PHONETICS LABORATORY
LUND UNIVERSITY

WORKING PAPERS

9·1973

PHONETICS SYMPOSIUM — Postgraduate students Stockholm - Lund

SIDNEY WOOD — Speech tempo

PHONETICS LABORATORY
LUND UNIVERSITY

WORKING PAPERS

9 · 1973

PHONETICS SYMPOSIUM	Postgraduate students Stockholm - Lund
SIDNEY WOOD	Speech tempo

Previous numbers and contents:

Correspondence concerning this series should be addressed to:

Robert Bannert
Phonetics Laboratory
Lund University
Kävlingevägen 20
S-222 40 LUND
Sweden

Phonetics Symposium 1973

Stockholm – Lund

On May 24–25 a group of postgraduate students from the Institute of Linguistics at the University of Stockholm visited our laboratory. The visit was combined with a symposium. The program and some of the contributions appear in this number of our Working Papers.

Thanking the organisers on behalf of the guests Björn Lindblom suggested that similar meetings should be held also in the future and include students of phonetics from the other Swedish universities as well.

PROGRAM

24 maj

Varje halvtimma beräknas omfatta både föredrag och diskussion.

Ordförande: 24 maj före lunch: E. Gårding
efter lunch: B. Lindblom
25 maj före lunch: K. Johansson
efter lunch: B. Brodda

On the whistle language of Gomera

Per Lindblad

This is a brief report on work in progress, carried out in collaboration with Jens Allwood, University of Gothenburg.

Gomera is one of the Canary Islands. The primary language of the Canary Islands is Spanish. On Gomera, a secondary whistle language is used by some persons in some circumstances, viz. by men only, mainly outdoors for communication over distances. The whistle used is of high intensity and has a very simple acoustic structure, which is not exposed to distortion over considerable distances as is ordinary speech - the whistle signal either reaches the addressee unmodified or does not reach him at all.

In December, 1972, we visited Gomera and made tape recordings of three informants speaking and whistling lists of minimal pairs, phrases given by us, and spontaneous phrases. Profile X-ray pictures were also made of one of the informants when he whistled some of the vowels.

The whistle may be produced in different ways. Two main methods of production can be distinguished. First, the whistler may form a small longitudinal groove in his tongue blade while pressing it against his upper teeth. Second, the whistler may put one or two fingers into his mouth, pressing it/them against his tongue while forming a narrow passage with his tongue in approximately the palatal position. There are many variants of this second type. Our present X-ray pictures are not sharp enough to give adequate information about conditions further back in the vocal tract. In both types, the lips and the front part of the tongue are constantly fixed, while the posterior part of the tongue and the articulators further back are free to vary the size and shape of the vocal tract.

The whistle signal is essentially a sine tone. Its intensity is not varied for phonemic distinctions. The information-bearing variations of the signal are onset and offset of whistle tone, and duration and frequency variations. Frequency is changed by varying the size and shape of the vocal tract. Classe (1957), who described the Gomera whistle language in the fifties, claimed that the whistler modifies the signal by approximating the articulatory position of ordinary speech, the fixation of the lips and the front part of the tongue hampering the normal execution of the commands, however. We agree that there is a close and natural relation between the Gomera speech and whistling, but propose a more perception-oriented view of the relation. There are strong arguments for this standpoint. One of these is that labial consonants are generally realized by a negative transition of the whistle tone. This feature corresponds of course to the negative F2 transition in ordinary speech, typical of labial pronounciation. This F2 transition is caused by labial articulation, whereas the negative whistle transition must be caused by some quite different articulatory movement, as the lips are always fixed during whistling.

Acoustic analysis of the recorded material is now in progress. It is based mainly on F_o curves of the whistles (See Fig. 1.) The recordings have been lowered eight times in speed in order to match the whistle tone frequency to the Fonema analysing equipment of the Phonetics Department in Lund.

Some important features of the secondary whistle language as compared to the primary Spanish will be mentioned here:

1. The whistle language does not have a code relation to the primary language but stands in a direct, natural relation to the Gomera Spanish. By this is meant that the whistler with his simple whistling instrument produces acoustic signals that share some essential features with the corresponding speech signal. The whistle mechanism produces a one-dimensional signal, lacking the rich redundancy

of speech signals. It may be supposed that the choice of distinguishing features in the whistle signal - with due regard to the limits set by the instrument itself - will illuminate essential perceptual cues of Gomera Spanish and of languages in general. Our study may thus contribute to the elucidation of the general phonetic problem of speech perception cues.

2. The whistle signal gives an auditory impression well corresponding to the syllabic structure of the same message as spoken. Roughly speaking, this is attained by realizing vowels as frequency levels or relatively slow frequency variations, whereas consonants are characterized by rapid variations of the frequency combined, for plosives and fricatives, with silent intervals.

3. Word and sentence stress as well as sentence intonation are all at least to some extent realized in the whistle signal. Being a one-dimensional signal, the whistle tone thus manifests segmental as well as suprasegmental features in the some dimension. We think that the study of the weighting of these different factors in the whistle signal may be of interest also in a more general phonetic context.

4. In the whistle segmental paradigm there is a far-gone reduction of oppositions compared to the Gomera Spanish system. Five vowel phonemes of Spanish are reduced to three whistle categories. Insteed of 17 consonant phonemes the whistle language uses perhaps only four categories. It is interesting to note that we have heard and seen whistlers communicate complicated messages rich in new information under controlled circumstances. Thus, the Gomera whistle language is a rich domain for research also from the information theory point of view.

Reference

Classe A. 1957. Phonetics of the Silbo Gomero, Archivum lingvisticum, Vol. 9, I

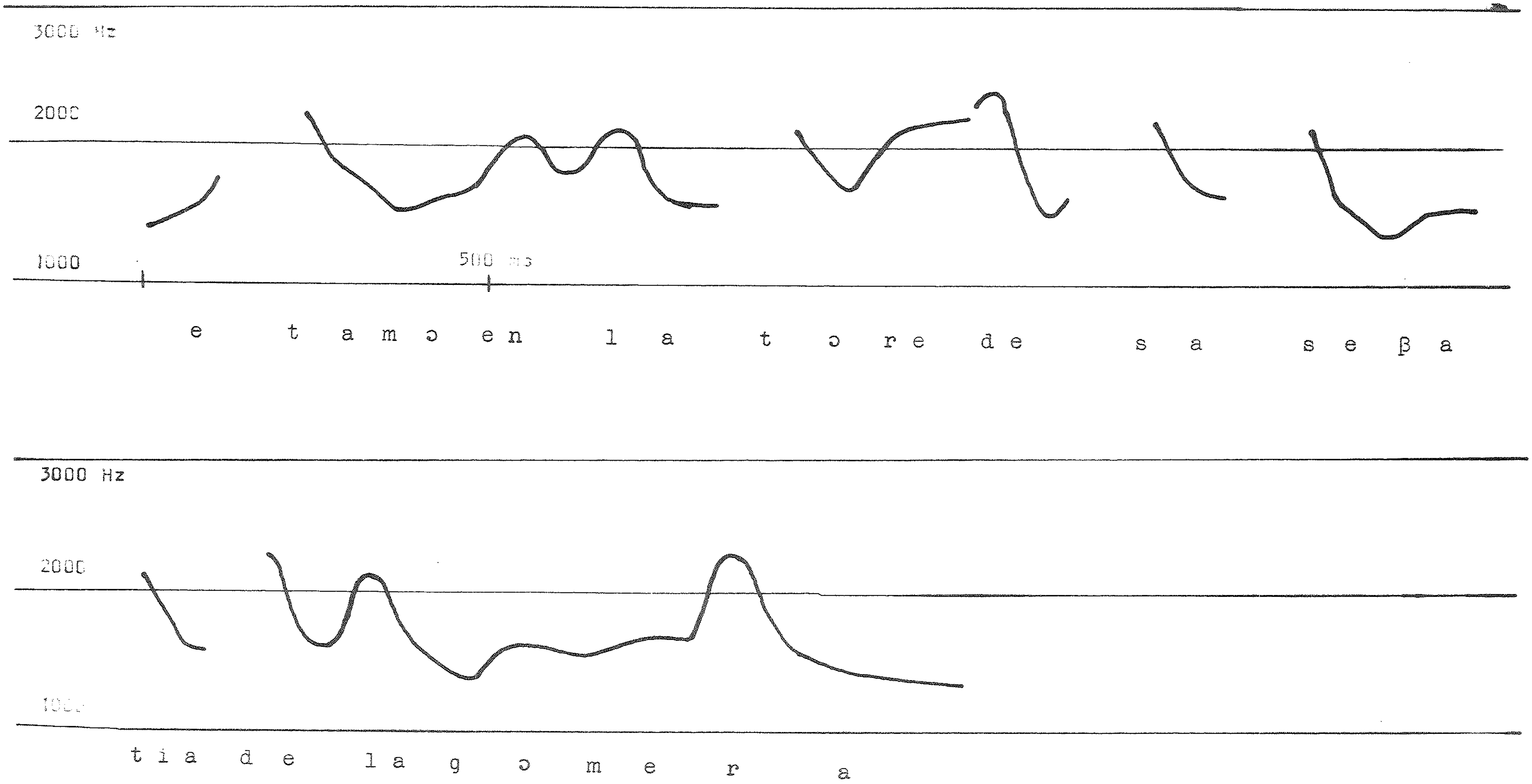


Figure 1. Tracing of the whistle tone from a sonagram. The informant whistled "Estamos en la torre de San Sebastian de la Gomera".

PERCEPTION OF THE FIRST FORMANT CONSIDERING THE HARMONIC STRUCTURE IN THE SPECTRUM

Rolf Carlson and Björn Granström

Perception of the first formant is normally thought of as an estimation of the peak in the trasfer function of the vocal tract. But how could a human estimate this peak when most times there is no energy at that special frequency because of the harmonic structure of the voice source?

Three different hypotheses about the perception of the first formant could be proposed.

a) The listener can reconstruct the peak of the envelope from the perceived harmonics irrespective if there is energy at the peak or not.

b) The listener computes the weighted mean of the excitation in different low frequency areas along the basilar membrane.

c) The listener selects the largest peak of the auditory excitation pattern. At medium and high pitch a single harmonic is picked out.

In an identification test using isolated vowels with different F1 and F0 values we searched for the phoneme-boundary between [i] and [e]. The result showed that hypothesis c) had to be rejected. We also used different weighting techniques in order to get a parameter representing the most important frequency (MIF) in the F1 domain. Only one of the methods tried could be accepted, taking the result of the identification test into account, namely the weighted mean of the two loudest harmonics in the loudness (sone) space. This method gave an MIF quite close to F1.

Reference:

Carlson R., Fant G., and Granström B. 1973."Two-formant models, pitch, and vowel perception", to be presented at the Symposium on Auditory Analysis and Perception of Speech, Leningrad, Aug. 1973.

WHAT HAPPENS TO VOWELS AND CONSONANTS WHEN WE SPEAK FASTER?

Sidney Wood

October 1973

SUMMARY

The subjects of this article are Kozhevnikov and Chistovitch's finding regarding the constancy of relative syllable durations and the inconstancy of relative speech sound durations for different rates of speech, and their proposal that vowels are elided because the necessary minimum duration of a consonant consumes all the time that happens to have been assigned to that syllable.

The relative consonantal durations of my seven informants (six languages) varied with speaking rate but mostly not as predicted by Kozhevnikov and Chistovitch. Instead of rising continuously at faster rates, the relative consonantal durations decreased again (Figs. 2 and 3). The variation was so small however, that a linear model, assuming a constant consonantal proportion, provides an excellent approximation to the results (Figs. 4, 5, and 6). A spot check on the syllables in two selected words in the German sample L revealed that the relative syllable durations were not less variable than the relative consonantal durations there (Fig. 9).

Several problems and difficulties related to speech reduction are discussed in general terms (§ 4.2). Are segments squeezed out when temporally constrained, or are they deliberately omitted? If segments or gestures are suppressed during production, does this occur peripherally or centrally? The view is expressed that most examples of segment synkope and syllable contraction in everyday speech are regular and habitual, and are not necessarily caused by increasing rate although their occurrence has the effect of accelerating the message.

1. INTRODUCTION

1.1. Aim

Kozhevnikov and Chistovitch (1965) found that "upon the increasing of the duration of a phrase, there is a clear lessening of the portion of the time occupied in it by the consonants" (p. 87).

The converse decrease of the relative vowel duration at faster rates was accompanied by the total elision of a vowel. In consequence of this, they went on to formulate a hypothesis of reduction, whereby a consonant must be executed in a necessary minimum duration, possibly leaving no time over for the vowel in that syllable. In their own words, the loss of a vowel occurs "when the interval between the syllables assigned in the programme is too small in order to accomplish both the closing and the opening of the organs which articulate the consonant in the case of a rapid rate of speech. As a result, there is simply not sufficient time for the vowel" (p. 89), and again "the changes of rate significantly change the relative durations of the consonant and vowel within a syllable. In the case of a rapid rate of speech the vowel can disappear completely (full reduction of the vowel) and in the case of a considerable slowing of the rate the duration of the consonant practically does not change and the prolongation of the syllable occurs at the expense of the vowel" (p. 89).

They had previously concluded that the variation of the relative duration of syllables was random. In view of the apparent inconstancy of phoneme relative durations, they decided in favour of the syllable as the basic programme unit of speech:

"If we examine the syntagma as a sequence of sounds of speech, we cannot find any constancy in its time figure. However, if we turn to the syllables and consider the syntagma as a sequence of syllables, its rhythmic figure is an invariant and completely independent of the rate of speech. From this conclusion naturally follows that in the programme of a syntagma the syllable commands are rhythmically controlled" (p. 89).

I wish to query the following two points:

Firstly, is it true that the relative durations of vowels and consonants vary with speech rate, and, if so, in the way described by Kozhevnikov and Chistovitch? This is important for their conclusion that syllables are the programme unit.

Secondly, are segments "squeezed out" for want of sufficient time for their execution? This is important for theories of reduction.

Nooteboom and Slis (1969) recognized that if Kozhevnikov and Chistovitch were right about the relative durations, it would be a "very important finding for the research on the control of timing and articulation". They therefore performed a similar experiment, but unlike Kozhevnikov and Chistovitch (who had aggregated the consonants and the vowels through the test sentence) they compared individual consonants and vowels in their respective positions in two nonsense words <u>mamamám</u> and <u>mamamám</u>. They were unable to confirm that segment relative durations were entirely correlated with speaking rate. At fast rates, they tended to remain constant (except in the internal weak syllable for which they accepted Kozhevnikov and Chistovitch's reduction hypothesis). At slow rates, the relative vowel durations did increase, more so in the long vowels <u>ɑ</u> and less so in the short vowels <u>a</u>. There was also a difference between the subjects. They conclude by suggesting that "a slow speaking rate distorts in some way or other the rhythmic figure of a word in the brain of the speaker".

Gaitenby (1965) found that the relative durations of words, syllables and phonemes remained constant at different rates. However, the rate differences were those between the normal speech of "fast", "medium" and "slow" speakers. This does not exclude the possibility that there are differences when each speaker varies his own rate of speech.

1.2. Some theoretical difficulties

1.2.1. In any investigation of tempo, it is necessary to distinguish between gross and net measures of rate (Clevenger and Clark [1963], Goldman-Eisler [1958, 1961], Kelly and Steer [1949] and Wood [1973:§1]).

The distinction between gross and net measures can be obscured by the varying degrees of abstraction of the speech units counted in the analysed sentence. It is possible to count segments actually represented in the speech wave, or segments deemed to have been present in some idealized underlying construction. The "phoneme" has been a notoriously vague concept in this respect, with much controversy between schools. But even "syllables" and "words" are not always as concrete as we might like to imagine. I have discussed this problem in greater detail in my 1973:§3. The difficulty of the degree of abstractness is especially relevant to the present problem. Kozhevnikov and Chistovitch had a 7 syllable test sentence Tonya topila banyu, where the second vowel disappeared at faster rates (i.e. at shorter sentence durations), yielding (I assume) a 6 syllable rendering similar to [toɲ topila banyu]. If this is so, then by taking the utterance duration to indicate tempo, they would have been using a gross type of measure for rate, disregarding the reduction – the sentence was uttered in a briefer duration hence the message was transmitted more quickly. On the other hand, if we were to take account of the fact that at shorter sentence durations there were fewer syllables generated, then we would be using a net type of measure that paradoxically might show a slower articulation rate after the reduction.

Suppose a sentence were uttered in an unbroken phrase of n syllables in t_1 seconds, and then repeated more quickly in the briefer duration t_2 seconds but now in (n - 1) syllables following reduction at the "faster" rate. It is given that $t_1 > t_2$ seconds, but any of the following relations between the net syllable rates are possible:

$$\frac{n}{t_1} > \frac{n-1}{t_2}, \quad \frac{n}{t_1} = \frac{n-1}{t_2}, \quad \frac{n}{t_1} < \frac{n-1}{t_2} \quad \text{sylls/sec}$$

For example, in the case where n = 8 syllables and $t_1 = 2$ seconds, the following values of t_2 will make each of the three relations true: $t_2 > 1.75$ seconds, $t_2 = 1.75$ seconds and $t_2 < 1.75$ seconds respectively. If the 8 syllables are uttered in 2 seconds, we have a net rate of 4 syllables/second. If the message is transmitted "more quickly" in, say, 1.8 seconds but with contraction of 1 syllable, we find a net rate of 7/1.8 syllables/second (3.9) which is "slower". Clearly, it is necessary to choose between the two measures with care, when a "faster" rendering can be articulated "more slowly".

It might be argued that it is not a gross measure but a net measure that indicates how hard the articulators are working during an utterance and is therefore more relevant for discussions of reduction. In the paradoxical situation cited above, the load on the articulators seems to have been lightened by the syllable contraction, despite the briefer utterance duration. Consequently, any experiment designed to relate segment reductions to gesture durations and frequencies ought to be based on a concrete measure rather than an abstract measure. Fortunately, this difficulty can be circumvented by avoiding the type of sentence in which syllable contractions occur, so that there is always the same number of syllables generated at any rate of utterance throughout the entire experimental series. I have followed this policy for the experiments to be described below, and test sentences were carefully selected to avoid the possibility of syllable elision. In so doing, I am of course implying that elisions are not the result of increased speaking rate, and any attempt to prove that from the experiments would inevitably end in circularity. Instead, I shall put forward in the conclusion my belief that elisions are largely habitual and amount to rejection of redundancy, enabling the speaker to increase his message transmission rate (if he wishes) without having to speed up articulation to the same degree. The English sentence Mary wants to go to Swansea can hardly be uttered in anything but 8 syllables. In contrast, the sentence perhaps he will go to Brighton contains two possible syllable contractions, reducing to p'raps e'll go to Brighton, with the added complication that the final syllable might contain either a syllabic n or a vowel.

1.2.2. We all know intuitively what a vowel is but it has never been one of the more easily defined concepts, which can lead to segmentation difficulties. It is wiser to avoid test sentences containing other syllabic sonorants than vowels.

Similarly, while it is often profitable in phonology to describe long vowels and diphthongs as simple vowels followed by semivowels, there is no phonetic boundary within a long vowel or diphthong that could be unambiguously used for measuring purposes. All post-vocalic semivowels (if indeed they have concrete reality outside of phonology) have therefore been included in the vowel aggregates in my test sentences. I could have restricted the set of test sentences by excluding long vowels and diphthongs, but my possible test sentence structure is severely constrained as it is, and some of my subjects already had difficulty in helping me compose suitable sentences in their languages.

Prevocalic semivowels have been included in the consonantal aggregates, but they have been avoided as much as possible as the acoustic wave form does not always permit conclusive segmentation. Post-vocalic nasals, especially in a ..VNC.. combination, revealed a tendency to disappear in anything but slow careful speech, leaving only nasalization on the vowel. This is in itself an interesting observation of a possible phonetic universal, but it is nevertheless a disturbing factor in this type of investigation and several of my repetition series had to be discarded because it was impossible to resolve this difficulty.

The price of these precautions might be decreased generality of the results, but we hardly have any use for test sentences that we do not know how to measure.*

1.2.3. Kozhevnikov and Chistovitch's relative consonant duration refers to an average for the whole sentence, not for selected individual consonants. All consonant durations and all vowel durations were aggregated separately and the consonant durations expressed relative to the vowel durations for the entire utterance, unity being assigned to the vowels. In their own words:

"Since the fluctuations of the relative durations of the sounds of speech

* These problems suggest that the categories of "vowels" and "consonants" may be too comprehensive for this type of investigation. It might be more profitable to investigate the temporal structure of segments in greater detail by looking at the individual sub-classes, in specified syllable structures, as at Figure 9.

in individual syllables were considerable, use was made of an integral index consisting of the ratio of all the consonant intervals to the sum of all the vowel intervals" (p. 87).

Their Figure 3.9 (p. 88) showed that the value of the $\Sigma C/\Sigma V$ ratio increased as utterance duration shortened, indicating that the consonantal portion of the utterance increased at faster rates. With the help of their Figure 3.9, these ratios can easily be transformed into proportions of the utterance duration, which are more immediately comprehended (this was also the form in which Kozhevnikov and Chistovitch expressed their finding in their statement quoted in my opening paragraph). The consonantal proportions of the sentence fell roughly from 35 % to 29 % as the utterance duration lengthened from 0.8 to 3.0 seconds. The vowels are the complement of the consonants with respect to the sentence and we consequently find the vocalic proportions increasing from about 65 % to about 71 % as the sentence duration lengthened.

Incidentally, their Figure 3.3 (p. 81) gives an example of the oscillographic write-out of the articulator electrodes and the speech wave for one rendering of the test sentence. For this illustration, the $\Sigma C/\Sigma V$ ratio was about 0.58:1, the vowels occupying about 62 % of the utterance. These values are typical for short utterance durations according to their Figure 3.9. However, the duration of this rendering was about 2.6 seconds, which is near the longest of the series. This rendering indicates that considerable variation is to be expected in this type of experiment.

2. PROCEDURE

2.1. Methods

2.1.1. The experiment consists in having a suitable test sentence repeated several times at different rates and then making suitable numerical analysis of the measured consonant, vowel and sentence durations.

Kozhevnikov and Chistovitch's Russian sentence had consonants specially chosen to contain labial or coronal gestures that would close electrical circuits at electrodes suitably placed in the mouth, thereby providing automatic segmentation. Nooteboom and Slis also used lip electrodes to facilitate segmentation in their mVmVmVm test words. I have used natural sentences with consonants that are not too difficult to distinguish from vowels on an oscillogram of the wave form, but at the same time sufficiently varied to avoid monotony or articulatory difficulty.

2.1.2. Some experimenters investigating temporal phenomena in speech have found it desirable to dictate a rate to their subjects, especially to hold tempo constant. For the present type of experiment, the aim is the opposite, to elicit a wide variation of tempo. Kozhevnikov and Chistovitch nevertheless chose to dictate rates to their subjects, the closure of the first t initiating a time signal that interrupted the speaker after one of 7 set sentence durations from 0.8 to 3.0 seconds. The speaker had to learn to adjust his rendering to the given interval and as Kozhevnikov and Chistovitch report, "the duration of the first pronunciations, as a rule, differed significantly from the assigned durations as the subjects sought the required rate of speech". Unsuccessful attempts were then rejected. Nooteboom and Slis simply asked their subjects to speak slowly, normally and quickly. The three rates were freely chosen by the subjects and data was obtained on a vast range of rates of speech. I have followed a similar procedure.

2.1.3. Kozhevnikov and Chistovitch's two informants made 25 to 30 attempts at each of the assigned sentence durations and for three different stress positions, yielding a total of 1260 attempts. Each speaker's five closest attempts at each sentence duration for a given stress position were selected for further processing. The point at each of the 7 utterance durations on the graph at the authors' Figure 3.9 presumably represents the $\Sigma C/\Sigma V$ quotient for the average of five renderings. Nooteboom and Slis had at least 20 repetitions each of slow, normal and fast rates. The slow rate was subsequently defined as test word durations in excess of 700 ms (corresponding to net rates of less than 4.3 sylls/sec). Their results were then averaged within each of the three broad rate classes. I have sought to obtain a total of 20 to 30 repetitions of each sentence to cover all rates. Some informants provided up to 40 repetitions.

2.2. Measurement and treatment

The data resulting from the experiment are absolute consonant, vowel and sentence durations, which were measured on oscillograms of the wave form and intensity, the paper running at 125 mm/sec to line up with Voiceprint spectrograms (Fig. 1). The spectrograms were used to help solve segmentation difficulties. The aggregated consonant durations are then expressed as a proportion of the sentence duration.

2.3. Possible outcomes

We can expect any of several outcomes from the experiment when speech is accelerated, for example that the consonants occupy (a) a progressively larger proportion of the sentence (as described by Kozhevnikov and Chistovitch), or (b) a progressively smaller portion of the sentence (the opposite to their predicted tendency) or (c) the same portion of the sentence (relative durations remaining constant, contrary to the belief that they vary).

Such outcomes will be indicated by the degree and sign of the correlation between consonantal proportion and speaking rate.

Alternatively, these outcomes can also be revealed by the linearity of the regression of absolute consonant durations for different rates. If (a) is true, the consonant durations will fall more steeply as sentences become briefer. If (b) is true, the consonant durations will fall less steeply as sentences become briefer. If (c) is true, the consonant durations will fall uniformly as sentences become briefer. There will thus be non-linear regressions for (a) and (b) and a linear regression for (c). The steepness of the slope depends on the proportion of the sentence occupied by the consonants, a larger proportion giving a steeper slope. The changing slope of the non-linear regression will therefore reflect changing relative consonant duration. If one linear regression is steeper than another, then the constant relative duration will have been that much larger.

2.4. Informants and test-sentences

Seven informants took part - B (Southern British English), I (Chinese), J (Polish), K (Southern Swedish), L (German), M (Southern Swedish) and N (Egyptian).

The sentences were pronounced with normal sentence intonation (M also recorded a contrasting series with an alternative intonation):

B1 Mary wants to go to Swansea

B2 Spiker's kidney pies are bigger

I Shuo hua shi ren lei chuan you de xing wei

J Krakowski pociąg czeka tam

K Sven köpte det gamla huset

L Ich kaufte zwei gelbe Tische

M1 Sture fick ett tåg på teknis

M2 Sture fick ett tåg på teknis

N Sidīd katab qiṣṣagdīda

None of these sentences contained syllables that were likely to be contracted, which ensured a constant number of syllables throughout the experiment.

3. RESULTS

The consonantal proportions are plotted against sentence duration for each subject at Figures 2 and 3. Sentence duration can easily be transformed into rate (the number of syllables uttered is constant in each series, and syllable rate is the reciprocal of syllable duration). The net articulation rate has therefore been given alongside the sentence duration scale. The consonantal proportions have also been given in the table at Figure 4 for different rate classes, with a class interval of 1 syll/sec. The absolute aggregated consonant durations have been plotted against sentence duration (rate) at Figures 5, 6 and 7.

3.1. Relative consonant durations

Sample M2 (fig. 3) supports Kozhevnikov and Chistovitch's finding that the consonantal proportion continues to increase with speaking rate.

In samples B2 and N (Figs. 2 and 3), the consonantal proportion appears to vary randomly with speaking rate, suggesting it remains constant independently of rate variation.

For B1 (Fig. 2), the consonantal proportion also appears to vary randomly, although it is possibly a little smaller at slower rates. If so, this would agree with Nooteboom and Slis's finding that the relative consonant durations varied at slower rates.

The remaining samples I, J, K, L and M1 show a very different tendency. The consonantal proportion increases from slow to medium rates - as found by Kozhevnikov and Chistovitch and by Nooteboom and Slis - but then decreases again for fast rates (where Kozhevnicov and Chistovitch expect a continued rise). This type of outcome was not foreseen and could not have been expected from the previous investigations.

The table at Figure 4 highlights smaller differences than can be detected visually from the graphs at Figures 2 and 3.

The table not only confirms that M2 supports Kozhevnikov and Chistovitch's finding (the consonantal proportion rose continuously from 53 % to 58 %). It also shows that B1 follows the same tendency, althought with very small increments.

The table suggests that B2's may not vary as randomly as was thought and that this sample may follow the unforeseen tendency. But the variations between rate classes are extremely small.

The table shows that N's consonantal proportion may not vary randomly. It seems to fall slightly as rate increases (there are only examples of medium and fast rates for this speaker).

The tendency for the consonantal proportion to increase from slow to medium rates, and then decrease again to fast rates, was not foreseen as a possible outcome in § 2.3 above. This tendency was exhibited in most samples – I, J, K, L and M1. It is just possible that B2 and N also belong to this group.

3.2. Absolute consonantal durations

The aggregated consonantal durations are plotted against sentence duration (and rate) at Figures 5, 6 and 7. The tendency for relative consonantal durations to vary with rate will appear as a non-linear regression in this presentation (§ 2.3), but the relationship revealed in Figures 5, 6 and 7 seems very linear. The variations of relative consonantal duration indicated by the table at Figure 4 are so small (a few percent only) that they can hardly be detected on Figures 5, 6 and 7. The straight lines drawn on these graphs represent the constant consonantal proportions stated there and have been fitted by eye to pass through origo and the set of points.

The table at Figure 8 gives the product moment correlation coefficients for consonant and sentence durations and vowel and sentence durations. These are so close to 1 that they underline how close these regressions are to linearity (this correlation coefficient assumes a linear relationship between the variables).

4. DISCUSSION

4.1. Variations of relative consonantal durations

Practically all of the cases illustrated at Figures 2 and 3 and tabulated at Figure 4 show that there were some variations of relative consonant duration at different speaking rates, although a large majority do not follow the tendency described by Kozhevnikov and Chistovitch. The range of variation was small. Figures 5, 6 and 7 show that such small variations hardly deviate from a virtually linear relationship between absolute consonantal durations and sentence durations. If we look very closely at I, J, K, L or M1, we can just see that the absolute consonantal durations do rise a little more steeply at faster rates, and do flatten off a little at slower rates. But the changes of slope are very slight and I am sure they can only be spotted because we know the answer in advance from Figures 2 and 3. Had we been looking for a truly linear relationship in any other experiment, we would have been overjoyed to find a set of plots like those of Figures 5, 6 and 7. The proximity of the coefficients to 1 is an indication that a linear model is an excellent approximation to these results (Fig. 8). The vowels are equally intimately correlated with speaking rate.

The problem is, are the very small departures from the excellent linear model nevertheless sufficiently large to warrant the conclusion that relative consonantal durations are invariant? Kozhevnikov and Chistovitch believed so. The degree of constancy of their relative syllable durations (that they concluded were invariant) may provide a yardstick to judge this by.

Their Figures 3.6, 3.7 and 3.8 illustrate the syllable relative to its word. I am not convinced that the variation they show is as random as they suggest. A definite pattern can be discerned, especially in their word <u>topila</u>. These variations appear to be as large as the variations of their consonantal proportions in the sentence. For comparison, I have taken two words from my German subject <u>L</u>, <u>kaufte</u> and <u>gelbe</u>. Figure 9 shows the variations of the word durations relative to the sentence, the syllables to the words and the consonants to the syllables. The consonantal proportion of the sentence has been included for reference. Figure 9 shows that the relative duration of none of the units was invariant. The words show the least variation - they decrease slightly from slow to moderate rates, and then rise again at faster rates. The stronger syllables <u>kauf</u> and <u>gel</u> show greater variation than the words. The weaker syllables are the complement of the stronger syllables in their respective words and have not therefore been included on the graphs - <u>te</u> falls from 32 % to 27 % and then rises again to 33 %, while <u>be</u> rises from 31 % to 40 %. This quick spot check does not support the idea that the relative durations of syllables should be less dependent on speaking rate than are the relative durations of the consonants. It might be worth while to investigate other syllables and words from my other informants and languages.

Even if Kozhevnikov and Chistovitch had been mistaken about the relative durations of phonemes and syllables, it can still nevertheless be true that the syllable is the programme unit. Many intuitively feel that the syllable is a basic unit of speech production. But some other type of argument and evidence may be necessary to confirm it.

4.1. <u>Synkope and syllable contraction</u>

There are two basic problems that need to be solved before a final theory of reduction can be built. Firstly, if a particular speech unit disappears in some renderings, has it been squeezed out because there is insufficient time to accomplish it, or has it been deliberately omitted? Are the necessary gesture commands extinguished at some point during the production process, or are they never initiated? Secondly, are reductions peripheral or central from the neuromotor point of view? There are possibly occasions when any of these alternatives may be true. I doubt whether a complete answer can be provided yet, but some aspects can be discussed briefly.

It is not easy to define an increase in tempo that can be related to reductions of the synkope type. The observed speaking rate, computed by counting how many of a given speech unit are produced in the measured time, represents the sum of the various influences acting on the temporal characteristics of speech segments. It reflects the consequences of reductions rather than the drive that is postulated to have occasioned them.

There are two phenomena that contribute to a shorter utterance duration and hence an accelerated message - shorter segment durations and coalescence or loss of segments. The first is the area observed in investigations of articulatory undershooting. The second is the area of Kozhevnikov and Chistovitch's hypothesis of elision.

4.2.2. Undershooting of gesture targets has been observed and related to the time available for their execution (e.g. by Stetson et alia 1940, Lindblom 1963, Gay 1968, Kent and Moll 1972). This is doubtless peripheral in character - a body travelling at a given velocity or accelerating at a given rate will not move so far if the duration of the movement is shortened. So long as there is a gesture to measure, it is a simple matter to relate undershooting and speech rate. But what happens if the observed segment is elided?

The loss of a segment is perceived because the rendering can be compared with a well known complete form. It would be tempting to consider such an elision as a case of maximum undershoot and relate observed reductions to the rate of transmission of the complete forms. For example, I found in a sequence of General American speech (recorded from a radio interview) the five syllables of <u>the Americans</u> spoken in 0.38 seconds, a rate of 13.2 syllables/second. We cannot articulate syllables at that speed, the expected maximum being about 8, perhaps 9, syllables/second (Wood 1973: § 4). In fact, he uttered three syllables, [ð i m e r k n̥ z] at 7.9 syllables/second, a very plausible fast rate of articulation. But to relate the reductions to the transmission rate of the complete forms would take for granted that elisions are of the squeeze-out type, and ignore the possibility that some might instead be deliberate omissions from the message. The first type of elision can be attributed to temporal constraints on the articulatory programme (i.e. encoding constraints), including Kozhevnicov and Chistovitch's hypothesis, the second type concerns the composition of the underlying message that is to be encoded. These are two possible ways in which the brain can work in this situation and must be born in mind during any discussion or speculation about the planning or programming that precedes the articulation of speech. The two cannot be distinguished in the speech output.

4.2.3. Simple experiments like Lehiste's (1970:7) comparison of the frequency of free apical vibrations in a trilled <u>r</u> (28 per second) with the voluntary apico-dental gesture of <u>t</u> (7 per second) suggest that the limiting factor for voluntary gestures lies within the nervous system, the bottleneck being the rate at which some higher motor centre can transmit sequences of different commands to the articulators. In syllable repetition experiments (Kaiser 1934, Hudgins and Stetson 1937, Sigurd 1971) this limiting rate appears peripherally as the rate at which the speaker has switched between different gestures, for example t-a-t-a-t-a... where the maximum rate appears as 7 or 8 syllables/second. Lehiste's experiment indicates that the articulators can move more quickly than the motor centres ever require for a deliberate gesture. Where synkope is a case of gesture commands being extinguished, it seems likely this would be due to the inability of a motor centre to pass on or switch between the necessary coordinated commands at a sufficiently fast rate, rather than to the inability of the articulators to respond properly owing to mechanical constraints.

It is clear from my wording of the previous paragraph that I imagine individual gestures, rather than whole segments, suffering from that type of neuromotor constraint. In test sentence B1, examples of the following reductions occurred in <ins>...wants to...</ins>: [w ´̃ n t s t ŭ], [w ɔ̃ n s t u], [w ɔ̃ s t u]. Only a few gestures were lost at a time and not the entire packages of gestures needed for <ins>n</ins> and <ins>t</ins>. Note particularly that when the dental occlusion vanished, the <ins>s</ins> remained voiceless without assimilating to the voiced <ins>n</ins>; similarly the nasal cavity remains open during the vowel even when the <ins>n</ins> has vanished. In contrast I would expect the sequence <ins>wands to</ins> ("wands to do conjuring tricks with") to reduce as follows: [w ɔ̃ n d z t u], [w ɔ̃ n z t u], [w ɔ̃ z t u]. If it is assumed that these are cases of gesture commands being extinguished, then it must also be admitted that the occlusions were discarded subsequent to the assignment of voicing assimilation to the inflexions <ins>s</ins>. I usually hesitate to agree that the ordered rules of a transformational generative model for phonology necessarily always reflect cerebral processes that are part of speech production, but the parallel is undoubtedly striking in these examples.

4.2.4. Supposing the Kozhevnikov and Chistovitch hypothesis were true, that vowels are omitted when the consonant happens to need all the time available for the syllable. Segment elisions ought then to occur at random, whenever there was a momentary shortage of time. But we know that vowels do not disappear at random in speech. Elisions and syllable contractions are largely non-random and are instead habitual and predictable from the environment. This is a view that is supported by the regularity of such phenomena synchronically in everyday speech and diachronically in sound change. For example, it is typical of many languages that certain weak vowels are omitted between obstruents and liquids, as in French app'ler or English delib'rate. The conventional, rather than tempo-dependent, character of such reductions is underlined when hyper-correct forms appear with spurious vowels as in Latin saec(u)lum. If the reduced forms lose contact with their complete form, and ultimately become established as the normal form, evidence of a lost segment may only appear in morphological alternations as in English hist'ry-historical, vict'ry-victorious.

Habitual reductions become part of the common speech code shared by all members of a definable speech community, and comprehension is not endangered. I suspect that these reductions are not the consequence of a rapid speech rate but rather that they permit more rapid rates. The articulatory programme is simplified and shortened by dispensing with some of the redundancy in the speech signal, enabling the linguistic message to be transmitted in a briefer period of time, that is, more rapidly. It was demonstrated in § 1.2.1. above that syllable contraction leads to a slight reduction of the articulation rate while the utterance becomes briefer and the message is transmitted more quickly. This reorganization of the articulatory programme permits more rapid message transmission rates while a comfortable rate of articulation is retained.

BIBLIOGRAPHICAL REFERENCES

Clevenger T. and M.L. Clark. 1963. Coincidental variation as a source of confusion in the experimental study of rate. Language and Speech 6: 144–150

Gaitenby J.H. 1965. The elastic word. Status Report of Speech Research, Haskins Laboratories, SR-2:3.1–3.12

Gay T. 1968. The effect of speaking rate on diphthong formant movements. Journal of the Acoustical Society of America 44:1570–74

Goldman-Eisler F. 1958. Speech analysis and mental processes. Language and Speech 1:59–75

Goldman-Eisler F. 1961. The significance of changes in the rate of articulation. Language and Speech 4:171–174

Hudgins C.V. and Stetson R.H. 1937. The relative speed of articulatory movements. Archives Néerlandaises de Phonétique Expérimentale 13: 85–94

Kaiser L. 1934. Some properties of speech muscles and the influence thereof on language. Archives Néerlandaises de Phonétique Expérimentale 10: 121–133

Kelly J.C. and Steer M.D. 1949. Revised concept of rate. Journal of Speech and Hearing Disorders 14: 222–226

Kent R.D. and Moll K.L. 1972. Cinefluorgraphic analysis of selected lingual consonants. Journal of Speech and Hearing Research 15:453–473

Kozhevnikov V.A. and Chistovitch L.A. 1965. Speech, articulation and perception. Joint Publications Research Service, Washington

Lehiste I. 1970. Suprasegmentals. MIT Press

Lindblom B. 1963. Spectrographic study of vowel reduction. Journal of the Acoustical Society of America 35: 1773–1781

Nooteboom S.G. and Slis I.H. 1969. A note on rate of speech. IPO Annual Progress Report 4: 58–60. Eindhoven

Sigurd B. 1971. Maximal rate and minimal duration of repeated syllables. Papers from the Institute of Linguistics of the University of Stockholm 3

Stetson R.H., Hudgins C.V. and Moses E.R. 1940. Palatograms change with rates of articulation. Archives Néerlandaises de Phonétique Expérimentale 16: 52–61

Wood S. 1973. Speech tempo. Working Papers of the Phonetics Laboratory of Lund University 9.

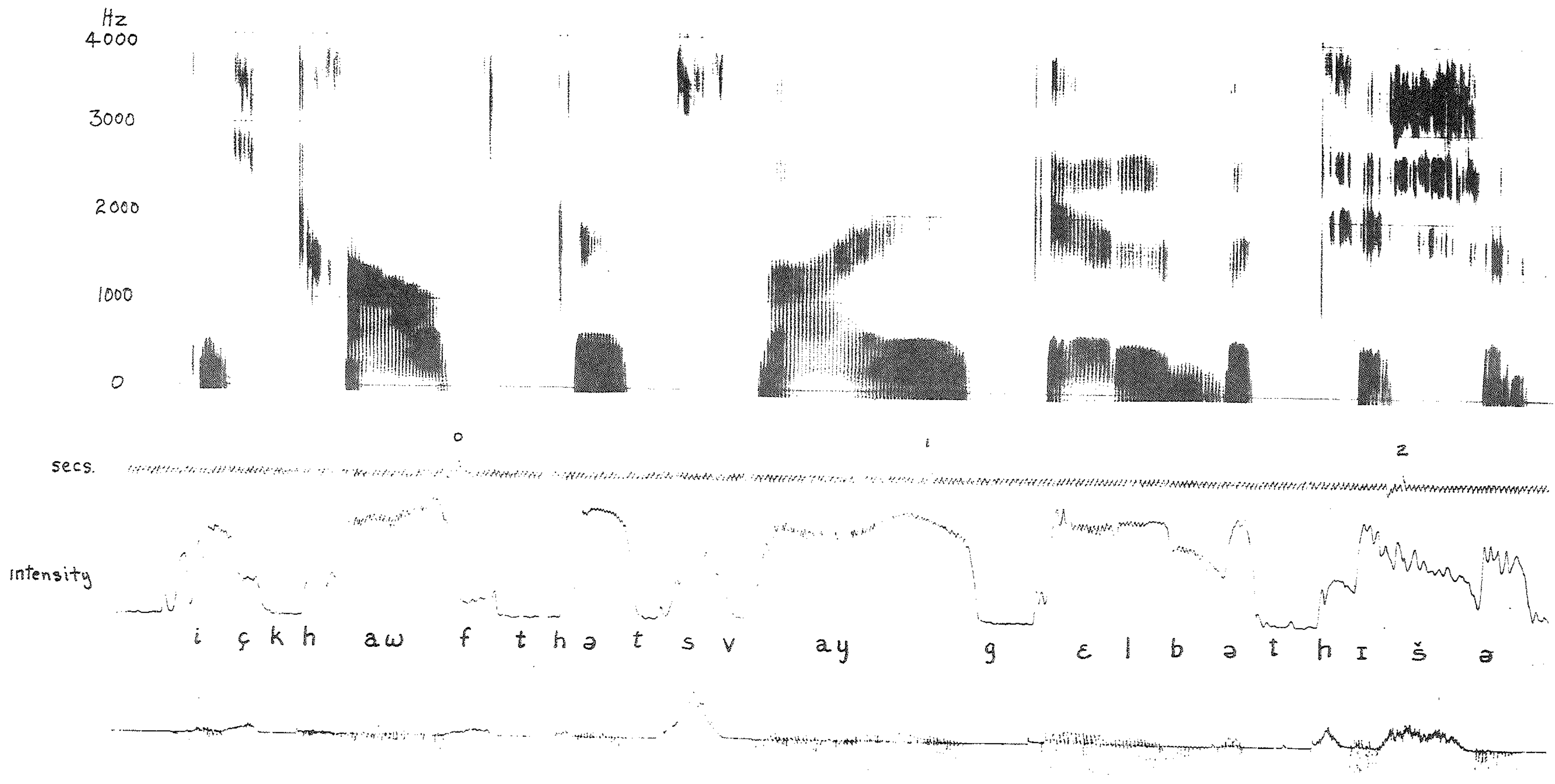


Fig.1a. A slow rendering (2.8 sylls/sec) of the sentence "Ich kaufte zwei gelbe Tische" (L).

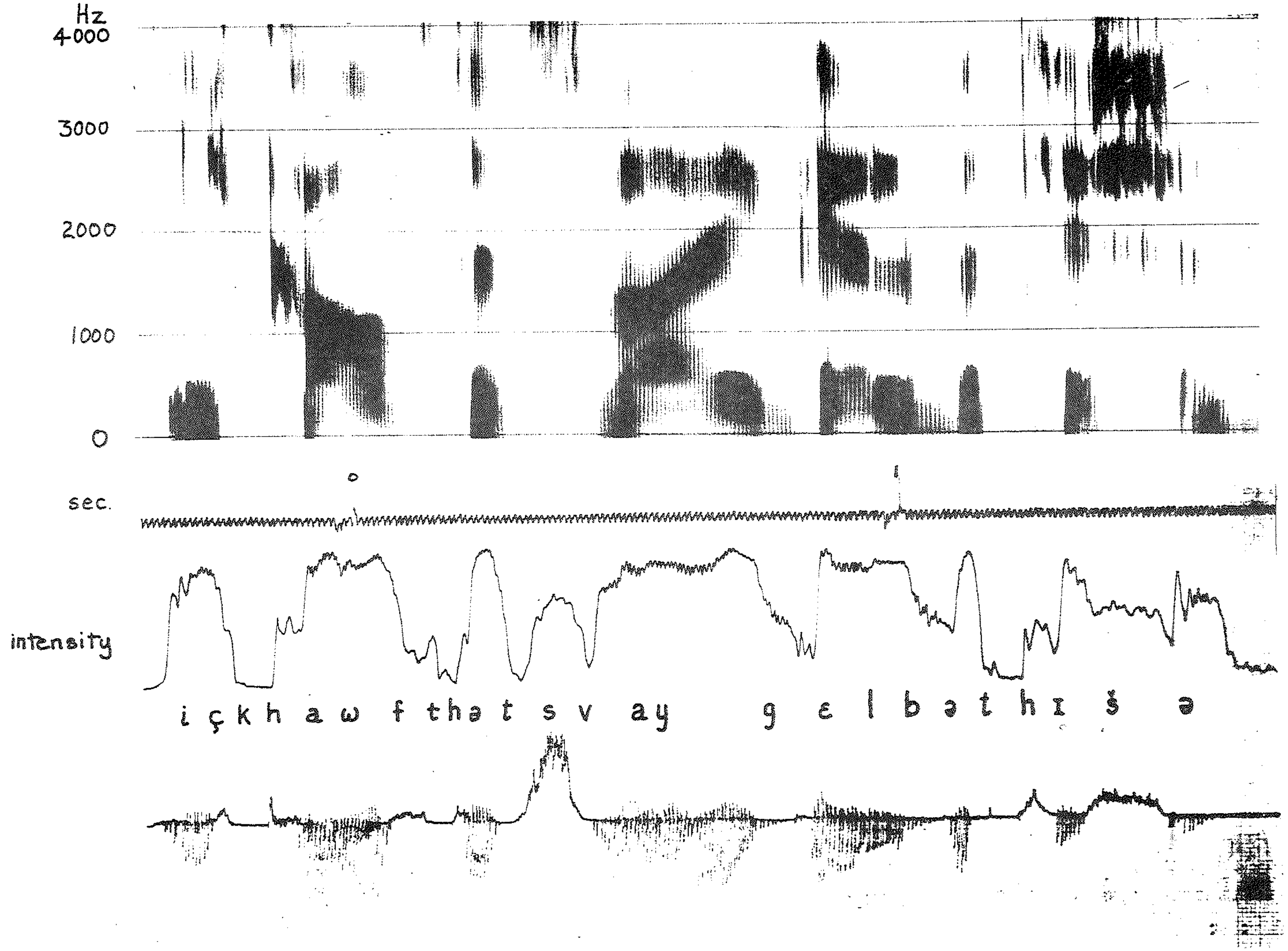


Fig.1b. A medium rendering (4.2 sylls/sec) of the sentence "Ich kaufte zwei gelbe Tische"(L).

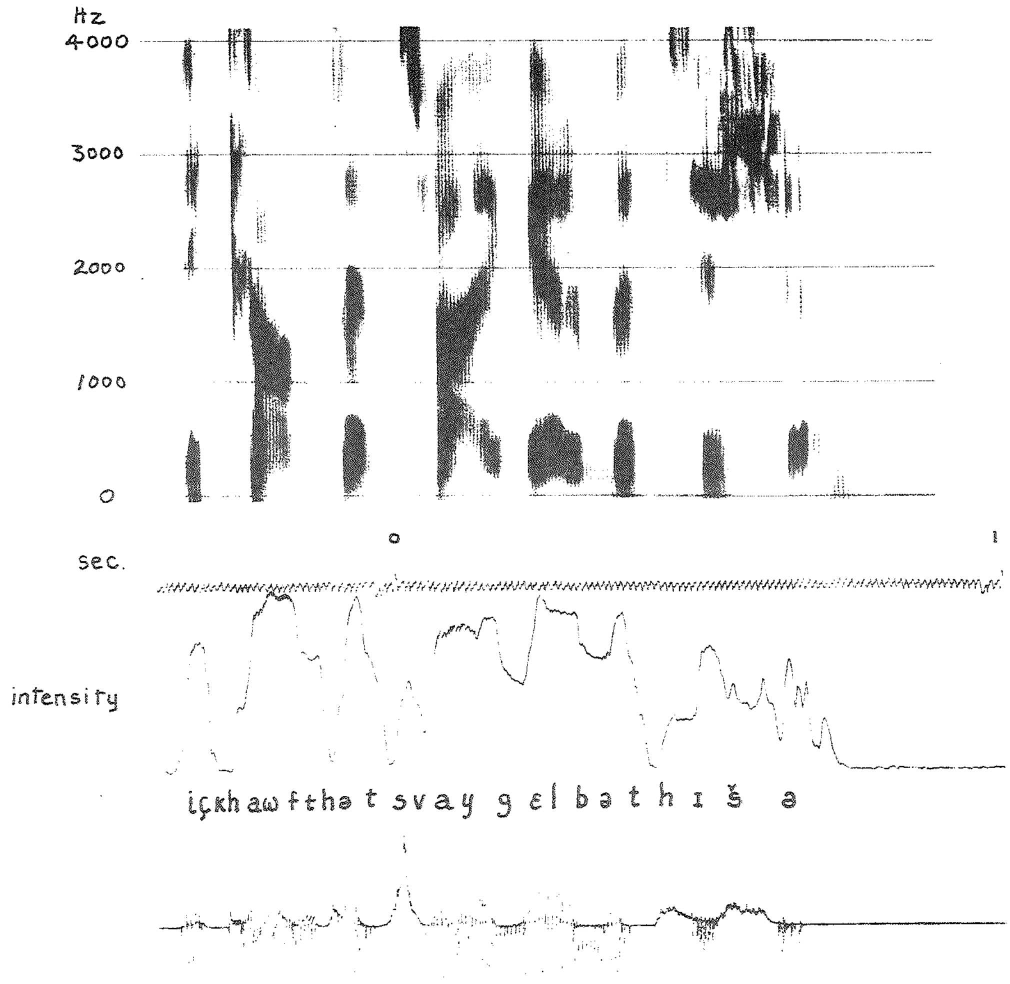


Fig.1c. A fast rendering (7.2 sylls/sec) of the sentence "Ich kaufte zwei gelbe Tische" (L).

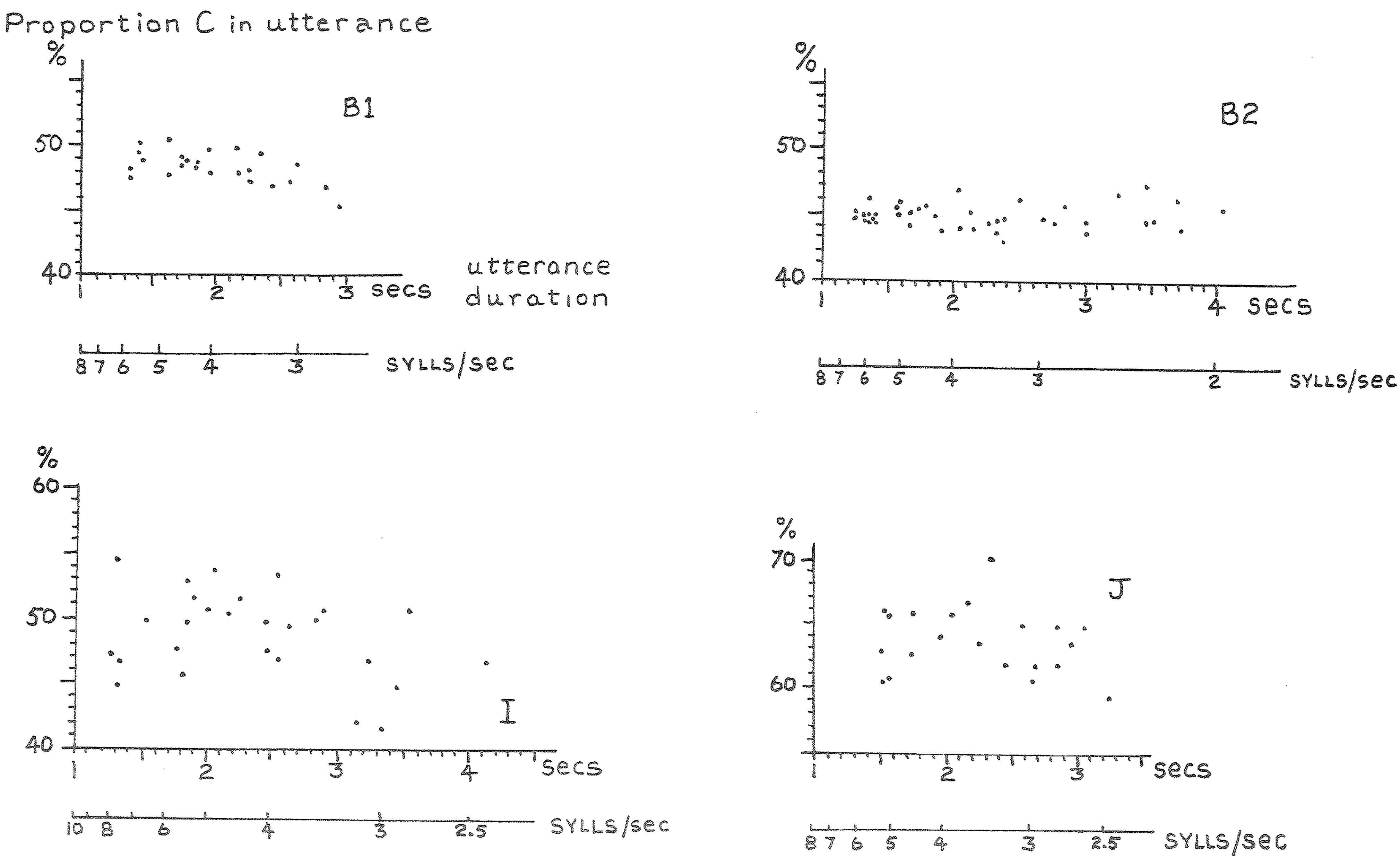


Fig. 2. The proportion of sentence duration consumed by the consonants in the sentence, for series B1, B2, I, J.

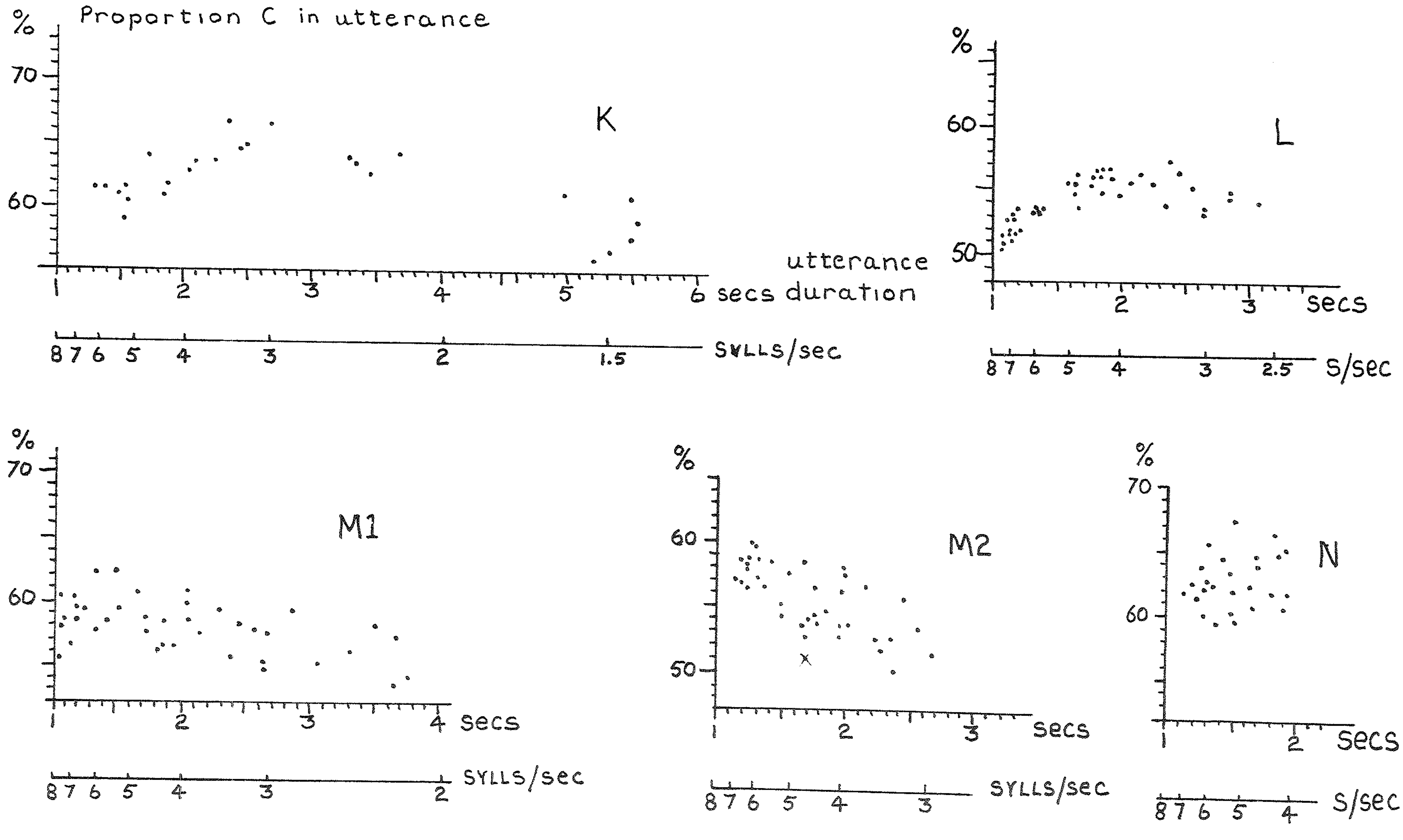


Fig. 3. The proportion of sentence duration consumed by the consonants in the sentence, for series K, L, M1, M2, N.

RATE CLASSES	CONSONANTAL PROPORTIONS (%) AT DIFFERENT RATES								
(Sylls/sec)	B1	B2	I	J	K	L	M1	M2	N
1.01–2.00					58.2				
2.01–3.00	45.5	44.7	45.8	58.7	61.9	54.8	56.3	53.0	
3.01–4.00	48.1	44.6	47.8	64.1	64.2	55.4	57.7	55.0	63.6
4.01–5.00	47.9	44.7	50.4	63.8	62.1	55.6	57.9	56.1	62.7
5.01–6.00	48.7	45.7	51.6	63.4	60.6	54.1	60.1	58.1	62.4
6.01–7.00		45.4				52.6	59.4		
7.01–8.00			48.1			51.6	57.8		

Fig. 4. The average consonantal proportions of the utterance duration for each speaker at different articulation rates.

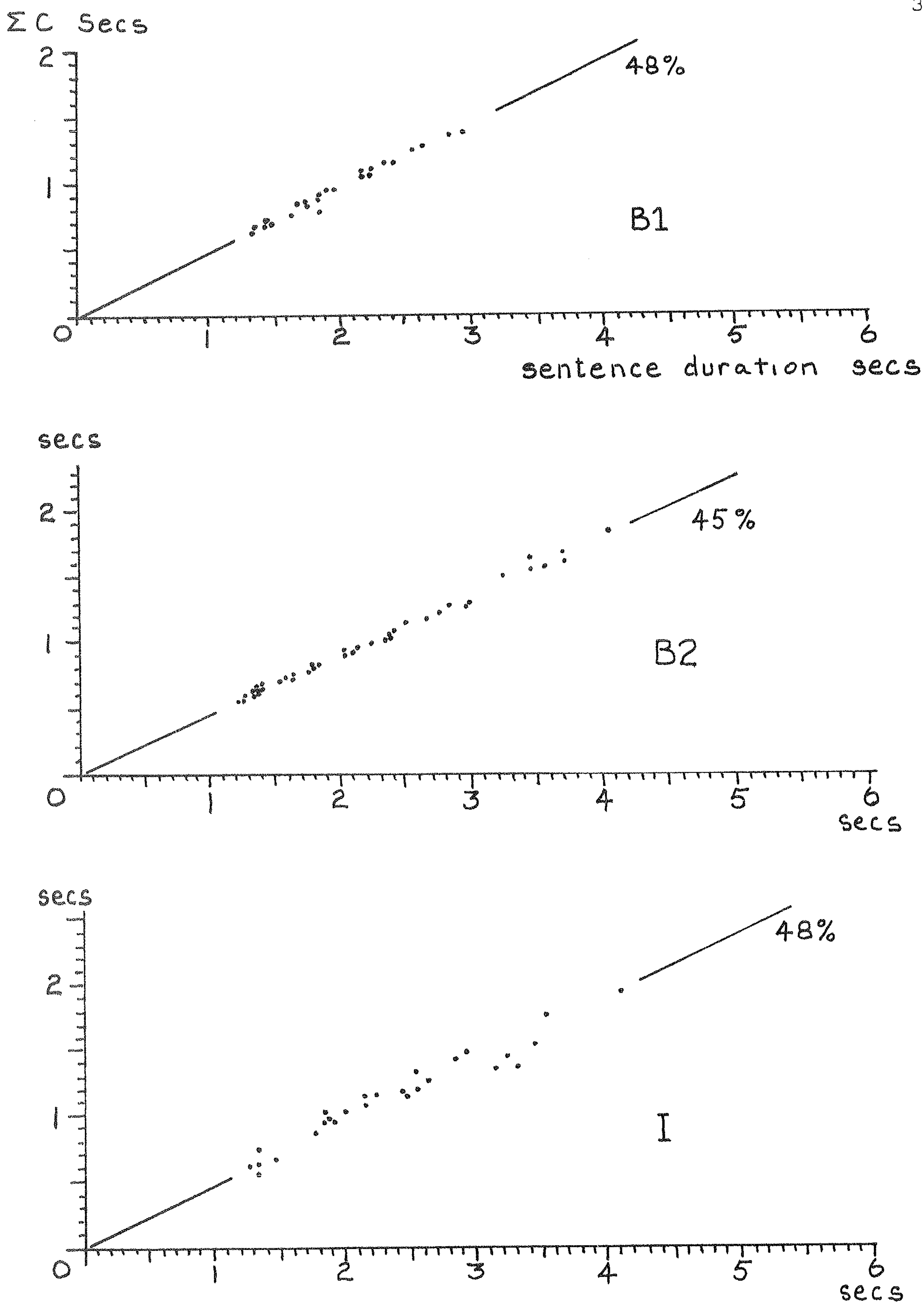


Fig. 5. The relationship between aggregated consonantal durations and sentence durations in series B1, B2 and I.

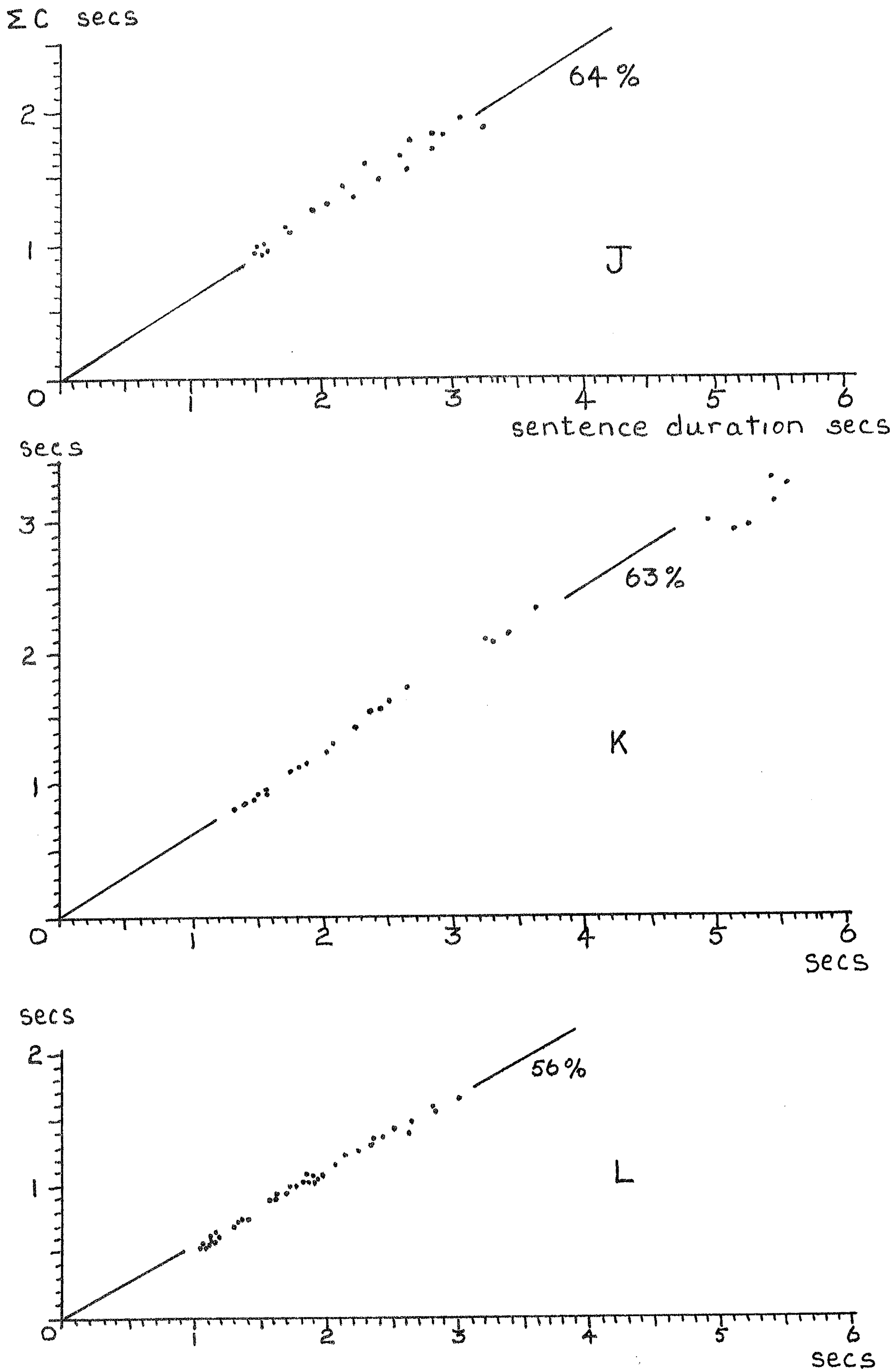


Fig. 6. The relationship between aggregated consonantal durations and sentence duratons in series J, K and L.

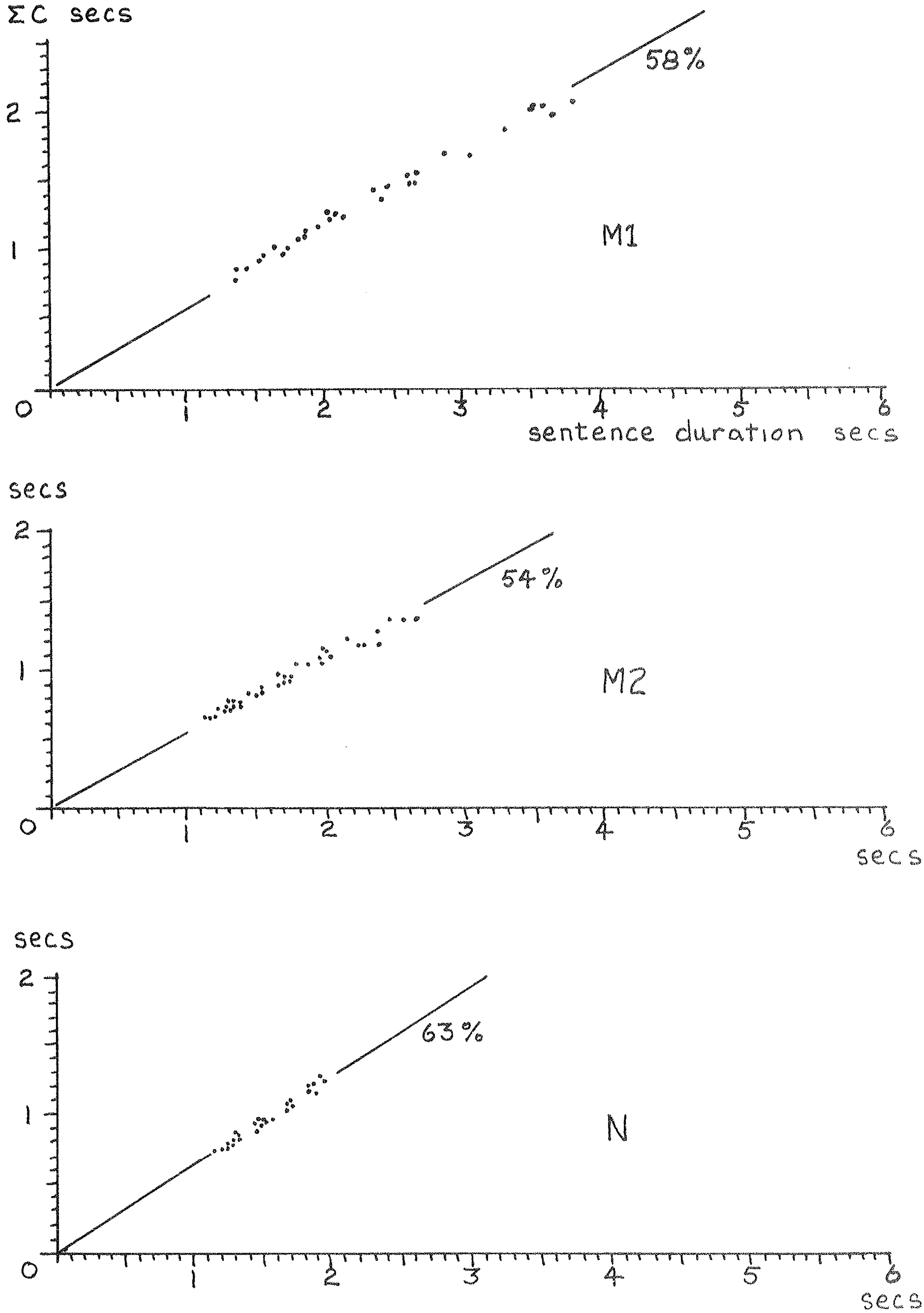


Fig. 7. The relationship between aggregated consonantal durations and sentence durations in series M1, M2 and N.

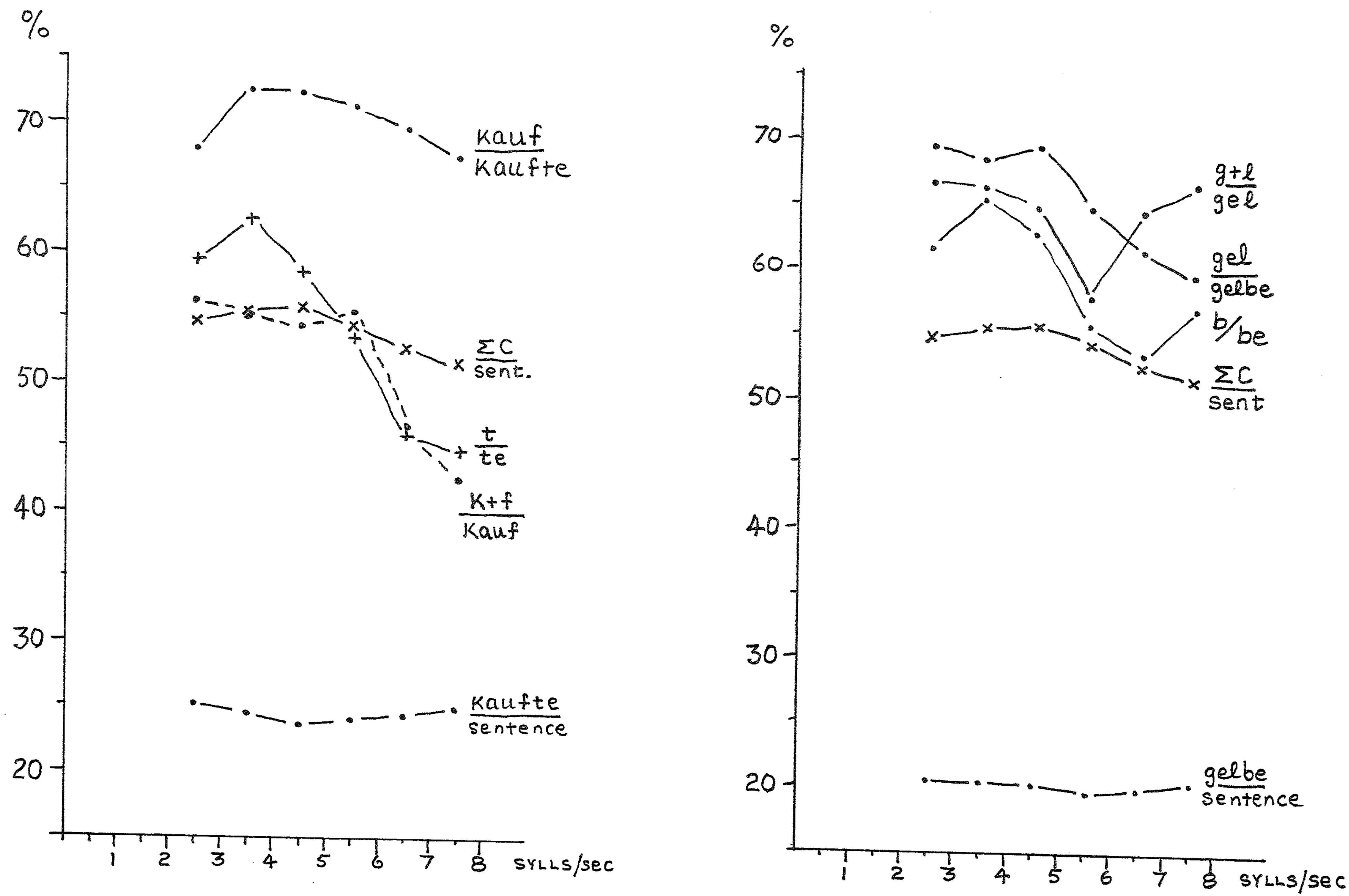


Fig. 9. The relative durations of selected words, syllables and consonants from L's series.

Speaker	Correlation coefficients for sentence duration and aggregated consonantal duration	aggregated vowel duration
B1	0.974	0.996
B2	0.998	0.998
I	0.967	0.982
J	0.984	0.965
K	0.954	0.960
L	0.982	0.994
M1	0.996	0.993
M2	0.985	0.986
N	0.981	0.942

Fig. 8. Coefficients for the correlations between sentence duration and the aggregated consonantal and vowel durations respectively in each rendering (cf. Figs. 5, 6 and 7).

COARTICULATION IN SOME SWEDISH STOP SYLLABLES

Per-Erik Nordström

This investigation was started because results from an earlier study of coarticulation needed to be tested and, hopefully, verified. From them, the following working hypothesis was formulated: Initial and final F_2-frequencies in symmetrical CVC-syllables with stops can be described in terms of an inherent feature (formant frequency) of the vowel segment.

Those earlier results indicate that the initial and final F_2-frequencies in dVd-syllables can be reasonably well predicted from the F_2-frequency in the vowel segment. The functions derived (hyperbola - like functions) are not easily understood, and articulatory correlates would be hard to find. If the initial and final F_2-frequencies are related to the vowel's F_1, the results are easier to grasp. F_1 plotted as a function of F_2 initially and finally turns out to be an area, similar in shape, but not in size, to the acoustic vowel space (F_1 as a function of F_2 for the vowels in the language).

A general observation in that dVd-study is that the F_2-frequencies have a very limited range of variation initially, whereas the final values are much more like the vowel's F_2. This fact is interpreted as an apical occlusion before the tongue body movement has been completed.

The present study includes all Swedish stops in symmetrical CVC-syllables with the vowels /i: e: ɛ: ɑ: o: u:/. The previous results are borne out for /d/ and /t/, but new formulations are neeeded to capture all places of articulation.

<u>Figure 1:</u> Plotted against the vowel's F_2, both the initial and final F_2-values form the same groups (with the exception of finally dipthongized /i:/): Velars with front vowels, dentals and labials with front vowels, dentals with back vowels, and velars and labials with back vowels.
<u>Figure 2:</u> Plotted against the vowel's F_1, exactly the same groups can be identified.

The investigator's interpretation is that there is <u>a basic tongue position</u> for the "typical" case of coarticulation (dentals/labials with front vowels and velars/labials with back vowels) and a <u>fronted tongue position</u> for velar stops with front vowels and for dental stops with back vowels. Thus labial articulation (with unaffected tongue body) becomes the "norm". A coarticulation space in the F_1-F_2-plane, similar in shape and constitution to the acoustic vowel space, can be specified for each place of articulation.

At the time of this symposium, results from F_3-measurements have not yet been evaluated.

Fig. 1 a and b: Initial and final (a and b) coarticulation categories in the F_2- plane. Initial = at the instant of release, final = at the instant of vowel offset.

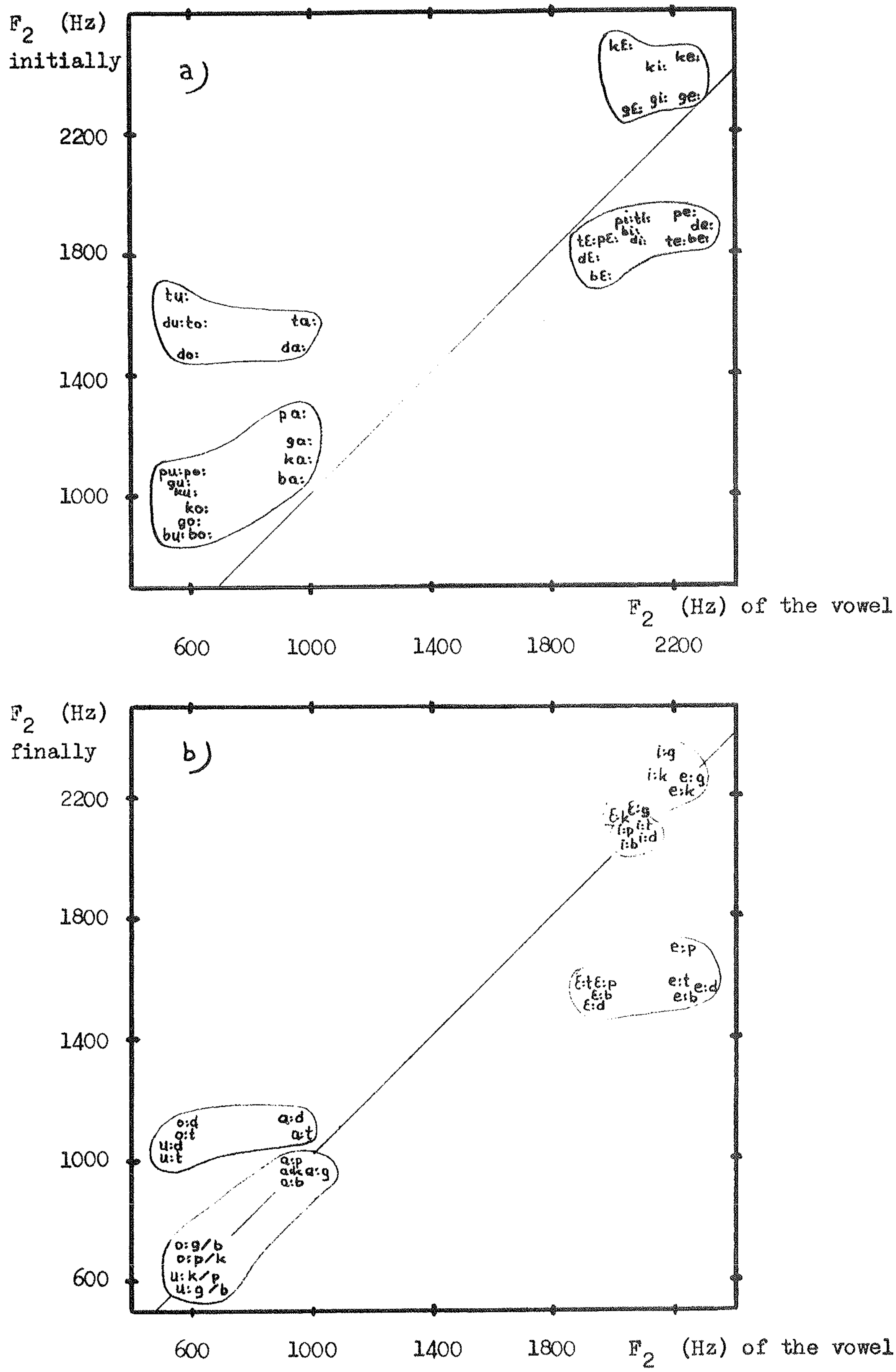
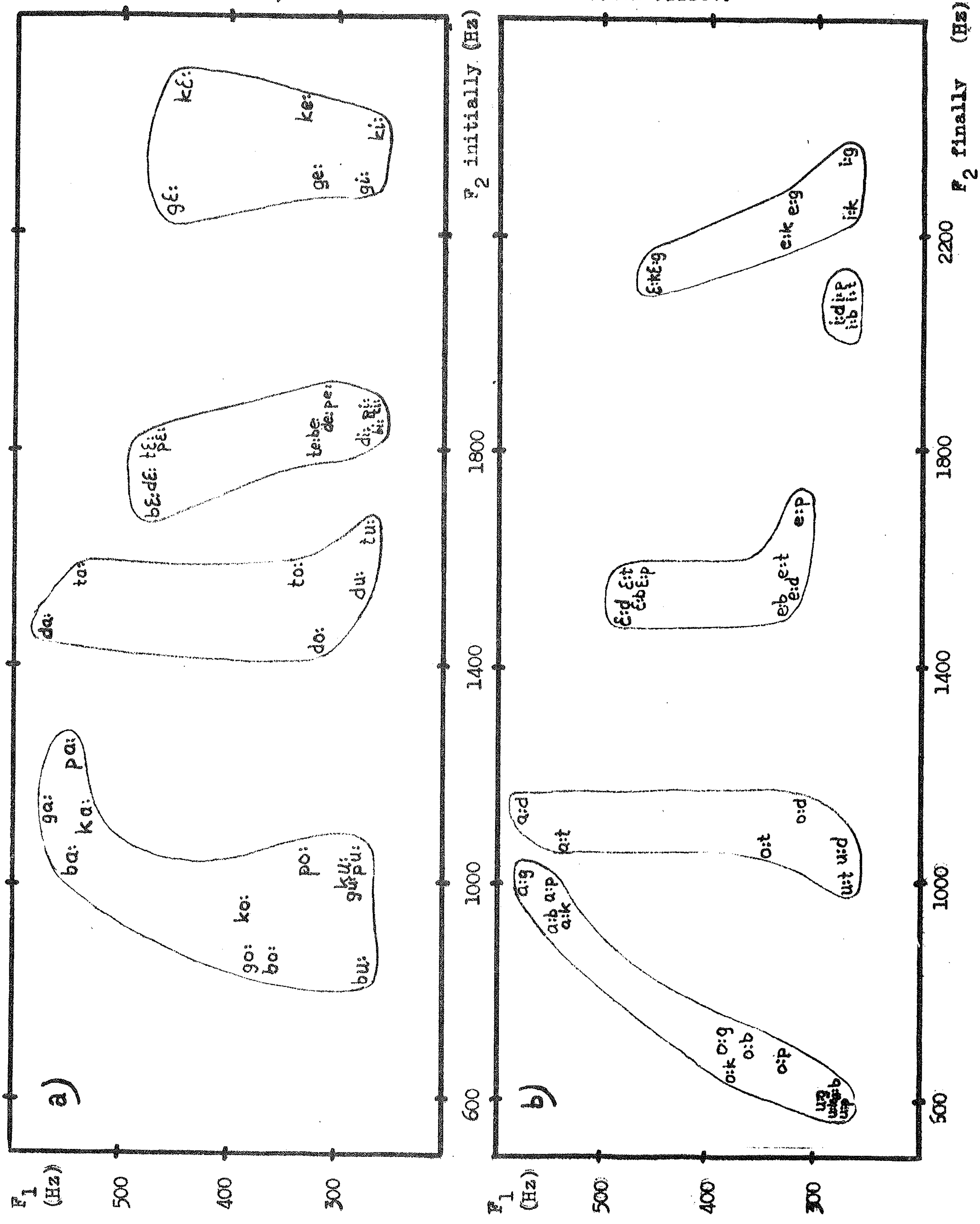


Fig. 2 a and b: Initial and final (a and b) coarticulation categories in the F_1- F_2- plane. Initial = at the instant of release, final = at the instant of vowel offset.



DUAF – DATORBASERAD UNDERVISNING I AKUSTISK FONETIK: "FORMANTS WITHOUT TEARS"

Stefan Pauli

PROJEKTETS BAKGRUND

Att lära ut fonetik erbjuder alldeles speciella problem, emedan ämnet innehåller många moment med tydlig förankring i matematik och fysik och de flesta studenter saknar bakgrund i dessa ämnen. Ett exempel på ett område, där förkunskaper och fonetikundervisning är speciellt illa anpassade är den akustiska fonetiken, som behandlar talet som ljudvåg. Den matematiska och fysikaliska beskrivningen av talvågen är minst sagt svårsmält för en humanist. Även för den matematiskt skolade fonetikern kan det vara svårt att få en känsla för den fysikaliska verkligheten bakom matematiska formler.

Men faktum är att fenomenen och egenskaperna som sådana inte är svåra att förstå. Det är egentligen bara formlerna som är besvärliga. Det gäller att hitta på en pedagogisk metod att kringgå dessa, dvs. att med en "formellös" metod lära ut vad de beskriver.

PROJEKTETS PROBLEMSTÄLLNING

En del av fonetikstudierna syftar till att bibringa eleven kunskaper om hur talljud är akustiskt uppbyggda och hur deras egenskaper ändras när utalet varieras. Detta mål sammanhänger med att studenten i de flesta tillämpningar av fonetiken måste känna till grundläggande begrepp som formant, bandbredd, källa mm. och ha en intuitiv förståelse av hur dessa begrepp fungerar som talljudens byggklossar. Målet för vårt projekt är att utveckla en pedagogisk "formellös" metod med vars hjälp man kan lära ut de fysikaliska och fonetiska sammanhang som komplicerade matematiska formler beskriver: Vi kan sammanfatta innebörden i vår metod med mottot: FORMEL I BILD.

Vi har utvecklat ett dator program med vars hjälp de matematiska formlerna kan åskådliggöras grafiskt med en kurva som presenteras av datorn i rörlig eller stillastående form på en TV-skärm, eller automatiskt skrivs ut på ett papper som eleven kan ta med sig hem.

Vi skall se detta pedagogiska grepp mot bakgrunden av det alternativ som läraren hittills huvudsakligen har baserat sin undervisning på, nämligen svarta tavlan.

BESKRIVNING AV PROGRAMANVÄNDNING

I den typiska undervisningssituationen befinner sig eleven framför datorns tangentbord och kan samtidigt betrakta den till datorn anslutna TV-bildskärmen och X-Y-skrivaren, se Figur 1.

Ett exempel på bildpresentation visas i Figur 2. Översta raden visar källans, stämbandens, spektrum; mittersta raden visar de akustiska överföringsegenskaperna från stämbanden till lyssnaren. Den understa raden slutligen, visar det resulterande talljudsspektrat. De tre kolumnerna svarar mot tre olika höga stämbandstoner: 100, 200 respektive 400 Hz (svängningar per sekund). För varje uppritad bil erhålles också en numerisk specifikation av bildinnehållet. Med en kontinuerligt varierbar manuell kontroll, en s.k. "joystick" kan en eller två egenskaper i bilden (sam-)varieras. Kopplingen mellan handrörelse och bildändring är omedelbar. Denna snabba återkoppling innebär, att eleven får en intuitiv känsla för de komplicerade akustiska sammanhang som ovan nämnda matematiska formler beskriver.

FRAMTIDA PLANER

Vi planerar att bygga ut datorprogrammet med en modul för analys av elevens eget tal. Talmaskinen OVE III skall kopplas in så att eleven genom ett tryck på en knapp kan lyssna till det talljud, som han studerar på skärmen. Eleven kan på detta sätt i viss omfattning utföra analys och syntes av sitt eget tal.

Presentationsmetoden på TV-skärm lämpar sig också mycket väl för framställning av undervisningsfilmer.

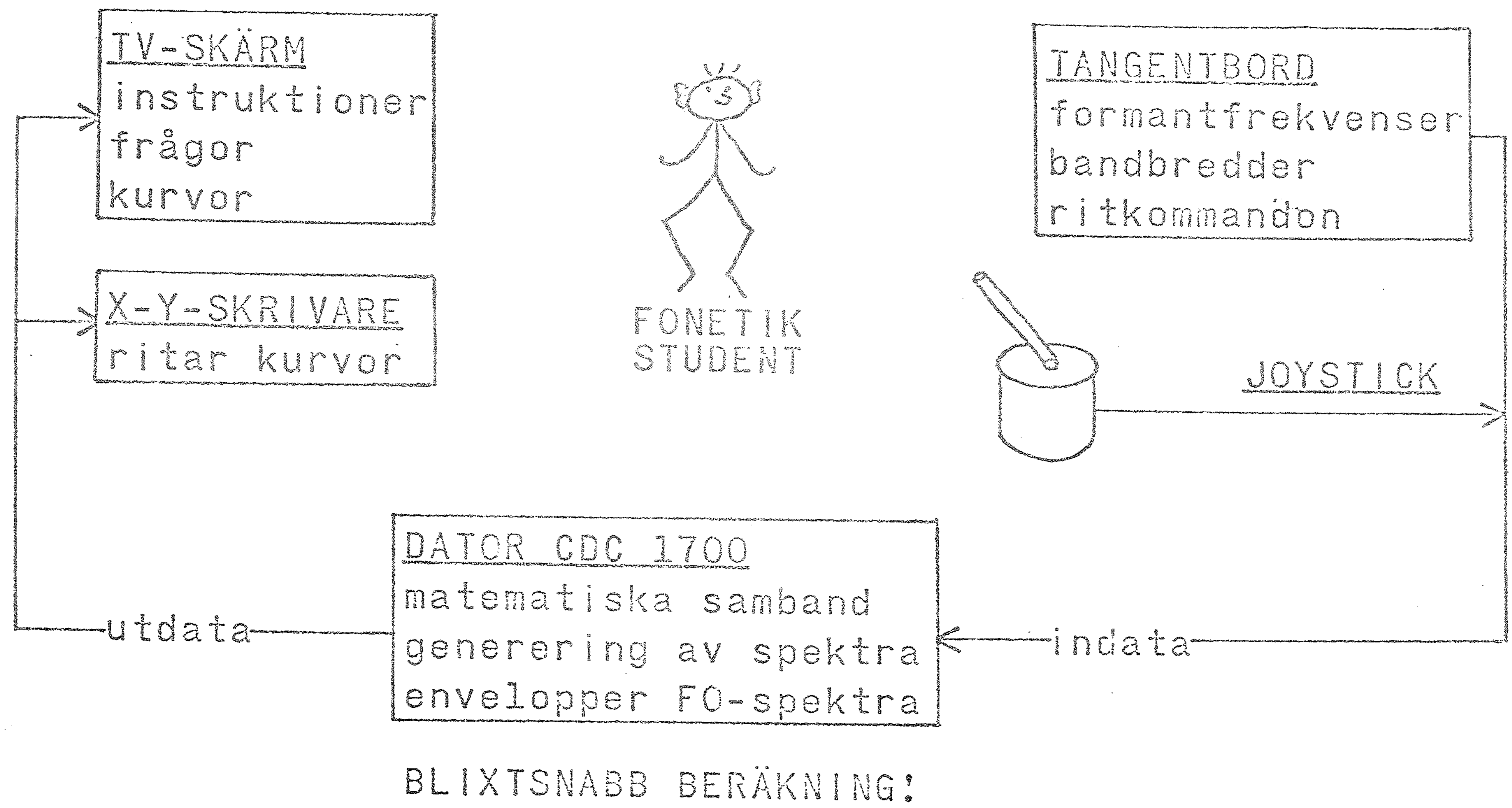


Fig. 1 Fonetikstudent i vårt datorbaserade undervisnings-system DUAF.

DATÙM 730404
S PI

UNDVIS 24

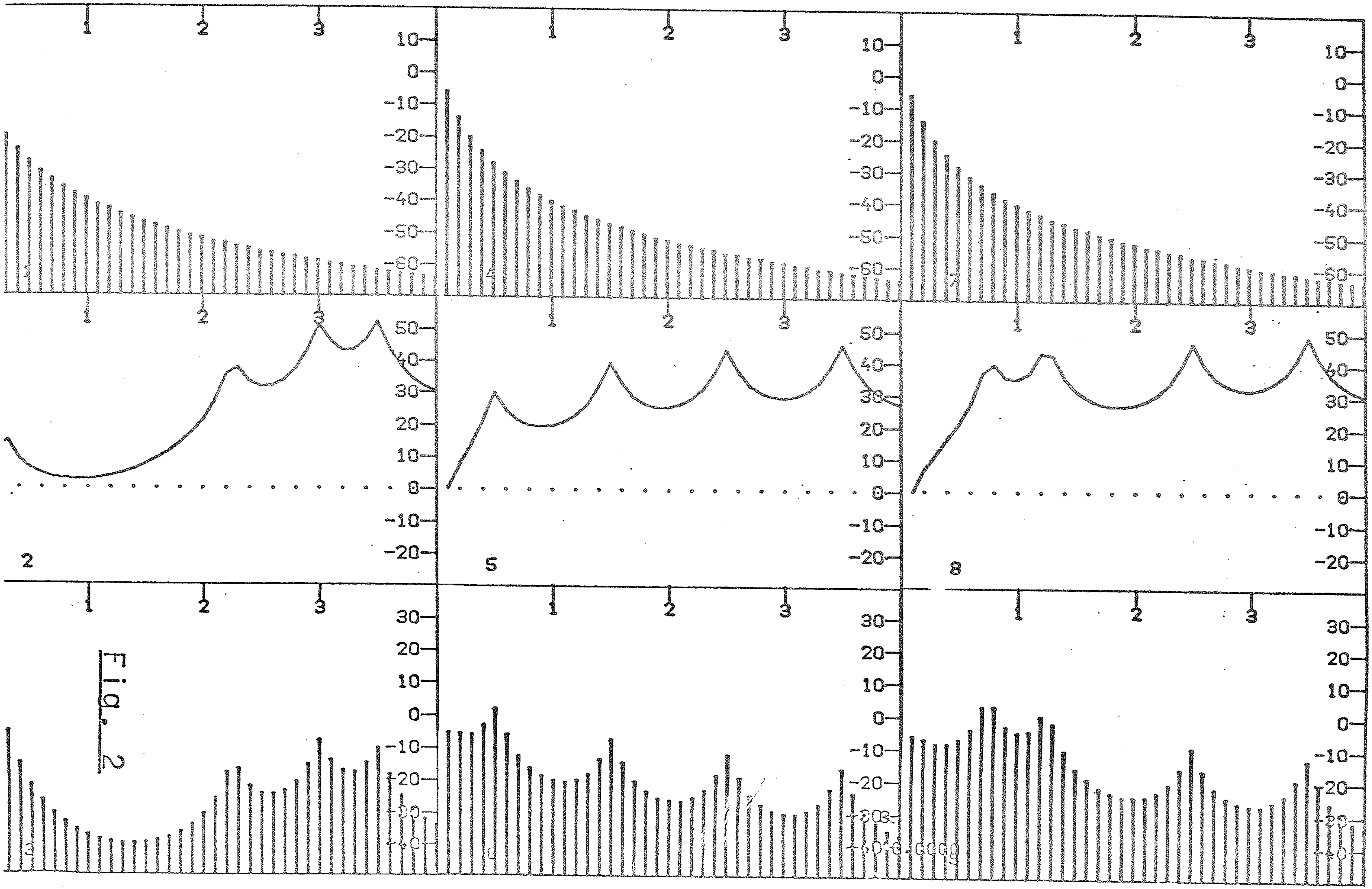


Fig. 2

Dialectal variation and constancy in Swedish word accent patterns

Per Lindblad

On the basis of E.A. Meyer's data (1937, 1954) on Swedish word accent F_0 patterns in isolated two-syllable words, E. Gårding (1970) has made a typology of the Swedish word accents, dividing them into four types according to the number and time location of F_0 maxima within the words. Choosing two informants from each of four Swedish dialects representing these four types, namely the Stockholm, Scanian, Gotland, and Västergötland dialects, E. Gårding and the present author have made an acoustic investigation of a number of one, two, three, and four-syllable words, pronounced both in isolation and within a sentence frame. The whole material was pronounced as neutral and emphatic statements, and neutral and emphatic questions. Mingograms have been made of the whole material, spectrograms of part of it.

In a first report (1973), we give data on the time and frequency positions of F_0 maxima and minima in isolated words for the four dialects. Further, on the basis of these data, we give a fairly formalised description of the F_0 contours for each of the dialects. The description is given as a number of rules, each generating the F_0 curve for part of the given test words. The frequency dimension is quantified into four pitch levels. The rules apply in a strict order and are subject to the principle that a later rule does not change the output of an earlier one. As concerns their function, there are three types of rules. First, there is a rule that caters for the terminating sentence intonation - the words are isolated, as you may remember. Secondly there are rules that take care of the accent distinction by allocating F_0 maxima and minima to specific time postitions and pitch levels. Thirdly, there is a rule that connects by the shortest way the parts of the F_0 shape of the word generated by earlier rules.

For all the four investigated dialects these types of rule are used, applied in the same order. As concerns the contents of the rules, the dialects differ. In the main, the F_0 contours of the two speakers of each special dialect may be generated by the same formulation of the rules.

For all the investigated dialects, the F_0 contour of a grave accent word may be regarded as a contracted F_0 contour of a minimally contrasting acute word plus an additional initial F_0 segment, called the pre-contour. This interdependence between the acute and grave accent seems to be an interdialectal constancy in the Swedish word accent opposition. Another constancy in the realization of the opposition found in our material can be formulated thus: The grave accent is contrasted to the acute either in having two F_0 peaks - acute has always one peak only - or in having a single peak later in the word than the acute accent.

References

Gårding E. 1970. Word tones and larynx muscles. Working Papers 3, Phonetics Laboratory, Lund University

Gårding E. and P. Lindblad. 1973. Constancy and variation in Swedish word accent patterns. Working Papers 7, Phonetics Laboratory, Lund University

Meyer E.A. 1937. Die Intonation im Schwedischen, I: Die Sveamundarten. Studies Scand. Philol. Nr 10, Univ. Stockholm

Meyer E.A. 1954. Die Intonation im Schwedischen, II: Die norrländischen Mundarten. Studies Scand. Philol. Nr 11, Univ. Stockholm

SENTENCE STRESS AND TONAL ACCENTS

Gösta Bruce

The present paper raises the question: how are the tonal accents in Swedish manifested when they occur in stress positions other than sentence stress position (cf. Gårding 1967)?

The inspiration of this preliminary study is the proposal of a binary suprasegmental system for English by Vanderslice & Ladefoged (1971), where sentence stress - 1-stress - is analysed as [+ ACCENT], [+ INTONATION], while other non-primary stresses are equated: [+ ACCENT], [- INTONATION], i.e. a sentence stressed syllable is distinguished from other stressed syllables by the contribution of sentence intonation - the tonal manifestation of statement or question - and not by degree of stress. The aim of this paper is to test the validity of a corresponding treatment of Swedish accentual phenomena. The hypothesis is that one should be able to subtract what depends on sentence intonation by looking at the manifestations of the two contrasting tonal accents - accent 1 and accent 2 - in stress positions other than sentence stress position and there find the 'true' tonal accents, i.e. the tonal accents unaffected by sentence intonation.

In order to test this hypothesis the following test sentences were devised:

1. Neutral statement: answer to the question "What were you doing?"

$$\text{vi lira'} \begin{Bmatrix} \text{bándy} \\ \text{kùla} \end{Bmatrix} \text{vi'} \begin{Bmatrix} \text{vállen} \\ \text{vàllarna} \end{Bmatrix} \left(\text{we played} \begin{Bmatrix} \text{bandy} \\ \text{marbles} \end{Bmatrix} \text{near the} \begin{Bmatrix} \text{slope} \\ \text{slopes} \end{Bmatrix}\right)$$

2. Statement, game emphasized; answer to the question "What game were you playing there?"

$$\text{vi lira'} \begin{Bmatrix} \text{BÁNDY} \\ \text{KÙLA} \end{Bmatrix} \text{vi'} \begin{Bmatrix} \text{vállen} \\ \text{vàllarna} \end{Bmatrix}$$

3. Statement, place emphasized; answer to the question "Where were you playing the game?"

vi lira' {bândy / kûla} vi' {VÀLLEN / VÀLLARNA}

The dialect under examination is the Stockholm dialect.

From earlier investigations we know that in the Stockholm dialect the Fo-curve of the grave accent has two peaks with a valley in between. The first peak is located in the stressed syllable and the second peak is in the last syllable of the word. The acute accent is usually characterized by a rise in the stressed syllable, often preceded by a small dip (Öhman 1967). The Fo-data of the present study show that in sentence stress position the manifestations of accent 1 and 2 are as described above. But in other stress positions we find a different picture: the grave accent has only one peak, in the stressed syllable, while the second one is missing. The acute accent is manifested merely as a valley.

Now how do we interpret the data? It may be suggested that the second peak of the grave accent, which is present only in sentence stress position, is the tonal manifestation of statement intonation, so that the first peak of the grave accent is what actually embodies the grave word accent. But what about the acute accent? One interpretation is that the valley, which is the manifestation of the acute accent in non-primary stress position constitutes the 'true' acute word accent. In sentence stress position this valley will be smoothed out by the sentence intonation peak, so that we will get the picture of a mainly rising acute accent. It has been shown that the Fo-pattern characteristic of the acute accent recurs in the later part of the grave accent, i.e. the second peak (Gårding 1970). So the same statement intonation pattern is found both in the grave and the acute accent in sentence stress position, but occurs earlier for the acute accent, so that the acute valley is affected by the sentence intonation peak, while the grave peak is left unaffected by it.

BIBLIOGRAPHY

Gårding E. 1967. Prosodiska drag i spontant och uppläst tal. Svenskt talspråk (G. Holm red.), 40-85, Stockholm

Gårding E. 1970. Word tones and larynx muscles. Working Papers 3, Phonetics Laboratory, Lund University

Öhman S. 1967. Word and sentence intonation: A quantitative model STL-QPSR II-III. Stockholm

Vanderslice R. and P. Ladefoged. 1971. Binary suprasegmental features. Working Papers in Phonetics 17. UCLA, Los Angeles

AN EMPTY SPACE ON THE MAP

Anne-Christine Bredvad-Jensen

Under this heading I talked about the twofold aim of the laboratory work the students are obliged to attend in order to carry out the basic courses in phonetics in Lund. 1. In the first place we want to create stimulating tasks for the students. We consider it meaningful for them to "discover new lands" in phonetics. 2. In the second place we are interested in the description and classification of dialects by prosodic criteria. This term, spring 1973, the combination of these aims has resulted in one group's concentration on the F_o pattern in the accent 1 and accent 2 words of the inhabitants in Kronoberg county in Småland (in the south of Sweden). This work is an offshoot especially from Gårding and Lindblad's work (1973). See also Lindblad's paper on Dialectal variation and constancy in Swedish word accent patterns in this volume (p. 49). Our starting point was the tonal dialect map of Sweden given by Gårding and Lindblad (p. 48) and Gårding's tentative tonal typology for accent 1 and accent 2 words in Swedish (p. 46). The southern part of the map is shown in Fig. 1. As shown on the map, the tonal data are unevenly distributed over the country. For example, north of Skåne (in the south of Sweden) there is an "empty space on the map". This area has traditionally been regarded as belonging to the South-Swedish dialect area. The tradition is based mainly on non-prosodic phonetic features. There is also some evidence regarding perceptual data that people in this area "need a fundamental frequency rise to perceive a grave word and a fall to hear an acute" in the stressed syllable just as the Skåne inhabitants do, see Johansson (1970, p. 70).

The material used consisted of disyllabic minimal pairs with tonal accent contrast and with the first syllable stressed, and the corresponding monosyllabic word, e.g. pålen - Polen, Pål. The informants were five phonetically naïve persons from Kronoberg county. Each informant was recorded twice and at different sessions.

Contrary to what has often been assumed our data showed no conformity to the Skåne prosodic patterns. The outstanding feature which all the five informants had in common was a two-peaked accent 2 pattern with one peak in each syllable. Fig. 2 shows the typical accent 2 pattern. This pattern resembled both the Svea dialects (type 2A in Gårding's tonal typology) and the Göta dialects (type 2B). The timing of the peak within the syllable seemed to be irrelevant because no consistent pattern was found. The acute accent showed a still more varying F_0 pattern, probably due to varied intonation, but the tonal peak was always in the stressed syllable and it was mostly late in the syllable also, as shown in Fig. 3. Consequently there is no similarity with the acute pattern in type 2B, as this pattern has its tonal peak in the post-tonic syllable. The character of the tonal accent in the dialect of these five informants thus seems to fit fairly well with type 2A in Gårding's typology, that is the Svea dialects (in Central Sweden).

Needed: more informants from Småland.

Volunteers may address themselves to: Phonetics Laboratory, Lund University.

References

Gårding E. and Lindblad P. 1973. Constancy and variation in Swedish word accent patterns. Working Papers 7. Phonetics Laboratory, Lund University

Johansson K. 1970. Perceptual experiments with Swedish disyllabic accent 1 and accent 2 words. Working Papers 3. Phonetics Laboratory, Lund University

Lindblad P. 1973. Dialectal variation and constancy in Swedish word accent patterns. Working Papers 8. Phonetics Laboratory, Lund University

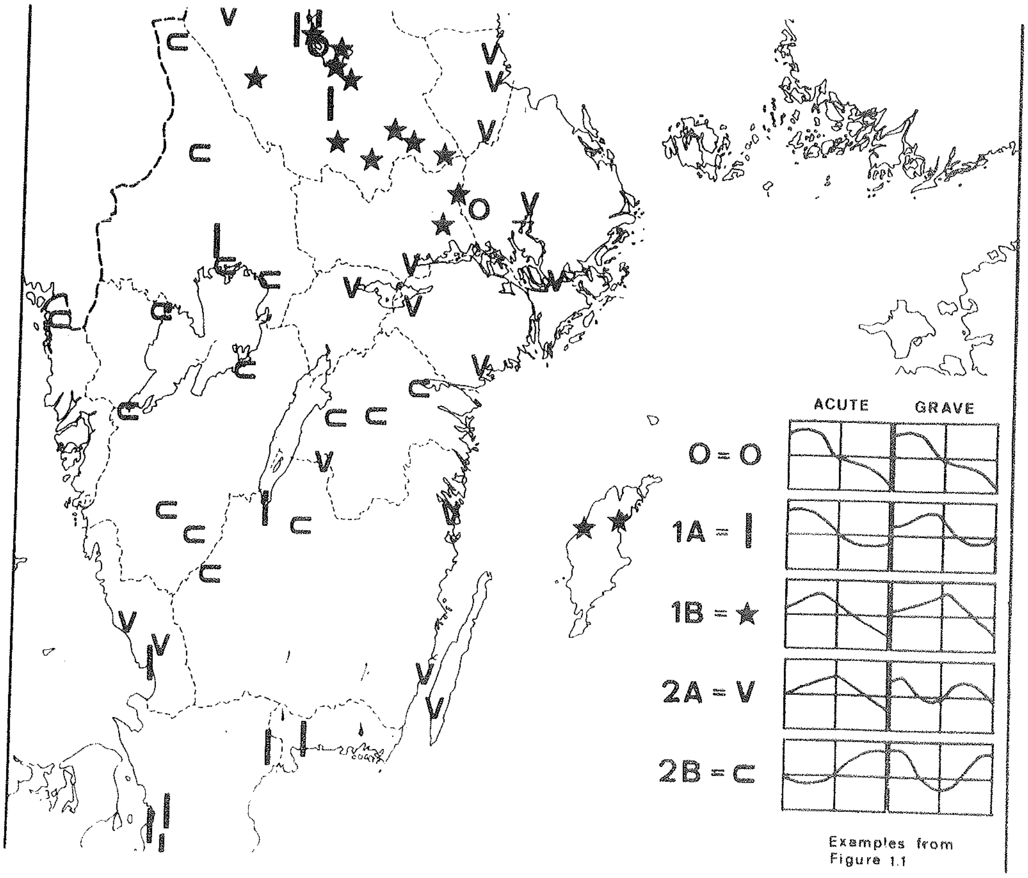


Fig. 1.

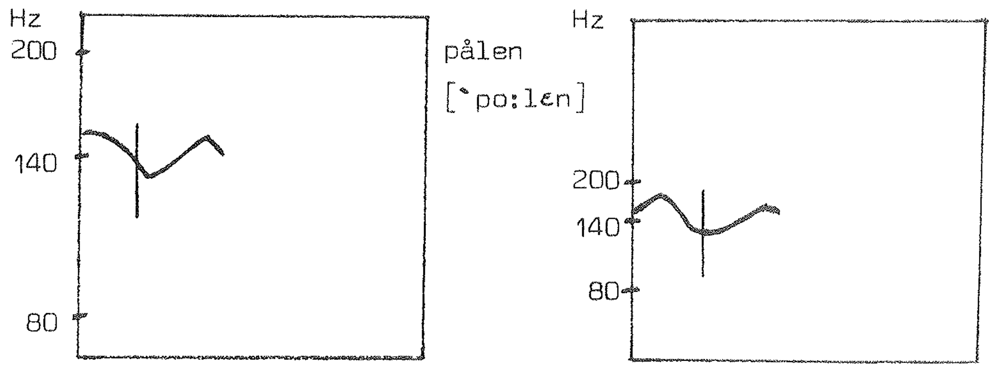


Fig. 2. Accent 2. Speaker: CJ, Alvesta Speaker: PT, Ryd

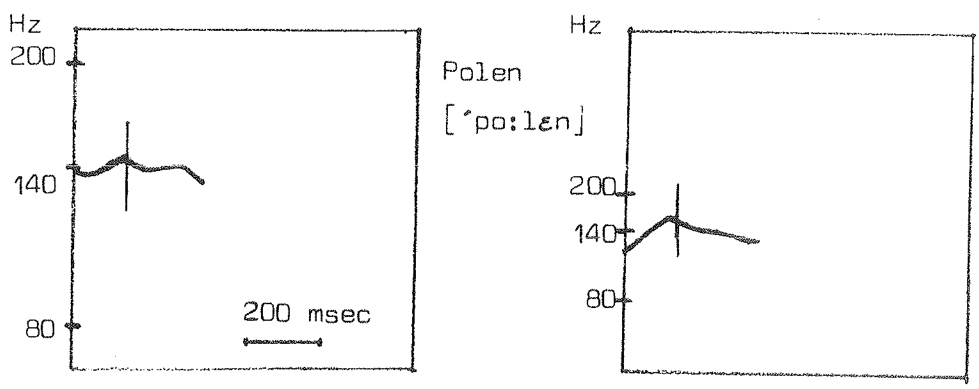


Fig. 3. Accent 1. Speaker: CJ, Alvesta Speaker: PT, Ryd

COORDINATION OF FUNDAMENTAL FREQUENCY AND ARTICULATION IN SWEDISH ACCENT II WORDS

Ylva Erikson

Fundamental frequency correlates of the Swedish word accent were investigated with special attention to the coordination of F_0 and articulation. A previous study found only minor variations bewteen F_0 contours from grave accented syllables belonging to words differing in syllable number. A vowel in such a syllable has a F_0 maximum in the beginning preceeded by a slight rise and followed by a relatively steep fall. To certify the importance of the final fall to the percieved accent type, vowel duration was varied by a systematic manipulation of phonological vowel length, voicing of the following consonant and number of syllables. It was shown that a decrease in vowel duration did not bring about a reorganization of the F_0 pattern, but resulted in a final truncation. An attempt was then made to vary the temporal conditions in the beginning of a given F_0 contour, by changing the number of intervocalic consonants in compound words with stress and accent pattern 32. The F_0 contour in syllables with stress level 2 is often described as rising because of the F_0 rise in the vowel. As the number of consonants between the two vowel onsets increased, the temporal distance increased accordingly. It was found that temporal perturbations of this kind did not bring about a reorganization of the contours. There was no truncation of the initial rise, instead the entire contour was displaced forward in time as the second vowel onset was delayed. These findings suggest that truncation is only possible at the final part of the contour and that F_0 is coupled to the vowel onset.

Litterature:

Alstermark M. and Erikson Y. 1971. Swedish word accent as a function of word length. STL-QPSR1, 1-13

Erikson Y. and Alstermark M. 1972. Fundamental frequency correlates of the grave word accent in Swedish: The effect of vowel duration. STL-QPSR 2-3, 53-60

Erikson Y. Forthcoming

ON INTRINSIC AND EXTRINSIC F_0 VARIATIONS IN SWEDISH TONAL ACCENTS.[1]

Anders Löfqvist

The importance of fundamental frequency variations for the distinction between the Swedish tonal accents is well established. This paper deals with the F_0 in the stressed vowel of words with different tonal accents and examines how it is affected when the consonantal environment of the vowel is changed from voiceless to voiced consonants and when the vowel itself is changed from long to short. The speech material consisted of nonsense words and was read by three speakers representing two Swedish dialects.

As expected from other investigations the duration of the vowel was longer before voiced than before voiceless consonants, the duration in the latter case being 85-90% of the duration in the former case. After voiced consonants the peak F_0 of the vowel was about 15 Hz lower than after unvoiced consonants. The influence of the preceding consonant on the fundamental frequency of the following vowel was, however, not confined to the beginning of the vowel but was still present when the F_0 peak was located near the end of the vowel. The interval from the onset of the vowel to the frequency peak increased about 25 msec as the preceding consonant was changed from voiceless to voiced. These variations probably reflect universal processes and were the same for all speakers and accents.

When the F_0 curves for long and short vowels were compared the location of the frequency peak was found to remain constant relative to the nearest boundary of the vowel and thus varied for different speakers and accents.

The variations are discussed in relation to the underlying mechanisms which cause them to occur and to their rôle in the speech communication process.

---

1. The full text will appear in POLA Reports. This work was supported in part by a grant from the National Science Foundation, GS 2386 A1, to Phonology Laboratory, University of California, Berkeley.

TONE ACCENT PATTERNS OF CHILDREN AND FOREIGNERS

Kurt Johansson

As quite a few of the papers presented at this symposium deal with prosody, I want to take the opportunity to report some results from investigations carried out primarily by second-term students of phonetics in connection with some seminars I led in 1971.

A. One of the investigations concerned the tone accents of children, 2–6 years of age[1]:

Some bisyllabic words (13 in all, 6 with the acute and 7 with the grave tone accent), known to belong to the vocabulary of the children, were illustrated by simple pictures and shown to 35 children from Malmö (Scania)[2]. Their pronunciations of the words were recorded on tape. At the same time one of the investigators made an interpretation of the pronounced words as acute or grave. Further interpretations by the investigators, and by a Scanian phonetician, in most cases confirmed this first interpretation. So did fundamental frequency analysis.

Results

The accent distinction is acquired rather late (fig. 1), with the age of 4 as a rather sharp boundary. There are, however, great variations among the younger children. Some learn the distinction early, some considerably later. There also seems to be a gradual acquisition of the acute accent, which is the one totally missing from the beginning. As could perhaps be expected, and at least judging from this investigation, children do not apply a certain rule consistently at once, but the accents

ACUTE ACCENT

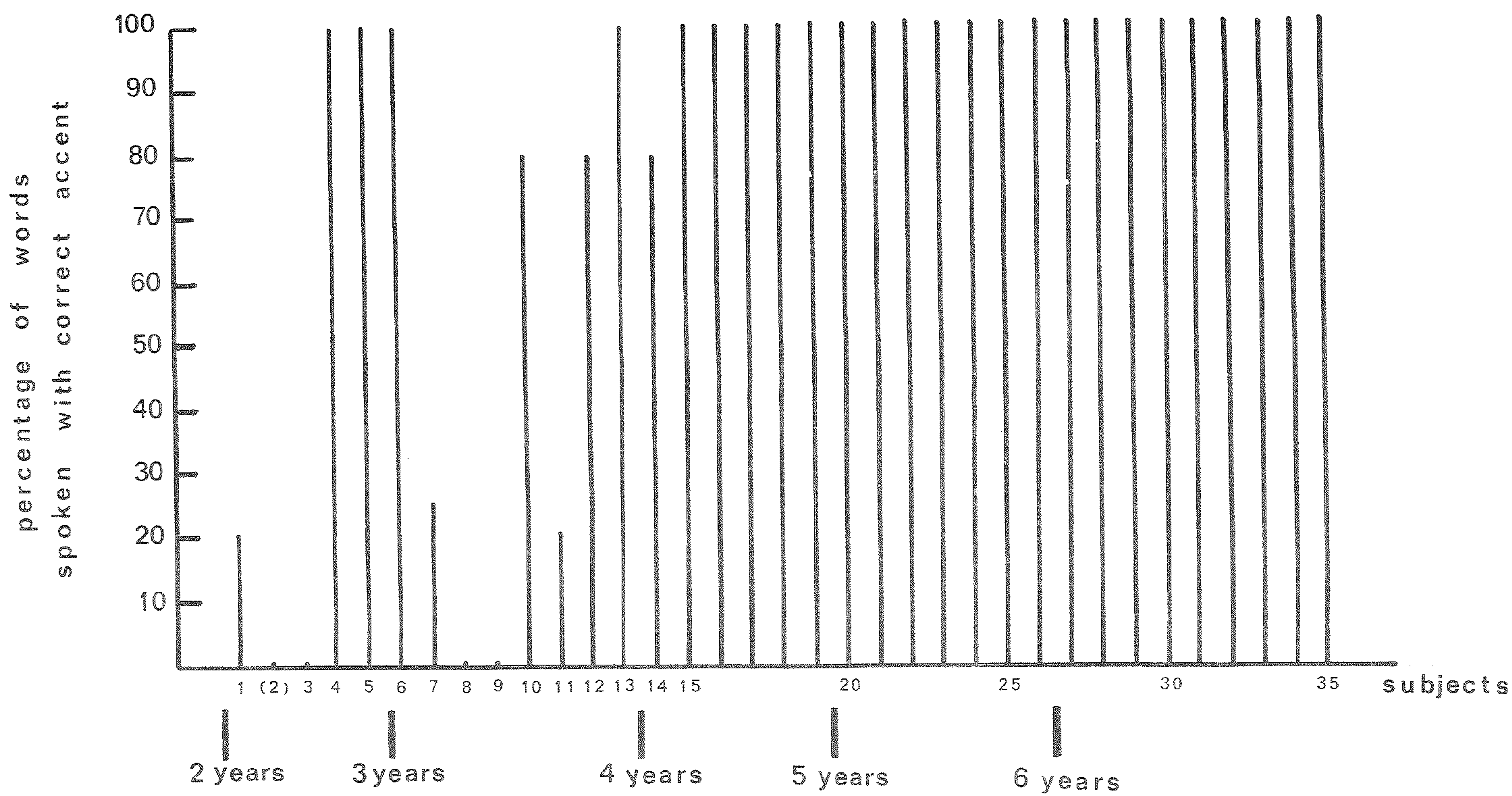


Fig. 1. The diagram shows how the acute accent is often infrequent in the speech of subjects younger than 4 years.

of common everyday words are evidently learnt word by word, not type by type. The test was not, however, designed to examine this.

B. Another investigation concerned the realisations of the Swedish tone accents by foreigners and the interpretation by Swedes of the tone accent patterns produced by these foreigners[3]:

A preliminary investigations of Polish and Czechish speaking people, learning Swedish as a foreign language at voluntary evening classes in Malmö, revealed that none had acquired the distinction between the acute and the grave accents, although they were said to belong to the most advanced of the students. Nor was the distinction clear for some American students living in Lund.

The investigator then looked for informants who had been living in Sweden for a long time, and who had tried to learn the language also in its more advanced forms. A small but rather homogeneous group was tested. The group consisted of three people only: one man and two women, all 35–40 years of age, two of them born in England, one in Scotland, but all three grown up in England. Two of them were teachers at the English Institute, Lund University, and the third was a phonetician. All three had been living in Sweden for 13–17 years and were married to Swedish (Scanian) people.

Twelve minimal pairs, consisting of everyday Swedish words, were recorded.[4] As preliminary tests had revealed considerable difficulties with respect to the tone accents, the acute and the grave accent words always appeared together in the reading list, sometimes with the acute, sometimes with the grave accent word first, like this:

Vinden viner i trädtopparna. viner – viner – viner

Många viner kommer från Frankrike. viner – viner – viner

Dessa strider var mycket jämna. strider - strider - strider

Detta strider mot reglerna. strider - strider - strider

Många gifter sprids i naturen. gifter - gifter - gifter

Många gifter sig mycket unga. gifter - gifter - gifter

In this way the speakers were given every help to understand what the test was about and thus to realize the differences.

The second of the three isolated versions of the words were listened to by the investigators. For every speaker those four words which were considered to be the most successfully pronounced grave examples[5] were used in a listening test together with four acute words (not minimal pairs in order to avoid a discrimination choice). What is said below is thus valid only for these "best" representations, and only for isolated words.

10 listeners from Southern Sweden (as all speakers were expected to have been influenced by the Scanian accent patterns) listened to these words appearing twice in a randomized test.

Results

As could be expected the grave accent presented the greatest difficulties. Speaker A (fig. 2) was more successful with this accent than the others. It seemed to be the other way around with the acute, strangely enough, but this speaker was more conscious of the purpose of the study, which may have caused him to overemphasize the grave type. Had, however, the selection of test words been made quite by chance, this would certainly have resulted in still fewer grave responses. Once more it should be stressed that all the speakers are linguistically conscious people with time enough to be able to realize linguistic rules. In running speech there ought to be less time to remember these rules, particularly as they only apply in stressed positions.

	A	B	C
Acute	69 %	79 %	76 %
Grave	80 %	43 %	40 %
Together	74 %	61 %	58 %

Fig. 2. Percentage of words where listeners' judgments of accent agreed with the intentions of speakers A, B and C.

Acoustic measurements of the words produced and judged as grave showed that the fundamental frequency contour in all cases was characterized by two frequency maxima. Furthermore, the second syllables of these words had greater intensities and longer durations. The fundamental frequency contour with two maxima is certainly not typical of the Scanian dialects, which was the type one might have expected from these speakers. However, it is the pattern of the standard central Swedish pronunciation commonly used as a model for language teaching.

Almost at the same time as these investigations were carried out a recording was made in the department of an interview with a German-speaking phonetician.[6] He had been living in Sweden for 6-7 years. His grave words, too, revealed two frequency maxima. (They generally sound, according to my own oppinion, very Swedish).

This could be compared with the results of an investigation concerning how Swedes and foreigners try to describe how they perceive the Swedish tonal accents.[7] Contrary to what was the case for the acute accent, 50

Swedes (of different dialects) and 24 foreigners chose, both for a Scanian speaker (with only one frequency maximum in the grave words) and for a Central Swedish speaker (with two maxima), to illustrate the grave accent with symbols indicating two maxima, e.g.  for the acute accent and  for the grave. To me, this seems to be an indication that some sort of loudness feature is involved here.

Notes

1. Investigators were: E. Holmberg, S. Lawrischin, and J. Sörensson.
2. Number of children in the different age classes:

2–3 years	–	5
3–4 years	–	8
4–5 years	–	7
5–6 years	–	6
6–7 years	–	9

3. Investigators were: K. Ebrelius Nilsson, B. Epps, A. Jönsson, and E. Stéen.
4. The following minimal pairs were used:

   bíten (the piece, the bit): bìten (bitten),

   Pólen (Poland): pålen (the pole),

   stégen (the steps): stègen (the ladder),

   tánken (the tank): tànken (the thought),

   gífter (pres. of verb marry): gìfter (poisons [noun pl.]),

   ánden (the duck): ànden (the ghost),

   världen (the world): värden (values [noun pl.]),

   strider (pres. of verb fight): strìder (fights [noun pl.]), Óskar (Oscar): àskar (pres. of verb thunder),

slågen (battles [noun pl.]), slàgen (beaten up),

våken (the hole in the ice): vàken (awake),

víner (pres. of verb whizz): vìner (wines [noun pl.]).

5. The grave accent caused most problems.

6. Investigators were: K. Karlsson and I. Svensson. (The investigation was carried out under the guidance of R. Bannert.)

7. Investigators were: M. Elevant, M. Hamrén, and A.-C. Johansson. (The investigation was carried out under the guidance of E. Gårding.)

ON SEGMENT DURATIONS IN WORDS AND PHRASES

Karin Rapp

Two experiments have been conducted investigating durations of segments on the word and on the phrase level. A limited speech material consisting of acute accent words made up by nonsense syllables was used. On the word level the number of syllables and the position of the stressed syllable were varied. The following observations were made regarding stressed vowel durations:

1. There is a compensation for word length in terms of number of syllables such that the vowel duration decreases as a function of increasing word length. However this compensation is not complete.
2. Vowel duration is increased towards the end of the word i.e. there is a final lengthening effect.

The same observations were noted for unstressed vowels and for stressed and unstressed consonants. The number and position of words in a phrase gave similar results for the vowel durations but to a lesser degree.

The above observations are summarized by the following rule:

$$\text{Duration} = \frac{k \cdot D}{n_p^{\frac{m_p}{n_p}} \alpha_p \cdot n_w^{\frac{m_w}{n_w}} \alpha_w}$$

With a correction for increased duration in phrase initial position this rule has been applied in producing synthetic speech.

Litterature:

Lindblom B. and Rapp K. 1971. Reexamining the Compensatory Adjustment of Vowel Duration in Swedish Words. STL-QPSR 4, 19-25

Carlsson R., Granström B., Lindblom B., and Rapp K. 1972. Some Timing and Fundamental Frequency Characteristics of Swedish Sentences: Data, Rules, and a Perceptual Evaluation. STL-QPSR 4, 11-19

PERCEPTUAL EVALUATION OF PROSODIC RULES

Rolf Carlson and Björn Granström

In another paper to this symposium rules have been given that describe timing regularities in a limited set of Swedish sentences consisting of acute accented words. (1)

Based on observations on a F0 analysis of the same sentence material tentative rules for intonation have been designed. These rules presuppose a falling sentence intonation contour upon which are superimposed positive excursions initiated in stressed vowels and generally reaching their peaks during unstressed segments. These maxima are proportional to the duration of the initiating vowels.

The F0 rules are hence applied after the duration rules which have been supplemented with a rule taking account of observed lengthening in initial positions.

To look into the generality and communicative relevance of the rules arrived at they were used in producing synthetic speech. This synthesis was compared with human speech and with an alternative rule synthesis basically devoid of prosodic rules except a sentence final modification.

The sentences chosen did not contain any sentence from the observed material but had a less restricted structure. Listeners were asked to report the stress pattern in terms of stressed and unstressed syllables and also to indicate word boundaries. The result shows that the described rules used in speech synthesis gave the most accurate stress pattern identification and also contributed in word boundary judgement and hence they appear to serve a definite perceptual function.

Litterature:

(1) Rapp K. 1973. On Segment Durations in Words and Phrases. This symposium

Carlson R., Granström B., Lindblom B., and Rapp K. 1972. Some Timing and Fundamental Frequency Characteristics of Swedish Sentences: Data, Rules and a Perceptual Evaluation. STL-QPSR 4, 11-19

MUTUAL COMPLEMENTATION OF VC- SEQUENCES IN CENTRAL BAVARIAN

Robert Bannert

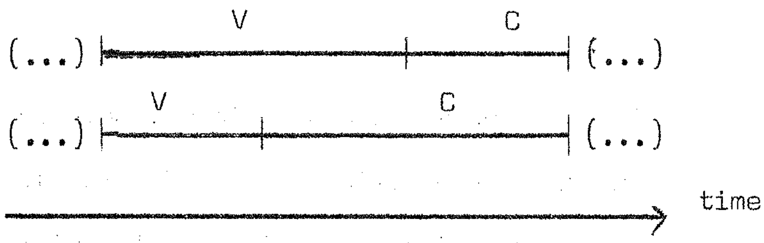
Ratios of duration and the domain of quantity

As a rule investigations of duration and quantity contain not only absolute values of segment durations but also several kinds of ratios due to the multitude of factors affecting the duration of a given segment (Elert 1964:51 f). Based on minimal pairs, contrasting optimally only in different durations of the identical segment, the vowel-to-vowel ratio (V/V:) and the consonant-to-consonant ratio (C/C:) are calculated. These two ratios being paradigmatic, the vowel-to-consonant ratio (V/C) refers to syntagmatic relations.

The ratio of V/C ratios (V:/C : V/C:), a double ratio, is considered basic to the perception of quantitative contrasts in a language (Elert 1964:171 ff).

Dealing with the phonological concept of quantity and its phonetic manifestations as duration (physiologically and acoustically) and length (perceptually) it seems necessary to start from the domain of quantity. In those languages which utilize durational contrasts distinctively, the domain of quantity may be different: the sound or sequence of sounds over which quantity is manifested may be only one segment (only the vowel, only the consonant, the vowel and the consonant), two segments (stressed vowel and the following consonant), or larger units (Lehiste 1970:42).

In the Nordic languages of Standard Swedish, Norwegian and Icelandic a reversed relationship, called mutual complementation (Lehiste 1970:49), between the duration of the stressed vowel and the following consonant is observed. Thus a long vowel is followed by a short consonant and a short vowel is followed by a long consonant:

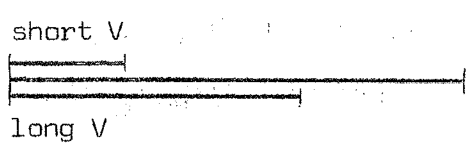


A similar reversed relationship between vowel and consonant is reported for the Upper German dialect of Central Bavarian (CB) (Kufner 1957, 1961. Bannert 1972). The domain of quantity in the Nordic languages as well as in CB is the VC-sequence which does not seem to be the case for Standard German.

The distribution of mutual complementation in CB is in general restricted to words longer than one syllable: $\# \; (\ldots) \; C^3 \; V \; C \; \ldots \; \#$, where the vowel is stressed and the second syllable contains a sonorant (vowel, lateral or nasal).

The vowel-to-sequence ratio V/VC

Due to the fact that distinctive durational (= quantitative) contrasts in CB always are manifested over the VC-sequence I assume that these two segments together constitute a timing unit at a certain level of the language. Instead of relating vowel duration to the duration of the following consonant, the duration of which varies considerebly after long and short vowel, by calculating the V/C ratio, it might therefore be more relevant to calculate the vowel-to-sequence ratio (V/VC ratio), thus expressing the duration of the vowel with reference to the whole VC-sequence.



VC-sequence (V+C) = 100 %

Material

To describe the durational relationships of VC-sequences in CB, some preliminary measurements based on mingograms (operation speed 100 mm/sec), the V/VC ratio and some observations pertaining to one informant G will be presented. One portion of the material was collected in December 1972, the other in March 1973. The pair Gegal [g̊e:gal] versus Gekal [g̊ekal] (name of a hill vs. grilled chicken) appeared in both recordings. This material is compared to measurements from Informants G and A of January 1972. The stressed vowel, the following (medial) consonant (C_m), and the prosodic patterns (prosodies) (cf. Gårding and Lindblad 1973) were varied. In order to affect segment duration heavily, four prosodies were chosen:

(1) neutral isolated statement (isolated words),

(2) neutral sentence statement, e.g. / i hab s Bupal gseng/, (I saw the little doll), where the test word is stressed,

(3) fast isolated statement (isolated words),

(4) emphatic sentence statement with topicalization, e.g. / s Bupal hab i gseng/.

The testword and the variables of the investigations are summarized in the table on the following page.

Results

Tables 1, 2, and 3 give the following values: the mean segment duration (msec) of the vowel and the consonant, the mean duration of the VC-sequence, the decrease (or increase, marked with +) in % of the observed means for prosodies (2), (3), and (4) relative to prosody (1) (neutral isolated statement), the V/C and V/VC ratios in %, the change of these ratios in % with reference to prosody (1), and the range of their variation in %.

Test-word	Prosody	Mean duration (msec) V	C	V+C	Decrease from prosody (1) in % V	C	V+C	Ratios % V/C	V/VC	Change of ratios % V/C	V/VC	Range of variation % V/C	V/VC
Biabal (long vowel)	(1)	138	50	188	0	0	0	276	73	0	0	119	8
	(2)	128	39	167	7,2	22,0	11,2	328	77	52	4		
	(3)	100	43	143	27,5	13,5	23,9	233	70	−43	−3		
	(4)	[illegible]	[illegible]	176	0,7	22,0	6,4	352	78	76	5		
Sibal (long vowel)	(1)	116	56	172	0	0	0	207	67	0	0	95	9
	(2)	106	45	151	9,4	19,6	12,2	236	71	29	4		
	(3)	94	51	145	19,0	8,9	15,7	184	65	−23	−2		
	(4)	117	42	159	+ 0,9	25,0	7,5	279	74	72	7		
Supal (short vowel)	(1)	79	150	229	0	0	0	53	35	0	0	27	9
	(2)	66	121	187	16,5	19,3	18,3	55	35	2	0		
	(3)	59	88	147	25,3	41,3	35,8	67	40	14	5		
	(4)	84	105	189	+ 6,0	30,0	17,5	80	44	27	9		
Feda (long vowel)	(1)	143	58	201	0	0	0	246	71	0	0		
	(3)	101	38	139	29,4	34,5	30,8	266	73	20	2		
Feta (short vowel)	(1)	104	169	273	0	0	0	62	38	0	0		
	(3)	83	124	207	20,2	26,6	24,2	67	40	5	2		

Table 1. Mean duration (msec) of the vowel V, the following consonant C, and their sum (V+C) of five testwords and different prosodies from the recordings of December 1972, medial consonant C_m = /b/ and /d/, informant O. Decrease (or increase in a few cases, marked by a +) in % of the durations of V, C, and V+C from prosody (1) = the isolated testword as a neutral statement. V/C and V/VC ratios in %. Change of these ratios with reference to those of prosody (1). Range of variation in % of these ratios.

Test-word	Prosody	Mean duration (msec) V	C	V+C	Decrease from prosody (1) in % V	C	V+C	Ratios % V/C	V/VC	Change of ratios % V/C	V/VC	Range of variation V/C	V/VC
Gekal	(1)	93	155	248	0	0	0	60	38	0	0	25	8
	(2)	83	123	205	10,8	21,4	17,3	68	40	8	2		
	(3)	66	94	160	29,0	39,4	35,4	70	41	10	3		
Dec 1972	(4)	92	108	200	1,1	30,3	19,3	85	46	25	8		
Gegal	(1)	123	64	187	0	0	0	193	66	0	0	52	6
	(2)	109	65	174	11,4	+ 1,5	6,9	168	63	−25	− 3		
	(3)	79	49	128	35,8	23,4	31,5	161	62	−32	− 4		
	(4)	112	53	164	8,9	17,2	12,3	212	68	19	2		
(1 Gekal	(1)	112	196	308	0	0	0	57	36	0	0	19	7
	(2)	84	131	215	25,0	33,2	30,2	64	39	7	3		
	(3)	93	123	216	17,0	37,2	29,9	76	43	19	7		
March 1973	(4)	89	120	209	20,5	38,8	32,1	74	43	17	7		
Gegal	(1)	173	87	260	0	0	0	199	67	0	0	72	8
	(2)	122	68	190	29,5	21,8	26,9	180	64	−19	− 3		
	(3)	118	65	183	31,8	25,3	29,6	182	65	−17	− 2		
	(4)	126	50	176	27,2	42,5	32,3	252	72	53	5		

Table 2. Same as Table 1, but only for the pair Gegal–Gekal from two different recordings (December 1972 and March 1973), informant G.

Test-word	Prosody	Mean duration (msec) V	C	V+C	Decrease from prosody (1) in % V	C	V+C	Ratios % V/C	V/VC	Change of ratios % V/C	V/VC	Range of variation % V/C	V/VC
Dafal (long vowel)	(1)	167	116	238	0	0	0	144	59	0	0	53	7
	(2)	137	79	216	18,0	31,9	23,6	174	63	30	4		
	(3)	133	75	208	20,3	35,4	26,4	178	64	34	5		
	(4)	138	70	208	17,4	39,6	26,4	197	66	53	7		
Nasn (long vowel)	(1)	177	111	288	0	0	0	159	61	0	0		
	(3)	113	72	185	36,2	26,1	35,8	157	61	–2	0		
nassn (short vowel)	(1)	132	190	322	0	0	0	70	41	0	0		
	(2)	98	148	246	25,8	22,0	23,6	66	40	–4	–1		
biassn (short vowel)	(1)	135	192	328	0	0	0	70	41	0	0		
	(3)	95	138	233	29,6	28,2	29,0	68	41	–2	0		
baissn (short vowel)	(1)	144	191	334	0	0	0	75	43	0	0		
	(3)	111	150	261	22,9	21,4	21,8	74	43	–1	0		

Table 3. Same as Table 1, different testwords (C_m = /f, s/), informant G.

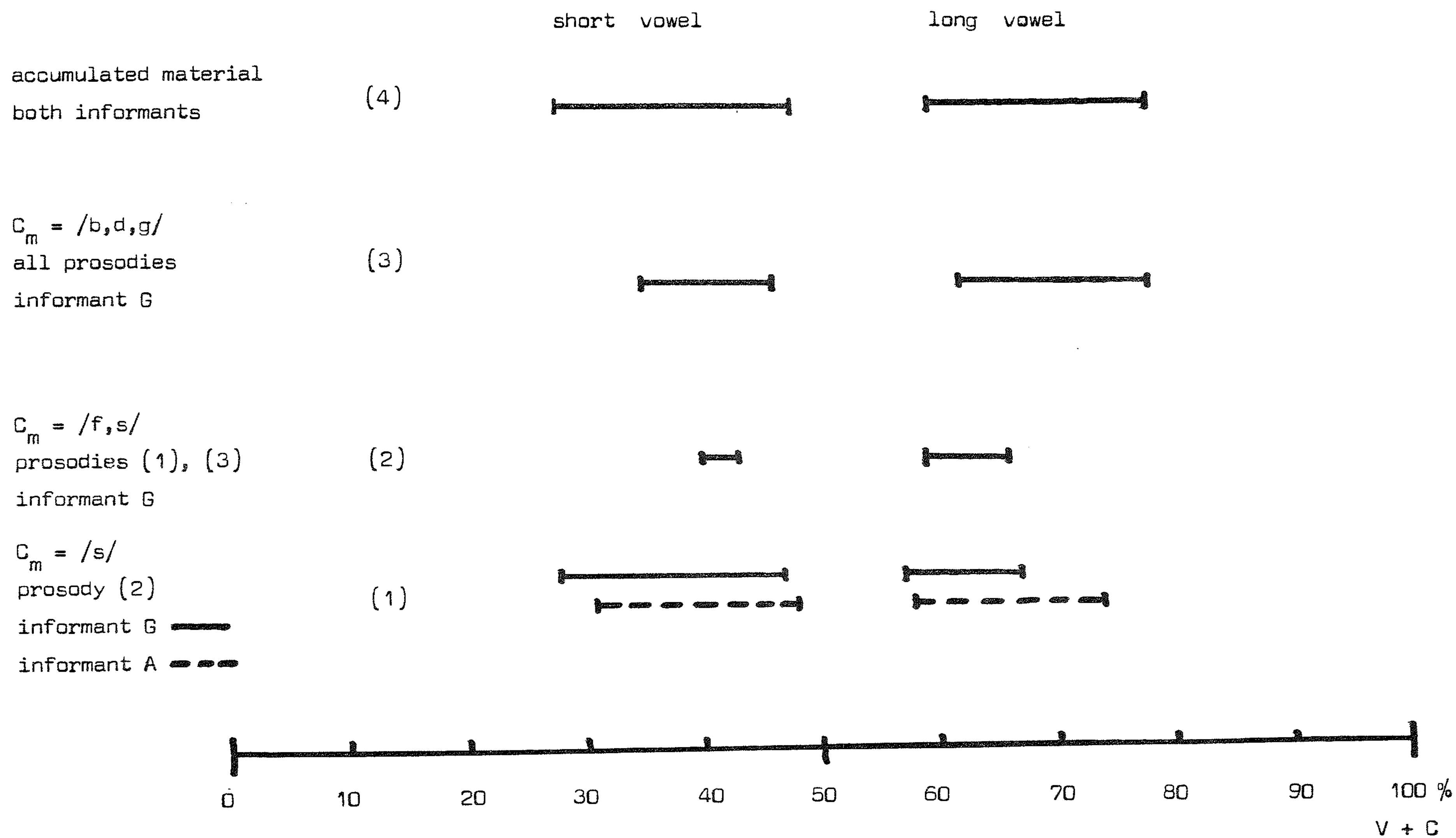


Diagram showing the V/VC-ratios of the accumulated material (4) and grouped according to certain variables (1-3). See text.

Date and Informants	Test words: longvowel short cons.	Test words: short vowel long cons.	Prosodic patterns (prosodies)
Jan 72 G, A	C_m = /S/: bäsa Wiesn Basl etc. (Bannert 1972:9)	grässa wissn Massl	(2)
Dec 72 G	Biabal Bibal Gegal	Bupal Gekal	(1), (2) (3), (4)
	Feda Nasn	Feta nassn biassn baissn	(1), (3)
March 73 G	Gegal Dafal	Gekal	(1), (2) (3), (4)

The diagram shows the V/VC ratios for the accumulated material, containing all variables for informants G and A, and the V/VC ratios grouped according to the following variables:

(1) C_m = /s/, informants G and A, January 1972, (2) C_m = /f,s/, informant G, December 1972 and March 1973, (3) C_m = /b,d,g/, informant G, December 1972 and March 1973, (4) the complete material reported, informant G.

The following observations can be made:

1. The sum of the absolute durations of short vowel and long (voiceless) consonant (V+C) is always greater than that of the corresponding long vowel and short (voiced) consonant. This is in agreement with Elert's findings for Standard Swedish (Elert 1964:160).

2. The value of the V/VC ratio is always largest in prosody (4), that is,

the vowel duration increases relative to the duration of the VC-sequence when a given word is produced at the very beginning of an emphatic sentence statement. The value of the V/VC ratio for short vowel and following long and voiceless stop consonant (Bupal, Gekal, Feta) is smallets in prosody (1) (isolated neutral statement). The order of the V/VC values of the three other prosodies does not vary systematically for long vowels. This might be explained, at least partially, by the fact that short consonants after long vowels are often, but not always, produced with voicing, thus obscuring potential tendencies.

3. The range of variation of the V/VC ratio for each test word and its prosodies is rather constant, about 8 %, while the V/C ratio varies differently for long and short vowels, from 52 to 119 % for the long vowel and from 19 to 27 % for the short vowel:

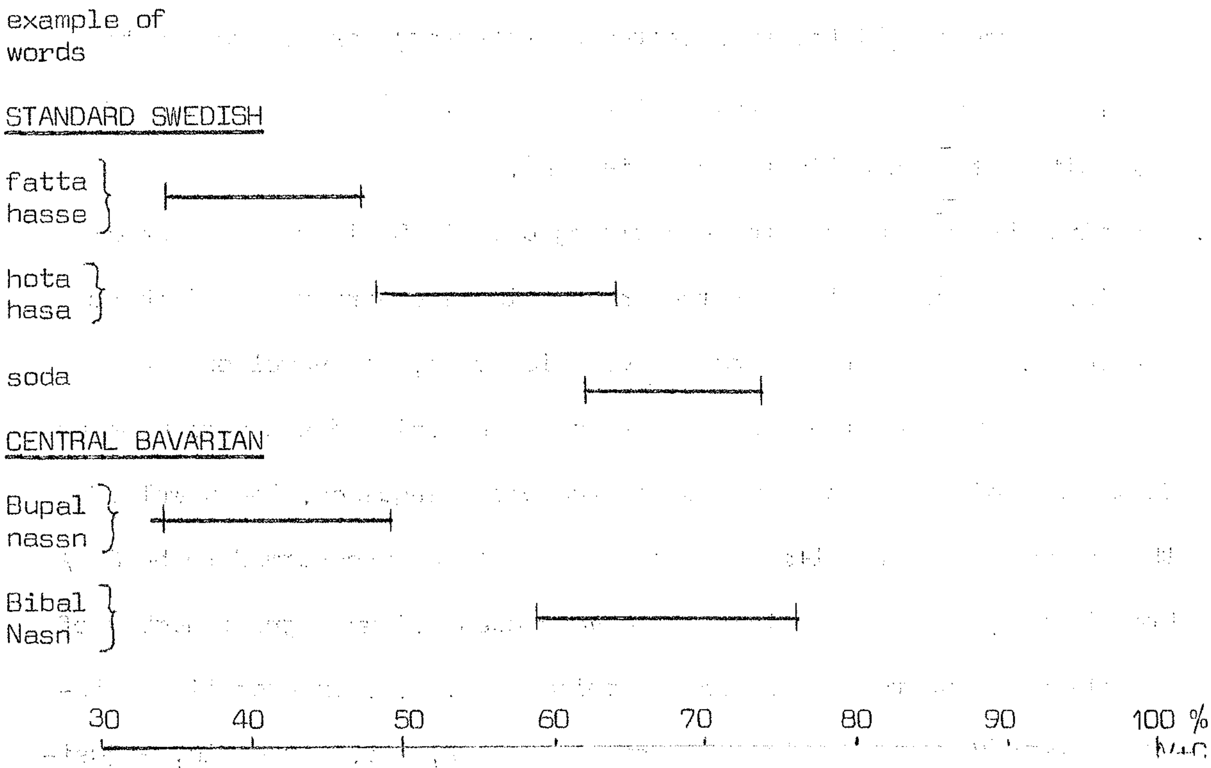
Test word	range of variation in % V/C	V/VC
Biabal	119	8
Bibal	95	9
Gegal	52	6
	72	8
Dafal	53	7
Bupal	27	9
Gekal	25	8
	19	7

4. The range of the V/VC ratios for the whole material of informant G (Jan and Dec 1972, and March 1973), all the segmental variations (vowel and consonant), and the four prosodies appears to be the same for short and long vowel, about 20 %. The V/VC ratio for the short

vowel of prosody (2) (neutral sentence statement) varies from 28 to 47 %, that for the long vowel from 57 to 77 %. All the other prosodies fall within the same limits (see diagram).

5. Assuming the VC-sequence to be a timing unit in CB it does not seem difficult to write rules for the programming and performing of these sequences. In order to program disyllabic words, the vowel of the sequence short vowel+long consonant would be assigned a duration being 33 % (one third) of the duration of the whole sequence. The vowel of the sequence long vowel+short consonant would be programmed to be 66 % (two thirds) of the duration of the VC-sequence. Then a great number of adjustments have to be made or are being made, e.g. in emphatic statements the V/VC ratio is to be increased by v %, etc. Automatic, adjustments not due to lexical conditioning will change the V/VC ratio of the acoustic output, e.g.
   (a) if the vowel is an open monophthong or a diphthong, the V/VC ratio will increase by x %.
   (b) if the vowel is closed, the V/VC ratio will decrease by y %.
   (c) if the short consonant is produced voiceless, the V/VC ratio will decrease by z %.

6. A comparison of V/VC ratios for the CB material and corresponding Standard Swedish (SS) material (70 words by 8 speakers) calculated from Elert's measurements (Elert 1964:91 ff) shows that the V/VC ratios for the short vowel in both languages coincide. Short stop consonants in SS may be either voiceless or voiced, thus the V/VC ratios for short voiced consonants in SS are higher than those for voiceless consonants. In CB short consonants, both stops and fricatives, are usually manifested voiced. Therefore the V/VC ratios for short consonants in CB are like those for short voiced consonants in SS, that is, they are higher than those for short voiceless consonants:



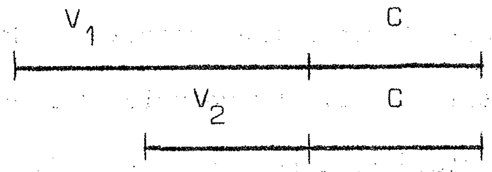
7. The V/C and the V/VC ratios change in parallel, reflecting syntagmatic relations. But as to my CB informants, the V/VC ratio appears to yield a simpler description of the data:

(a) the range of V/VC ratios for the informants is identical (about 20 %) for <u>both</u> long and short vowel,

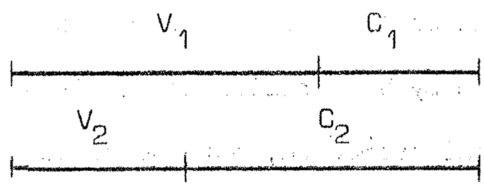
(b) the range of variation of the V/VC ratios is about 8 % for the four prosodies of each test word,

(c) there is good correspondance between the V/VC values of the two informants of CB (material of January 1972, Bannert 1972:53) and between CB and SS.

Calculating the V/C ratio, the duration of the vowel (the denominator of the ratio) is expressed as a function of the duration of the consonant (the numerator of the ratio). From a mathematical point of view this ratio is meaningful if the numerator is constant. Thus the V/C ratio should be an appropriate measure in cases where vowels, the durations of which vary considerably, are followed by a consonant with a rather constant duration as depicted by the following figure:



But if the V/C ratio is calculated for languages with mutual complementation where the numerator varies as much as the denominator this ratio consists of two variables indicated by the following figure:



As shown above, the duration of the vowel + consonant sequence functions as a constant numerator in the V/VC ratio.

The reported data are, of course, very limited. But further measurements providing a more systematic treatment of mutual complementation in CB are in progress.

References

Abramson A.S. 1962. The vowels and tones of standard Thai: Acoustical measruements and experiments. Intern. Journal of Am. Ling. 28

Bannert R. 1972 (a). Die deutschen Lautsysteme. Eine phonologische und instrumentalphonetische Untersuchung. Licentiat thesis

Bannert R. 1972 (b). Zur Stimmhaftigkeit und Quantität in einem bairischen Dialekt. Phonetics Laboratory, Lund University, Working Papers 6

Elert C.-C. 1964. Phonologic studies of quantity in Swedish

Gårding E. and P. Lindblad. 1973. Constancy and variation in Swedish accent patterns. Phonetics Laboratory, Lund University, Working Papers 7:36-110

Hadding-Koch K. and A.S. Abramson. 1964. Duration versus spectrum in Swedish vowels: Some perceptual experiments. Studia linguistica 18:94-107

Kufner H.L. 1957. Zur Phonologie einer mittelbairischen Mundart. Zeitschrift für Mundartforschung 25:175-184

Kufner H.L. 1961. Strukturelle Grammatik der Münchner Stadtmundart

Lehiste I. 1970. Suprasegmentals

Nooteboom S.G. and I.H. Slis. 1972. The phonetic feature of vowel length in Dutch. Language and Speech 15:301-316

A BINARY TREATMENT OF THE DISTINCTIVE PROSODIC FEATURE OF QUANTITY

Robert Bannert

The theoretical framework of generative phonology (Chomsky and Halle 1968) covers only the segmental part of the phonological aspects of the lexicon. Prosodic (or suprasegmental) features are not dealt with because "Our investigations of these features have not progressed to a point where a discussion in print would be useful." (Chomsky and Halle 1968:329).

Contributions to a development of a prosodic framework have been made since then. A binary treatment of lexical word tone was proposed by Wang (1967) and of stress and intonation by Vanderslice and Ladefoged (1972). In order to complete the phonological framework, both segmental and prosodic, quantity, the third prosodic feature (Lehiste 1970), should be treated as well.

Much work has been done on sound durations and a large variety of durational variations have been reported. The causes of the observed variations of segment durations might be labelled phonetic and phonological. Regularities of durational changes as a result of compensatory adjustments, both segmental and within words, during performance (i.e. not distinctively controlled) were studied by e.g. Lindblom and Rapp (1972), Nooteboom (1972), Slis (1972), and Klatt (1973) who tried to capture these regularities by rules. Other systematic changes of sound durations, however, are due to the speaker's voluntary control of the timing of these sounds in order to distinguish between words. Duration functioning as a controlled feature of the phonological system of a language, independent of e.g. segmental context or number of syllables in the word, is called quantity (Lehiste 1970:42). Thus quantity is a distinctive feature of lexical items, i.e. it is not predictable and should not be confused with duration or length which are properties of the phonetic manifestation of abstract entities and can therefore be predicted.

Being supplied with the set of universal phonological features, a certain language need not use all of them but chooses a subset out of the total number. There are languages where quantity is not distinctive, for example Russian. For other languages, the status of quantity is not generally agreed upon, for example English. Again, for Spanish, it has been suggested by Ladefoged (1971:50) that the opposition of the medial consonant in pairs like _caro_ vs _carro_, _pero_ vs _perro_, etc. is not one of quantity but of manner of articulation, i.e. tap vs drill. On the other hand, languages such as Swedish, Finnish, and Italian make use of the distinctive prosodic feature of quantity. It may be manifested in different ways. The sound or sequence of sounds, the duration of which is controlled by the speaker for distinctive purposes, that is the _domain_ of quantity, may differ from language to language. It can be analysed as one segment (vowel or consonant, only the vowel, only the consonant), two segments (vowel and consonant) or larger units (Lehiste 1970:42). Restrictions in the distribution of quantitative contrasts may occur, e.g. for consonants in word initial position. Besides existing minimal pairs one can usually find potential minimal pairs where one member has not received any meaning yet.

In agreement with current phonological theory and as a starting point for the outline of a binary treatment of quantity, I assume the phonological representation of _lexical items_ (morphemes, formatives) to consist of segmental specifications (number of segments and the redundancy-free specification of their distinctive features). In addition, I postulate the prosodic representation of lexical items to contain the three features of STRESS, QUANTITY, and TONE in languages which utilize some or all of them. While lexical items are part of the grammar, _words_ are to be found in the phonetic manifestation (the substance) of abstract structures. I propose the following scheme from which, since it is universal, all languages may choose a certain set of features. The formalization of the segmental as well as the prosodic parts of the phonological representation of a lexical item is shown in the following figure:

/lexical item/

distinctive features:

(a) segmental	VOCALIC	[| | ...]
	CONSONANTAL	
	SONORANT	
	CONTINUANT	
	⋮	⋮ ...
	VOICE	
(b) prosodic	STRESS	
	QUANTITY	
	TONE	

TONE functioning here as a general label is further developed into a set of distinctive features like CONTOUR, RISE, HIGH, etc. (Wang 1967).

I do not consider the prosodic feature of QUANTITY to correspond to the tense/lax opposition of vowels which I take to be segmental and which ought not to be analysed as the feature LONG among the segmental features.

In a language with quantity, both members of a pair exhibiting distinctive durational differences contain the feature QUANTITY. One member will be specified for QUANTITY with a plus sign (+).

In the following I will illustrate the working of my proposed scheme applying it to some quantity languages, especially to Standard Swedish.

The phonological system of Standard Swedish which is analysed as containing long and short vowels and consonants shows the following distribution of these segments (on the level of phonetic manifestation):

Long vowel appears:

(a) ö, gå, bedra (...) V #

(b) barn, bord (...) V r C

<u>structural condition</u>:

C = voiced dental

(c) rät, veta (...) V C (...)

Short vowel appears:

(a) list, klump (...) V C_x C_y (...)

<u>structural condition</u>:

$C_x \neq$ r <u>and</u>

$C_y \neq$ dental

(b) sport, fors (...) V r C

<u>structural condition</u>:

C = voiceless dental

<u>exceptions</u>: art, etc.

(c) rätt, vetta (...) V C (...)

There may be dialectal or individual variations. They pertain to cases (a) and (b) and will not be discussed here.

Translations:				
	(a)	'island' 'go' 'deceive'	(a)	'cunning' or 'border' 'lump'
	(b)	'child' 'table'	(b)	'sports' 'rapids' 'species'
	(c)	'straight' 'to know'	(c)	'correct' 'to face'

The length of a simple consonant following a stressed vowel (case c) is predictable from the relationship of mutual complementation (Lehiste 1970:49) between vowel and consonant: the consonant is short after a long vowel, it is long after a short vowel.

In Standard Swedish there seems to be a general stress placing rule which Linell (1972) calls Native Word Stress Rule, assigning stress to the first vowel of the word stem. Hence the final vowel of case (a) is manifested with stress.

Words of certain segmental structures (cases a and b) cannot exhibit quantitative contrasts since the length of the vowel and consonant can be derived by rule from the segmental structure of the item. The only possibility for distinctive durational differences is to be found in words with identical segmental specifications (case c). Since the feature of QUANTITY is prosodic, I specify the long member of such minimal pairs like e.g. rät vs rätt ('straight' vs 'correct') for QUANTITY.

When analysing quantity in Standard Swedish as the segmental feature LONG or TENSE, Linell, Svensson, and Öhman (1971) and Lindau (1970) do not consider QUANTITY to be a prosodic feature. Another kind of segmental analysis of quantity is suggested by Eliasson and LaPelle (1972) who derive the length of the short vowel by doubling the following consonant in the lexical representation. Thus the pair rät vs rätt, in their analysis, would differ in the number of segments, rät consisting of three segments (CVC) and rätt of four (CVCC), the post-vocalic consonants being identical.

The phonological representation of this pair in the lexicon of Standard Swedish, in my analysis, may be formalized as follows:

		rät			rätt		
		/ r	ä	t /	/ r	ä	t /
distinctive features:							
(a) segmental	VOCALIC	+			+		
	CONSONANTAL	+	−		+	−	
	SONORANT			−			−
	CONTINUANT			−			−
	⋮		⋮			⋮	
	VOICE			−			−
(b) prosodic	STRESS						
	QUANTITY		+				
	TONE						

Contrary to Lindau (1970), I do not assume every lexical item to have stress at all. I consider the rows for STRESS and TONE to be empty in this pair. Both prosodic features will be assigned to these items by phonological rules applying to the surface structure of some sentence.

For the present, I consider the short vowels to be the unmarked case. Therefore the vowel segment being manifested long is specified for QUANTITY by a plus sign.

It seems to me most appropriate to mark a segment and not a syllable or even a larger unit for QUANTITY because the smallest domain of this feature can be one segment as e.g. in Finnish. This solution is more satisfying from a universal point of view: The lexical columns in languages without any distinctive prosodic feature would end with the last row of the segmental features, i.e. VOICE. The columns in the lexicon of languages, however, utilizing one or more distinctive prosodic features are longer and a certain segment would contain the specification for the utilized prosodic feature. Furthermore, I avoid the problem of defining the syllable which is probably not a concept of the lexicon of the grammar.

My proposed scheme seems to capture certain other phonetic facts easily and adequately. Some languages use one or the other of the three prosodic features, which are obviously available to all humans, systematically in their phonological systems, others, however, only in a few lexical items.

While Standard Swedish uses QUANTITY to a great extent in the phonological part of the grammar, I postulate STRESS and (WORD)TONE to appear only in a limited subset of the lexicon, Some lexical items are differentiated by the placement of STRESS only, e.g. 'kaffe vs café ('coffee' vs 'café'), 'Japan vs ja'pan ('Japan' vs 'Japanese'), 'formel vs for'mell ('formula' vs 'formal'), etc. Some lexical items are manifested with STRESS placed on a syllable contrary to the general stress placing rules which I assume for the phonological component of Standard Swedish, e.g. ka'bin ('cabin'), ka'nel ('cinnamon'), to'mat ('tomato'), vä'sentlig ('essential'), etc.

In certain lexical items (WORD)TONE, which is not considered a feature of their phonological representation in the lexicon, is manifested contrary to word tone assigning rules (Öhman 1965, Elert 1972), e.g. blåbär ('bilberry'), trådgård ('garden'), riksdag ('parliament'), etc. manifested with Accent 1 (acute).

The above mentioned cases are real exceptions, that is stress and word tone are idiosyncratic properties of these items and have to be learned separately.

For the lexical representation of Standard Swedish items which utilize all three prosodic features, I consider the prosodic rows to be empty in all cases captured by rules. However, where a certain prosodic feature functions distinctively in a certain item, that feature is specified in the lexicon. As an example, the phonological representation of the pair *kaffe* vs *café* differing in the placement of STRESS is indicated as follows:

		kaffe				café			
		/k	a	f	e/	/k	a	f	e/
distinctive features:									
(a) segmental	VOCALIC								
	CONSONANTAL								
	⋮			⋮				⋮	
	VOICE								
(b) prosodic	STRESS								+
	QUANTITY								
	TONE								

The first item *kaffe* is not marked for any of the prosodic features. Stress is placed on the first vowel /a/ at some point of the derivation in accordance with the general stress assigning rules of Standard Swedish. This stressed vowel is not specified for QUANTITY and in the context C__C... it will be manifested short. Word tone (Accent 2 or grave) is predicted by rule.

The second item *café*, however, is specified for STRESS on the second vowel /e/ because stress occurs irregularly on this vowel. This stressed vowel preceding a word boundary needs not be specified for QUANTITY because it fulfills one of the structural conditions for long vowels (case a). The word tone assigning rules correctly generate the Accent 1 (acute) from the stress placement on the final vowel.

These deviations might be due to

(a) segmental influences (vowel hight, manifestation as monophthong or diphthong, manner and place of articulation of the following consonant, the voicedness of the consonant, etc.),

(b) position in higher level units (compounds, phrases, etc.),

(c) different prosodic patterns (statement, question; neutral or emphatic mood, etc.),

(d) speech tempo, etc.

By way of conclusion an outline will be given of the phonological specification for QUANTITY of some lexical items of a few languages exhibiting quantitative differences.

In Finnish, where the domain of quantity is a non-initial vowel or consonant, every long segment is marked. In Danish and Dutch, the long vowel has to be specified because the domain of quantity is the vowel. Italian, however, shows durational differences of the consonant. Hence the consonant is marked for QUANTITY. In languages with mutual complementation (Lehiste 1970:49), e.g. Standard Swedish, Norwegian, Bavarian, the domain of quantity is the sequence of vowel and consonant. A long vowel is always followed by a short consonant, a short vowel by a long consonant. As consonant duration does not seem to be a primary cue for perception of these contrasts (Hadding-Koch and Abramson 1964, Bannert 1972), the vowel will be marked for QUANTITY even in these languages. The dashes indicate the segments of columns of the items.

distinctive features:

Finnish	t	u	l	e	t	uu	l	e	t	uu	ll	e	t	u	ll	e
(a) segmental																
(b) prosodic																
QUANT	−	−	−	−	−	+	−	−	−	+	+	−	−	−	+	−

```
Danish                      m i l e   m i ll e
      (a) segmental
      (b) prosodic
                  QUANT     - + - -   - - - -

Dutch                       t ie n    t i n
      (a) segmental
      (b) prosodic
                  QUANT     - + -     - - -

Italian                     p o l o   p o ll o
      (a) segmental
      (b) prosodic
                  QUANT     - - - -   - - + -

Bavarian                    W ie s n  w i ss n
      (a) segmental
      (b) prosodic
                  QUANT     - + - -   - - - -
```

This paper should not be considered more than just a very preliminary and incomplete outline. It will be elaborated and extended later.

References

Bannert R. 1972. Zur Stimmhaftigkeit und Quantität in einem bairischen Dialekt. Phonetics Laboratory, Lund University, Working Papers 6

Bannert R. 1973. Mutual complementation of VC-sequences in Central Bavarian. Phonetics Laboratory, Lund University, Working Papers 9

Chomsky N. and M. Halle. 1968. The sound pattern of English

Elert C-C. 1964. Phonologic studies of quantity in Swedish. Stockholm

Elert C-C. 1972. Tonality in Swedish: Rules and a list of minimal pairs. Studies for Einar Haugen, Scherabon Firchow et al. (ed.), pp 151-173. The Hague

Eliasson S. and N. LaPelle. 1972. Generativa regler för svenskans kvantitet. Svenskans beskrivning 6:104-122

Hadding-Koch K. and A.S. Abramson. 1964. Duration versus spectrum in Swedish vowels: Some perceptual experiments. Studia Linguistica 18:94-107

Jakobson R. and M. Halle. 1970. Phonology in relation to phonetics. Manual of phonetics, B. Malmberg (ed.): 411-449

Klatt D.H. 1973. Durational characteristics of prestressed wordinitial consonant clusters in English. MIT Research Laboratory of Electronics, QPR 108:253-260

Ladefoged P. 1971. Preliminaries to linguistic phonetics. Chicago

Lehiste I. 1970. Suprasegmentals. MIT Press

Lindau M. 1970. Prosodic problems in a generative phonology of Swedish. Phonetics Laboratory, Lund University, Working Papers 2

Lindblom B. 1971. Phonetics and the description of language. Proceedings 7th International Congress of Phonetic Sciences: 63-97. Montreal

Lindblom B. and K. Rapp. 1972. Reexaming the compensatory adjustment of vowel duration in Swedish words. University of Essex, Occasional Papers 13:204-225

Linell P. 1972. Remarks on Swedish morphology. Reports from Uppsala University, Department of Linguistics 1

Linell P., B. Svensson, and S. Öhman. 1971. Ljudstruktur. Lund

Nooteboom S.G. 1972. Some timing factors in the production and perception of vowels. University of Essex, Occasional Papers 13:49-76

Nootebom S.G. and I.H. Slis. 1972. The phonetic feature of vowel length in Dutch. Language and Speech 15:301-316

Öhman S. 1965. Generativa regler för det svenska verbets fonologi och prosodi. Svenskans beskrivning 3

Slis I.H. 1972. The influence of articulatory effort on the timing of speech. University of Essex, Occasional Papers 13:128-150

Uttalsordlista. 1970. Skrifter utgivna av Nämnden för svensk språkvård 6. Stockholm

Vanderslice R. and P. Ladefoged. 1972. Binary suprasegmental features and transformational word-accentuation rules. Language 48:319-338

Wang W. S-Y. 1967. Phonological features of tone. International Journal of American Linguistics 33:93-105

SPEECH TEMPO

Sidney Wood

June 1973

1. INTRODUCTION

Tempo is not one of the more frequently explored areas of speech research, and any possible consequences of tempo variation for other phonetic phenomena have all too often been taken for granted. Unfortunately, the situation is complicated by pitfalls of definition and hazards of numerical treatment.

If we glance at some of the elementary textbooks we find the following. Jones (1967: §43) put the average conversational rate of native English speakers at 300 syllables a minute and recommended this as a convenient target for foreign learners. Gimson (1962:25) makes several brief observations in one short paragraph in a discussion of quantity and duration: (i) "the absolute duration of sounds or syllables will, of course, depend on the speed of utterance", (ii) "an average rate of delivery might contain anything from 6 to 20 sounds per second", and (iii) "lower and higher speeds are frequently used without loss of intelligibility". These simple statements alone disclose a number of fundamental problems. How is speaking rate to be measured? (In "sounds" or "syllables", neither concept being easy to define?) What is the range of variation of speaking rate? How far do durations of other physical phenomena depend on speaking rate? Or, conversely, how far is speaking rate a disturbing factor in investigations concerning physical quantities in speech? How are speaking rate and intelligibility related? Another author, Abercrombie (1967: 46), has the following to say about speech rate: (i) tempo (speed of speaking) is best measured by rate of syllable succession, (ii) tempo is variable, and (iii) "everyone who starts learning a foreign language has the impression that its speakers use an exceptionally rapid tempo". His third observation indicates a further area of interest – what is perceived rate? Yet another area is revealed by Heffner (1960: §8.1) who discusses in particular the maximum rate of articulation available to man.

"Tempo" is not one single, unambiguous concept, but has in fact been used to denote the speed of several different processes in speech production. And "speed" has a special sense when applied to speech tempo. It refers to frequency of repetition and the rate measures give the number of speech units occurring in a defined period of time (words, morphemes, syllables, phonemes, gestures etc.). It does not refer to velocity. But the question has been raised as to whether or not we do accelerate articulator velocities when we "speak faster".

It is customary to make a fundamental distinction between gross rates based on the total time of speaking (i.e. including pauses) and net rates based on the periods of actual utterance (i.e. excluding pauses). These two types of measure have received various names. Goldman-Eisler found it convenient to refer to talking rate as a measure of the entire cognitive and articulatory activity involved in the production of an utterance and articulation rate for the amount of speech produced in the time actively taken to produce it. Kelly and Steer (1949) have over-all rate (comprising "intentional pauses and unintentional pauses as well as meaningful words spoken in the elapsed time") and phrase by phrase sentence rate, excluding pauses. (A decision on "meaningful words" is important since sooner or later the investigator must face up to the problem of what to do with the hesitant repetitions and uhms and ahs of sponaneous utterances.) Clevenger and Clark (1963) define three measures based on total time, phrase time, and pause time. In addition to gross rate (making no distinction between pause time and phrase time) and intra-phrase rate (based on phrase time only), they suggest that percentage of pause (pause time as a proportion of total time) can also be a useful measure of rate.

The difference between these measures can be illustrated with some data for my informant F, a speaker of West Greenlandic Eskimo, who has read a page from a novel. His style was fairly casual. He uttered 333 syllables in 31 phrases in a total time of 74 seconds, a gross (talking) rate of 4.5 sylls/sec. His pauses amounted to 24 seconds (32 % of the time or roughly one third), which means that he actively produced the 333 syllables in 50 seconds, an average net articulation (or intra-phrase) rate of 6.7 sylls/sec. The gross rate indicates how fast he was communicating (i.e. composing and transmitting his message), but tells us nothing of how fast he was uttering speech (which might indicate the load on the articulators and possibly be related to the degree of coarticulation and reduction etc.). Fig. 1 shows the cumulative frequency of net, intraphrase, sentence or articulation rates in individual phrases for this informant (F). He varied between 4.7 and 8.7 sylls/sec in individual phrases, the average being about 6.5 sylls/sec. For most of the time (66 % of his phrases) he ranged within 5.8 - 7.6 sylls/sec. Another type of presentation is given at Fig. 2(a) which shows how his net articulation rate fluctuated phrase by phrase.

We should not now be surprised to discover that "tempo" or "speech rate" are used with different meanings by different authors. For example, Kozhevnikov and Chistovitch (1965) first defined tempo as the speed with which an articulatory programme is accomplished (p. 77) and subsequently as the speed of succession of individual commands as distinct from the speed of individual movements (p. 90). They accuse Stetson, Hudgins and Moses (1940) of confusiong the issue by failing to observe this distinction. These three authors were studying the ranges of temporally constrained gestures, which is itself a legitimate area of investigation. Their aim was to study factors influencing the interpretation of palatograms and their conclusions are consequently of relevance for experiment design. This is a very different area from Kozhevnikov and Chistovitch's interesting speculations regarding the programming of speech articulation. In yet another area, Karlgren's interest in the application of information theory to speech led him to place emphasis on the rate at which the content of the underlying message was transmitted.

Kozhevnikov and Chistovitch explicitly excluded an interpretation of speech rate as a source of interference distorting the data. It may seem tempting to regard the duration depency of many gestures or acoustic features (described by, for example, Stetson et alia [1940], Lindblom [1963] or Gay [1968]) as a form of transmitter noise, reflecting the inability of the transmitter components to function adequately when temporally constrained. But there is an alternative view. Karlgren expressed the opinion that the reductions often associated with rapid speech are a measure of coding efficiency, which is the very opposite to interference from noise. Liberman et alia (1967) have emphasized the necessity for restructuring phonemes to overcome the inability of the ear to resolve discrete elements arriving at the rates of phoneme flow customary in speech, or of the articulators to produce separate distinct gestures at such rates. They suggest that "dividing the load among the articulators allows each to operate at at reasonable pace, and tightening the code keeps the information rate high. It is this kind of parallel processing that makes it possible to get high speed performance with low speed machinery..."

And yet the range of definitions outlined above represents only a few of the possibilities. The treatment of pauses requires careful consideration since this determines the duration measured for the speech sample. Similarly, the speech units counted can be concrete or abstract in various degrees. Care must be given to the treatment of reduced segments. There is wide freedom for combining decisions on these few factors alone. Neither of the two quantities involved in the computation of speech rate - duration and amount of speech - is a priori defined and the number of possible definitions of speech rate becomes, theoretically, infinitely large. Nor have we yet tried to handle acceleration or retardation.

The literature reviewed below is not exhaustive - it represents what is accessible to me at present. The topics treated appeared to fall naturally into the following areas:

- measurement of duration (§ 2)
- suitable quanta of speech (§ 3)
- estimates of normal tempo (§ 4)
- cognitive activity, planning (§ 5)
- why tempo varies (§ 6)
- what happens when tempo varies (§ 7)
- information theory (§ 8)
- the perception of tempo (§ 9)
- experiment design (§ 10)

2. MEASUREMENT OF DURATION

Early investigators were faced with considerable measuring difficulties. For many years the typical design was to use the stopwatch for measuring duration (e.g. Roudet), subsequently supplemented by the gramophone for storing and reproducing the speech sample (Wijma, 1938). It was hardly possible to determine the articulation rate by these means – at best the estimation of pause durations was very rough, at worst it could only be a blind guess (Wijma quotes Bourdon [1892] evaluating a colon as two commas, a stop as four commas and a comma as 0.375 seconds). The improvement of spectro-analysers, ink oscillographs and magnetic tape recorders (nowadays all standard items of equipment in the phonetics laboratory) has eased the difficulty of measuring the duration of defined stretches of speech. However, the stopwatch is not entirely extinct and was used as recently as 1969 by Cook for determining the gross overall talking rate of his interview subjects.

But while there are hardly any technical problems today as to how to measure duration, no amount of hardware can identify and isolate the units of speech whose duration is to be measured. A possible exception is Verzeano's electronic analyser (1950) which had the ostensibly straight-forward task of measuring entire phrase and pause durations and accumulating and counting them in distributional cells. Yet such an analyser would be unable to distinguish hesitant repetitions and uhms and ahs from meaningful speech. There are also difficulties involved in the selection of a suitable minimum pause duration for such a machine to define the end of a phrase, as Verzeano himself subsequently discusses in (1951). The theoretical and practical difficulties of segmenting speech into smaller units remain. In Kozhevnikov and Chistovich's words (1965:81), there are two irreconcilable and mutually exclusive requirements for the segment to contain all the cues of the speech sound and yet be discrete. Karlgren's equipment (1962) chopped the speech signal into 1 second portions, but the analysis was done by hand. Kozhevnikov and Chistovitch (pp. 79-81) placed electrodes on the lips and roof of the mouth so that labial or coronal gestures would close electric circuits. Test sentences were then designed with consonants containing one or other of these gestures. Huggins (1968) has pointed out the arbitrary character of the decision to let one single gesture mark the segment boundary. A segment is composed of several coordinated gestures and an error of timing of one gesture does not necessarily indicate a displacement of the entire segment. Lisker, in a review of the literature on temporal aspects of speech (1973), also emphasises the arbitrariness of segmenting smaller units and points to the syllable or word as convenient units for by-passing this difficulty. The next section is devoted to the problem of selecting a suitable speech unit.

3. SUITABLE QUANTA OF SPEECH

Hegedüs (1957), Fônagy and Magdics (1960), Osser and Peng (1964), and Gårding (1967) counted phonemes. Osser and Peng give the following motive for rejecting the syllable in favour of the phoneme in a comparison of English and Japanese speaking rates -- they feared that the simplicity of Japanese syllable structure compared with the complexity of the English could bias their results.

In order to minimize the difficulty of segmenting the speech wave into individual phonemes, Kozhevnikov and Chistovitch used the sum of all consonants and the sum of all vowels in a test sentence of one of their experiments, thereby reducing the number of segment boundaries that had to be identified.

The syllable is probably the most customary unit for rate measures. It has been used, for example, by Roudet, Wijma, Goldman-Eisler, Meinhold, Grosjean and Deschamps and by Malécot et alia. Repeated syllables were also used for establishing maximum possible rates of articulation (or gesture repetition rates) by Stetson, Heffner, Lehiste and Sigurd.

Fairbanks et alia, Goldman-Eisler and Cook have counted words.

Kelly and Steer, confronted with this problem of choosing a suitable speech unit, observed that measures of speaking rate based on syllables or words were well correlated (0.84) in American English and concluded that either estimate would give substantially the same result. An important factor influencing such a correlation must be the word structure of the language in question. For example, in languages like Swahili or Eskimo, strings of morphemes form relatively long words whereas the corresponding morpheme sequence in English would yield a number of separate shorter words. The distribution of word lengths in the speech samples from four of my informants is given in Table 1 to demonstrate this point. The most frequent word length in the Eskimo text (F) was 3 syllables (32 %) while a further 50 % were 4 – 7 syllables long. The mean was 4.2 syllables. The most frequent Swahili (H) word length was 2 syllables (31 %) while a further 40 % were 3 – 5 syllables long. The mean was 2.8 syllables. The most frequent English word length was 1 syllable (70 % for the General American informant G and 68 % for the Southern British informant A), with means of 1.6 syllables and 1.6 syllables respectively. A high correlation is surely to be expected between word and syllable rate in English when as many as 70 % of the words used are monosyllabic. This does not detract from the validity of Kelly and Steer's conclusions regarding English, but it does complicate cross-linguistic comparisons. There is no correlation between word and syllabic rates in this Eskimo sample, where only 10 % of the words were monosyllabic.

An even larger unit, the entire phrase between pauses, was used by Henderson et alia and Verzeano.

Karlgren (1962) observing that the message is conveyed whatever the phonetic distortion of speech, suggested that the number of morphemes (in a reconstructed full version of the message underlying the utterance) transmitted per unit of time would be a suitable measure of the rate at which information is transmitted in speech.

The varying degree of abstractness or concreteness of the preferred unit does not facilitate quantification of the speech uttered.

The abstract or concrete character of the phoneme has always been a controversial matter. Nor is the syllable entirely free from differences of opinion about how concrete or abstract it is (cf. Malmberg 1955, 1966: chapt. 11). Further, now that generative models for phonology are attracting attention, it has become possible to set up underlying segments of phoneme or syllable size that never have physical correlates at the phonetic surface. Even the word is not free from a degree of abstractness - it is by no means uncommon for whole words to be reduced to zero in everyday speech, leaving the investigator with the choice of counting words actually uttered or words presumed to have been present in some ideal sentence form underlying the utterance.

The distributions in Table 1 were based on complete or ideal word forms. But many words were shortened by the informants in their speech, especially F who read in an informal style. The 333 syllables of his sample (already referred to above in the introduction) are the phonetic expression of complete phonemic word forms containing 395 possible syllables - a degree of reduction of 16 %. An extreme instance of reduction in Swedish is reported by Karlgren (1962), who observed how the pronunciation of naturligtvis ("certainly") ranged from [n a t ʉ: l i t v i s] to [n a ə s] in his speed samples.

The well known non-uniqueness of the phoneme (e.g. Chao 1934), reflected in the variety of solutions available for any one language, introduces a further element of arbitrariness to tempo measures based on this unit. While the phoneme, as traditionally conceived, is certainly a useful device for denoting distinctive contrasts or providing a non-redundant transcription of an utterance, there is a very real risk that seemingly conflicting experiment results based on rival phoneme solutions might reflect differences of linguistic creed between phoneticians rather than true variations of speech behaviour between informants. Another disadvantage of counting phonemes is that the rival phoneme systems proposed for a given language may not be mutually convertible, so that results obtained within the framework provided by one phoneme solution may be useless for a linguist preferring an alternative solution. King (1966) has proposed a set of rules designed to achieve maximum convertibility of data from one phoneme solution to another. The problem of convertibility of phonetic data between different investigations of the temporal aspects of speech is emphasised by Lisker (1973) who points out how decisions as to the character of segments etc are often made arbitrarily. He preferred the kind of definition that would yield segments appropriate to some stated goal.

In a sense, the net articulation rate is misnamed, since the exact concept it expresses will depend on the degree of abstraction of the speech units counted and on how reduced segments are treated. For example, if a sentence is repeated more briefly, we would intuitively say that it was uttered "more quickly". This could in fact mean two things - the semantic content has been transmitted more quickly, or the articulators have been working harder to produce the same gestures in less time. Only the first of these two interpretations is always true in every case, whereas the number of gestures performed will depend on the degree of phonetic reduction in each individual rendering. Only when there is no reduction do these two rate concepts - semantic transmission and physiological performance - coincide. The investigator must therefore decide whether to count the number of units in an ideal pronunciation (for example based on normal orthography) or the number actually uttered (based on a narrow phonetic transcription). The decision will depend on his purpose, but only the second concept is valid for discussions about articulatory behaviour. For example, my American informant G spoke the five "syllables" of the words the Americans in 0.38 seconds, a rate of 13.2 sylls/sec. Now we know that we cannot utter syllables at that speed (see section 4). In fact, he actually uttered three syllables [ð i m é r k n̩ z], at a rate of 7.9 sylls/sec. This is a very plausible fast rate which would not overtax the capabilities of the articulators. On the other hand, a phrase like speak faster can only be uttered in three syllables and the two concepts will coincide. It will be impossible to accelerate such a phrase beyond the usual maximum of 8 sylls/sec or so and it will certainly be impossible to attain an abstract semantic transmission rate of 13 sylls/sec in such a case.

I shall return to this point in section 7 in connection with theories of reduction. It is sufficient to underline here that there is a possible source of confusion if the difference between these two concepts is not respected.

4. NORMAL TEMPO RANGE

The interest for normal everyday tempo was at first largely the preserve of short-hand writing experts. Karlgren, in an unpublished thesis, and Nosz (1964) have collected and reviewed much of this material. The phonetics literature proper is only occasionally devoted to tempo. Typically, Roudet

(1910:228-9) quoted short-hand sources - stenographers at the French National Assembly reported that the rate (presumably gross or talking rate, i.e. including pauses) in the speeches of the deputies varied between 155 and 300 syllables a minute (2.6 - 5 sylls/sec) and he expected this to be exceeded in private conversation.

Meinhold (1972) reports average articulation (intra-phrase) rates in German of 3 - 4 sylls/sec for poetry and 5 - 6 sylls/sec for prose and news broadcasts, a difference he attributed to "subjective redundancy", suprasegmental expressive information and rhythmic structure.

Grosjean and Deschamps (1972) found the average articulation rate for French in 450 phrases was 5.29 sylls/sec/phrase. Their corpus consisted of 15 phrases for each of 30 speakers in spontaneous radio interviews, pooled. 82 % of the phrases were uttered between 4.4 and 6.0 sylls/sec. The slowest rate was about 3.5 sylls/sec/phrase, the fastest about 8 or 9 sylls/sec/phrase. About 3 % of the phrases were uttered faster than 7 sylls/sec. Malécot et alia (1972) have also investigated articulation rate in half-hour conversations with 50 members of the Parisian "establishment". Their grand average articulation rate for a total of 13 000 phrases was 344 sylls/min (5.7 sylls/sec), faster than Grosjean and Deschamps's result. The variation ranged mainly from 4 to 8.5 sylls/sec/phrase, with limits at 1.6 and 9.8 sylls/sec/phrase. 66 % of all phrases were uttered between 300 and 400 sylls/min (5 - 6.6 sylls/sec).

Grosjean and Deschamps also quote one of Goldman-Eisler's results, an average of 4.95 sylls/sec for 8 English interviews, and an earlier result by Grosjean, an average of 4.70 sylls/sec for 9 English informants.

At Fig. 1 I have given the cumulative frequencies of intra-phrase articulation rates for my four informants. The Eskimo and Swahili speakers F and

H (who were reading connected prose) spoke most quickly, averaging around 6.5 – 6.8 sylls/sec/phrase and mostly (the central 66 %) ranging between 6 and 7.5 sylls/sec/phrase. The Southern British informant A, a politician making a radio broadcast, was the slowest speaker, averaging about 4.9 sylls/sec/phrase and ranging mostly between 4 and 6 sylls/sec/phrase. His style was very deliberate and persuasive. The American G, a university teacher speaking spontaneously in a radio interview, averaged about 6 sylls/sec/phrase and ranged mostly between 5 and 7.5 sylls/sec/phrase. His style was frequently rapid (cf. his phrases faster than 6.5 sylls/sec with A's) while at times he weighed his words carefully (cf. his slower phrases with F and H). These stylistic factors probably account for the flatter distributions and generally slower intra-phrase rates of A and G as compared with F and H.

Figs. 2 and 3 show how the articulation rate varied between successive phrases for each informant. There are periods of acceleration (e.g. G, phrases 6 to 10) and retardation (e.g. F, phrases 20 to 23). Apart from the short term flucutation, there is a possible tendency to gradually increase the average rate over a longer stretch, one such period lasting about 20 phrases. Thus F gradually increased his intra-phrase rate from 5 – 7 sylls/sec to 6 – 8 sylls/sec through the first 18 phrases. This tendency seems to be less well defined in H's sample. However, it is clearly seen in G's with a periodicity of about 25 phrases and in A's with a periodicity of about 25 – 30 phrases.

Hegedüs (1957) calculated the phoneme rate word by word for Hungarian speakers. He found that tempo varied between 5 and 20 phonemes/sec/word and he related these word by word variations of phoneme flow to (i) semantic weight (the importance of semantic redundancy for perception as distinct from the acoustic information contained in the speech signal has also been discussed by Lieberman, 1963), (ii) to phoneme quantity (long phonemes would tend to have longer durations) and (iii) the number of phonemes in the sequence (this is related to the law of equalization, studied by Fónagy and Magdics, 1960). Similar rates of phoneme flow have been observed by other investigators and it has been noted that the faster phoneme rates exceed the capability of the ear to discriminate discrete events. This is of importance for theories about how speech is processed on reception.

The maximum possible articulation rates of repeated syllables, or maximum repetition rates for individual gestures, have been investigated from time to time, usually with a view to determining physiological and temporal constraints on speech. There is a history of this type of research extending beyond Stetson's "Motor Phonetics" (1928) back into the nineteenth century (Kaiser, 1934). Using mechanical and pneumatic devices recording on a smoked drum, Kaiser analysed the maximum repetition rates of gestures representing different speech muscle groups. She found maximum rates for various lip movements of 2.5 – 4 per second (the lower lip being more agile than the upper), for mandibular raising and lowering of 5 per second, for tongue tip gestures of 7.5 per second and for "voice fluttering (glottal interruption) of 10 per second. Hudgins and Stetson (1937) found faster lip repetitions than Kaiser, a grand average of 6.7 per second for 9 subjects (with the jaw fixed to neutralize the mandibular component of labial occlusion). However, the lips still had relatively lower repetition frequencies than other articulators, which were ranked from high to low as follows: tongue tip, mandible, tongue back, lips, velum. The rank of the mandible is interesting since this is sometimes referred to in the literature as a sluggish body. Such an observation is possibly not true, at least as regards the rate at which its movement can alternate.

We can only speculate as to the underlying cause of these limits, but evidence points to the central nervous system rather than the articulators themselves. For example, Lehiste (1970:7) found she could voluntarily repeat the syllable ta up to 8 times a second but that her tongue could vibrate freely in a trilled dental r 28 times a second, from which she concluded that it was not the mass of the tongue that imposed a limit but the properties of the nervous system. The repetition of some gestures involves alternation between agonist and antagonist muscles, others require periodic contraction and relaxation of muscles. This suggests that the limit is set by the maximum rate at which coordinated sequences of motor commands can be initiated from motor centres.

The maximum possible articulation rates are usually given as 7 – 8 sylls/sec for ta and progressively slower for other gestures (Hudgins and Setson 1937, Heffner 1960, Lehiste 1970, Sigurd 1971) or 8 – 9 sylls/sec (Meinhold) while isolated individuals have been found capable of repeating such a syllable at up to 10 times a second (Stetson 1945, Sigurd 1971). Both Hudgins and Stetson and Sigurd conclude their articles by speculating on the possible consequences of these physiological and temporal constraints for the structure of speech and the direction of sound change. For example, the most frequently used gestures in speech are apparently those that are most readily repeated, and sound changes such as fronting of back vowels or nazalization of vowels involve a shift from less favoured to more favoured gestures.

If we compare the data for maximum syllable repetition rates with the performance of my four informants, we can see in Figs. 1 – 3 how all four regularly approached the maxima. Fig. 1 shows how H uttered 35 % of his phrases faster than 7 sylls/sec, F 30 %, G 25 % while A uttered only 3 %

of his phrases at these rapid rates. A was instead hammering home his political message carefully and deliberately, which is reflected in the regularly recurring slow rates of 2 – 4 sylls/sec/phrase. Several of the authors investigating the maximum repetition rates of syllables observe that their subjects experienced difficulty in forcing up their repetition rates without becoming tongue-tied. In contrast, my informants seem to be producing speech at similar fast rates without undue discomfort. At the same time (with the exception of the Swahili informant whose sample is dominated by CV syllables), the syllables they are producing are more complex than the simple ta-ta-ta repetition. Sigurd included syllables of increasing complexity in his repetition experiments, and found drastically reduced repetition rates. This suggests a fundamental difference of neuro-motor behaviour between the experimental repetition situation and the conditions of everyday speech. My informants do not appear to be unique since the French samples investigated by Grosjean and Deschamps and by Malécot et alia also contained phrases uttered at these rapid rates.

A possible explanation for the faster possible articulation rates of everyday speech compared with repetition experiments is that successive syllables often contain different consonant articulations so that the same gesture would be repeated at a lower rate than the syllable rate. For example, labial and apico-dental gestures are repeated in every other syllable in words like <u>minimizing</u> or <u>mandibular</u>, permitting these gestures to overlap. (Indeed, it has been suggested, for example in the parallel processing theory of Liberman et alia [1967], that speech is only possible thanks to this overlapping of gestures during production, with a corresponding overlapping of acoustic features in the received speech wave.) If, then, sequences of syllables with different consonants can be uttered more quickly than repetitions of identical syllables, there must be another constraint determining the maximum possible articulation rate in speech. It could be the rate of 10 "voice flutters" per second recorded by Kaiser. The rate at which the glottal tone can be turned on and off would set the limit for the repetition of vowels. The maximum articulation rates observed in everyday utterances approach, but do not exceed, this rate.

The question has sometimes been raised whether there are differences of normal speaking rate between speakers of different languages - are some languages inherently faster than others? We must distinguish, as always, between talking rate and articulation rate. Differences of talking rate can be attributed to either of its two components pause time and articulation rate. Differences of articulation rate will be due to differences in production.

Osser and Peng (1964) invited 6 Japanese and 6 American students to speak for 5 minutes on student life, and then counted the number of phonemes produced in the final five minutes of each sample (i.e. this is a gross talking rate). The average number of phonemes produced per speaker in each language in this minute was then compared, and the difference found not to be significant.

Grosjean and Deschamps (1972) compared 30 French spontaneous radio interviews with 8 English interviews published by Goldman-Eisler and 9 English interviews analysed previously by Grosjean. They found significant differences between the gross talking rates of these samples, 264.37 sylls/min for the French against 197.25 and 221.49 sylls/min for the two English (4.4, 3.3, and 3.7 sylls/sec respectively). The French average articulation rate tended to be slightly faster than the two English groups (5.29 sylls/sec against 4.95 and 4.70 sylls/sec) but the difference was only statistically significant for the second English group only. In contrast, Malécot et alia (1972) found a faster average articulation rate of 344 sylls/min. (5.7 sylls/sec) for spontaneous French. Grosjean and Deschamps concluded that the difference in talking rate was mainly due to differences of pause time and this in turn they attribute to the generally shorter phrase lengths of the English speakers, which means that they paused more frequently than the French. It is impossible to be sure that French is articulated more quickly than English on this evidence since these differences could instead reflect differences in level of abstraction, amount of intellectual activity and degree of verbal planning in the interview task, as suggested by Goldman Eisler (see section 5). Grosjean and Deschamps recognised that it would be more satisfactory to compare samples collected under the same conditions and analysed according to the same principles.

There is one factor at least, syllable weight, which might make faster syllabic articulation rates more probable for some languages, as was feared by Osser and Peng, leading them to count phonemes for their comparison of Japanese and English. Swahili, for example, has virtually only CV syllables and it is possible these would be articulated more quickly than the heavier syllables of other languages. Indeed, of my four informants, the Swahili speaker ~~had the fastest average intra-phrase/~~ articulation rates, almost 7 sylls/sec/phrase/ (Fig. 1), and the speakers of English were the slowest, averaging about 5 and 6 sylls/sec/phrase. However it is not possible to decide this question with only four informants. The relationship between syllable weight and articulation rate is examined more closely below in section 6. Let us here merely note that the syllable structure of the Swahili sample was clearly simpler than that of English and Eskimo (Table 2). Yet while the Swahili speaker had a faster average articulation rate than the speakers of English, the Eskimo speaker was almost as fast as the Swahili speaker although his syllable structure was of similar complexity to the English. I am inclined to doubt that differences in syllable structure had any decisive effect on the differences of articulation rate between these speakers, although syllable simplicity is undeniably in favour of the Swahili speaker. The differences are probably due more to other factors, expecially style of delivery and the nature of the speech task (the rhetoric flourishes of A's political speech, the calm straight-forward reading of F and H and the impromptu conversation of G).

5. PLANNING

The nature of higher levels of linguistic planning can only be inferred from the structure of the speech output. It is not accessible by other means. Goldman-Eisler (1958), after repeating a quotation from Fournié that speech is "the only window through which the physiologist can view the cerebral life", adds her own view that it is also "the only window through which the psychologist may view the dynamic patterning woven of motivating, controlling and environmental forces". She observed the talking rate, the articulation rate, and the pauses (a component of the talking rate) for different speech tasks such as newly created or well practised utterances (repeating descriptions of pictures), different levels of abstraction (describing pictures or summarizing their content) and various types of conversation (discussions with adult academics, conversations with adolescents and phychiatric interviews with neurotics) (1958, 1961a, 1961b, 1961c). Her results indicated that talking rate (determined mainly by variations of pause length) reflects the degree of hesitancy and therefore of organization or automatism of speech, that breath rate indicates the strength of affect (emotional excitation) and the output of speech per breath (expulsion rate) reflects the degree of its cortical control. Fast fluent speech tended to be weighted with habitual, well organized sequences (automatic speech). The slow speech contained a good proportion of hesitation pauses implying that symbolic and structuring processes were in progress during the speech. She speculates on the possible significance of various combinations of speech rate and breath rate, that they might indicate various degrees of intellectual activity and emotion. In addition, she also found that while the articulation rate was a constant of such rigidity that it did not respond to changes in the level of verbal planning (to different degrees of abstraction when encoding information into speech) it did respond to practice. She believed that this would corroborate the idea that there is a more basic difference between speech sequences which are familiar and well learned and those that are spontaneous and organized at the time of utterance, than exists among spontaneous and newly organized speech sequences differing in levels of verbal planning.

Henderson et alia (1961) suggest that a pause and the following speech phrase form one unit of cognitive rhythm, and that fairly regular periods of planning and internal organization govern the final speech output for short periods ahead.

Kozhevnikov and Chistovitch's tempo experiments (pp. 76-90) were part of a series designed to test certain hypotheses (concerning what they called the time figure of a syntagma) that were formulated in the course of deriving a model of speech programming. Assuming that speech rate was not a source of interference distorting the data, that it was not itself programmed but was the speed of accomplishment of a programme, and that consequently only relative durations could be programmed, they proceeded to investigate the relative durations of words, syllables and sounds in a syntagma uttered at various rates. They found that relative durations of words fluctuated randomly at different rates (but there were considerable variations of duration between different word positions in the syntagma) and that relative syllable durations fluctuated randomly in words. However, relative consonant durations appeared to increase at faster rates. They concluded that the articulatory programme for a syntagma cannot be considered as a sequence of word sub-programmes, that the syntagma considered as a sequence of sounds has no constant time figure, and that the syntagma considered as a sequence of syllables has an invariant rhythmic figure independent of speaking rate. Consequently, they argue, it is the syllable commands that are rhythmically organized in the programme of a syntagma. Before proceeding to test implications of this, they reformulated their definition of speaking rate as the rate of syllable commands, independent of the speed of articulatory movements. However, Notteboom and Slis (1969) repeated this experiment, using lip electrodes and labial consonants, and found that the relative consonantal duration varied only at slow rates. I have also failed to find this regular variation of relative consonant and vowel duration (Wood 1973). On the other hand, the results of Wodarz-Magdics's (1972) investigation of the durations of Hungarian phoneme segments at slow and accelerated tempo appear to support Kozhevnikov and Chistovitch. Her result tables show that all the Hungarian vowels were shortened more than the consonants. The ratio of slow to fast stressed renderings varied from 1:0.57 to 1:0.67 for tense vowels and 1:0.76 to 1:0.86 for lax vowels except lax /i, y, u/ which varied 1:0.90 to 1:0.92. For consonants, the ratios varied mainly from 1:0.92 to 1:0.97, while voiced stops were hardly shortened at all and fricatives, especially voiceless fricatives, were shortened by about the same degree as lax /i, y, u/, 1:0.87 to 1:0.92.

Unfortunately, Wodraz-Magdics did not describe her experimental method or define "slow" and "accelerated" rate. If we take the liberty of adding some of the average consonant and vowel durations in her tables, we find for /ta/ a slow duration of 0.240 seconds and accelerated of 0.205 seconds, corresponding to 4.2 and 4.9 sylls/sec. For /pa/, /ti:/ and /pi:/ the similarly estimated ranges are 3.9 - 4.7 sylls/sec, 3.6 - 4.7 sylls/sec and 3.5 - 4.6 sylls/sec respectively. These estimates are all slower than the average normal rate encountered in everyday speech, and barely half the maximum possible. If the estimates are correct, they agree with Nooteboom and Slis's finding that the relative durations changed in the slow speech range, but they will not tell us anything about what happens in the normal and fast ranges. Gaitenby (1965) found that the relative durations of words, syllables and segments remained fairly constant in a sentence uttered at different rates. In this case the different rates are the variations between the normal rates of several different speakers and the result does not necessarily indicate what might happen if the same informant varies his own rate, as in the present problem. The question seems to be open at present and we are faced with the possiblity that there are differences in this respect between individual speakers or different languages, or that different experiment designs can yield different results. Kozhevnikov and Chistovitch, on the evidence of their findings regarding the relative durations of consonants and vowels (which suggested that the time figure of segments was not invariant), rejected the speech sound as a candidate for the programme unit. The syllable was all that remained. Now if they had been mistaken about the relative durations of consonants and vowels - as Nooteboom and Slis's and my own controlled experiments suggest - it could nevertheless still be true that the syllable is the programme unit. But some other type of argument would be necessary to decide conclusively in its favour.

6. WHY TEMPO VARIES

Roudet lists several causes - the temperament and character of the speaker, emotion (sadness was said to slow speech, anger and joy to accelerate it), habit and situation. Wijma also conjectured that emotion was involved. Verzeano pursued this factor more methodically - he argued that if emotion introduces periodic phenomena in mental processes (as had been reported) then periodicities should appear in the speech of normal subjects exposed to emotional stress as well as in that of psychoneurotic subjects; such periodicities would be a departure from randomness in the speech time pattern, so that phrase durations would no longer conform to a Poisson series (a model which appeared to describe their distribution adequately in normal speech, although he subsequently found [1951] that the randomness of the distribution in part depended on the magnitued of the minimum pause duration set on the analyser to define the end of a phrase). In this context we can also recall the work of Goldman-Eisler in partitioning the relationship between speech rate and the degree of habit, abstraction and emotion in the speech task and situation.

Cook (1969) has investigated the effects of anxiety on speech disturbances and speech rate. He points out that "anxiety" is a notoriously vague concept, having as many different meanings as it has measures and operational definitions. If we add to this the care that must be exercised when selecting and defining a suitable measure for "speech rate", it will be realized that this is not a very easy area to work in. Cook insists on a distinction that has previously been neglected, between permanent and transient anxiety. He also found that the literature revealed some doubt as to what effect anxiety should have on speech rate - to increase or decrease it or have no effect at all. He refers to a current hypothesis that anxiety is a drive that will energize any on-going behaviour. Just as an anxious rat runs faster, so an anxious person will speak faster. Recalling Goldman-Eisler's work, he points out that when a person talks faster he actually diminishes his pauses but not his articulation rate, whereas the rat actually runs faster. He concludes that if anxiety does increase speech rate (apparently a gross rate) it is because responses are more frequent not because they are faster, and the analogy with simple motor behaviour is misleading. Cook used two measures of permanent anxiety but the respective scores were not correlated, indicating that the two classifications are not the same. Transient anxiety was manipulated by switching between disturbing topics of discussion during an interview with the subject. The experiment revealed no clear effect of either transient or permanent anxiety on speech rate (i.e. gross talking rate in words/sec). However, subjects with high anxiety scores on one of the two permanent measures (Taylor Manifest Anxiety Scale) differ significantly from the low-scoring subjects in their reactions to transient anxiety - the high MAS subjects slowed down whereas the low MAS subjects spoke faster. Cook therefore rules out the drive theory since the permanently anxious did not speak even faster when transient anxiety increased, contrary to what the drive theory would predict. Other theories remain open.

Fónagy and Magdics investigated how far the law of equalization applied in speech (that sequences of different syntagmatic length tend to be uttered with the same duration, tempo changes compensating variations of length). This is related to what has been called the stress-timing theory - that stress groups tend to be uttered with the same duration, irrespective of the number of intervening weak syllables. Sweet, Sievers and Jesperson had all intuitively perceived that longer sequences were pronounced more rapidly than short sequences. Rousselot and Laclotte (1913: 87–90) had reported that a given articulation becomes briefer as the group it is in becomes longer for example <u>â</u> in <u>pâte</u>, <u>pâté</u>, <u>pâtisserie</u>. Lindblom (1968) investigated the same effect using Swedish nonsense utterances, and found he could express the expected duration of a syllable as a function of several factors including the position of the syllable and the number of syllables in the utterance. Fônagy and Magdics found that smaller stress groups were uttered more slowly, but that the dependence of tempo on stress group size weakened in sequences longer than three syllables. Exponential functions described the dependence of speed on the size of the stress group.

Goldman-Eisler showed that talking rate depended on pause duration which is itself, she argued, associated with the amount of cognitive activity involved in planning. In 1961a she describes how, in contrast, the rate of articulation is invariant for different degrees of abstraction, but that it increases in proportion to the amount of practice (or degree of familiarity with the utterance) – articulation rate is faster for well learned sequences, clichés, jargon etc. The rate of articulation also appeared to be a strong personality constant of the individual speaker.

Kozhevnikov and Chistovitch (p. 114) found that a sequence CCVCV...CV had a longer duration than a sequence CVCV...CV. The extra consonant in the first syllable delayed the subsequent syllables, and the difference in duration did not diminish upon an increase in the number of syllables. Their conclusion was that the change of duration caused by the additional consonant is not peripheral in origin, but that the neural commands for the subsequent syllables are delayed by a definite magnitude. An implication of this for speech tempo must be that syllable structure can influence the articulation rate when measured in syllables – the longer duration for the same number of syllables yields a slower syllable rate.

Sigurd (1971) systematically varied syllable complexity in his repetition experiments, and tested 51 different syllables such as ba, spa, stra etc. The grand average maximum repetition rates for all nine subjects were 6.7 sylls/sec for all single stops with a, 5.0 sylls/sec for pla and ëra, and 3.6 sylls/sec for stra and skwa, demonstrating the lower repetition frequency (or longer duration) of more complex syllables.

This syllable complexity effect is the cause of Osser and Peng's concern that the different syllable structures of English and Japanese might bias their experiment. English has a heavier syllable structure than Japanese and might therefore be spoken with slower syllable rates. The different articulation rates of my own informants (Fig. 1) might also be related to this effect. We have already noted in section 4, Table 2 and Fig. 1 that the Eskimo and Swahili speakers F and H tended to have faster net articulation rates than the two speakers of English dialects A and G and that Swahili had a very simple syllable structure. Let us examine more closely the possible effects of syllable structure in speaking rate in these languages. We can look for a within-speaker effect and a between-speaker effect indicating individual speaker performances. With only four informants it is not possible to analyse these effects systematically, but some tendencies can be found.

Each informant's phrases have been divided into classes according to articulation rate, and the average number of consonants uttered per syllable in each rate class has been calculated. If the articulation rate is negatively related to syllable complexity, as expected, we should find a smaller number of consonants per syllable in the phrases uttered at faster rates. The 100 phrases of A's sample made possible a class interval as small as 0.5 syll/sec. However, the smaller samples of F, G and H gave many classes with only three or fewer phrases if the same interval was used. A larger interval, 1 syll/sec, has therefore been preferred for these informants. The results are given in Table 3.

Values in parenthesis are for classes where only 1 - 3 phrases occurred, the results appearing to be more stable where there were at least 4 or 5 phrases in a rate class. The table shows that there was indeed a tendency for all four informants to have fewer consonants per syllable in faster phrases. For A, whose sample covered the whole rate range from 2 - 8 sylls/sec/phrase, the overall difference was at least from 2.3 to 1.3 consonants per syllable (corresponding to 10 consonants in a phrase of ten syllables) between slow and fast phrases. G's sample also covers the full rate range and shows a similar overall difference of from 2.0 to 1.3 consonants per syllable (corresponding to 7 consonants per phrase of 10 syllables). F's and H's samples do not cover the full rate range, the slower half not being well represented. Their overall differences are consequently smaller than A's or G's. F has a difference of about 0.5 consonants per syllable (corresponding to about 5 consonants per phrase of 10 syllables) between medium and fast phrases. The difference is least for H due to the rarity of more complicated syllables than CV in his language, Swahili.*

---

* The most productive type of consonant cluster in Swahili is C+semivowel. There are also a few examples of s+t. An occasional stop+liquid occurs in loans. Finally there is the unique nasal+obstruent group which is always traditionally held to be syllable-initial, thereby preserving the constant open syllableness of Swahili. My own impression of the informant's speech was that he instead produced such nasals together with the preceding vowel. My perception may admittedly have been prejudiced by my "English ear". However, the spectrograms did frequently indicate a nasalized preceding vowel in this situation (for example a word like mambo, traditionally /ma+mbo/ with open syllables, which I heard as [mã(m)-bo]).

The results given in the table indicate that within the speech of individual speakers, syllable weight is negatively associated with articulation rate, phrases uttered at faster rates in everyday speech tending to have less complex syllable structure. The magnitude of the contribution of syllable structure to articulation rate is difficult to assess, but the product moment correlation coefficient (r) for mean syllable duration per phrase and the mean number of consonants per syllable per phrase over the first 20 consecutive phrases in A's sample was 0.5, which means that at least 0.75 of the articulation rate variance* must be attributed to other factors than this correlation. Beyond this, a correlation cannot tell us anything about cause and effect.

So much for the tendencies within each individual's speech. But what of differences between the individual speakers? Table 3 demonstrated the simple syllable structure of H's Swahili (0.9 - 1.5 consonants/syllable/phrase) while F's is very similar to A's and G's (1.2 - 1.8 consonants/syllable at faster rates). Syllables with one or two consonants were most frequent for A, F and G, in nearly equal proportions (Table 2). The grand averages for the entire material were 1.57 and 1.58 consonants/syllable respectively for the speakers of English A and G**, and 1.49 for the Eskimo speaker F. Clearly, the differences in average syllable weight between these three speakers are relatively small (0.1 consonants/syllable or 1 consonant in a phrase of 10 syllables), especially when compared with the strikingly simple syllable structure of the Swahili sample, which had

---

* The residual variance: Var $(1 - r^2)$

** The English averages, 1.57 and 1.58 consonants/syllable are remarkably close considering the differences of dialect and style. Gerber and Vertin (1969) found good rank correlations between different phoneme frequency counts in English and they concluded "the statistical constraints upon a given language are so severe that variations in time, place and form are of little consequence...".

a grand average of 0.99 consonants/syllable (virtually only CV syllables). It is hardly likely that the small difference between F's syllable structure and A's and G's could account for F's much faster rate. Moreover, the difference between the average Eskimo and Swahili syllable was as much as 0.5 consonants/syllable or 5 consonants in a phrase of ten syllables. This is much greater than the difference between A or G and F, and yet F and H have very similar distributions of intraphrase articulation rate (Fig. 2), both averaging nearly 7 sylls/sec/phrase.

Table 4 gives the average number of consonants per syllable in three classes representing the faster rates (6.5 - 7 sylls/sec/phrase, 7 - 7.5 sylls/sec/phrase and faster than 7.5 sylls/sec/phrase). This shows how the English and Eskimo speakers were capable of the same fast rates as the Swahili speaker despite the heavier syllable weight of their languages.

7. WHAT HAPPENS WHEN TEMPO CHANGES?

Some work has been concerned with the changes appearing in the speech signal due to variations of tempo.

Lindblom (1963) investigated the association between vowel duration and spectral neutralization of vowels, using variations of tempo as one means of inducing changes of vowel duration. The extent to which formant frequencies in the vowels failed to reach target values at a given vowel duration could be described by a continuous monotonic exponential function. Similarly, Gay (1968) found that the formant frequencies of the terminal spectrum of a diphthong glide depended on the duration of that glide, the rate of spectral change being constant during it. Another example of undershooting, by apico-dental and dorso-velar consonants, was given by Stetson et alia (1940). The contact area diminished and the locus of contact was displaced as speaking rate increased. Kent and Moll (1972) also found a tendency for lingual consonants to undershoot. In contrast, Gay and Hirose (1973) found that for labial consonants an increase in speaking rate was not accompanied by undershoot or any corollary change in lip closure duration.

As mentioned in section 5, Kozhevnikov and Chistovitch found that when the same test sentence was repeated at different rates proportion of time consumed by all the consonants tended to increase as tempo increased. It was also noted there that Nooteboom and Slis (1969) and I (Wood 1973) had failed to reproduce this tendency. Kozhevnikov and Chistovitch based two hypotheses on this observation, (i) that the time figure of a syntagma considered as a sequence of sounds is not constant, and (ii) that the gradual disappearance of a vowel is due to the consonant requiring a necessary minimum duration, sometimes leaving no time for the vowel at faster rates. The first was discussed in section 5. Regarding the second, that vowels are "squeezed out" for want of sufficient time for their execution, this ought to occur at random, whenever a momentary shortage of time occurs. While it is possible that some reductions occur in this way, I would also suggest that vowel elision and syllable contraction are largely non-random and are instead habitual and predictable from the environment, a view supported by the regularity of such phenomena synchronically in daily speech and diachronically in sound change. Such reductions become part of the common speech code and comprehension is not endangered. Some of the redundancy in the speech signal is discarded, with the result that the message is transmitted in a briefer period of time, that is, more rapidly.

The message can be accelerated by shortening segment durations down to a minimum (which, it has been suggested, is determined by the properties of the nervous system limiting the rate at which coordinated sequences of motor commands can be initiated, see section 4). This is the region in which undershoot phenomena are investigated. Acceleration of the message is also achieved by merging or omitting segments. It was demonstrated in section 3 that syllable contraction is only possible in some messages, and the examples that occur are, as suggested above, largely habitual. The difficult, perhaps impossible, point for theories of reduction to decide is (i) whether segments are casually dropped because accelerated speech happens to leave no room for them, or (ii) whether they are deliberately omitted, for whatever reason, with consequent acceleration of the message. Is the restructuring of articulation accidental or deliberate? Possibly both situations are true. It is typical of many languages that certain weak vowels are omitted between obstruents and liquids, for example Latin vestib'lum, French app'ler, English delib'rate. The conventional, rather than tempo-dependent, nature of such reductions is underlined when hypercorrect forms appear with spurious vowels, as in Latin saec(u)lum.

It is not easy to observe an increase in tempo that can be related to reductions of the synkope type. A measured articulation rate represents the sum of the various influences acting on the temporal characteristics of speech and reflects the consequences of reductions rather than the drive that is postulated to have occasioned them. There is a grave danger that arguments and proofs may become circular. The articulation rate follows increased tempo so long as segments are shortened only. But when segments disappear, a paradoxical slowing of the physiological articulation rate can occur if syllables are contracted, as in perhaps ——— p'raps*. One way

* Suppose a sentence were uttered in n syllables in t_1 seconds at a rate of n/t_1 sylls/sec, and then repeated more quickly in the briefer duration of t_2 seconds but now in $(n - 1)$ syllables following reduction. The new

round this difficulty is to relate reductions to the rate of transmission of the unreduced forms (cf, section 3, where a rate of 13 sylls/sec was possible for the message <u>the Americans</u> thanks to reduction to [ð i m é r k n̩ z]); but this procedure takes for granted that the first of the two interpretations outlined above for the tempo/reduction relationship is always true.

The question has sometimes been raised whether the articulators actually move faster when tempo increases. The interpretation usually placed on the undershoot phenomenon is that the articulator does not travel so far when time is restricted. But this tells us nothing about articulator velocity except that it has not accelerated sufficiently to compensate for the shorter duration. In the case of labial closure, however, Gay and Hirose (1973) found both an increase in the activity level of the muscles and an increase in the velocity of these articulators for increased speaking rate. Kozhevnikov and Chistovitch (1965: 180) had found that peak lower lip velocity during a closing movement varied with the initial magnitude of opening and this was subsequently confirmed by Ohala et alia (1969), Ohala (1970: 138), with respect to the mandible. Kent and Netsel (1971) found that velocities increased with degree of stress, which they attributed to the greater distances travelled in stressed syllables and the generally higher level of activity associated with stress. The converse should be expected in accelerated speech where stress levels are ge-

---

rate is $(n - 1)/t_2$ sylls/sec. It is given that t_1 t_2 seconds, but any of the follwoing relations between the net syllable rates is possible:

$$\frac{n}{t_1} > \frac{(n-1)}{t_2} \qquad \frac{n}{t_1} = \frac{(n-1)}{t_2} \qquad \frac{n}{t_1} < \frac{(n-1)}{t_2}$$

For example in the case where n = 8 syllables and t_1 = 2 seconds, the following values for t_2 will make each of the three relations true: $t_2 > 1.75$ secs, $t_2 = 1.75$ secs and $t_2 < 1.75$ secs respectively.

nerally lower. Kent and Moll (1972) found that the mandible travelled shorter distances and did indeed tend to move more slowly at faster speaking rates. They found that the tongue body accelerated during the closure gesture to g so that its peak velocity depended on the distance travelled, which was largely determined by the preceding vowel (furthest from a, and so on). Again, the peak velocities were slightly slower at the faster rate due to the shorter articulator excursions. Changes of rate had little effect on tongue tip velocity, but the tip certainly did not move faster at faster speaking rates. The general picture therefore seems to be that the lips compensate for a shorter segment duration by increasing their velocity at faster rates, whereas the tongue and mandible accelerate during gestures and the peak velocity therefore depends on the distance travelled, which is usually slightly less at faster rates and the peak velocity correspondigly lower.

8. INFORMATION CONTENT

The interest in the application of information theory to linguistics in general and to speech in particular was probably at its height about a decade ago. It was quite natural that speech rate, being concerned with how fast we compose, transmit and receive spoken messages, should also have been viewed from this angle.

Karlgren (1962, and an unpublished thesis) aimed to investigate the efficiency of the linguistic communication process and the appropriateness of information theory models for quantitative treatment of language. He aspired to adapt the existing theory to the special needs of linguistics – to build a linguistic communication theory.

Goldman-Eisler examined the function of hesitation pauses in the light of information theory (variations of pause durations had been found to be largely responsible for variations of the talking rate). She regarded hesitation pauses as manifestations of the general blocking of activity occurring when organisms are confronted with conditions of uncertainty. The speaker being faced with an act of choice in the organization of his utterance. This hypothesis was tested by relating the incidence of pauses in spontaneous utterances to the information content of the words constituting them. The experimental evidence showed that "the close relation found to exist between pauses and information on the one hand and fluency of speech and redundancy on the other, seems to indicate that the interpolation of hesitation pauses in speech is a necessary condition for such an increase. Delay is thus an important element in the production of information" (1958, 1961 d).

Picket and Pollock (1963), considering the relative intelligibility of small messages pronounced slowly and well versus large rapid and garbled messages, found a balanced trade-off between the two, the former providing more acoustic clues to identification, the latter more semantic clues.

Verzeano, investigating disturbances to normal speech patterns, explained the higher frequency of shorter phrases in the speech of post-lobectomy patients as follows: the operation limits the complexity of the concepts the subject may form and thus the length of the phrases in which these concepts are expressed, and in order to transmit the same amount of information in shorter phrase units he has to increase his phrase frequency.

The idea that the relationship between speaking rate and information content can be seen as a consequence of the speaker's brain constituting the narrowest section in the production channel has also been expressed by several authors. There is a physiological functioning limit of the speech organs but their capacity is nevertheless adequate for any message the brain can create, so that the bottle-neck appears to be central rather than peripheral. This is demonstrated by Lehiste's experiment comparing voluntary tongue tip gestures with the vibrations of a trilled r (see section 4). The receiving side would also seem to have ample processing capacity - Sigurd (1971) quotes several sources to the effect that perception is faster than production (reading than writing, listening than speaking). Electronic methods for accelerating speech recordings have been devised, enabling, for example, the blind to increase their rate of information intake as compared with normal speech rates (Schroeder et alia 1955). Speech can be followed, though with difficulty, at rates approaching 400 words a minute (Orr et alia 1965, quoted by Liberman et alia 1967, and Fairbanks et alia 1957).

9. THE PERCEPTION OF TEMPO

Goldman-Eisler (1958) found that what is commonly perceived as the speed of talking is determined by the halts and pauses which interrupt the flow of speech, rather than the speed at which the actual speech movements are performed.

Osser and Peng investigated the often expressed belief that foreign languages are uttered more quickly than one's own native tongue. They found no significant differences in the rate of phoneme production between five American and five Japanese speakers. They speculate that a listener, being unfamiliar with the relevancy of acoustic features of the foreign speech, is overwhelmed by the flow of information whereas he has learned to disregard features that are irrelevant in his own language.

Picket and Pollock tested the relative intelligibility of sequences of equal duration consisting either of short carefully pronounced messages or longer rapid and garbled messages. They formulated their notion of a trade-off between acoustic and contextual clues at different rates (this can be related to the view outlined previously that the reductions of normal continuous speech are concerned with codig efficiency, enabling the message to be successfully transmitted at faster rates).

10. EXPERIMENT DESIGN

In the literature referred to, the experimental situation has mostly been concerned with evoking variations of tempo. Sometimes informants have been asked to speak slowly, normally or quickly, at others a target rate has been dictated by a signal. In one case, the differences between normal personal constant rates of several individuals were exploited.

There are occasions, howver, where investigators have sought to eliminate speech rate effects by holding it constant in the experiment or by suitable statistical treatment of the results.

Most textbooks of phonetics only mention tempo in passing, merely noting that it exists, perhaps with a warning attached that it can influence experiment results. Stetson et alia (1940) showed how consonant palatograms vary with changes of duration induced by varying the speaking rate - contact areas diminished and the contact locus was gradually shifted as durations shortened. Lindblom's and Gay's investigations are also a warning that vowel spectra are related to duration. This must be borne in mind, for example, when comparing physiological or spectral data collected under different conditions for cross-linguitic or dialect comparisons. Observed differences may be temporally rather than linguistically conditioned. Differences between phonetic and linguistic constraints on the temporal properties of speech have been discussed at length by Lehiste (1970: chapt. 1).

The influence of speaking rate has probably been most keenly dreaded in studies of presumed intrinsic durations of phonemes. Lehiste and Peterson (1960: 17) were troubled by the need to eliminate the influence of tempo in order to discover the presumed intrinsic durations of vowels. Their test items were embedded in a test word placed in a constant sentence frame. On completing the experiment, they tested various hypotheses concerning the possible effect of variation of frame duration and stress timing. They decided that variations in rate of utterance had little effect on the duration of the syllables studied, but they remained uneasy and concluded that further investigation of the relation of tempo to segment duration was needed before this variable could be specified adequately.

Gregorski et alia (1971) compared temporal variations in recordings made with controlled speech rate (dictated by a metronome) and uncontrolled (normal) speech rate. Differences of variation were not significant and they concluded that, for the conditions investigated, the two methods produced comparable results.

The experience of Kozhevnikov and Chistovitch (p 90) was that the normal rate of individual speakers remained constant over longer periods of time (they repeated recordings over a period of months). Goldman-Eisler also indicates that the rate of articulation is normally constant for an individual. We ought therefore to expect an informant's rate to remain stable during a recording session, so that variations can be ignored and error. When the informant's task is to read from lists, possible/ referred to experimental/disturbances due to fatigue or nervousness ought to be countered by the customary randomization of the order of the items. For more exacting types of recording situation, it is known that the rate of articulation is not affected by the degree of abstraction of the speech task while the talking rate is, and some interview situations can influence the temporal organization of speech (Goldman-Eisler).

A far greater source of variation of speaking rate is to be found in the differences between the normal rates of different individuals (Gaitenby 1965, Goldman-Eisler 1961b and the differences between my informants as reported above). The influence of variations of tempo should only become acute when material from more than one informant is to be processed. Lisker (forthcoming) fears that "with the discussion of speech rate, the focus of interest tends to become the individual speaker and thus peripheral to the central concerns of phonetic research". However, since between-speaker variations of tempo are so large, it would be desirable to treat each informant individually rather than pool data from several informants. Nor is this necessarily the case for tempo only, but is probably true wherever absolute physical measures are concerned - spectrography, radiography, electromyography etc. Data from several informants should not be indiscriminately mixed. A speaker organizes his own physical characteristics into an individual system and the absolute magnitudes of physical quantities signalling particular phonetic features will function within his own speech only. It is probably more meaningful to compare tendencies and behaviour abstracted from several individual cases and generalise from this, rather than from pooled data.

BIBLIOGRAPHY

Abercrombie D. 1967. Elements of general phonetics. Edinburgh

Bourdon B. 1892. L'expression des émotions et des tendences dans le langage. Paris

Chao Y.R. 1934. The non-uniqueness of phonemic solutions. Reproduced in Joos M. (ed.) 1957, Readings in linguistics I

Clevenger T. and M.L. Clark. 1963. Coincidental variation as a source of confusion in the experimental study of rate. Language and Speech 6: 144-150

Cook M. 1969. Anxiety, speech disturbances and speech rate. The British Journal of Social and Clinical Psychology 8: 13-21

Fairbanks G., N. Guttman and M.S. Miron. 1957. Effects of time compression upon the comprehension of connected speech. Journal of Speech and Hearing Research 22: 10-19

Fónagy I. and K. Magdics. 1960. Speed of utterance in phrases of different lengths. Language and Speech 3: 179-192

Gaitenby J.H. 1965. The elastic word. Status Report of Speech Research, Haskins Laboratories, SR-2: 3.1-3.12

Gårding E. 1967. Prosodiska drag i spontant och uppläst tal. In Holm G. (ed.), Svenskt talspråk, Stockholm

Gay T. 1968. Effect of speaking rate on diphthong formant movements. Journal of the Acoustical Society of America 44: 1570-1573

Gay T. and H. Hirose. 1973. Effect of speaking rate on labial consonant production. Phonetica 27: 44-56

Gerber S. and S. Vertin. 1969. Comparative frequency counts of English phonemes. Phonetica 19: 133-141

Gimson A.C. 1962. An introduction to the pronunciation of English. London

Goldman-Eisler F. 1958. Speech analysis and mental processes. Language and Speech 1: 59-75

Goldman-Eisler F. 1961 a. The significance of changes in the rate of articulation. Language and Speech 4: 171-174

Goldman-Eisler F. 1961 b. Continuity of speech utterance, its determants and significance. Language and Speech 4: 220-231

Goldman-Eisler F. 1961 c. The distribution of pause durations in speech. Language and Speech 4: 232-237

Goldman-Eisler F. 1961 d. Hesitation and information in speech. Information Theory: Proceedings of the 4th London Symposium on Information Theory

Gregorski R., L. Shockey and I. Lehiste. 1971. Comparison of controlled and uncontrolled normal speech rate. Working Papers in Linguistics 9: 132-144. Computer and Information Science Research Centre, Ohio State University

Grosjean R. and A. Deschamps. 1972. Analyse des variables temporelles du français spontané. Phonetica 26: 129-156

Hegedüs L. 1957. Sprechtempoanalysen im Ungarischen. Zeitschrift für Phonetik und allgemeine Sprachwissenschaft. 10: 8-20

Heffner R.-M. S. 1960. General phonetics

Henderson A., F. Goldman-Eisler and A. Skarbeck. 1965. Sequential temporal patterns in spontaneous speech. Language and Speech 9: 207-216

Hudgins C.V. and R.H. Stetson. 1937. Relative speed of articulatory movements. Archives Néerlandaises de Phonétique Expérimentale 13:85-94

Huggins A.W.F. 1968. The perception of timing in natural speech I: compensation within the syllable. Language and Speech 11:1-11

Jones D. 1967. An outline of English phonetics. 9th edn. reptd. Cambridge

Kaiser L. 1934. Some properties of speech muscles and the influence thereof on language. Archives Néerlandaises de Phonétique Experimentale 10: 121-133

Karlgren H. 1962. Speech rate and information theory. Proceedings of the 4th International Congress of Phonetic Sciences

Karlgren H. Information measures and speech rate. (Thesis, Department of Phonetics, University of Uppsala.)

Kelly J.C. and M.D. Steer. 1949. Revised concept of rate. Journal of Speech and Hearing Disorders 14:222-226

Kent R.D. and K.L. Moll. 1972. Cinefluorographic analysis of selected lingual consonants. Journal of Speech and Hearing Research 15: 453-473

Kent R.D. and R. Netsel. 1971. Effects of stress contrasts on certain articulatory parameters. Phonetica 24:23-44

King R.D. 1966. On preferred phonemicizations for statistical studies: phoneme frequencies in German. Phonetica 15:22-31

Kozhevnikov V.A. and Chistovitch L.A. 1965. Speech, articulation and perception. (Translation by Joint Publications Research Service, Washington.)

Lehiste I. 1970. Suprasegmentals. MIT Press

Lehiste I. and G. Peterson. 1960. Studies of syllable nuclei, part 2. Speech Research Laboratory, Ann Arbor

Liberman A., F. Cooper, D. Shankweiler and M. Studdert-Kennedy. 1967. Perception of the speech code, Psychological Revue 74:431-461. Reproduced in Pribram (ed.) 1969, Brain and Behaviour Vol. 4, Penguin Books

Lieberman P. 1963. Some effects of semantic and grammatical context on the production and perception of speech. Language and Speech 6: 172-187

Lindblom B. 1963. Spectrographic study of vowel reduction. Journal of the Acoustical Society of America 35:1773-1781

Lindblom B. 1968. Temporal organization of syllable production. STL-QPSR 2-3

Lisker H. 1973. On time and timing in speech. In Seboek (ed.), Current Trends in Linguistics XII, The Hague

Malecot A., R. Johnstone and P.A. Kizziar. 1972. Syllabic rate and utterance length in French. Phonetica 26:235-251

Malmberg B. 1955. The phonetic basis for syllable division. Studia Linguistica 9:80-87

Malmberg B. 1966. Nyare fonetiska rön. Lund

Meinhold G. 1972. Allgemeine phonetische Probleme der Sprechgeschwindigkeit. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 25:492-505

Nooteboom S.G. and I.H. Slis. 1969. A note on rate of speech. IPO Progress Report 4:58-60, Eindhoven

Nosz G. 1964. Redegeschwindigkeit in Silben und Worten pro Zeiteinheit. Statistical Methods in Linguistics 3, Stockholm

Ohala J. 1970. Aspects of the control and production of speech. Working Papers in Phonetics 15, UCLA

Ohala J., S. Hiki, S. Hubler and R. Harshman. 1969. Transducing jaw and lip movements in speech. Journal of the Acoustical Society of America 45:1324 (A)

Orr D.B., H.L. Friedman and J.C. Williams. 1965. The trainability of listening comprehension of speeded discourse. Journal of Educational Psychology 56:148-156

Osser H. and F. Peng. 1964. A cross-cultural study of speech rate. Language and Speech 7:120-125

Picket J.M. and I. Pollock. 1963. Intelligibility of excerpts from fluent speech: effects of utterance and duration of excerpt. Language and Speech 6:151-164

Roudet L. 1910. Éléments de phonétique générale. Paris

Rousselot P.J. and F. Laclotte. 1913. Précis de prononciation française. Paris

Schroeder M.R., B.F. Logan and A.J. Prestigiacomo. 1962. Proceedings of the Stockholm Speech Communication Seminar

Sigurd B. 1971. Maximum rate and minimum duration of repeated syllables. Pilus 3, Institute of Linguistics, University of Stockholm

Stetson R.H. 1928. Motor phonetics. Archives Néerlandaises de Phonétique Expérimentale 3. 2nd ed 1951 by the North Holland Publishing Co., Amsterdam

Stetson R.H., C.V. Hudgins and E.R. Moses. 1940. Palatograms change with rate of articulation. Archives Néerlandaises de Phonétique Expérimentale 16:52-61

Verzeano M. 1950. Time patterns of speech in normal subjects. Journal of Speech and Hearing Disorders 15:197-201

Verzeano M. 1951. Time patterns of speech in normal subjects 2. Journal of Speech and Hearing Disorder 16:346-350

Wijma H.A. 1938. Contribution à l'étude de la vitesse du débit et de la lecture dans le néerlandais. Archives Néerlandaises de Phonétique Expérimentale 14:103-116

Wood S. 1973. What happens to vowels and consonants when we speak faster? Working Papers 9. Phonetics Laboratory, Lund University.

Wozart-Magdics K. 1972. Lautdauerverhältnisse im Ungarischen bei ruhigem und beschleunigtem Sprechtempo. Phonetica 26:217-226

Number of syllables per word	Informant A	G	F	H
1	68.1 %	70.3 %	1.2 %	22.7 %
2	21.8 %	11.1 %	10.8 %	30.9 %
3	7.6 %	8.2 %	32.0 %	16.5 %
4	2.3 %	7.2 %	14.4 %	16.5 %
5	0.2 %	1.8 %	14.5 %	7.2 %
6		0.4 %	9.3 %	4.1 %
7			12.0 %	2.1 %
8			3.5 %	
9			1.2 %	
10				
11			1.2 %	
Average syllables	1.5	1.6	4.2	2.8

Table 1. The body of the table gives the proportions of words with the stated number of syllables in each informant's sample of continuous speech. The count is based on unreduced forms.

A S. Br. English — G Gen. Am. English

F W. Greenlandic Eskimo — H Swahili

Number of consonants per syllable	Informant A	G	F	H
0	7 %	4 %	7 %	12 %
1	44 %	46 %	46 %	77.5 %
2	38 %	42 %	40 %	10 %
3	10 %	7 %	7 %	0.5 %
4	1 %	0.5 %		
5		0.5 %		

Table 2. The body of the table gives the proportions of syllables with the stated number of consonants, as uttered in 250 – 350 syllables of continuous speech.

<table>
<tr><th>Rate classes</th><th colspan="4">Average number of consonants per syllable in all phrases in the stated rate classes</th></tr>
<tr><th><u>sylls/sec/phrase</u></th><th><u>A</u></th><th><u>F</u></th><th><u>G</u></th><th><u>H</u></th></tr>
<tr><td>2.00–2.49</td><td>2.34</td><td rowspan="2"></td><td rowspan="2">(2.00)</td><td rowspan="2"></td></tr>
<tr><td>2.50–2.99</td><td>2.00</td></tr>
<tr><td>3.00–3.49</td><td>1.79</td><td rowspan="2"></td><td rowspan="2">(2.00)</td><td rowspan="2">(1.50)</td></tr>
<tr><td>3.50–3.99</td><td>1.70</td></tr>
<tr><td>4.00–4.49</td><td>1.60</td><td rowspan="2">(1.80)</td><td rowspan="2">1.64</td><td rowspan="2"></td></tr>
<tr><td>4.50–4.99</td><td>1.62</td></tr>
<tr><td>5.00–5.49</td><td>1.56</td><td rowspan="2">1.58</td><td rowspan="2">1.67</td><td rowspan="2">1.06</td></tr>
<tr><td>5.50–5.99</td><td>1.44</td></tr>
<tr><td>6.00–6.49</td><td>(1.57)</td><td rowspan="2">1.59</td><td rowspan="2">1.63</td><td rowspan="2">1.00</td></tr>
<tr><td>6.50–6.99</td><td>(1.30)</td></tr>
<tr><td>7.00–7.49</td><td>(1.35)</td><td rowspan="2">(1.62)</td><td rowspan="2">1.30</td><td rowspan="2">0.98</td></tr>
<tr><td>7.50–7.99</td><td>(1.00)</td></tr>
<tr><td>8.00–8.49</td><td>(1.33)</td><td rowspan="2">1.24</td><td></td><td rowspan="2">(0.90)</td></tr>
<tr><td>8.50–8.99</td><td></td><td>1.30</td></tr>
<tr><td>9.00–9.49</td><td></td><td></td><td></td><td></td></tr>
<tr><td>phrases:</td><td>200</td><td>30</td><td>60</td><td>40</td></tr>
</table>

<u>Table 3</u>. The body of the table gives the average number of consonants uttered per syllable for all phrases in the stated rate classes. Values in parenthesis are for classes where only 1–3 phrases occurred.

Informant	Intraphrase articulation rate sylls/sec/phrase 6.50–6.99	7.00–7.49	7.50–
A (S.Br.Eng.)	1.30(2)	1.35 (3)	1.17 (2)
G (G.Am.Eng.)	1.68 (7)	1.44 (5)	1.25 (12)
F (W.Gr.Esk.)	1.65 (8)	1.70 (1)	1.39 (6)
H (Swahili)	1.00 (15)	0.98 (12)	0.96 (4)

Table 4. The body of the table gives the average number of consonants per syllable in phrases uttered in the stated rate ranges. The number of phrases is given in parenthesis.

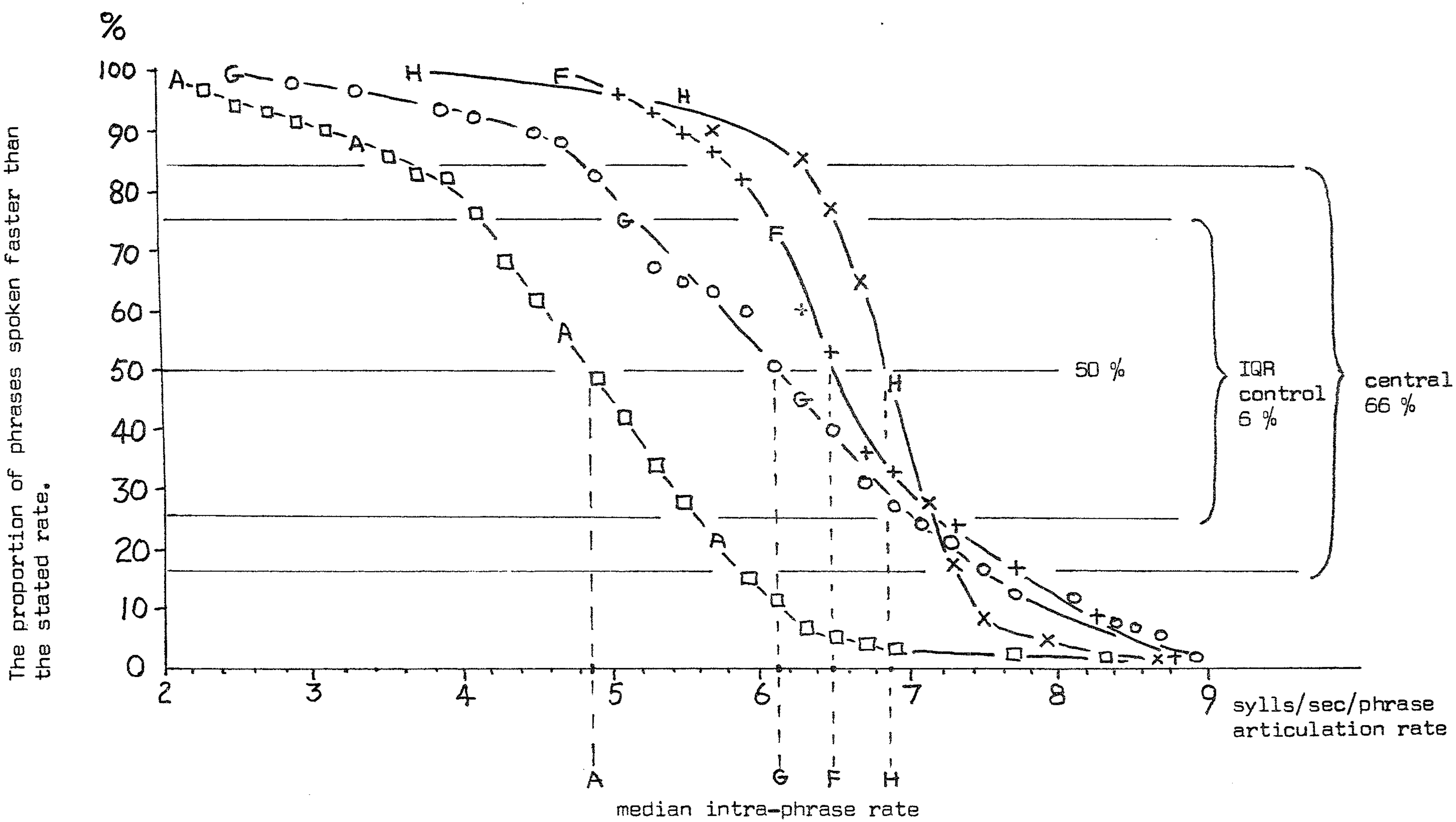


Fig. 1. Cumulative frequencies of intraphrase articulation rate in the continuous speech of 4 speakers: S.Br. English (A), Gen. Am. English (G), W. Greenlandic Eskimo (F) and Swahili (H).

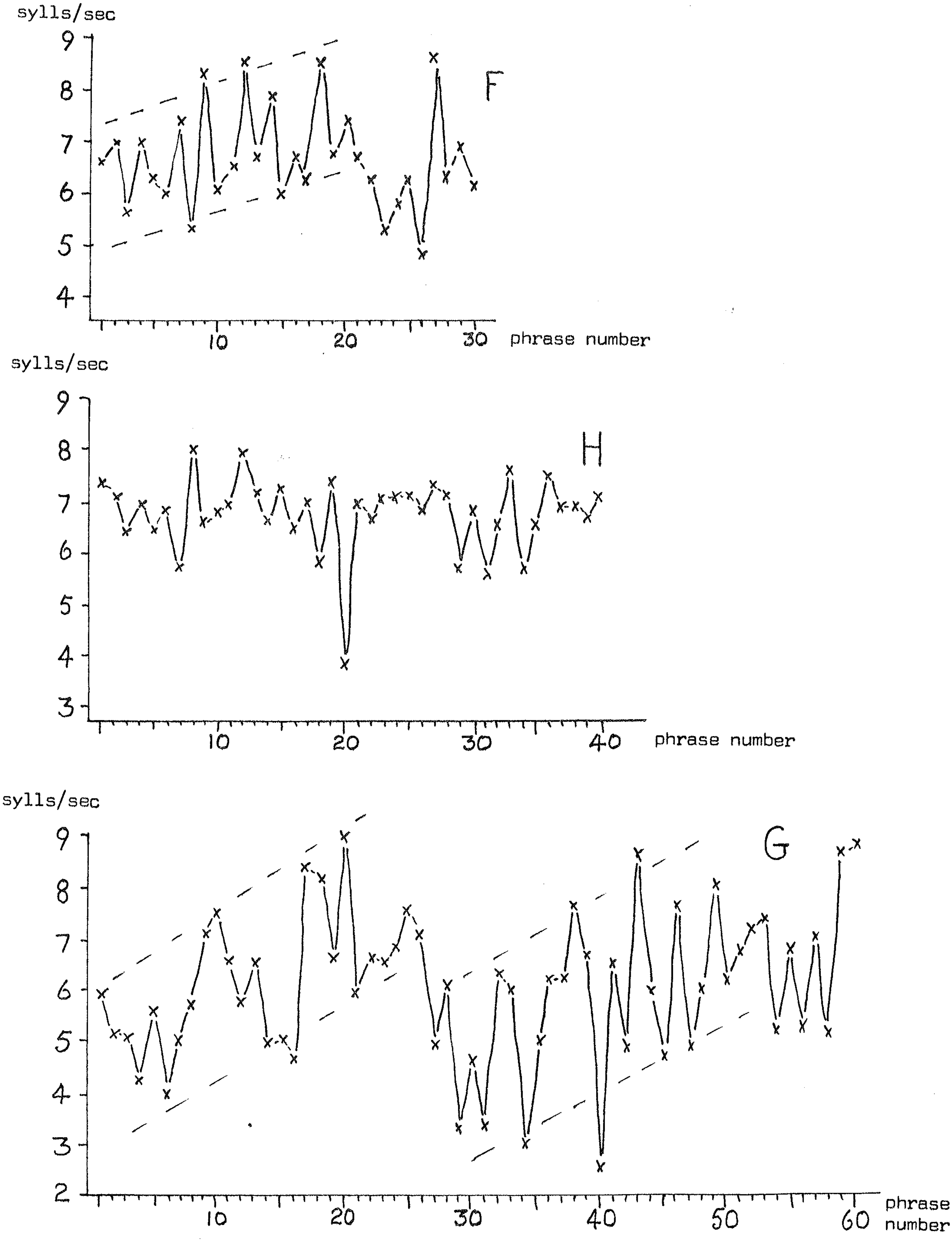


Fig. 2. Intraphrase articulation rate in successive phrases in the continuous speech of three speakers.

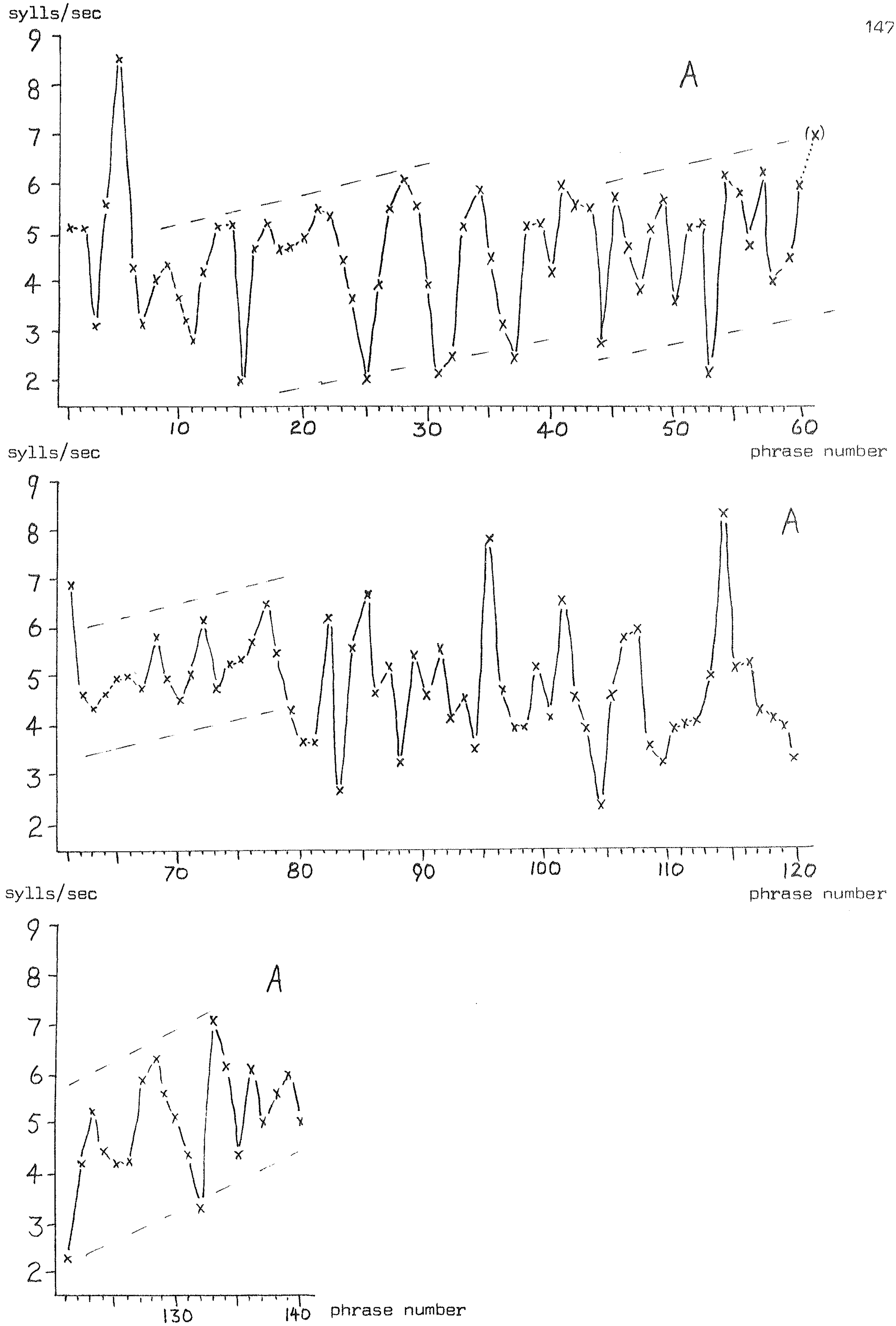


Fig. 3. Intraphrase articulation rate in successive phrases in the continuous speech of one speaker.

AV-Centralen i Lund, 1974