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), Fónagy 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 1.5 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 ∂ 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 <u>F</u> 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 relexation 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 minimizing or mandibular, 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 every-day 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 â in pâte, pâté, pâtisserie. 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 the Americans thanks to reduction to [ð i m ê 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. Clerk. 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 Phonêtique 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 Phonêtique 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><td>Rate classes</td><td colspan="4">Average number of consonants per syllable in all phrases in the stated rate classes</td></tr>
<tr><td><u>sylls/sec/phrase</u></td><td><u>A</u></td><td><u>F</u></td><td><u>G</u></td><td><u>H</u></td></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>

Table 3. 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.

	Intraphrase articulation rate sylls/sec/phrase		
Informant	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.

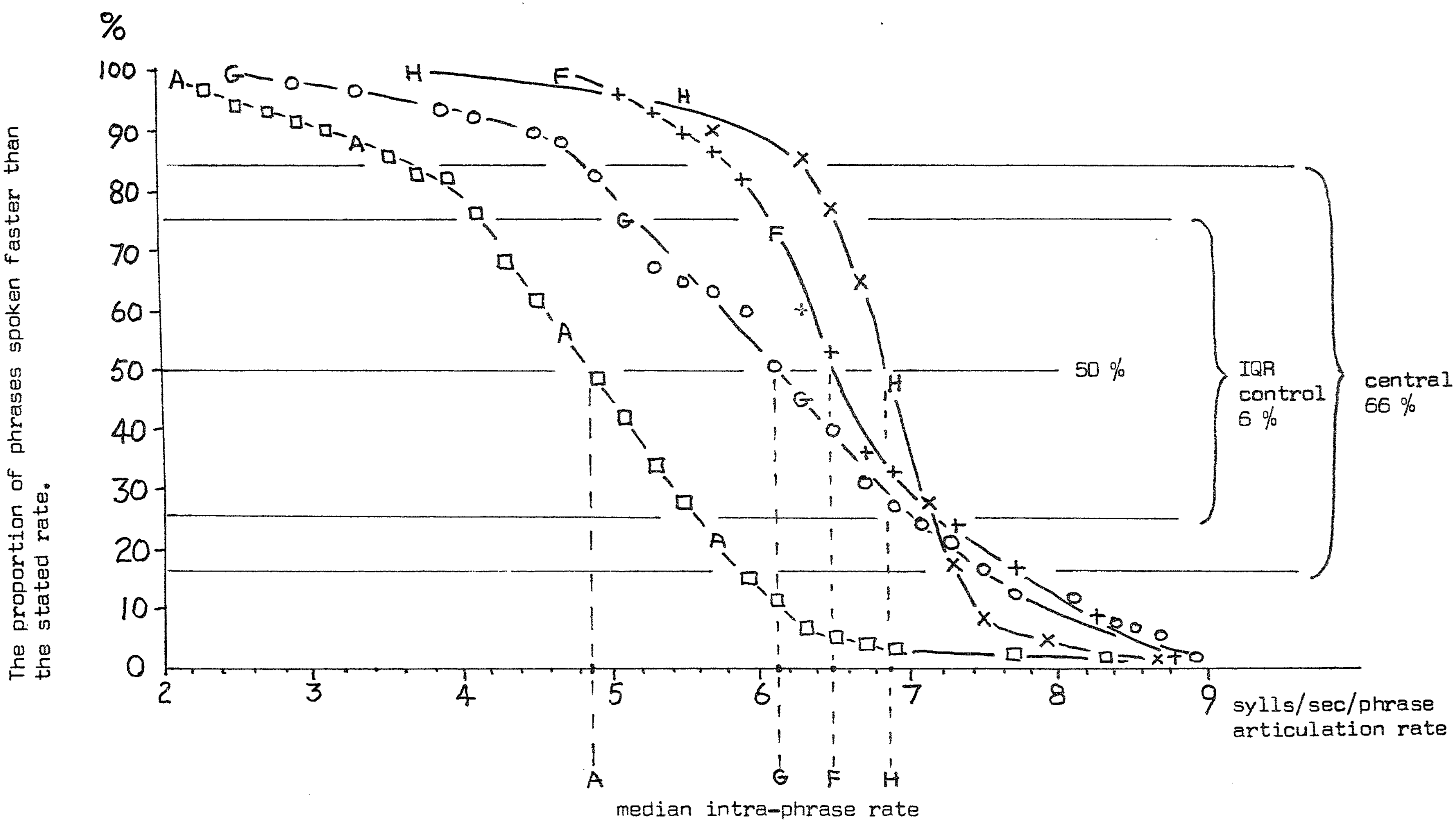

Fig. 1. Cumulative frequencies of intraphrase articulation rate in the continuous speech of 4 speakers: S.Br. English (A), Gen. Am. English (G), W. Greenlandic Eskimo (F) and Swahili (H).

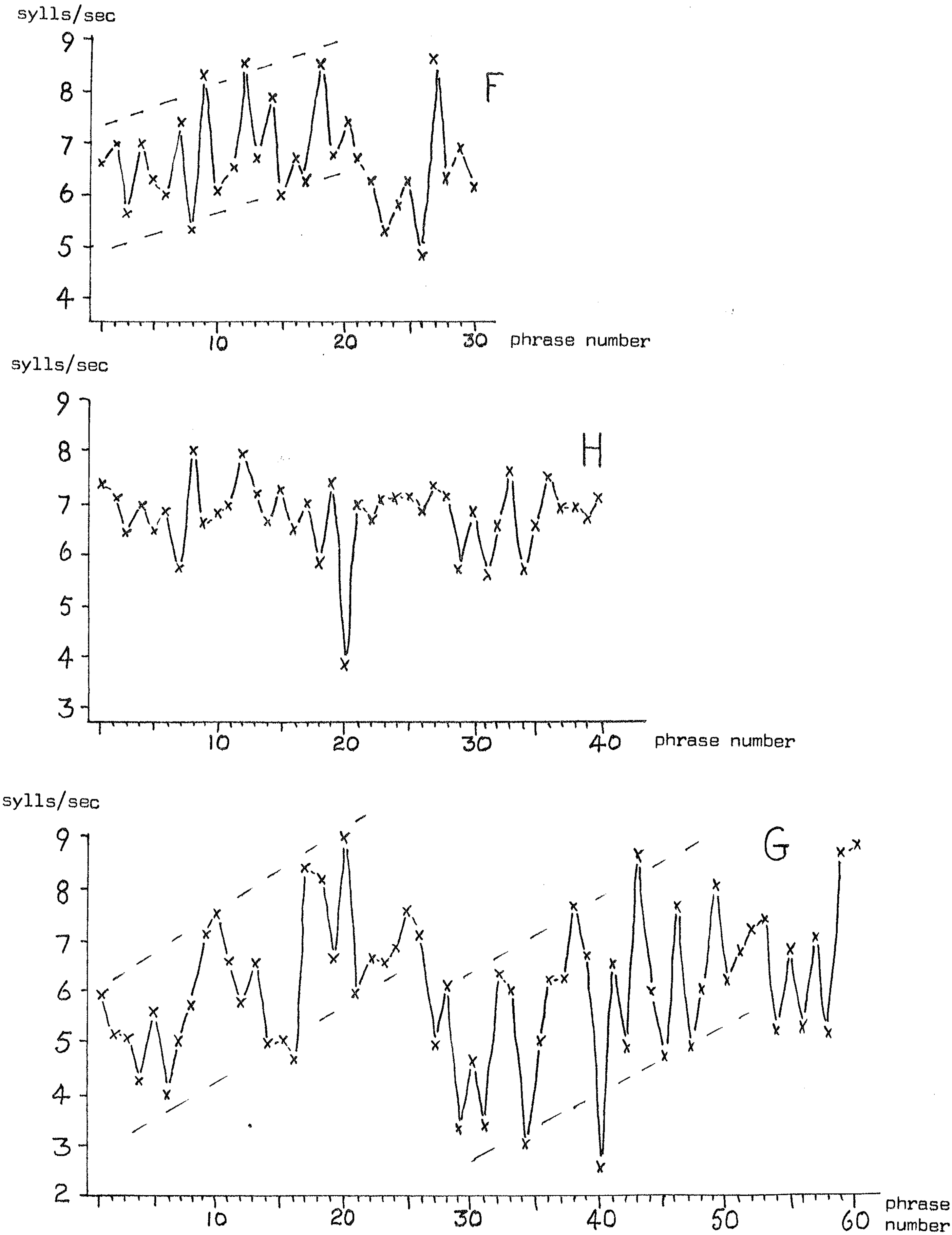

Fig. 2. Intraphrase articulation rate in successive phrases in the continuous speech of three speakers.

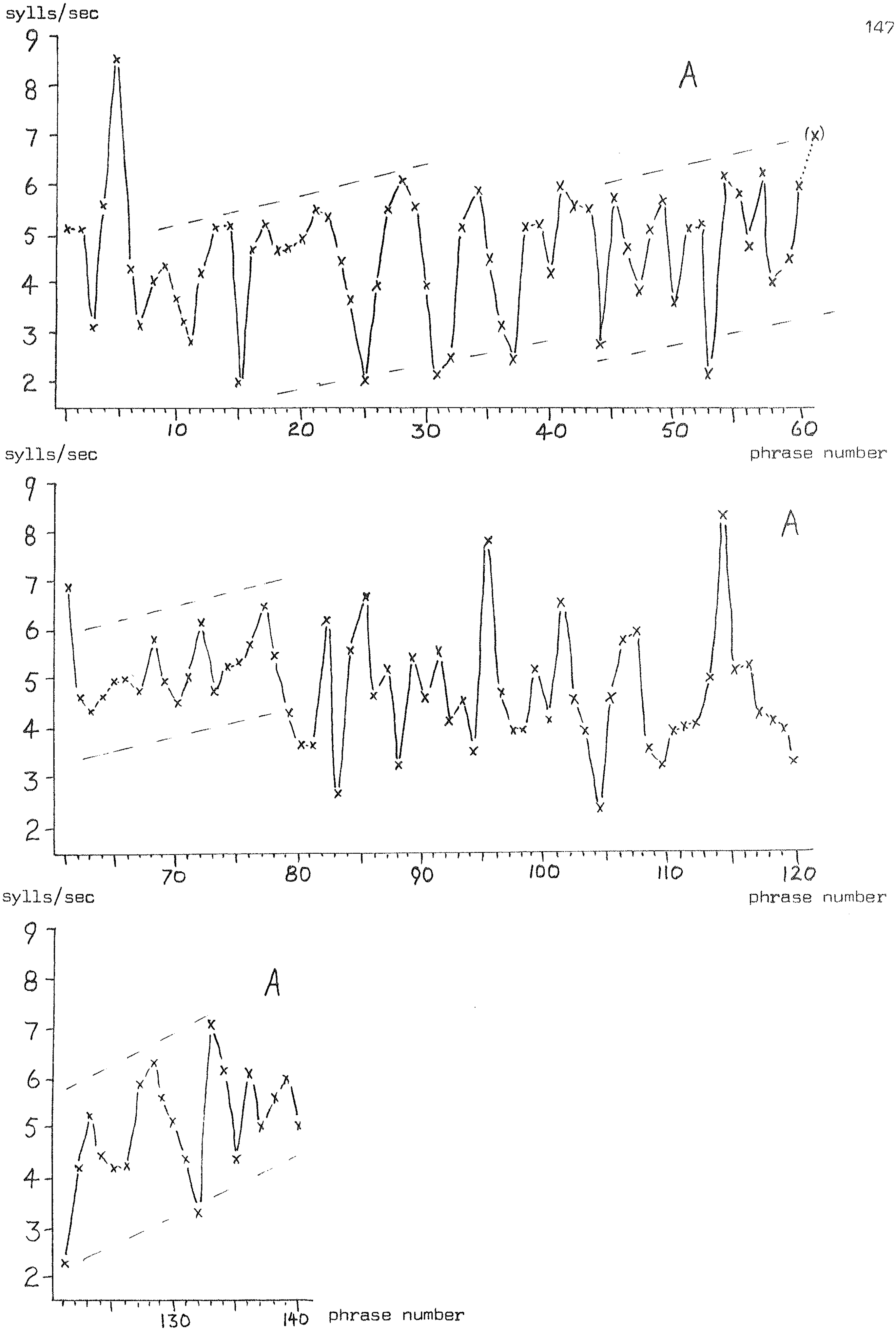

Fig. 3. Intraphrase articulation rate in successive phrases in the continuous speech of one speaker.