MAJOR DETERMINANTS OF SPEECH RHYTHM:
A PRELIMINARY MODEL AND SOME DATA

EVA STRANGERT

Department of Phonetics
Umeå University
S-901 87 Umeå, Sweden

ABSTRACT

In the speech signal rhythm manifests itself in the temporal structure of stressed and unstressed syllables. This structure differs between languages and seems to be the basis for perceived rhythmic differences. At the same time there is evidence of temporal adjustments towards regularity which seem to occur irrespective of the language spoken.

It is assumed that these characteristics of the speech signal reflect two major determinants of rhythm - language structure and speech production constraints, respectively.

Some predictions based on this model are tested on three rhythmically different languages, Swedish, Spanish, and Finnish.

INTRODUCTION

That something is rhythmic means that it is temporally constrained. The impression of rhythm seems to depend on the impression of temporal regularity. In speech this regularity concerns syllables, stressed and unstressed.

While temporal regularity seems to be at the base of rhythm, different languages seem to have different kinds of temporal regularities, that is, they often sound rhythmically different. To capture such differences Pike /1/ introduced the stress-timing/syllable-timing dichotomy implying two different rhythmic principles. In a language with stress-timing, then, the regularity concerned the stressed syllables, while in a language with syllable-timing the regularity concerned all syllables, stressed and unstressed alike. Temporal regularity also implied the strongest possible temporal constraints, isochrony. Thus, in a stress-timed language stressed syllables were assumed to recur at equal intervals irrespective of the number of intervening unstressed syllables, and in a syllable-timed language all syllables, stressed and unstressed alike, were assumed to have equal duration.

Isochrony seems to be an important aspect of the perception of speech. For example, Pike /1/ based his distinction between stress-timed and syllable-timed languages entirely on the listener's impression of temporal regularity. Observations of the production of speech, on the other hand, do not support any strict regularity in the sense implied by either stress-timing or syllable-timing. Intervals between stressed syllables, and syllable durations, seem to differ within fairly wide ranges in both allegedly stress-timed and syllable-timed languages.

However, in measurements of the speech signal tendencies to temporal regularities have been found. The duration of segments and syllables seem to be inversely related to the number of unstressed syllables between stressed ones, implying a weak tendency to stress-timing . Most of these observations have been based on English but also, to a certain extent, on other languages including so-called syllable-timed languages (see /2/, p. 3-5, for a survey). There are, on the other hand, several studies in which any tendencies to temporal regularities are denied. One example is a study by Lehtonen /3/ examining the temporal structure of Finnish.

All these aspects of rhythm have to be accounted for within a general model of speech rhythm. As first step to such a model I will outline a conceptual frame for studying rhythm in speech.

A CONCEPTUAL FRAME FOR STUDYING SPEECH RHYTHM

Three basic concepts all contribute to the complex of rhythm in speech as well as in other types of rhythmic behavior: (a) grouping, (b) alternation, and (c) temporal regularity.

Grouping is the most fundamental concept. All kinds of activities seem to be organized by grouping the elements of which they are made up. Grouping occurs in both production and perception, as shown in experiments by Fraisse /4/ and Woodrow /5/. In complex activities there may be several levels of organization. One group at a higher level may contain two or more groups at a lower level. Such hierarchical grouping is very obvious in music but it seems to be a characteristic also of speech and other kinds of human activities. Thus grouping may be seen as a general means for structuring information, and therefore what we perceive as rhythm may be a consequence of a natural way of handling information.

Alternation often characterizes a sequence of elements. Normally some elements in a sequence are marked from the others, for example by being long-

er or more intense. The marked elements will then alternate with the unmarked ones. Alternation is an important basis for grouping as groups are delimited from one marked element to another. However, grouping also occurs when there is no alternation at all in a sequence of elements. In this case grouping may be achieved by marking some of the originally unmarked elements.

Grouping may lead to temporal regularity. Grouping elements together means that there are temporal constraints on how they are processed. Related elements have to be kept together and will accordingly constitute a unit in the temporal domain, also. The basis for this temporal unity of groups might be a cyclic and regular processing of information. This means that in groups with many elements there has to be a temporal compression of these elements, while in groups with few elements no such compression will be needed. Such compression in longer groups has been reported in several studies. It even seems to be a tendency to adjust differently-sized groups towards an intermediate or average duration /4/.

Grouping, alternation, and temporal regularity all contribute to the complex of speech rhythm as outlined in the following simple input-output model:

MESSAGE STRUCTURE → ARTICULATORY PLANNING → RHYTHMIC STRUCTURE
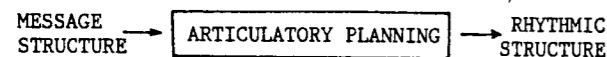
For a more detailed description, see /2/, p. 19-27.

The message structure

By message structure I refer to all structural information which is needed for uttering a short sentence or a phrase: phonologic, prosodic, syntactic, semantic, as well as pragmatic specifications. Thus, it is a situationally coloured language structure.

The message structure differs widely between languages. The most important differences as far as rhythm is concerned include stress, quantity distinctions, and syllable and word structure. These characteristics all have to be preserved throughout the production process in order to signal the intended message.

The characteristics of the message structure form the basis for grouping. Stressed alternating with unstressed syllables would be such a basis giving groups of one stressed and a number of unstressed syllables. Such a stress group is commonly defined as one stressed syllable and all following unstressed ones up to the next stressed syllable irrespective of word and syntactic boundaries.

However, both word boundaries and syntactic boundaries might be alternative bases for grouping. Thus, there may be words or word groups beside stress groups. Support for such alternatives may be found in /6/. Other characteristics of the message structure may be used too, and it seems most reasonable to suppose that different languages use different kinds of bases for groupings. Also, in each specific language there may be a certain optionality in the choice of what to base

the grouping on. Different alternatives may "compete" with each other. What determines the specific kind of grouping may be the situation as a whole and the specific intentions of the speaker.

Articulatory planning

In the articulatory planning grouping is a means of structuring information. Grouping is assumed to occur at two levels at least. At the first one the input string is restricted to contain units about the size of a short sentence or a phrase constituting the message structure as described above. The basis for this may be intonation characteristics coinciding with syntactic boundaries and delimiting semantically coherent units.

At the next level this string is scanned for elements to base further grouping on, for example stressed syllables or different kinds of boundaries as suggested above. Thus there will be stress groups or possibly words or word groups.

If the groups contain more than just a few elements there may be further subgrouping. Most reasonably, this would be the basis for rhythmic alternation of unstressed syllables. In this case grouping seems to be achieved by strengthening some in a string of several unstressed syllables /7/.

The planning system converts the elements in the input string into articulatory coded units. These units are then converted into commands to the motor execution system and eventually transformed into acoustic events in the speech signal.

The rhythmic structure

By rhythmic structure, the output of articulatory planning, I refer to those aspects of the temporal structure of the speech signal which are the basis for the impression of rhythm. Rhythmic structure is a result of both the articulatory planning and the message structure, as effects of both will be mixed in the speech signal.

However, within the conceptual frame as given above, together with careful analysis of the specific data, the two effects may be separated. Furthermore the strength of each may be predicted in each specific case.

TESTING SOME PREDICTIONS OF THE MODEL

Data

I will present some data from Swedish, Spanish, and Finnish, chosen so as to represent rhythmically different languages. Referring to the frequently-used rhythmic dichotomy Swedish would be a stress-timed and Spanish a syllable-timed language. Finnish, though difficult to categorize in rhythmic terms, was chosen for the complexity of its quantity system. This would make it possible to test the interplay between temporal constraints of the planning process and the constraints of the input structure.

The material consisted of sentences which were syntactically and semantically similar in all three languages. They all had an invariant frame in which test words with different numbers of syllables were inserted. The sentences were read in a neutral manner without giving prominence to any specific word.

The data are more thoroughly accounted for in /2/, p. 117-146:

The predictions against data

1 There will be similar temporal adjustments to regularity irrespective of the language spoken.

If grouping is a natural means of structuring information and tendencies to temporal adjustments are a consequence of grouping, then temporal adjustments should occur in languages in general. Also, as grouping occurs hierarchically, there should be adjustments on several levels, for example, (a) the phrase and (b) the stress group.

The effects of articulatory planning, then, will be temporal adjustments of segments and syllables so that they are more compressed the more elements there are in a unit.

However, there are no claims regarding isochrony and in effect, no timing rules at all are implied. The temporal adjustments are seen simply as a consequence of the assumed tendency to cyclic and regular processing as outlined above.

The data support the prediction. There are similar temporal adjustments decreasing the temporal differences between stress groups with different numbers of syllables in all three languages. In stress groups with one, two, and three syllables the duration of the first (stressed) syllable decreased successively upon the addition of the second and third syllable. Figure 1 gives an example from Spanish. Thus, temporal adjustments associated with stress-timing seem to occur also in languages assumed to be syllable-timed. Assuming that the temporal adjustments are related to articulatory planning, it seems that the stress-timing/syllable-timing distinction does not reflect different planning strategies.
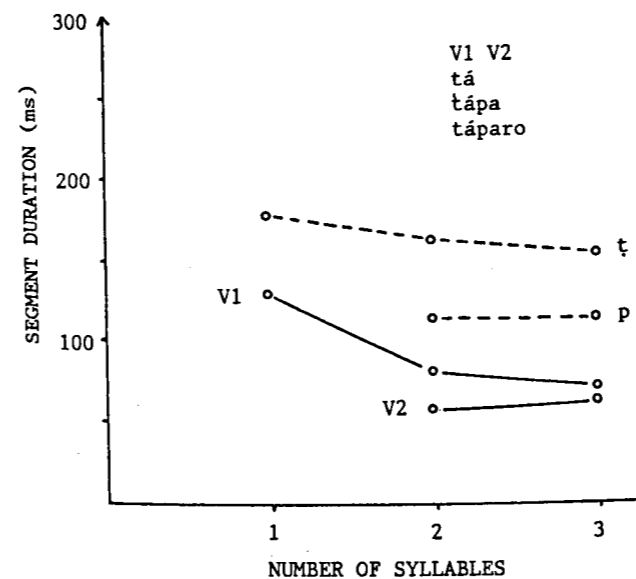

Figure 1. Duration of vowels and consonants in three Spanish test words as a function of the number of syllables (1-3) in the test word. The words were inserted in a sentence frame. N=6. From /2/.

2 Differences in rhythmic structure between languages is a consequence of structural differences.

The message structure, without doubt, is an important determinant of rhythmic structure. The reason naturally is that many of the characteristics of the input structure carry the burden of functional distinctions which have to be preserved in order to convey the intended message to the listener. Especially important characteristics are (a) stress, (b) quantity distinctions and (c) syllable structure.

It seems most reasonable to assume that such differences form the basis for the stress-timing/syllable-timing dichotomy. So-called stress-timed languages, for example, seem to have a clear distinction between stressed and unstressed syllables. In general, the stressed syllables have a more complex structure than the unstressed syllables. In so-called syllable-timed languages, on the other hand, stressed and unstressed syllables are structurally more alike /8/.

There is empirical support also for the second prediction. Language-specific structural characteristics and their temporal manifestations differ widely in the three languages. The greater similarity, structural and temporal, of stressed and unstressed syllables in Spanish as compared to Swedish contributes to making rhythmic structure quite different in the two languages. And the elaborate quantity system in Finnish contributes to the characteristics of rhythmic structure in this specific language.

3 Temporal adjustments to regularity will only occur insofar as functionally important structural features are not destroyed.

The planning mechanism is sensitive to the specifications in the message structure. Therefore, general characteristics of articulatory planning will be temporally reflected only when the temporal aspects of planning and these specifications do not conflict. When in conflict, the constraints of the message structure take precedence over, or simply obscure, temporal constraints of planning. Such conflicts may arise more often in elaborated than in neutral renditions of speech. The maintenance of certain structural distinctions may also produce such conflicts. They may occur, for example, in languages with elaborated quantity systems.

An analysis of the Finnish data point to a complex control of articulatory planning. Obviously there is complex interplay between the constraints of the input structure and the planning mechanism. This interplay seems to be conditioned by the demands of the quantity distinctions in Finnish. As is well known, both vowels and consonants are phonologically either long or short in Finnish, and this distinction has to be made in both stressed and unstressed syllables. In the speech output the phonological distinction is reflected in segments of longer or shorter duration. What happens in Finnish is that there are temporal adjustments to a certain regularity of stressed syllables in some cases but not in others. Phonological length seems

to be an important conditioning factor, as only long vowels and consonants are compressed to any significant degree. A second conditioning factor seems to be whether the temporal adjustments will obscure important quantity relations or not. Thus, compression only occurs insofar as it will not affect the quantity relation between the first and second syllable in a word. Figure 2 showing two different cases, one with (a) and the other without (b) temporal adjustments, illustrates this conditionality.

This may be the reason why Lehtonen /4/ found no compression effects in Finnish. His study was based mainly on such quantity patterns in which compression would be very restricted.
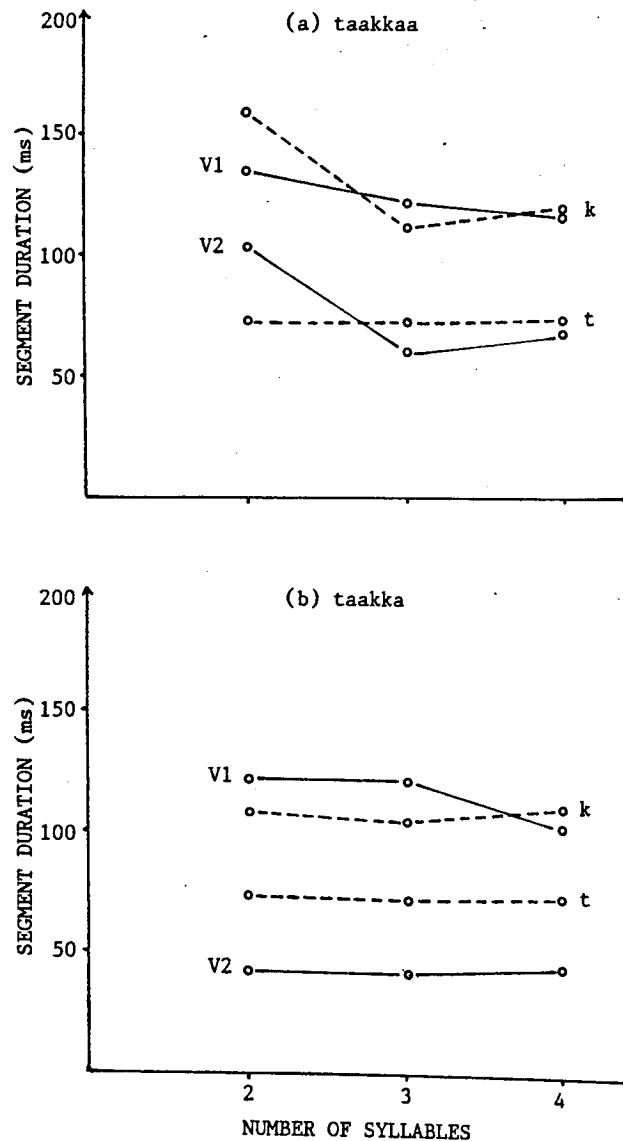

Figure 2. Duration of vowels and consonants in Finnish two-syllable sequences as a function of the number of following syllables. The test words contained 2-4 syllables: (a) taakkaa and (b) taakka followed by nsa and nsa + han in words with three and four syllables, respectively. The words were inserted in a sentence frame. N=6. From /2/.

CONCLUDING REMARKS

The conceptual frame as outlined fits well to the observations in the three languages which were chosen so as to represent different kinds of rhythm. The data reveal the expected differences as well as the similarities between the three languages. Thus the frame may be used as a starting-point for further research on speech rhythm.

A more detailed account of the contents of this paper is given in /9/.

REFERENCES

/1/ Pike, K.L. 1946. The Intonation of American English. Ann Arbor: University of Michigan Press.
/2/ Strangert, E. 1985. Swedish Speech Rhythm in a Cross-Language Perspective. Umeå Studies in the Humanities 69. Stockholm: Almqvist & Wiksell International.
/3/ Lehtonen, J. 1974. Word length and sound durations. Virittäjä, 78, 160. (English summary of a paper in Finnish).
/4/ Fraisse, P. 1982. Rhythm and tempo. In Deutsch D. (Ed.) The Psychology of Music, 149-180. New York: Academic Press.
/5/ Woodrow, H. 1951. Time perception. In Stevens, S.S. (Ed.) Handbook of Experimental Psychology, 1224-1236. New York: Wiley.
/6/ Fischer-Jørgensen, E. 1982. Segment duration in Danish words in dependency on higher level phonological units. Annual Report of the Institute of Phonetics, University of Copenhagen, 16, 137-189.
/7/ Bruce, G. 1984. Rhythmic alternation in Swedish. In Elert, C.-C., Johansson, I. & Strangert, E. (Eds.) Nordic Prosody III, 31-41. Umeå Studies in the Humanities 59. Stockholm: Almqvist & Wiksell International.
/8/ Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. Journal of Phonetics, 11, 51-62.
/9/ Strangert, E. 1987. Speech rhythm: Data and preliminaries to a model. Forthcoming.

Se 27.1.4