SPEECH RATE AND INFORMATION THEORY

H. KARLGREN

1. STUDY OF SPEED IN SPEAKING

Phoneticians all over the world, to make life a little easier for themselves, have to a very large extent been working with utterances provoked in controlled experimental situations. When test sentences are carefully uttered, the speed of this performance is not very essential. Naturally, the relative duration of vowel vs. consonants and the like has attracted attention, but the average tempo over longer segments has by most writers been mentioned only in passing.

2. LIFE-LIKE SITUATIONS: SPEED VARIATIONS AND FORM DISTORTIONS

But when we have to do with spontaneous talk, we cannot overlook the greatly varying speech tempo. And, closely associated with variations of speed, we find in any material recorded in life-like situations an overwhelming mass of what we may feel inclined to characterize as "distortions" of the "correct" word forms, as well as numerous omissions, insertions, repetitions and exchanges of one form for another. These modifications of what we naturally think of as the normal form, the one people believe they are saying and listening to, are no rare exceptions; they occur abundantly and – as I see it – necessarily in any conversation, only we cannot observe them until listening very attentively over and over again to the same tape. In fact our practical training to understand our language has taught us to abstract from such irrelevant features. That is why one always gets more or less normalized texts when taking down directly what an informant says.

Now, study of genuine everyday Swedish talk has been carried on lately at the Phonetics Department of Uppsala University and by the Research Group for Quantitative Linguistics in Stockholm. Variations of speech rate and the phenomena accompanying them has, then, demanded attention.

The study of speed variations in speech also offers special theoretical interest because it opens a field of research where communication theory models can reasonably be applied to linguistic and even to purely phonetic problems.

3. DEFINITION OF SPEECH RATE – HOW TO AVOID A VICIOUS CIRCLE

Speed must reasonably be defined as a relation between *time* and distance or volume or mass or something corresponding. What kind of mass unit can we define for the speech process?

Identification of morphemes

When listening to somebody speaking in one's own language one can as a rule readily identify the wording, the sequence of morphemes intended, independently of the precise number or nature of sounds uttered. Within wide limits of intelligibility, the "same" thing can be said in many various ways. Different versions of the same morpheme sequence – whether genuine talk or read from a written page – will certainly differ as to the specific sounds employed and as to the number of sounds produced.

The number of sounds, therefore, is no suitable mass measure. There is little help in counting larger units as long as they are of a phonetic nature: whole syllables and chains of syllables may "fail".

I shall give a few Swedish examples: Sw. *det är*, "it is", normally when pronounced slowly becomes [de: e:] and when pronounced rapidly [de:], that is exactly like the fuller form of the one word *det*. The word *naturligtvis*, "naturally, of course", has depending on context many different normal realizations, ranging from [natɯ: ɭitvis] over [natɯ:s] to [naəs].

Normally, as I said, we can nevertheless unambiguously identify what the man on the tape meant to say. Or rather the normal speaker and listener can – but we as linguists, working consciously from definite clues, are very poor at the job.

It seems to make sense, then, to talk about the time required to transmit the "same" morphemes.

Practical procedure

When working as we have been doing in the work mentioned with our own familiar language, the unsophistical listening and identification can in practice be done by ourselves, except for rare doubtful cases. The whole thing, then, boils down to writing down what the man on the tape says, in the very way we would have done, hadn't we known anything about phonemes, morphemes and the rest of the set and hadn't we been aware of the variety of forms for the "same" thing. But we have to listen many times with utter care to each recording to avoid "editing" the text.

Phonematic transcription

Writing down the morphemes identified is not equivalent to making a phonematic transcription. A phonematic transcription is based on phonematic analysis, and so far I have done no more than identified the meaningful units.

Normally phonematic analysis is based on distinctly uttered words. I believe no

conventional method for analysis will get to grips with the blurred sound mass directly observable in recorded spontaneous conversation. But I can, starting from the morphemic rendering, reconstruct the "full" pronunciation of the same text, namely the pronunciation that would have been used by a person belonging to the same language community if he were pronouncing the same morphemes slowly and emphatically. I can, in practice, do this reconstruction of the full form – or rather this construction – without testing informants. This is so, at least in the majority of cases, because I know which full form corresponds to the form I observed. But when in doubt – and I would be often enough if I worked with a foreign language – I can resort to the more elaborate method of asking a native person, preferably the original speaker himself, to repeat the same thing in slower tempo.

On this *constructed* form I can carry out the analysis and get a phonematic rendering, which is likewise independent of speech rate in my sense of the word. Actually, for my study I have chosen as mass units syllables, phonematic syllables which can be operationally defined fairly unequivocably, once you have the phonematic transcription.

Diagram

morphemic	2	{naturligtvis}
phonematic	4	/n/ /a/ /t/ /u:/ /r/ /l/ /i/ /t/ /v/ /i/ /s/
phonetic	3	[n] [a] [t] [ʉ:] [l] [i] [t] [v] [i] [s]
		[n] [a] [t] [ɯ] [s]
	1	[n] [a] [ə] [s]

4. REDUCTION PHONETICS

Explicitness

Thus, from listening to the observable sounds (line 1 in the diagram) we identify the morphemes (2), construct the fullest pronunciation (3), which in its turn on phonematic analysis yields the phonematic text (4).

On the phonetic level we have a wide range of forms, of different explicity. Just because we have defined speed independently of the performance of the speech organs, it is now very interesting to see how the reduction of the "full" forms takes place. We will ask:

a) where, at which points in the phonematic sequence, does reduction set in?

b) how much does the explicitness of one segment vary compared to that of others, i.e. which is the amplitude of the fluctuations?

c) along which lines does the transformation take place: shortening, assimilation, dissimilation, omission, random variation? Probably we can, in a Reduction phonetics' grammar, formulate rules for this, though certainly not for the reverse. The reverse, to tell the explicit form corresponding to a given reduced form, is not generally possible, except via consideration of long range context.

If the explicitness can be quantitatively stated, the two first questions can possibly be given at least an approximate answer by means of an information theory model. It is – I think – possible to formulate general rules, based on the conditional probability of occurrence and context analysis, for this variation of explicitness. I have for my part concentrated on duration; I know it is only one aspect of explicitness. But I don't know how to *count* nor define phonetic entities on the lower levels of explicitness, and it is, after all, natural to use time as an approximation to cost of transmission.

Also, it goes without saying that the degree of explicitness is correlated with many other factors as well, not only the "technical" demands that communication theory can account for. There are other reasons for emphasis than the need to transmit the message unambiguously, and this technical factor is not necessarily even the most powerful of those influencing speech rate.

Recoding and physiological prestanda

Thus my way of looking at the problem is that the variation of the form used to signal the same morpheme is not random nor merely individual or haphazard but contains an element of rational adaption of the signal to the need of the situation. The seemingly careless pronunciation is thus in terms of information theory an efficient coding to fit the channel, a recoding of the phonematic message under consideration of conditional probabilities. This indeed is much better than what any technical transmission system so far has accomplished.

It follows that the physiological limit of functioning speed for the speech organs – as enquired into by Heffner, Stetson & al. – does not necessarily determine the upper speed in the sense I am now talking about. Speech rate depends on the information content, among other things, and empty talk can be transmitted very rapidly indeed. After all, the brain is the narrow sector, not our speech organs.

Technological view of language

Languages are surprisingly often rationally built even where we are not aware of anything intentional or even regular.

This, to one with an engineer's view of language, is highly satisfactory. It is also encouraging, for it enables us to form good hypotheses on the linguistic communication system in *other* respects. Thus we might reason: how should a language be if I constructed one from the best of our knowledge about communication systems and with certain elementary requirements? Sometimes this reasoning yields quite good hypotheses on the functioning of real languages.

5. INFORMATION VALUE COMPUTATIONS

Starting now from information theory with its theoremes of how efficient coding

for technical purposes should be done, we may ask if the speed and form variations do fit in with the technical conditions for optimal transmission. From data on the statistical structure and through psychological experiments we compute quite theoretically information value numbers to correlate with our physical observations. Apart from minor refinement the information theory part of the work has been carried out on the lines suggested by Shannon. Especially word frequency data and prediction tests have been used.[1]

6. SPEECH RATE MEASUREMENTS

Over-all speed

The speed over longer ranges of text, say half a minute to a minute, can very easily be measured.

First I wanted to see if the relative long-windedness of some languages has been compensated for by a spontaneous adjustment, so that "short" languages were spoken more distinctly and slowly and long languages relatively carelessly and fast, other things being kept equal. I still believe it is so, but it is almost impossible to find even approximately equal circumstances.

A comparative over-all speed test is not practicable even to reading from written material. I have made extensive statistical studies about the relative length of translation of the same text from and into different European languages. The differences appear to be so small, that I cannot hope that the variation due to them will come to the surface in an experiment to test my hypothesis. They will certainly be drowned by all inter-individual variation which is due to temperament, style, attitude and so on.

The same applies when I take different longer passages of the *same* language: the style influences the readers more than the information value seems to be doing.

Short range variations

Next we turn to the fluctuations of speed over intervals of a second or so.

Methods for measurement

The obvious method is to measure the time of the natural segments of utterances, the phrases delimited by pauses at either end. This is less easy than it may at first appear, there being no sharp demarcation of pause and speech. Also this method is unsatisfactory because pauses are often used as it were for the very purpose to lengthen neighbouring units: they appear before, after or enclosing an unexpected or emphatic word.

A better way of doing it is to measure the duration of equivalent lengths of text, 5-syllable and 25-syllable ranges. Or inversely – that amounts to very much the same

[1] A full account of these data and these computations, which might be useful for other purposes as well, will be given in the forthcoming journal *Statistical Methods in Linguistics*, publ. by Språkförlaget Skriptor, Fack, Stockholm, 40.

thing but has proved to save labour – I count the number of phonematic syllables in segments of constant length. An electronic relay has been constructed for this purpose at the Phonetics Department at Uppsala. This instrument chops off the recording into one-second mouth-fulls which are fed alternatively into two tape-recorders. When replaying, only one segment is listened to at a time.

Short range variation turns out to be much more characteristic, when comparing different persons, different texts or different languages.

Interpretation

a) One way to interpret the results is to seek to explain in terms of information theory why certain segments are fast and others slow, thus answering the question *where* reduction sets in.

b) Another way is to compute the speed variability as a whole of different texts and languages and correlate this with known facts about their statistical structure. This in a way is an attempt to derive the *amplitude* of the fluctuations from general optimalization rules.

The variability would be expected to be greater in e.g. English than in Italian, judging from the subjective impression of unevenness and smoothness respectively and from known data about the language structures. A still more striking pair of contrasting languages are the two in use in this country, Finnish and Swedish. As a matter of fact, the variation does show marked difference in samples from these languages, and this fits in well with statistical computations. Professor Antti Sovijärvi has given me permission to continue these investigations at the Phonetics Laboratory in Helsinki. Thus I hope soon to be able to verify my hypotheses on more extensive material.

7. CONCLUSIONS

It is a challenging thought that general optimalization rules could be formulated for the relation between speech rate variation and the statistical structure of a language. Judging from my experiments, there are reasons to believe that there is an equilibrium between information value on the one hand and duration and similar qualities of the realization on the other.

This evidently does not imply that speed variability is so great as it is because of the statistical structure nor vica versa – how language once attained its present equilibrium is a matter for diachronical linguistics.

If the supposed relation can be proved, it should be possible to predict the pronunciation of a language from structural data and even to "predict" the pronunciation of dead languages.

Naturally accent, considered as relative prominence and realized concurrently

as variation of duration and in other ways, comes in when we are treating very short range speech rate variations.

Stockholm/Uppsala

REFERENCES

Heffner, R. M. S., *General Phonetics* (Madison, The University of Wisconsin Press, 1952).
Shannon, C. E., "A Mathematikal Theory of Communication", *Bell Syst. Techn. J.*, 27 (1948), 379—423.
Stetson, R. H., *Bases of Phonology* (Oberlin College, 1945), p. 55.

DISCUSSION

I should like to emphasize that a candid, naturally spoken text can hardly be the starting-point for phonemic analysis nor do I think it could ever have been. In "phonological oppositions", "minimal pairs", "commutation tests", pronunciations are used which are particularly neat and "complete", or so to say, *ideal* forms of isolated words.

Hammarström