SOME OBSERVATIONS ON THE TEMPORAL ORGANISATION AND RHYTHM OF SPEECH

S.G. Nooteboom

**Research Institute for Language and Speech,
Utrecht University.**

ABSTRACT

This paper pleads for quantitative models incorporating many interacting factors controlling the temporal organization of speech, and for testing such models in statistical studies. The paper also warns that such studies tend to obscure real regularities and should be supplemented with classical laboratory experiments.

Isochrony and stress groups are rejected as useful notions, words are proposed as important units. Evidence is shown that within words there is rhythmical alternation of unstressed syllable durations, and that vowel shortening due to change of tempo not necessarily leads to vowel reduction. There is an urgent need for studying the acoustic/phonetic characteristics that distinguish spontaneous from prepared speech.

1. INTRODUCTION

Speech is a slippery phenomenon. Physically speaking, at each moment in time the sound of speech is nothing more than a momentary air pressure perturbation. One moment it is there, the next moment it is gone. The sound of speech has only extension in time as far as we, in our role of listeners, can hold it in memory, or in as far as we, in our role of researchers, can transform it into oscillograms or spectrograms where time is transformed into space. In such registrations we observe rapid discontinuities in intensity and spectral structure, delineating fragments where changes seem to be less rapid.

Such changes or discontinuities in the sound of speech are caused by movements of the sound generating vocal organs. They delineate fragments of speech that can be associated with vowels and consonants realized by the speaker, and perceived by the listener.

Because such fragments with measurable durations can be associated with consonant and vowel realizations, we can observe that realizations of one and the same phoneme can vary tremendously in duration. Durations of realizations of one and the same vowel phoneme may vary from practically zero to many hundreds of ms.

Such variation is not random, but rather rigorously controlled by many factors and their interactions. The result is what we call the temporal organization or temporal patterning of speech.

It is a major task of phonetic research to account for temporal patterns functioning in speech communication. This is not an easy task. Part of the complexity of the problem stems from the fact that there are so many factors involved, on different levels of speech production, and that these factors often strongly interact.

In this presentation I intend to present some opinions, observations, and experimental results that may help to further the ongoing discussion of some, mainly prosodic, aspects of the temporal organization and rhythm of speech.

2. QUANTITATIVE MODELS AND THEIR LIMITATIONS

Let me begin with the following statement:

(1) The systematic effects on speech sound durations of any one particular factor can only reliably be assessed when we take the effects of many other factors into account.

This point is illustrated in Fig. 1, containing some nearly twenty years old data of mine [31]. Here we see the effect of compensatory shortening of the lexically stressed vowel of a word as a function of the number of following unstressed syllables in the word. The data show that this effect strongly interacts with vowel identity, postvocalic consonant and tempo. Interactions are found both in the absolute and in the relative durations.

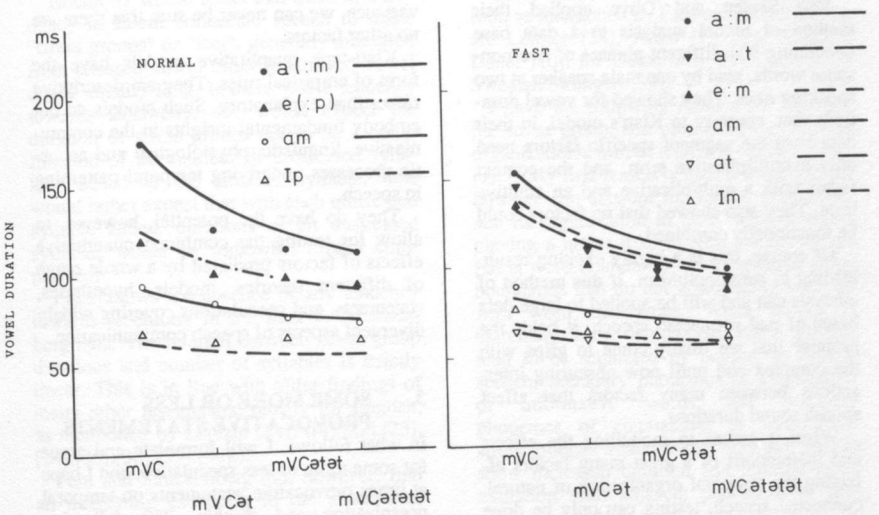

Fig.1. Vowel duration in the initial stressed syllable as a function of number of following unstressed syllables in the word. The parameter is syllable rhyme. Left: normal speech rate, right: fast speech rate (data from Nooteboom, [31]).

Now this is only one example of a great many such interactions that can be demonstrated on the basis of available data in the literature. The existence of such strong interactions can largely explain seemingly contradictory findings by different researchers. An effect that is strong in one speech tempo or one position, may virtually vanish in another speech tempo or another position. As long as we do not take into account such quantitative interactions between different factors, we will not know where to expect an effect of any one particular factor, and how big this effect will be. This remains true despite a recent and interesting demonstration that interactions become less strong, and that some interactions perhaps even vanish, when durational variations of phoneme segments are described in terms of the durational variance of the phoneme type concerned [5].

The upshot of this is that we can only reliably assess the effect of any one particular factor when we take the interactions of this factor with many other factors into account. This leads to my second statement:

(2) There is a real need for quantitative models accounting for speech sound durations, and ways of testing these models.

Interactions of the type demonstrated can be modeled by equations combining additive terms with multiplicative terms. A well known example is the empirical rule proposed by Klatt [21], which in its simple form can be written as:

$$\mathbf{DUR = k(D_{inh} - D_{min}) + D_{min}}$$

in which **DUR** is the segment duration to be calculated, **k** is a parameter describing a context effect, or any combination of such parameters, $\mathbf{D_{inh}}$ is a table value standing for the segment specific inherent duration, and $\mathbf{D_{min}}$ is a table value standing for the segment specific minimal duration.

In this model context parameters provide a multiplicative term, and segment specific parameters provide additive terms. Furthermore, context parameters are functionally combined, under the implicit assumption that the order of the joint effects of these parameters is unaffected by other factors.

Klatt's model was until recently never rigorously tested. It is the merit of Van Santen and Olive [43], that they show how to generalize models of this type mathematically, and how such models can be tested by analyzing the covariances between subarrays of a multifactorial data matrix.

Van Santen and Olive applied their method of model analysis to a data base containing 304 different phrases of two nonsense words, read by one male speaker at two speaking rates. They showed for vowel durations that, contrary to Klatt's model, in their data base the segment specific factors need only a multiplicative term, and the context factor both a multiplicative and an additive term. They also showed that no factors could be functionally combined.

Of course, this is a highly exciting result, leading to some optimism. If this method of analysis can and will be applied to large data bases of real connected speech, it holds the promise that we finally come to grips with the complex and until now obscuring interactions between many factors that affect speech sound durations.

When it comes to modelling the effects and interactions of a great many factors affecting the temporal organization of natural, connected speech, testing can only be done on data obtained in statistical studies based on extensive corpuses of connected speech. There are already a number of such studies available in the current literature. Examples are Barnwell [2], Harris and Umeda [19], Umeda [41], Crystal and House [6],[7],[8], Fant and Kruckenberg [14],[15], Fant, Nord and Kruckenberg [16],[17], and Van Santen [42].

Tuning quantitative models to such data bases has not yet been done. It will be exciting to watch the outcome of such an enterprise, and see how far research tools as provided by Van Santen and Olive will bring us. We should be aware of the fact, however, that going back and forth between quantitative models of this kind and statistical data bases has its limitations as a research tool, among other things, for the following reason:

(3) Statistical studies on corpuses of connected speech obscure real regularities: there remains a need for testing specific ideas with well controlled materials in laboratory experiments.

The point is, of course, that factors we have never thought of will not show up in such studies, except in their contribution to the remaining variance. If such factors have big effects, but are not very frequent in the data base, this contribution will be only marginal. Even if the factors investigated account for a high percentage of the overall variance, we can never be sure that there are no other factors.

Klatt-type quantitative models have the form of empirical rules. They are descriptive rather than explanatory. Such models do not embody fundamental insights in the communicative, linguistic, physiological and acoustic processes underlying temporal patterning in speech.

They do have the potential, however, to allow for testing the combined quantitative effects of factors predicted by a whole range of different theories, models, hypotheses, statements, and speculations covering widely divergent aspects of speech communication.

3. SOME MORE OR LESS PROVOCATIVE STATEMENTS

In what follows, I will formulate and argue for some more or less speculative, and I hope at times provocative, statements on temporal organization and rhythm. The following statement concerns the age honoured question of isochrony, and is perhaps less provocative now than it would have been twenty years ago:

(4) There is no tendency towards isochrony in speech production.

The absence of isochrony is illustrated by data for Swedish read aloud text, taken from Fant and Kruckenberg [15] and presented in Fig.2.

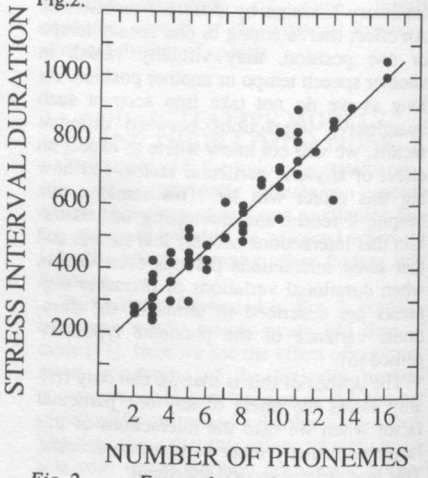

Fig. 2. Free foot duration versus number of phonemes per foot (data from Fant and Kruckenberg, [15]).

Isochrony would mean that there is a tendency in fluent connected speech to make "stress groups" or "feet", generally measured from stressed vowel onset to stressed vowel onset, equally long. If there was a tendency towards isochrony, one would expect the duration of stress groups not to be a linear function of the number of unstressed syllables added to the stressed syllable. One would rather expect that with each unstressed syllable added, durations of all unstressed syllables would be somewhat further compressed.

Fig.2 shows the absence of any tendency towards isochrony in the Fant & Kruckenberg data. The relation between stress group durations and number of syllables is strictly linear. This is in line with older findings of many other researchers, mostly for English, as mentioned by Lehiste [24]: ([4],[22],[23], [34],[35],[36],[39],[40],).

Fant and Kruckenberg did, however, find an interesting exception to the absence of isochrony. They were able to show that the duration of a stress group containing a speech pause is predictable from the number of phonemes in this stress group plus the duration of an average embedded stress group. Apparently, the moment in time a speaker continues after pausing seems to be determined by some rhythmical measure derived from the average time interval between stressed vowel onsets in the preceding stretch of speech.

Of course there remain some questions here. How important is the effect to speech perception and how sensitive are listeners for deviations from predicted durations of pause containing stress groups?

Also one would like to know whether there is any other measure to be derived from the preceding stretch of speech, from which pause duration could be predicted equally well. It is likely, for example, that average word duration is closely correlated with average stress group duration. This leads up to my following statement:

(5) Words are important units for the temporal organization of speech, stress groups are not.

This statement runs counter, for example, to the following statement by Fant and Kruckenberg: "The stress group (....) is a major constituent of durational structure. As an organizational unit of connected speech it overrides the word".

Other proponents of the stress group or foot, as mentioned by Fant and Kruckenberg, are Lehiste [24], Allen [1], Lea [22],[23], Dauer [10], for Dutch Den Os [11], and for Swedish Strangert [38].

I have the following reasons for rejecting the stress group, and promoting the word as organizational unit of temporal organization:

(a) My first reason is this: it is hard to see how we can account for speech production, and its organization in time, without words playing a major role. Speech pauses always fall at word boundaries, never at stress group boundaries that do not accidentally coincide with word boundaries. When speech is extremely slow, words, not stress groups, tend to be separated by pauses. In normal speech, boundary phonemes of emphasized or informative words, not boundary phonemes of emphasized or informative stress groups (whatever that may be), tend to show increased duration and reduced coarticulation with adjacent phonemes of surrounding words.

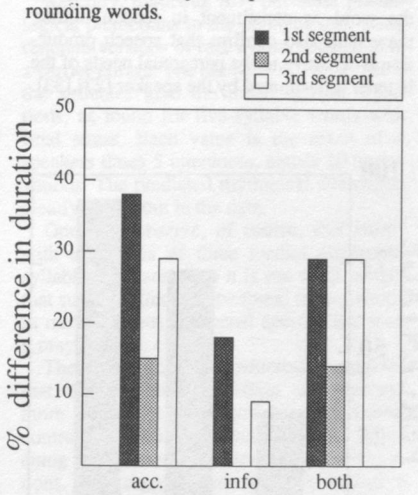

Fig. 3. Percent difference in duration between two realizations of C_1, V, and C_2 in embedded CVC words. Left: the effect of plus versus minus accent (100% is plus accent), middle: the effect of new versus given (100% is new), right: the effect of plus accent new versus minus accent given (100% is plus accent new).(Data after Eefting [13]).

An illustration of the last point is given in Fig.3, containing data from Eefting [13] for

Dutch. She, among other things, compared the temporal structure of within-sentence realizations of the same one-syllable CVC words in three different comparisons: Accented versus unaccented for informative words (i.e. words containing new information), informative versus not informative (new versus given) for unaccented words, and both accented and informative versus unaccented and not informative. The figure plots the relative differences in percent of the accented or informative values, for C_1, V, and C_2.

My interpretation of these data is that the temporal structure of these words is affected by two factors, accent and informativeness. Accent leads to increased vowel duration and some increase in prevocalic and postvocalic consonant durations, informativeness leads to increased durations of word boundary segments, in a tendency to disconnect the word somewhat from its preceding and following context. Apart from providing evidence for the word as constituent in speech timing, these data also confirm that speech production is sensitive to the perceptual needs of the listener as estimated by the speaker [27],[33].

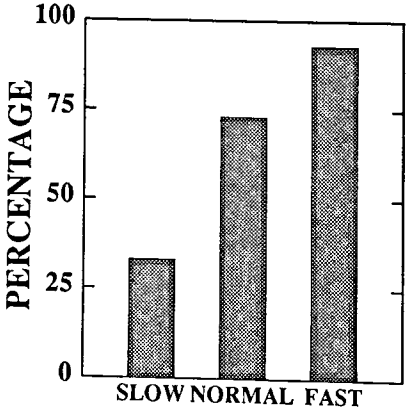

Fig. 4. Percent cases in which the clusters /p#b/ and /t#d/, containing a word boundary, are perceived as /b/ and /d/ respectively, due to assimilation and degemination, as a function of tempo. The deletion of /t/ or /d/ in each case gave an existing alternative word before the boundary (data from Menert, [24]).

Fig.4 gives another illustration of the same point. The figure shows data from Menert [29] who experimentally studied the frequency of perceptual ambiguity resulting from assimilation and degemination in /td/ and /pb/ clusters across word boundaries as a function of speech rate. Word boundaries coincided with potential phonological phrase boundaries. Such data reflect at the same time the relevance of phonological phrases and the relevance of words as constituents of timing in speech. Similar effects for stress groups have yet to be shown.

(b) Secondly, Beckman and Edwards [3], in a carefully controlled experiment examining relations between vowel duration and prosodic constituency, found that there are two different prosodic boundary effects, phrase-final lengthening and word-final lengthening. The word-final effect could not be explained in terms of isochronous intervals of some sort.

(c) Thirdly, Van Santen [42], in his earlier mentioned statistical study on a data base derived from 2,262 sentences read aloud by a single male speaker, finds considerable effects on durations of both stressed and unstressed vowels of the number of syllables and position of the stressed syllable for words, but fails to find similar effects for stress groups. This study is exceptional, because the author explicitly compares words and stress groups as potential constituents of temporal structure. In most other comparable studies, either words or stress groups are chosen as units and we can not judge which of the two explains most of the variance in the data, or whether one of the two would be superfluous.

(d) My fourth and last reason for rejecting the stress group is one of economy. We should not introduce more units than necessary to account for our data. The question is of course, whether there are durational data concerning normal fluent speech that cannot be explained without recourse to stress groups. Perhaps there are. But I know of no publications where that is convincingly shown. As long as that is the case, we should hesitate to accept the stress group as a necessary control factor of the temporal organization of speech.

Let me add two remarks here. One is that the fact that some phenomena can, pragmatically, be easily described in terms of stress groups is not sufficient evidence that stress

groups are part of the mental control of durational variation in speech. The other is that I do not wish to deny that some phenomena in speech appear to be controlled by a rhythmic principle, as exemplified by the rhythm rule in accent structures. There may be similar phenomena in durational control. But these can be accounted for without recourse to the beginnings and endings of stress groups. In this sense stress groups are dispensable, words are not.

Consequently, my next statement is concerned with the temporal pattern of words, and runs as follows:

(6) Within-word sequences of medial unstressed syllables follow a pattern of rhythmic alternation,'short'- 'long'-'short, etc.

Syllable duration depends among many other things on lexical stress, and on position in the word. There is an interesting difference between observations by phoneticians, and predictions made by word level phonologists. This difference particularly concerns temporal patterns in sequences of unstressed syllables. An example in case is shown in fig. 5, showing the abstract rhythmical pattern of a five-syllable Dutch word, with final stress. The example is taken from Kager [20].

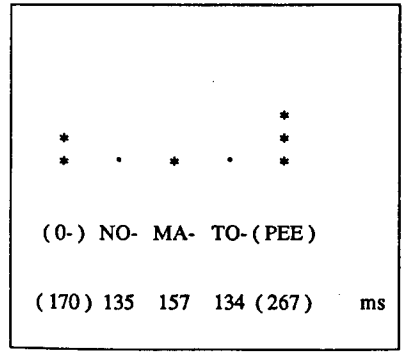

Fig. 5. Stress pattern of the Dutch word **onomatopee** *according to Kager [20], and vowel durations as measured in reiterant versions of such words by Slootweg [37].*

The number of stars above each vowel represent the relative prominence level of the syllable. Kager predicts that there is a hierarchy of reducibility, in which syllables with the lowest stress level, having no stars, are most easily reducible to schwa, etc. This agrees with intuitions about reducibility. He also predicts that in nonreduced realizations, the syllable durations follow an underlying pattern represented in the columns of stars. Similar predictions are made for English.

The pattern shown correctly predicts of course, that the lexically stressed syllable is most prominent, and has the longest syllable duration. Also the prediction that the unstressed initial syllable is somewhat longer than medial unstressed syllables is in agreement with phonetic measurements and earlier proposed empirical rules [32]. The rhythmical alternation, however, in the sequence of three unstressed medial syllables, following a 'short'-'long'-'short' pattern, is not part of commonly proposed empirical rules for temporal patterning, such as proposed for Dutch by Nooteboom [32] for Swedish by Lindblom and Rapp [25], and for English by Klatt [21].

Recently, Slootweg [37] put these phonological predictions to the phonetic test, using reiterant versions of real words, embedded in a carrier phrase. The numbers in Fig. 5 below the syllables give the mean syllable durations, as found for five-syllable words with final stress. Each value is the mean of 4 speakers times 5 utterances, equals 20 observations. The predicted rhythmical alternation clearly stands out in the data.

One may observe, of course, that words with sequences of three medial unstressed syllables are rare. Also it is not at all certain that such an effect can be found in real words in normal fluent connected speech. But then it may.

The connection with reducibility suggests that this and similar effects will become more prominent as speech becomes faster, contrary for example to compensatory shortening of stressed syllables. These observations also suggest that there is indeed a rhythmical component to speech production, albeit different from the kind that would lead to isochrony of stress groups. Note also that this kind of rhythmical pattern has no recourse to units that are orthogonal to linguistic structure, such as stress groups. The relation between vowel duration and spectral vowel reduction is the topic of my next statement:

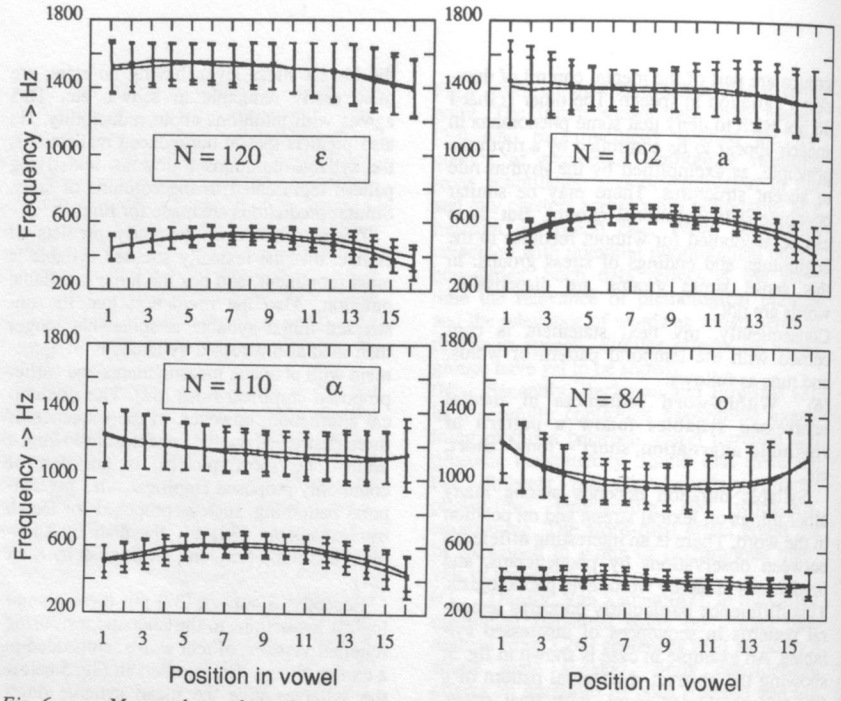

Fig. 6. Mean value and standard deviation for first (F_1) and second (F_2) formant frequency of fast and normal rate vowels. Vowels as indicated (data from Van Son and Pols [44]).

(7) Other things being equal, vowel shortening due to a higher speech tempo, does not lead to vowel reduction.

More than a quarter of a century ago, Lindblom [26] proposed the "Target-Undershoot" model, predicting that, when durations of a certain vowel get shorter and shorter, the articulatory movement keeps farther away from the vowel target, and there will be increasingly more spectral reduction. Essential to this model is that the articulatory movements do not move faster when vowel duration is decreased.

Although Lindblom confirmed his prediction with some simple nonsense utterances, ever since attempts to find further confirmation have led to controversial results [45]. Of course, in testing the model, it is important that speed of articulatory movement and target value are not different due to other factors than vowel duration alone.

Some recent results on this issue are exemplified in Fig. 6, taken from Van Son and Pols [44]. They showed in an acoustic study that if everything is kept equal except speaking rate, the speed of articulatory movement is adapted to vowel duration and there is no spectral reduction.

Their data were obtained from two versions of the same text, once read at a normal speaking rate and once at a fast speaking rate by a single highly experienced professional speaker. They paired normal and fast realizations of the same vowels in the text only when prosodic conditions were the same. Fig. 6 shows average tracks of the first two formants each time of the same vowel, sampled every millisecond, after normalization of vowel duration. Obviously, there is no vowel reduction at all. And if after normalization of vowel duration the formant movements are identical, then before normalization they were of course different, being slower in slow speech and faster in fast speech.

Of course we should be aware of the fact that in cases where durational differences depend on differences in stress level, or on position, the relation between duration and articulatory movements reflected in speed of formant changes can be quite different. Macchi, Spiegel, and Wallace [28] convincingly

showed that the effect of position in the word on vowel duration leaves formant transitions virtually intact, and lengthens and shortens specifically the steady-state portions of vowels. It is yet to be assessed how important such phenomena are for perception. A recent study by Drullman and Collier [12] showed that using reduced instead of full diphones in appropriate positions in synthetic Dutch speech, does not improve speech quality as soon as segment durations are optimized.

Admittedly, the data of Van Son and Pols shown in Fig.6 are based on text read aloud by a highly competent professional speaker. We do not know whether perhaps other speakers do behave according to Lindblom's predictions, or whether perhaps vowel reduction as a function of speech tempo does not occur in prepared speech but does in spontaneous speech.

Unfortunately, this is true of practically all effects discussed. We know preciously little about temporal organization and rhythm of spontaneous speech. It is a source of constant amazement to me that when I turn on the radio or television, and hear someone speak, I seem to need only a few syllables to determine whether I listen to spontaneous speech or prepared speech. The acoustic-phonetic correlates of this difference are unknown. There is room here for the following plea:

(8) There is an urgent need for studying the acoustic/phonetic characteristics that distinguish spontaneous speech from prepared speech.

4. CONCLUSION

I conclude this presentation with the following remarks. Somewhat oversimplifying the situation, we can say that research on temporal organization and rhythm in speech is either descriptive, or directed towards understanding the mechanisms underlying observable timing in speech.

The descriptive type of research seems to hold the promise that we may account for the combined effects of many factors of widely different origins in connected prepared speech. Such an account of necessity will have the form of empirical rules that can be very useful for speech synthesis and perhaps for speech recognition. But this approach will not tell us where the many effects and their interactions come from. It will not satisfy our scientific curiosity, nor will it lead to the detection of completely new phenomena, phenomena we have never thought of before.

In the research directed at the underlying mechanisms of timing in speech often only a single aspect or a few aspects of timing control are studied within the same theoretical framework. This may be unavoidable, given the highly complex and multifaceted nature of speech, but it is also unsatisfactory. There is, according to Osamu Fujimura, a need for "an integrated understanding of linguistic and behavioral as well as physiological and pathological processes involved in speech production" [18], and, I like to add, speech perception. Such integrated understanding will not come fast. But it will not come at all, if we do not make a conscious effort to bring together insights from different areas, and study how the predicted effects interact in actual speech production and perception.

The descriptive approach proposed by Van Santen and Olive [43] can then perhaps offer a thorough testing ground for our predictions.

5. REFERENCES

[1] Allen, G.D. (1975): "Speech rhythm: its relation to performance universals and articulatory timing", Journal of Phonetics 3, pp.75-86.

[2] Barnwell, T.P. (1971): "An algorithm for segment durations in a reading machine context", MIT-RLE Technical Report No 479.

[3] Beckman, M.E. and Edwards,J. (1987): "Lengthenings and shortenings and the nature of prosodic constituency", Paper presented at the First Laboratory Phonology Conference, Columbus, Ohio.

[4] Bolinger, D.L. (1965): "Pitch accent and sentence rhythm". In: Forms of English: Accent, Morpheme, Order, p.163ff. Cambridge Mass., Harvard Unive. Press.

[5] Campbell, N. (1990): "Evidence for a syllable-based model of soeech timing". Proceedings of the First International Congress on the Processing of Spoken Language, Acoustic Society of japan, Kobe, pp. 9-12.

6] Crystal, T.H. and House, A.S. (1982): "Segmental durations in connected-speech signals: Preliminary results".

Journal of the acoustical Society of America 72, 705-716.

[7] Crystal, T.H. and House, A.S.(1988a): "Segmental durations in connected-speech signals: Current results". Journal of the acoustical Society of America 83, 1553-1573.

[8] Crystal, T.H. and House, A.S.(1988b): "Segmental durations in connected-speech signals: Syllabic stress". Journal of the acoustical Society of America 83, 1574-1585.

[9] Crystal, T.H. and House, A.S. (1989): "Articulation rate and the duration of syllables and stress groups in connected speech", unpublished manuscript, Institute for Defense Analysis, Princeton.

[10] Dauer, R.M. (1983): ""Stress-timing and syllable-timing reanalyzed", Journal of Phonetics 11, pp.51-62.

[11] Den Os, E. (1988): "Rhythm and tempo of Dutch and Italian", Doctoral dissertation, Utrecht.

[12] Drullman, R. and Collier, R. (1990) "On the combined use of full and reduced diphones in speech synthesis", unpublished manuscript, Institute for Perception Research, Eindhoven.

[13] Eefting, W. (1990): "The effect of "information value" and "accentuation" on the duration of Dutch words, syllables, and segments", unpublished manuscript, accepted for publication. Institute for Language and Speech, Utrecht University.

[14] Fant, G. and Kruckenberg, A. (1988): "Some durational correlates of Swedish prosody", In: Proceedings of the Seventh FASE Symposium, Vol.2 (SPEECH '88), Edinburgh.

[15] Fant, G. and Kruckenberg, A. (1989): "Preliminaries to the study of Swedish prose reading style", Speech Transmission Laboratory, Quarterly Progress Report No 2/1989, pp. 1-83.

[16] Fant, G., Nord, L., and Kruckenberg, A. (1986): "Individual variations in text reading. A data-bank pilot study", Speech Transmission Laboratory, Quarterly progress 4/1986, pp. 1-7.

[17] Fant, G., Nord, L. and Kruckenberg, A. (1987): "Segmental and prosodic variabilities in connected speech. An applied data-bank study", Proceedings of XIth International Congress of Phonetic Sciences, Vol. 6, Estonian Academy of Sciences, Tallinn, USSR, pp. 102- 105.

[18] Fujimura, O. (1989): "Articulatory Perspectives of Speech Organization", to be published in the Proceedings of the Bonas Conference on Speech Production.

[19] Harris, M.S. and Umeda, N. (1974): "Effects of speaking mode on temporal factors in speech: vowel duration", Journal of the acoustical Society of America 56, 1016- 1018.

[20] Kager, R.W.J. (1989): "A metrical Theory of Stress and Destressing in English and Dutch", doctoral dissertation, Utrecht.

[21] Klatt, D.H. (1976): "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence", Journal of the acoustical Society of America 59,(5), 1208-1221.

[22] Lea, W.A. (1974) "Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding", Univac Report No. PX10791, St.Paul, Minnesota, Sperry Univac, DSD.

[23] Lea, W.A. (1980) "Prosodic aids to speech recognition", In: "Trends in Speech Recognition", Prentice Hall Inc, London.

[24] Lehiste, I. (1977): "Isochrony reconsidered", Journal of Phonetics, 5, 253-263.

[25] Lindblom, B. and Rapp, K. (1973): "Some temporal regularities in spoken Swedish", Papers from the Institute of Linguistics 21, Stockholm University, Stockholm.

[26] Lindblom, B. (1963) "Spectrographic study of vowel reduction", Journal of the acoustical Society of America 35, 1773-1781.

[27] Lindblom, B.E.F. (1989): "Phonetic invariance and the adaptive nature of speech", In: Working Models of human Perception, edited by B.A.G.Elsendoorn and H.Bouma, Academic Press, London, pp. 139-173.

[28] Macchi, M.J., Spiegel, M.F., and Wallace, K.L. (1989): "Using dynamic time warping duration rules for speech synthesis", unpublished manuscript, Bell Communications Research, Morristown.

[29] Menert, L. (1989): "Perceptual ambiguity as an indicator of voice assimilation", In: "OTS Yearbook 1989", edited by P. Coopmans, B. Schouten, and W. Zonneveld, Research Institute for Language and Speech, Utrecht University.

[30] Nooteboom S.G. and Cohen, A. (1975): "Anticipation in speech production and its implication for perception", In: "Structure and Process in Speech Perception", edited by A.Cohen and S.G.Nooteboom, Springer-Verlag, Berlin, pp.124-142.

[31] Nooteboom, S.G. (1972a): "The interaction of some intrasyllable and extrasyllable factors acting on syllable nucleus duration", Institute for Perception Research Annual Progress Report 7, 30-39.

[32] Nooteboom, S.G. (1972b): "Production and Perception of Vowel Duration, a Study of durational Properties of Vowels in Dutch", Utrecht doctoral dissertation.

[33] Nooteboom, S.G. (1985): "A functional view of prosodic timing in speech", In: "Time, Mind and Behavior", edited by J.A.Michon and J.L.Jackson, Springer-Verlag, Berlin, pp.242-252.

[34] O'Connor, J.D. (1965): "The perception of time intervals", Progress Report 2, Phonetics Laboratory, University College, London, 11-15.

[35] O'Connor, J.D. (1968): "The duration of the foot in relation to the number of component sound segments", Progress Report 3, Phonetics Laboratory, University College, London, 1-6.

[36] Shen, Y. and Peterson G.G. (1962): "Isochronism in English", Occasional Papers 9, Studies in Linguistics, University of Buffalo, 1-36.

[37] Slootweg, A. (1988): "Metrical prominence and syllable duration", In: "Linguistics in the Netherlands", edited by P. Coopmans and A. Hulk, Foris Publications, Dordrecht, pp. 139-148.

[38] Strangert, E. (1985): "Swedish Speech Rhythms in a Cross-Language Perspective", Almqvist and Wiksell Int., Stockholm.

[39] Uldall, E.T. (1971): "Isochronous stresses in R.P.", In: Form and Substance: Phonetic and linguistic papers presented to Eli Fischer-Jörgensen", edited by L.L. Hammerich, R. Jakobson, E. Zwirner, Akademisk Forlag, Copenhagen, pp. 205-210.

[40] Uldall, E.T. (1972): "Relative durations of syllables in two-syllable rhythmic feet in R.P. in connected speech", Work in Progress 5, Edinburgh University Department of Linguistis, 110-111.

[41] Umeda, N. (1975): "Vowel duration in English", Journal of the acoustical Society of America 58, 434-445.

[42] Van Santen, J.P.H. (1989): "Modeling contextual effects on vowel duration. I. Description of individual factors", unpublished manuscript, ATT Bell Telephone Laboratories, Murray Hill.

[43] Van Santen J.P.H. and Olive J.P. (1990): "The analysis of contextual effects on segmental duration", Computer, Speech and Language, 4, 359-390.

[44] Van Son R.J.J.H. and Pols L.C.W. (1989): "Comparing formant movements in fast and normal rate speech", In: "Eurospeech 89, the European Conference on Speech Communication and Technology", edited by J.P.Tubach and J.J.Mariani, CEP Consultants, Edinburgh, Vol 2., pp.665-668.

[45] Van Son, R.J.J.H. (1990): "Formant frequencies of Dutch vowels in a text, read at normal and fast rate", manuscript accepted for publication, Institute for Phonetic Sciences, University of Amsterdam.