liaison avec /t/ était plus fréquente que la liaison avec /s/. La catégorie grammaticale avait aussi un effet significatif. Finalement, la liaison se faisait plus souvent après les mots très fréquemment utilisés qu'après les mots peu fréquemment utilisés. Des résultats comparables ont été obtenus pour le Corpus de Montréal.

3. LES FACTEURS EXTRALINGUISTIQUES
La fréquence d'emploi de la liaison est aussi significativement influence par plusieurs facteurs extra-linguistiques. Nous avons examiné le rôle de la classe sociale, de l'âge et du sexe, pour Orléans et pour Montréal. Les principaux résultats sont résumés dans les figures 2a, 2b et 2c, qui montrent que l'emploi de la liaison décroît avec la classe sociale, augmente avec l'âge, et que les femmes utilisent plus de liaison que les hommes. Ces données montrent aussi, contrairement à ce qui est dit dans Encrevé (1988: 50), que la liaison variable n'est pas limitée aux classes sociales supérieures, mais se retrouve dans toutes les classes sociales. Ces figures montrent aussi que la liaison se comporte tout à fait comme les variables sociolinguistiques décrites dans Labov (1972), et absolument pas comme 'une variable socio-linguistique inversée' (Encrevé 1988: 45). Finalement, ces figures montrent que celles qui varient aussi bien pour Orléans que pour Montréal (voir aussi [2,4]).

4. DIFFERENCES ENTRE ORLEANS ET MONTRÉAL
À part des ressemblances, il y a aussi des différences importantes entre Montréal et Orléans. Par exemple, à Montréal, la liaison après mais ne se fait souvent non avec /s/, mais plutôt avec /t/. La fig.2 montre que l'emploi du /t/ après mais est fréquent dans les trois classes inférieures, mais presque absente dans la classe supérieure.

5. NOTES
1. Je veux remercier Pierrette Thibault, David Sankoff et Henrietta Cederberg pour m'avoir donné accès au corpus de Montréal. Je tiens également à remercier le Département de Linguistique et de Philologie de l'Université de Montréal, où j'ai pu écrire cet article, financer ma professeur invitée. Les recherches pour cet article ont été rendues possibles grâce à l'Académie Royale Néerlandaise des Sciences et des Lettres.

6. REFERENCES
VERB STRESS IN SPANISH

S. Alcoba

Universidad Autónoma de Barcelona.

ABSTRACT

In this study, extrametricality is put forward and the hierarchy of stress markedness is precisely identified among verbal forms in the Thematic Vowel (TV) affix, keeping the stress rules parameters in a pcancategorical sense, in order to avoid some incoherences which we shall refer to in the solution of J.W.Harris, the most solid and best directed proposal.

The hypothesis in connection with the morphological determination of the verbal stress in Spanish is in a state of crisis. What is mentioned in [4], page 84, that "Segmental phonological representation and morphological identification are jointly necessary and sufficient to determine placement of stress for all verb forms" is nowadays considered to be a challenge which needs to be and can be overcome: [5], [7] and [8] point in this direction.

1. EXTRAMETRICITY OF THE FINAL METRIC ELEMENT.

1.1. Established generalisations.

The generalisations in connection with the stress in Spanish words which I consider to be established are the following:

First. Spanish words, as far as the accent is concerned, are hierarchically divided into three classes: on the one hand, a wide selection of words with regular stress on the penultimate syllable, or Type B ('sa'bana's, 'comi'cas, 'arbo'les, 'utiles), and the other of words with the stress on the last syllable, or Type C (far'a'tus, ca'jies, far'cies, marro'ques, domi'nos, ban'tues). The roots or affixes of type B are marked in the lexicon with a responsible diacritic feature that expresses the markedness of the stress in Type B words. The unmarked lexical entries will be of Type A/C.

Second. All morphological words, of whatever category, fall in with the so-called window of three syllables restriction (WTS) which makes the existence of words (*)-**) with the stress on one syllable further back than the third from the end impossible in Spanish.

Third. The Spanish stress is sensitive to quantity (Branching Condition of [4]). An stress of Type B is not possible if the penult syllable has a branching rhyme: (*)-**(V)-, (*)-(V)-, (*)-(V)-. Nor is a Type B stress possible in words ending in a final rhyme GV, (*)-(V)-GV) (*cont'ina, cont'ina, cont'ina). Under these circumstances the window for stress is reduced to the last two syllables.

Fourth. The domain of stress assignment is the morphological word. The rules or parameters for the assignation of the stress explore the whole word (theme and non-cyclic or inflectional affixes) and establish the stress in accordance with the foregoing conditions.

The hypothesis of the word as the domain of the stress rules presupposes that the parameters explore lexical forms in a derivative stratum in which the non-cyclic affix constituents (word markers and paradigmatic constituents: TIP, TIN, TITMA and TIPN) are specified.

Fifth. The representation of the stress is placed within the framework of the theoretical model developed in [3].

1.2. Harris's stress rules.

The stress rules in [7] to generate the stress grids of Spanish words are collected in (1) and illustrated (2) where, for simplicity's sake, only line 0 of the stress grid is shown.

(1). Stressassignable elements are syllable nuclei (rhyme heads).

2. The rightmost stressable element is extrametrical if word-final or followed by an inflectional consonant.

3. Forms constitute(s) on line 0 and mark head(s) on line 1:

Parameter settings:

a. unbounded, right-headed (general case).

b. binary, left-headed, right-to-left (special case).

4. Form constituent(s) on line 1 and mark head(s) on line 2:

Parameter settings:

a. unbounded, right-headed.

5. Conflate lines 1 and 2 ( = remove asterisks in columns that have no line 2 asterisks).

(The).

1.3. Limits and drawbacks to Harris's hypothesis.

In the manner in which extrametricality is formulated in (1.2), the distinction is maintained between words of Type A (ba'clones, a'zu's), and of Type C (polo'nos, ban'tues) and these last remain pending an exceptional explanation (cfr. [7]: 257 and n. 4). Among the verb forms the final resolution of the oxion forms of the "weak preterite" and the "future" also remains outstanding. (cfr. [7]: 257).

Furthermore, it is obvious that the difficulty in generating the infinitive itself consists by reason of the incoherence deriving from calling the final segment a derivative. In addition, if the paradigmatic morphemes -BA-, -RASE-, -RE- and -STE- are considered as being Type B to explain the accent on forms such as contes'ta.BA.<mos>liz>, how can the stress of contes'ta.<BA(s/n)> be explained? In the same way as occurs in contes'ta.STE.<u>'s> as against contes'ta.<STE>.

2. EXTRAMETRICITY OF FLEXIBLE METRICAL ELEMENTS.

2.1. Hierarchy of markedness in verb forms.

I suggest that it is the outermost cyclic affix of the verb forms, the various forms of TV which carry the diacritic of markedness. The morpheme TV (a, e, l) of the Theme of the Preterite (cfr. [1] and [2]) will be of the type A: weak preterite, imperfect indicative, imperfect and future subjunctive, gerund and participle. The morpheme TV (A, E, D) of the Theme of the Present will be of type B: present indicative, present subjunctive and imperative. And the morpheme TV (a, e, l) of the Theme of the Future will be of Type C: future indicative and conditional.

I suggest furthermore, that extrametricality be understood in terms of (3):

(3). Extrametricality (replaces (1.2)).

A stressable element is extrametrical if it matches an inflectional constituent.

The formulation of (3) does not have a higher theoretical cost than (1.2): "It is perfectly straightforward to distinguish between «inflexional» and «non-inflectional» morphemes in Spanish... the set of «inflexional» morphemes contains exactly class markers and the plural morphemes in non-verbs plus tense/mood/aspect and person/number suffixes in verbs... the set of «inflexional» morphemes corresponds exactly to the noncyclic affixes in the Halle-Vergnaud theory of phonological organisation" (cfr. [7]: 253).

If (1.2) is a pcancategorical formulation, so is (3); but (1.2) has to treat as exceptions words of Class C, which is not necessary with the formulation of (3): words which lack fractional constituents will not have extrametrical elements.

Now, given the formulation of (3), if it
is postulated that the domain scanned by
the stress rules (1.3-5) is the derivative
the stress rules (1.3-5) is the derivative
the stress rules (1.3-5) is the derivative
type: (3) is specific to Spanish.
A Study of Vowel Coarticulation in British English

James L. Hieronymus
Centre for Speech Technology Research, Edinburgh University
80 South Bridge, Edinburgh EH1 1HN, Scotland

Abstract
Coarticulation in continuous speech causes vowel formant frequencies to be affected by nearby phonemes. Generally continuous speech causes the vowel formant targets to be centralized relative to their isolated word counterparts. The present study concentrates on 660 phonetically hand labelled sentences from one male talker of the RP accent of British English. This allows the study of coarticulation without the confounding effects of accents, speech habits and differing individual formant targets. The 12 monophthongal vowels of RP British English, i, i, e, a, ʌ,ʊ, o, u, ʌ,ʊ, ɔ, æ, ɒ, have been studied using formant frequency and amplitude tracks and duration, and sentential stress (sentence stress as opposed to lexical stress). Generally the vowels are most affected by nearby semi-vowels /ɪ, ʏ, w/. No simple relationship between adjacent phoneme place of articulation and the vowel target change has been found when all the vowels are treated together. However, the data shows the presence of "robust vowels" which are not greatly affected by nearby semi-vowels. These vowels are not simply stressed vowels, but depend on duration and other factors being studied. The weak effect of duration is that the pre-pausal lengthened vowels are in the "robust" category, but shorter vowels can either be robust or ordinary. The categories of function word and content word do not account for robustness.

Introduction
Most coarticulation studies have considered isolated words. An early study by Sherwood and Holmes [1] showed that vowels in continuous speech, very seldom had steady states and often did not overlap the Peterson–Barney [2] 95 percentile contours in any part of their frequency trajectories in time. Generally the vowels are much more centralized in continuous speech and the vowel formant regions overlap considerably due to coarticulation. Kirehara [3] found a renormalization technique based on the theory of Lindblom and Studdert-Kennedy [4] which disambiguates Japanese vowels in continuous speech.

Hieronymus and Majarksi [5] tried this technique on American English vowels and found that it did not work well. It has been proposed that the stress structure of English causes this method to fail. The presence of "robust vowels" as found in this study would cause this technique to fail, because the renormalization is applied uniformly to all vowels.

This is a report of an ongoing study of vowel properties and coarticulation in British English. The present approach is to study the speech of one talker at a time in detail to find the underlying mechanisms in coarticulation. Thus coarticulation can be studied without the confounding effects of regional accent, speaking styles, and formant ranges due to different talkers. Then speech data from other talkers will be studied and the pooling of the data explored to achieve speaker independent results later in this study.

It is postulated that some sort of hierarchical structure of linguistic factors modifies the effect of nearby phonemes such that the same vowel in the same phonetic context will have markedly different formant frequency trajectories in time. Some possibilities for factors which have been explored are sentential syllable stress, duration, and word identity. Originally it was thought that sentential stress would be the determining factor of vowel precision of production. Previous studies by us [5, 6] for American English have shown that sentential stress is not a determining factor, based on automatic stress labelling. The present study uses hand labelled stress and shows that, on average, sententially stressed vowels are more precisely produced than their unstressed counterparts.

Method
The data was read by one male talker of a near RP dialect of British English in a sound isolation booth. The microphone was a Shure SM81. The speech was digitized directly using a 16 bit a-to-d converter at 20 kHz sampling frequency with an anti-aliasing filter at 8 kHz. The talker was told to speak the sentence as if he was saying it in conversation and was prompted with the sentence on a computer screen. The speech was hand labelled by graduate phoneticians as a broad phonetic level with syllable stress marked using a PC based labelling workstation. The labelers were presented with a spectrumogram and could play the segments. Subsequently the sentences were parsed by hand to provide loose bracketing of phrase boundaries, so that syntactic effects could be studied. Of the 660 sentences designed for the CSTR/ATR database project to collect and label speech for speech technology studies. The other 460 sentences were labelled versions of the TIMIT compact sentence database designed by the MIT Speech Group. Each vowel formant is characterized by three values for each hand labelled vowel. The values are the first and second formant frequencies at points 10 %, 50 % and 90 % of the duration of the vowel. These values were chosen to minimize an effect of formant tracking errors. Formant tracks are obtained from a centroid based formant tracker developed by Crowe [8].

Except for low formant frequency values in the nasalized vowels the formant tracker seems to have a lower error rate. These values are then fed into the APS system developed at CSTR by Watson [9] providing an interface to the S package to allow statistical studies of the data.

Discussion of the Data
Figure 1 shows a scatter plot of the first and second formant values measured at the temporal center for the long British English vowels in extracts from 358 sentences with ellipses representing 66 percent of the data (a/b is a long vowel durationally in this data even though it is phonologically lax). Figure 2 shows data for short vowels. The normal range for a male talker is 200–1000 Hz and 2000–2300 Hz for the second formant. The minimum perceptible differences (DL) in formants were measured by Planagan [11] and found to be 4.5–50 Hz for F1 and 4.5–75 Hz for F2. Thus a measure of precision of production is how large the standard deviation of the data is relative to the DL. The cross hatched area in each vowel region is the ellipse for the sententially stressed vowels. The formant regions for most vowels are as expected except that this talker has a very fronted /ʌ/. The vowel /u/ is the highest back vowel for this talker with a median second formant of approximately 800 Hz. The long /u/ is more precisely produced than the reduced vowel schwa (not plotted because of us large standard deviation) with the long form having a significantly lower second formant.

While the stressed vowels are more compact in the 66 percentile ellipses, there are a considerable number of wide ranging outliers. Secondly the distribution of data points towards the outer edge of these ellipses. These are the "robust" vowels as will be shown.

The short vowels have more scatter and thus seem to be produced with less precision. Once again the stressed vowels are statistically more compact than the unstressed vowels. A cross hatched dot plot shows a considerable overlap between vowels in the tense–lax pairs. Duration plots show that the durations of the tense vowels are greater than the lax vowels. The presence of "robust" vowels is shown in Figure 3 which shows the stylized formant trajectories for /e/ in the environment of preceding /m/ and following /l/. The smaller font characters are the preceding context and the large character represents the following context. The question to be answered is why do some examples of the vowel /e/ have second formant "targets" above 2100 Hz, even in this environment? The primary stress vowels in /eme/ show up in a circle and the secondary stressed vowels are highlighted by a square.
see, all of the stressed vowels and one unstressed vowel are the ones with high second formant. Consider the context /w i k/ in the figure. Two of the tracks are for the stressed vowel and appear in the robust region. One of these vowels in the same context (shown by the arrow) is not robust, and is unstressed. The /w i k/ context shows two unstressed examples, one of which is robust and one not. Thus stress does not seem to be a reliable correlate of "robustness" for vowel production. Similar plots of prevocalic /f/ show an even greater tendency for unstressed vowels to be robust but are more difficult to see at this scale. The vowel /f/ was also examined for the presence of robust exemplars in the presence of /f/ the palatal glide, and they were found. There was a weak correlation between stress and robustness.

Two possible factors, the type of word and duration, were examined to determine whether or not they determined robustness of the vowel. The longest, pre-pausal lengthened vowels are all robust. However for shorter vowels, duration is not a good correlate of robustness. Both function and content words were found to contain robust vowels. The word "between" was found to have a robust vowel on one occasion and a coarticulated vowel in another.

Acknowledgement

The programs written by Alan Crowe and Gordon Watson were used in this study. The speech labelling by Marie Alexander, Cathy Bennett, Delth Davies, Imogen Cohen, Keith Edwards, Hazel Syderiff, Ivan Uemlinian and Christina Barr were instrumental in this study. The collection and labelling project was supported by ATR, Japan. This work was supported by the Alvey Integrated Speech Technology Demonstrator project in collaboration with Marconi Speech and Information Systems and HUSAT, Loughborough University. The demonstrator is supported by the UK Department of Trade and Industry and the British Science and Engineering Research Council under grants D/29611, D/29628, D/29638, D/29640 and F/110309.

References

Three Experimental Sources of Evidence’

Bruce L. Derwing and Terrance M. Nearey
University of Alberta, Edmonton, Canada

ABSTRACT
Data are reported from three independent sources, involving active word-manipulation tasks, substitution-identification tasks and both active and passive syllable boundary tasks. All show a consistent tendency for glides in English to adhere more closely to vowels than to nasals or post-vocalic obstruents. Cross-linguistic studies are under way to test the universality of these findings, as well as formal models being planned to account for these results.

1. BACKGROUND
We use the term ‘vowel-stickiness’ to refer to the tendency for some segments to adhere more closely to vowels than to nasals or obstruents [1]. Though much of the evidence for this phenomenon was conducted under the rubric of ‘syllabic structure’ or ‘intra- or sub-syllabic units,’ these terms imply a sharply delineated or ‘hierarchical’ view of syllables that is less well supported by the facts. Experimental evidence for the ‘stickiness’ notion comes from three distinct sources: production experiments and pattern-identification studies that were focussed on questions of the internal structure of syllables, plus a combination of production and judgment tasks that were directed at the question of syllable boundaries.

2. EXPERIMENTAL WORD GAMES (PRODUCTION TASKS)
Treiman [2,3] used a variety of experimental word games (notably word-bending) to explore the internal structure of English syllables, and Dow strengthened these findings, using primarily a unit-substitution (or deletion) task [4,5]. What all this work demonstrated was that there was more to a syllable than a simple linear sequence of (phonemic) segments: it also purported to show that well-defined ‘units’ were also involved (such as the onset, the rime, the nucleus/peak and the coda) and that the structure of syllables was not only hierarchical but also (at least for English) right-branching. One disquieting fact emerged from this early work, however, to complicate the picture. Specifically, in one series of studies [6], Treiman found that the boundary between the nucleus and the coda was less than firm and, in fact, tended to shift in response to the sound class of the post-vocalic consonant involved. Specifically, subjects tended to break VCC syllables before the first consonant if that consonant was an obstruent, but after it if the consonant was a liquid (i.e., /l/ or /r/), whereas the two tendencies were of about equal strength if the first consonant was a nasal. Thus liquids (L) tended to stick with their original vowel in these tasks and obstruents (O) to split away, with the nasals (N) holding an intermediate position. In terms of their general tendency towards vowel-stickiness, therefore, the order L > N > O was observed.

3. SUBSTITUTION-PATTERN IDENTIFICATION TASKS
In order to circumvent the slow and laborious production data-collection methods of these early production studies, we experimented with a new forced-choice judgment technique called the ‘substitution-pattern identification task.’ In this task, rather than asking subjects to actively replace some part of a syllable (such as the vowel, or an all-obstruent onset or post-vocalic coda) with a substitute segment or string, as Dow had done, subjects were trained instead merely to identify such a replacement. Thus, in a training session, subjects were orally presented with a dozen or so examples of a particular substitution pattern (e.g., replace the vowel by /l/; or delete the onset; or replace the coda by /ps/); then, in the testing phase, the subjects were asked to respond to new word pairs, merely by indicating whether the substitutions involved were the same (‘YES’) or different (‘NO’) in kind to the particular patterns that they were trained on. Reinforcement items from the training set were also regularly interspersed among the test items, in order to remind subjects of the nature of the pattern that they were looking for (see [1,7,8] for details).

What was critical about the test items in this last study was that they all contained either pre- or post-vocalic sonorant consonants, and these were sometimes replaced along with the units in question and sometimes not. Thus, having been trained to replace an all-obstruent coda by /ps/ (as in /avr/ → /apr/ or / felony → felony/), a subject might now be asked whether the nonsense-pair /eisl/ → /eip/ illustrated the pattern (where all post-vocalic consonants were replaced) and, somewhere else on the test, also asked whether the pair /elg/ → /elp/ did (where only the post-vocalic obstruents were replaced, leaving the sonorant — in this case /l/ — ‘snuck to the vowel.’) Using a slightly modified form of the χ2 statistic from signal detection theory, the relative tendency of the various sonorant consonants to adhere to vowels was then calculated, taking into account not only correct HITS (involving the nominally correct pattern, where all sonorants were treated as part of consonantal clusters) and MISSES (where such nominally correct substitutions were rejected), but also CORRECT REJECTIONS (where all but the nominally correct substitutions were rejected) and FALSE ALARMS (where nominally incorrect pairs were accepted, i.e., pairs that kept the vowel and associated sonorant stuck together). On the basis of a large body of experimental data for such a task, the following differences in backwardness were observed, adding the categories G (for the English glides /w,y/) and R (for English /r/) to the ones already discussed, and where data for O came from reinforcement items from the training session: G > R > N > O. (Other tasks, such as onset deletion and vowel substitution, showed a similar tendency in this study, though the absolute differences were not in all cases statistically significant.)

4. TESTS FOR SYLLABLE BOUNDARIES
Similar effects can also be extracted from the more recent work done on the problem of syllable boundaries by Treiman & Danis (T&D). Relying primarily on a production task of syllable inversion, T&D [10] investigated the problem of where common English syllabic words were broken that contained only a single intervocalic consonant. Their results (largely confirmed by an associated forced-choice written task) showed that the position of the break depended on a number of factors, including (1) the quality (tense vs. lax) of the vowel in the first syllable, (2) the position whether the pair /elg/ → /elp/ did (where only the post-vocalic obstruents were replaced, leaving the sonorant — in this case /l/ — ‘snuck to the vowel.’) Using a slightly modified form of the χ2 statistic from signal detection theory, the relative tendency of the various sonorant consonants to adhere to vowels was then calculated, taking into account not only correct HITS (involving the nominally correct pattern, where all sonorants were treated as part of consonantal clusters) and MISSES (where such nominally correct substitutions were rejected), but also CORRECT REJECTIONS (where all but the nominally correct substitutions were rejected) and FALSE ALARMS (where nominally incorrect pairs were accepted, i.e., pairs that kept the vowel and associated sonorant stuck together). On the basis of a large body of experimental data for such a task, the following differences in backwardness were observed, adding the categories G (for the English glides /w,y/) and R (for English /r/) to the ones already discussed, and where data for O came from reinforcement items from the training session: G > R > N > O. (Other tasks, such as onset deletion and vowel substitution, showed a similar tendency in this study, though the absolute differences were not in all cases statistically significant.)
of stress (on first vowel or second vowel), (3) the way the medial consonant was spelled (i.e., with one letter, as in melon, or two, as in gallon) and (4), most interesting from our current standpoint, the quality of the consonant itself. Most notably, in the case of consonants with singlet spellings in words with initial stress on lax vowels (such as melon, lemon and seven), L showed the strongest tendency to be treated as part of the first syllable, and O the weakest, with N, once again, taking the intermediate position.4

Finally, in the attempt to extend this work to typologically diverse languages (see [11]) in these proceedings for some initial results for Korean, Derwing sought to develop a simpler technique for syllable division that could be performed by subjects who were not necessarily literate, as well as administered to large groups of subjects simultaneously. The results were a so-called ‘pause-break’ task, in which subjects were asked to choose which of two or three alternative ‘breakings’ of a word sounded the ‘most natural’. In the case of the English word melon, for example, the following three alternatives were offered (where ... indicates the location of the pause):

(a) /mələn/ (where /ə/ is treated as the onset of the second syllable),
(b) /mələn/ /ən/ (where /ən/ is the coda of the first syllable), or (c) /mələn/ /ən/ /lən/ (where /ən/ is ambisyllabic).

In the English pilot study, 95 speakers were presented with a word-set much like T&D's. All four of T&D's main effects re-emerged, as well as a new factor of the morheme division. Of chief interest to us here, however, is the now-familiar four-way distinction among R, L, N and O, which the table below displays for words like herald, melon, lemon and seven:

<table>
<thead>
<tr>
<th>Sound Class</th>
<th>S1/Co</th>
<th>S2/On</th>
<th>Amb</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>.76</td>
<td>.07</td>
<td>.18</td>
</tr>
<tr>
<td>L</td>
<td>.62</td>
<td>.19</td>
<td>.19</td>
</tr>
<tr>
<td>N</td>
<td>.52</td>
<td>.37</td>
<td>.12</td>
</tr>
<tr>
<td>O</td>
<td>.29</td>
<td>.61</td>
<td>.09</td>
</tr>
</tbody>
</table>

Once again we see the same familiar differential tendency towards ‘vowel-stickiness’ as before, in this case realized as a tendency for singlet-spelled consonants to stick together with a lax, stressed vowel as part of the first syllable of a word: R > L > N > O.

5. CONCLUSIONS

In sum, the ‘vowel-stickiness’ phenomenon now seems to be quite firmly established, as it has been shown to be manifested in a consistent way across three different methodologies originally conceived for quite different purposes: in productive word-blends, in substitution-pattern judgments, and now in both production and judgment tasks for syllable divisions. Two major questions now remain: (1) to ascertain whether the same pattern holds for other, typologically diverse languages; and, if so, (2) to find a satisfactory explanation for the phenomenon. (It is worthy of note that a tantalizingly similar ordering - variously referred to as the ‘sonority’ or, inversely, ‘strength’ hierarchy - has emerged from descriptive linguistics, based on the investigation of both synchronic and diachronic data.) Extensive cross-linguistic work is now underway in our laboratory in search of an answer to question (1), combined with theoretical modeling and testing efforts suitable to satisfy the needs of (2).

NOTES

1. The research reported here was supported in part by a research grant from the Social Sciences and Humanities Research Council of Canada (No. 410-88-0256), awarded to the first author.

2. Note that L here now refers to English /l/, alone, as the distinct term R has been applied to /r/.

3. Using this same technique, the L > N portion of this hierarchy was re-confirmed in a later study [7] for post-vocalic sonorants, which also demonstrated the effect on ‘stickiness’ of both vowel and consonant quality, much along the lines suggested by Selkirk [9].

4. In this study, both English /l/ and /r/ were again tested as members of the same class (‘liquids’) and analyzed together.

5. Except that the list was modified to include separate samples for both /l/ and /r/, which, as already noted, were collapsed in T&D and treated together as ‘liquids’. A few new words (olly vs. doily, sailor vs. molar, foaming vs. moment, etc.) were also added to check on the effect of the morpheme boundaries.

6. S1Co = coda of first syllable. S1On = onset of second syllable, Amb = both (ambisyllabic). Response proportions are shown for each, with majority responses in boldface.

7. These include the construction of Markovian and neural network models of our own design, as well as alternatives proposed elsewhere (e.g., [12]).

REFERENCES


MARGINAL VOWELS IN HUNGARIAN

P. Siptár

Hungarian Academy of Sciences, Budapest, Hungary.

ABSTRACT

This paper suggests a variety of ways in which the number of categories needed for characterizing the surface phonetic vowel inventory of Hungarian (Table 1) can be reduced until eventually a minimal underlying system (Table 4) is reached. From 'marginal vowels' (parenthesized in Table 1) are discussed in particular, [a], [o], and [e] are argued not to be necessary in the underlying system; on the other hand, nonround /a/ turns out to be one of the most loaded Hungarian vowels: one that surfaces as [o] in the regular case, due to an independently motivated rule of the language.

1. INTRODUCTION

In surface phonetic classification the Hungarian vowel system is shown in Table 1. The system has fourteen 'full members' plus four additional (parenthesized) whose phonological status will be considered in this paper (Section 2). There will be a (flexical) alternation appearing in Table 1 involves five heights, three points of articulation along the sagittal axis, plus the rowed/unrowed distinction. Obviously, a number of phonetic details can be filtered out of this representation on grounds of predictability. Height 1 is conventionally labelled 'High'; the rest of the heights might be called Upper Mid, Lower Mid, Upper Low, and Lower Low, respectively. The difference between Upper Mid and Lower Mid might be taken to be a matter of Tense/Lax; but even that is predictable (redundant) on the basis of Long vs. Short (alternatively, VV vs. V in terms of timing slots). On the other hand, the two Lows may be simply taken to be the same height phonologically: the exact height of [a] is, as well as their centrality, is a matter of phonetic implementation since in the (morpho)phonological pattern of Hungarian [a] behaves as a low back vowel (e.g. with respect to vowel harmony, long/short alternations, etc.). Hence, the simplified pattern in Table 2 emerges; this classification will serve as the general framework within which the phonological status of the four 'marginal vowels' will be discussed in Sections 2.1-2.3 below. In Section 3, some general conclusions will be drawn and further simplification of the system will be proposed.

2. DATA AND DISCUSSION

2.1. Unrounded short [a]

This vowel appears on the surface (apart from regional dialects) in the following cases: (i) In nonfinal closed syllables it is the normal (colloquial) realization of /a/ as in dicsi̞nos 'salt-pan' "general", edősvíz 'water-vapor' 'market town'; in certain phonetic contexts with vacillation (where the optional shortening rule concerned is optional / rate-dependent): [atsk[incl]t̸̂] = [atsk[incl]t̸̂[e]nt] 'persecution'. (ii) Also with [a] ~ [a] free variation in words like spóz 'larder', Sóvic 'Switzerland', Mont; here, however, 'free variation' means inter-speaker variability rather than intra-speaker vacillation. (iii) On the other hand, [a] ~ [e] (inter-speaker) variation is found in words like govvez 'gavotte', hátvzer 'hardware', Csajkovskýi 'Tschaikowsky', and in hallo [hal] 'hallo' as used in phone calls (where classical minimal pairs can also be found for both [a] and [e]: hold [h] 'dying' vs. hallo [hal] 'hallo' vs. hold [h] 'set').

In this discussion, then, there are two phonological status of all these [a]'s should be. (From now on, I use the symbol /a/ to refer to the rowed / nonrounded types - short back low vowel with no roundness specification intended; the choice of symbol is motivated by considerations of clarity, i.e. I wanted a symbol that is distinct from both a and e.) There are a number of convincing arguments to the effects that /a/ behaves morphophonologically as a nonround vowel (cf. the length alternation /a/ ~ /a/ and the vowel harmony alternation /a/ ~ /e/). In both cases an intermediate nonround low back vowel is derived that surfaces via an /e/-realization rule. Since the rounding of /a/ is phonologically irrelevant (post-dissociative) and phonetically rather moderate as opposed to mid and especially high back vowels (though this does not weigh much in phonology), it is at least possible to claim that /a/ is in general (= not only in the alternating cases) underlyingly nonround. It was pointed out in Section 1 above that the centrality of [a] and the fact that in terms of tongue height it is lower than [e] or [o] are just as redundant phonologically as the surface roundness of [a] is. Hence, the /a/ ~ /e/ alternation will fit the rest of the pattern where alternates only differ in length (cf. 2.3. on /i/ ~ /e/).

Now if we accept this reasoning, the following can be said about the three groups of surface /a/ exemplified above:

(i) In addition to the morphophonological rule /a/ ~ /a/ (symp ~ Topp 'sum-mer') non-loc., followed by rounding adj.-ment /a/ ~ /o/ there is also a surface (postlexical) shortening rule that will of course apply (much) later than rounding adjustment and will therefore remain unrounded. (ii) For speakers who say /pa/ etc., underlying nonround /a/ will be (al)locaus 'general', unrounded adjustment in this words; for other speakers, the lexical representation will be /pa/ etc. to which shortening or rounding adjustment is inapplicable. (iii) The word hallo and other similar items (the exact range of which varies from speaker to speaker) are exceptional in that they will be (optionally or categorically) exempt from rounding adjustment /a/ ~ /e/. Alternatively, in terms of underspecification theory, garden-variety /a/ will be underly-ingly unspecified for rounding whereas the vowel in hold etc., as well as spóz etc. for /a/ speakers, will be specified as /e/-round; rounding adjustment would then be a 'free' rule in that it cannot change feature specifications but only fill in blanks; the desired result then follows without recourse to any exception.

In sum: If these conjectures are on the right track, nonround /a/ is not marginal: in fact, it is one of the most loaded members of the Hungarian vowel system; what is marginal is the range of cases where the surface unrounded.

2.2. Short mid [e]

The case of this vowel is in some respects similar to that of [a], in others it is quite different. On the surface it appears with regional/cultural restrictions (i.e. in certain regional varieties); its use is much wider than that of /e/ - dialectal - [a], but does not include standard Hungarian in the strict sense. (The postlexical shortening of /e/ as in the second syllable of kéroms 'hard-ness' results in a vowel tense than [e], just like that of /o/ and /o/; that is, as was pointed out in Section 1 above, [e] and [o] and [o] are not only in length but also in tenseness).

If, in standard Budapest Hungarian, [e] does not appear even to the limited ex-tent that [a] does, why do we mention it here? The reason is that Hungarian mor-phonology works as if there was an "a"-less version of the vowel /a/ (a different member of the alternation s ~ e (at the level of the immediate output of the rule) is, whereas the front member of the alternat-
tion $\varepsilon \sim \varepsilon$ and the long member of $e \sim e$ (left $\varepsilon$ left 'brush sound') are low (at the same level), hence an $e/e$-adjustment (reduction) rule is needed to convert back the derived $e$'s into a low, and derived $e$'s into a mid (and tense) vowel. (Alternatively, Street's derivation might produce the same effects without an explicit adjustment rule.) These facts, however, are still not sufficient to identify an underlying $e$, unless the ambiguous behaviour of $i$ in vowel harmony could be explained by positional mid /i/ as well as low /i/...
LOCAL PROMINENCE OF ACOUSTIC AND PSYCHOACOUSTIC FUNCTIONS AND PERCEIVED STRESS IN FRENCH

P. Hertens
K.U.Leuven, Linguistics Department, Leuven, Belgium.

Abstract

Syllable duration, pitch, loudness, pause length, pitch change, and local difference values for the first 3 parameters, were studied for their ability to predict perceived stress as measured in a listening task. The best cues were duration increase relative to preceding and following syllables, followed by nuclear duration.

1. Introduction

Syllable stress is a linguistic attribute realized in various ways, with or without prominence. It cannot be observed directly. A measure of perceived prominence has to be established in order to classify the syllables. A brief review of terminology will clarify this point. (1) A syllable is prominent when it stands out from its context due to a local difference for some prosodic parameter. Prominence is continuous (not categorical) and contributions of multiple parameters interact. (2) Stress is an abstract linguistic category, which can be realized by several types of prominence, in a way which is language-specific. (a) In French, an intra-syllabic pitch glide of a given interval suffices to signal stress. Prominence by duration or loudness will be functionally redundant although very common. (b) For static syllables prominence will result from an inter-syllable change of a parameter. (c) Finally, stress can result from tone level itself, as the basis of tone distribution [3,6]. (3) Word stress (lexical stress) indicates the syllable in a word which can receive stress. (4) French has two stress types: final (word stress position) and initial stress (emphatic), with a different distribution. In a listentask, the stress judgment will be based on a mixture of heterogeneous factors: acoustic, structural, lexical. Subjects may focus on an isolated factor, or on many; they find it very difficult to separately rate the stress properties. The task can show how untrained subjects judge stress, and whether they agree. Given the continuous nature of 40ms, stress scores, the number of listeners who perceived a syllable as stressed [1,2], allows for a classification in min. 3 categories: stressed, unambiguous. Because of space limitations, previous studies on stress perception and stress cues can not be reviewed here.

2. Methods

Six extracts (277 syll.) were selected from a corpus [3] in such a way that the test contained at least 2 occurrences of each stressed tone. A male and a female speaker each provided 3 extracts. The mean length of 46 syll/test was suggested by [2] where it was found that the proportion of syllables judged stressed decreases as the length of the carrier sentence increases. For lengths above 40 syll, the ratings are similar to those for continuous speech. The passages were very different in terms of prosodic complexity (table 1), which can be defined in terms of (1) rate of speech (without pauses), (2) proportion of stressed syllables, (3) of emphatic stresses, (4) of pauses, (5) of glides, and (6) rhythm structure.

2.1. Perceptual experiment

The 20 untrained subjects heard each passage once (with 25s silence) and 6 times (with 6s intervals) during which they had to indicate the stressed syllables on the test sheet. Each syllable was judged either stressed or unstressed; so, it was assigned to one of 3 groups. The nominal scale calls for a non-parametric test: the kappa statistic [9] was used. (P(A), the proportion of agreements in that the raters agree, and P(E), the proportion of estimated chance agreement, are determined. The kappa coefficient is the ratio of P(A) to the maximum proportion of times that raters could agree, both corrected for chance agreement. A kappa 1 indicates complete agreement, a 0 indicates no agreement other than chance. Since only 2 categories are used here, chance agreement is high, and kappa rather low (table 2). The pooled data (P(A) = 0.7867, kappa = 0.38) show a moderate agreement among the raters, although significantly different from 0. The relation with prosodic complexity is obvious.

2.2. Acoustic measurements

For each syllable 5 primary attributes are obtained, using an interactive analysis program [3,5]: nucleus DURATION, PITCH peak, LOUDNESS peak, intra-syllabic GLIDE, PAUSE duration. The segmentation into syllabic nuclei [4] provides boundaries necessary for the parameter extraction and pitch contour stylization. PITCH is the peak and GLIDE the interval of the stylized contour, positive or negative according to slope. Pitch values are expressed in semitones (ST): the melodic (in mel) and harmonic (in ST) scales are almost identical in the FB-range of speech [10]. The results were hand-corrected where necessary.

The measurement of LOUDNESS [10,8] (in sound) accounts for frequency dependence, critical bands, frequency masking, level, but ignores the effect of syllable duration. Level values (dB SPL) for each critical band were obtained from the power spectrum (Stft/FFT, 3064 points, by summation of the components in the band range, and db-conversion). Prominence estimates were calculated for duration, pitch and loudness. Prominence is defined as the difference between the parameter value for a syllable and the parameter mean of the context, either left (L) or right (R), with length 1 and 2 syll., giving 4 relative values: resp. PL1, PL2, LR1 and LR2 for duration, PL1, PL2, PR1, PR2 for pitch, and LL1, LL2, LR1, and LR2 for loudness. This allows for a continuous scaling of prominence. A similar measure combining left and right contexts with length 1 was used in [7].
Table 2. Correlation between stress SCORE and parameters (above line) prominence measure (below line), for low word data.

<table>
<thead>
<tr>
<th>DURATION</th>
<th>PITCH</th>
<th>LOUDNESS</th>
<th>GLIDE</th>
<th>PAUSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>.477</td>
<td>.203</td>
<td>.199</td>
<td>.298</td>
<td>.206</td>
</tr>
</tbody>
</table>

DL1 DL2 DR1 DR2 PL1 PL2 PR1 PR2 LL1 LL2 LRI LL2 LL1 LL2 LRI GLIDE
| .48 | .46 | .41 | .47 | .44 | .45 | .46 | .21 | .30 | .43 | .26 |

Table 3. Mean values for 7 variables cross-tabulated with ranges for SCORE: N is the number of syllables in a group.

<table>
<thead>
<tr>
<th>RANGE</th>
<th>N</th>
<th>DUR</th>
<th>DL1</th>
<th>DR1</th>
<th>PL1</th>
<th>PL2</th>
<th>LRI</th>
<th>LL1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2</td>
<td>177</td>
<td>88</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0-3</td>
<td>144</td>
<td>72</td>
<td>26</td>
<td>19</td>
<td>-1.2</td>
<td>-0.5</td>
<td>-1.2</td>
<td>-1.0</td>
</tr>
<tr>
<td>4-5</td>
<td>50</td>
<td>198</td>
<td>46</td>
<td>25</td>
<td>1.6</td>
<td>1.7</td>
<td>2.1</td>
<td>1.6</td>
</tr>
<tr>
<td>6-10</td>
<td>33</td>
<td>156</td>
<td>89</td>
<td>77</td>
<td>2.8</td>
<td>4.8</td>
<td>2.8</td>
<td>2.0</td>
</tr>
</tbody>
</table>

(because of speaker's range, declination line, etc.) and so goes LOUDNESS. There are too few cases of glides and pauses to find a relation with SCORE. Although no clear linear relation was found, Pearson correlation coefficient r was used to estimate the amount of information that could be garnered from each variable (Table 2). r varies considerably from one passage to another: for DURATION, from .63 to .18. Test 4 (with high complexity) gives very poor correlation for all attributes and is to a large extent responsible for the low r in the pooled data.

DURATION is the only primary parameter with relatively high r; this can be explained by minimal syllabic duration, small variability for unstressed syllables and large for the stressed. The best prominences estimates are DL1 and DL2, indicating that syllables with high scores are generally longer than the preceding one(s). DL2 and DR2 give results close to DLI and DL1. LOUDNESS, LLI and especially LRI score quite good (r=.5) in some tests, but not on the average. Depending on the method used and the number of variables taken into account, multiple regression gives a correlation of .40 to .88 with the stress SCOREs.

Score stress can be used to classify the syllables in 3 groups: prominent, ambiguous, and prominent (Table 3), showing clear differences between groups. The choice of the ranges depended on the number of elements in each group.

Labeling according to the transcription by a phonetician gives a further classification (Table 4). Group means show that the stressed are twice as long as unstressed; they are prominent by duration (DL1,DL1) and, in the case of low stress, also by low prominence (LL1,LI1). PL1, PR1 and GLIDE reflect the tones used (H,L,HL,LH,L+). The values for emphatic stress are very close to those for unstressed syllables. The parameters do not reflect the evident phonatory effort of emphatic stress. Predictions by the intonation model are observed in the data: (1) syllables with extra-low tone (L-) can be short and weak because their stressed status is already indicated by tone level, (2) glides (HL,LH) lack loudness prominence because stress is already signaled by the glide.

4. Conclusion
A listening task provided ratings of perceptual prominence for 277 syllables. The relative agreement between the raters indicates the perceptual reality of prominence. The importance of acoustic parameters as well as of four prominence measures were studied. The stress scores by the listeners are best predicted by a model of relative prominence compared to the preceding 1 or 2 syllables, and by the total duration itself. When the transcription of intonation by a phonetician is used for syllable classification, the same order of importance for the studied parameters is found. Predictions by the intonation model on the relative importance of individual prosodic parameters depending on the tone used, are confirmed by the data.

References:
WORD STRESS IN GEORGIAN

P. McCoy

University of California, Los Angeles, USA

ABSTRACT

Word stress in Modern Georgian, the language spoken in the Soviet Republic of Georgia in the USSR, is known to be weak in nature, in fact it is not certain that there is stress in Georgian. An experiment was conducted to test words in isolation, in phrases and in complete texts to see if there were any common denominators. The parameters examined here were pitch and duration. For each phonological word, FO measurements found a single peak for the whole word; correlations between words in text and in isolation were fairly consistent, though not uniform. Greatest duration fell either in the syllable with the FO peak or in the initial syllable. The results indicate that although stress in Georgian is weak, it is clearly a word level phenomenon.

1. INTRODUCTION

One of the more confusing questions for the student of Georgian is the placing of stress. Although this may be elementary for languages with fixed stress on some syllable of the phonological word, for example Czech, where it is always on the first syllable, or for languages with mobile stress where for the most part it has to be learned, the question of stress in Georgian is one that is almost avoided. Part of the reason for this lies in the fact that it is not certain that there is stress in Georgian. Or if so, there is no consensus as to its location. In all the varying opinions, the one thing that Georgian is agreement on one point, that stress is weakly dynamic and has melodic tone. Because it is weak, it tends to defy both description and analysis. This paper takes as its point of departure two questions of import for stress in Georgian: (1) Is there a word level stress (or is it phrase or sentence level)? (2) If there is, then how is it implemented in Georgian?

The structure of the paper will be as follows. I will first review literature on the subject as it is useful to be fully aware of the variety of views there are available in stress in Georgian. Having done this, I will proceed to look at some data from a study examining minimal pairs of words representing two environments -- within the flow of continuous speech, a read text, and the same words read in isolation.

2. LITERATURE

Starting with the most impressionistic, we have two descriptions: 1) "Die Betonung gleicht dem geöffneten Meer nach dem Sturm." [1: 2] "...wie nuschelndes Wasser lauft die georgische Rede hin." [2] These would imply that there is a significant lack of perceptual cues with which to identify stress. This may well be true at the most impressionistic level where not much attention is focussed on the physical aspects of perceptual cues, but at a slightly more concrete level, there do seem to be enough cues to generate varying opinions on the nature of stress in Georgian.

As regards duration, a common indicator for stress in language, some sources say that vowels are of equal duration throughout the word, irrespective of the length of the word. These sources seem to focus more on the melodic nature of Georgian. Others however indicate that in addition to the melodic structure, duration may have a place in determining the place of stress. Tschekheli [3] in his grammar indicates a function of duration and Japaridze, a Georgian phonetician who has done work on perception, comments on the perception of Czech by Georgians. He proposes an element of duration as Georgians hear Czech (a language which has constant phonetic stress on the initial syllable, phonemically long and short vowels) as stressed in syllables with long vowels.

Table 1 [3]-[10] gives a summary of Georgian stress as described by various views, proposed in grammars and articles. What is interesting is that in addition to the differences expressed among the works, there also seems to be a lack clarity within a given description.

3. PROCEDURE

The speaker was male speaker of the literary norm as judged by colleagues at Tbilisi State University. He pronounced the sets of words in isolation, and in paragraphs, presented to him in a random order. A practice session was conducted in order that he be familiar with the words and his task. The speaker was cautioned to read at a set pace and to observe a fixed distance from the microphone. Each word was read twice in each environment for a total of six tokens per word across environments.

Recordings were made in a sound booth using a Marantz Audiotape cassette FDM 350 and a super-scope condenser microphone. Broad and narrow band spectograms were made of the tokens on a Kay Digital spectograph. Fundamental frequency FO was measured by tracing the 10th harmonic in the narrow band spectrograms. Duration measurements were taken from the broad band spectrograms.

4. RESULTS

Correlations of duration and pitch for words in isolation and in a text were fairly consistent, though not uniform. The greater duration measurements fell either in the syllable with the FO peak or in the initial syllable. There was only one peak in a word and the peak, taking the duration of any word as a whole, seemed to be a third of the way into the word from the onset of the word.
From these results one can infer that stress, though weak is a word level phonomenon. Further, the acoustic triggers for stress would seem to be a combination of duration and rising F0.
STRESS CLASH AVOIDANCE IN DUTCH: INVERSION OF STRESS PATTERN IN COMPLEX NOUNS

Vincent J. van Heusden
Dept. Linguistics/Phonetics Laboratory, Leyden University, The Netherlands

ABSTRACT

We tested the phonetic basis of a recent claim made by metrical phonologists that the stress pattern of di-syllabic Dutch words with initial stress is inverted to final stress in order to avoid "stress clash" when such words are embedded (as the right-hand element) in a compound noun. In one experiment, speakers produced crucial words both as simplex nouns and embedded in compounds; listeners were then asked where they perceived the stress in the targets after these had been excised from their spoken context. In a second experiment we presented digitised versions of the crucial word types with systematically varied stress patterns; listeners had to rate the acceptability of the range of patterns in various rhythmic contexts. Results indicate that listeners perceive no stress shift in naturally produced word tokens, and that they always disallow versions of such words with inverted stress patterns.

1. INTRODUCTION

Compound adjectives in Dutch and English, such as red hot, have final stress when used predicatively: the 'poker is red hot' (a single quote preceding a syllable marks strong stress). In attributive position, however, the final stress on these words is retracted: a 'red hot' poker. If the stress had not been retracted, the result would have been two strong stresses abutting one another, a situation called "stress clash": a red hot 'poker. It is generally claimed that an immediate succession of two strong stresses on the same prosodic level violates a basic rhythmic principle underly-ing languages such as Dutch and English. These languages have a strong preference for a so-called alternating stress pattern, i.e., a regular alternation of strong (stressed) and weak (unstressed) syllables. Native speakers of Dutch and English can easily be convinced that stress retraction occurs in compound adjectives. In the older literature we find numerous claims to the same effect (2,3,4). Moreover, laboratory experiments have shown that the inversion of stress pattern in Dutch compound adjectives is clearly audible and has robust acoustic correlates (1).

In the past few years Dutch phonologists have studied another class of rhythm stress adjustment phenomena, viz., the behavior of stress patterns in polyphonic nouns embedded in compounds (cf. 6,7). When a word like 'harnas' (armour), with lexical stress on the first syllable, is embedded in a compound noun, a situation of stress clash may arise as in 'borsteharnas' (breast armour). The authors concerned [6,7] claim that stress clash is resolved in these cases by inverting the stress pattern of the embedded word, yielding 'boretharnas', which would have the same stress pattern as 'keepsake' in English, (as shap's compass) of which the first syllable (compass) has lexical stress in final position. Moreover, stress pattern inversion is claimed to be applicable only when the embedded noun has initial stress on a closed syllable (a so-called non-branching rhyme). Therefore no stress adjustment is said to occur when the lexical stressed first syllable of the embedded noun is open as in, e.g., 'premie' (premium) 'jaarpremie' (annual premium)

Curiously enough, the older literature contains no allusions to this type of stress adjustment at all, and even since the claims were made, phoneticians have expressed their doubts whether these are indeed cases of stress adjustment. In the present study we tried to settle this issue in a series of experiments.

2. EXPERIMENT I: PERCEIVED STRESS IN NATURAL SPEECH

2.1. Method

The basic stimulus material consisted of three types of di-syl-la-ble Dutch nouns, each category filled with five exemplars:

1. initial stress on an open syllable ('premie-type').
2. initial stress on a closed syllable ('harnas-type').
3. final stress ('kom'pas-type').

These 15 words were used as simplex words as well as embedded word-finally in tri-syllabic compound nouns, e.g., 'jaarpremie', 'boretharnas', and 'kepsakkom'.

The recordings of 30 words were recorded four times onto audio tape by two male speakers of Dutch, who pronounced the target words twice in a fixed carrier phrase 'Heb Jij een [TARGET] ontdek?' (Have you a [TARGET] discovered?) with accent on the target and two more times in the fixed phrase 'Jij een [TARGET] ontdek?' (with a contrasting accent on Jij).

The 120 di-syllabic target word tokens were excised from their spoken contexts using a digital wave form editor, and presented twice in different random orders to 18 Dutch listeners. These were asked for each stimulus word to indicate along a scale from -5 to +5 what stress pattern they perceived. In this meaningful that the stress levels of the two syllables were exactly equal. "-5" was to be chosen if the subject felt that the lexical stress was much less stressed than the final syllable, "+5" had to be responded when the subject perceived much more stress on the initial syllable than on the final syllable. Intermediate values stood for less extreme differences in the distribution of stress over the two syllables.

2.2. Results and conclusions

Table 1 contains the results.

Table 1: Mean perceived stress distribution (see text) broken down by accentness of target, type of word (simplex vs. embedded in compound), and lexical stress type (each mean is based on 36 judgments nominally).

<table>
<thead>
<tr>
<th></th>
<th>targeted</th>
<th>accented</th>
<th>unaccented</th>
</tr>
</thead>
<tbody>
<tr>
<td>'premie'</td>
<td>3.7</td>
<td>1.6</td>
<td>1.9</td>
</tr>
<tr>
<td>'harnas'</td>
<td>3.7</td>
<td>1.7</td>
<td>1.5</td>
</tr>
<tr>
<td>'kom'pas'</td>
<td>-2.5</td>
<td>0.0</td>
<td>-0.5</td>
</tr>
</tbody>
</table>

The perceived stress distribution clearly differs for words with initial stress ('premie-type and 'harnas-type') and those with final stress ('kom'pas-type), F(3,4194) = 957.1 p<.001. The difference between initial stress and final stress is larger for simplex words than for the same words incorporated in a compound. F(1,415) = 55.8, p<.001 (this corresponds to the difference between primary versus secondary stress on the word level). The stress patterns are perceived as more extreme in accented simplex words than elsewhere. Crucially, however, none of the differences between the 'premie-type and the 'harnas-type are ever significant. But these two types always differ significantly from the 'kom'pas-type (Scheffé procedure p<.01).

So far, these results do not support the claims made by metric-
al phonologists, who predicted that the stress pattern of harnas would resemble that of 'premie in simplex words but that of kom'pas in compounds. One may argue somewhat peremptorily, however, that our speakers may have behaved a-typically, and that an (even) more proficient speaker would have displayed the predicted stress shift after all. In order to resolve this possibility we ran a control experiment with an ideal (synthetic) speaker who produced the desired stress shifts exactly the way we wanted. 

3. EXPERIMENT II: PREFERRED STRESS PATTERN IN SYNTHETIC SPEECH

3.1. Method

The lexical material underlying the stimuli were three word pairs:

(jaar)premie: initial stress, 1st syll. open (boret)harnas: initial stress, 2nd syll. closed (scheep)kom'pas: final stress

These words were embedded in final position in compounds; the resulting set of eight words were then synthesized from diphones (using the P303 diphone set; for details cf. van Rijnsoever, 1988) in the same two carrier phrases (i.e., once with and once without accent on the target) that were used in experiment I. Each utterance was given the same pitch pattern with appropriate accentuation and a 6 semitone rise-fall on the accent syllable. The duration of final two syllables in the targets was systematically varied in five steps, so as to create a continuum from stress on the penultimate syllable, via level stress, to stress on the final syllable (note that 80% of the original recording speed is the standard synthesis output rate).

<table>
<thead>
<tr>
<th>penultimate</th>
<th>rising</th>
<th>48%</th>
<th>88%</th>
<th>96%</th>
<th>falling</th>
<th>112%</th>
<th>48%</th>
</tr>
</thead>
<tbody>
<tr>
<td>level</td>
<td>80%</td>
<td></td>
<td></td>
<td></td>
<td>80%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>falling</td>
<td>112%</td>
<td></td>
<td></td>
<td></td>
<td>48%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The resulting set of 3 lexical words: premie, harnas, kom'pas (yes/no accent on target) × 5 (temporal stress patterns) = 60 stimulus types were presented to 19 native Dutch listeners in two different random orders, who had to indicate the acceptability of each item along a scale from 0 (unacceptable stress pattern) to 7 (completely acceptable stress pattern).

3.2. Results

From the acceptability scores of the five temporally different versions of a stimulus type we derived its preferred stress pattern for each individual listener. To this effect we devised an index such that negative values indicate stronger preference for initial stress (i.e., a relatively long first syllable), and positive values stronger preference for final stress (i.e., a relatively long second syllable); an index of 0 indicates that perfectly even stress is preferred. Table II summarizes the results.

<table>
<thead>
<tr>
<th>Table II: Mean preferred stress pattern broken down by accent type (yes/no accent on target), word type (simplex vs. embedded in compound), and lexical stress type.</th>
</tr>
</thead>
<tbody>
<tr>
<td>target</td>
</tr>
<tr>
<td>harnas</td>
</tr>
<tr>
<td>kom'pas</td>
</tr>
</tbody>
</table>

We notice that the effects are stronger for unaccented than for accented targets. Words with initial lexical stress are always towards the negative end of the scale, while words with final stress appear at the positive end of the scale. When the simplex words are embedded into compounds, there is a general preference for a stronger (i.e., more stressed, longer) final syllable. This effect is especially clear when the targets are accented, and somewhat instable for unaccented targets. Crucially, however, there is not the slightest preference for stronger final stress when 'harnas' is embedded. 2. Simple rhythmic inversion was predicted there. Moreover, counter to the linguists' prediction, there is no systematic difference between 'premie and 'harnas'.

4. CONCLUSION AND DISCUSSION

Our experiments have failed to support the predictions of metric-al phonology to the effect that embedding an initially stressed word in a compound noun would lead to an inversion of stress pattern. The same holds for the temporal organisation associated with it, of an embedded noun with initial stress remains completely distinct from the stress pattern of an embedded word with final stress. We therefore take the view that these phonological predictions are wrong, and suggest that the principle of stress clash avoidance be restricted to the class of compound adjectives (the "stress retraction"-cases in §1). Notice that compound adjectives receive their stress pattern through the phrasal stress rule, i.e., by a process that is intrinsically above the level of the word. Apparently, there is no stress clash when two lexical stresses become adjacent in a compound noun, i.e., no stress clash is felt at the word level. The duration of the first syllable in any di-syllabic word gets relatively shorter if this word is the final element of a compound (cf. table II). Three general (non-language-specific) low-level duration rules account for this phenomenon: (i) A syllable with main stress is longer than other syllables. When a word is embedded in a compound it loses its main stress, i.e., the lexically stressed syllable loses its pitch movement if shortened. (ii) Longer words are spoken faster than shorter words, therefore the syllables of the di-syllabic words will be shortened when they are embedded in a longer compound. (iii) A word-final syllable is lengthened so as to mark off the word (final lengthening). Since the result, a short syllable at the onset of the embedded word, is compatible with the desired stress pattern of kom'pas, the shortening is not picked up for this type of word. When a long, open initial syllable is shortened (as in 'premie'), the decrement in duration will be too small to reach the listener's awareness. But if a short, closed syllable is shortened by the same amount, the effect may be above threshold and the linguist will be tempted to interpret this as a shift in stress.

We take the view, of course, that the effects of such low-level duration rules should not be mistaken for stress effects; or else we would have to interpret the same effect as a stress shift in one case ('harnas') and as a sub-liminal duration shift in others ('premie, kom'pas').

NOTE

Experiments 1 and 2 were run by my students Ellen L. Bish and Ruben van de Vijver, respectively.

5. REFERENCES