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 a 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 complexity 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 serve as the marked ones. Alternatively, an important basis for grouping as groups are defined as the marked ones. However, grouping also occurs when there is no alternating stress. Thus, it is a situationally coloured language. The message structure may be used too, and it seems that many languages have different kinds of bases and different kinds of grouping types. This case grouping may be achieved by marking some of the unmarked elements.

Grouping is achieved by first regularity. Grouping elements together means that there are temporal constraints on how elements are related. 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 may be a cyclic and regular processing of information. This means that in groups with many elements there must 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 different elements to longer groups. This is done in order to maintain a certain regularity in the temporal domain.

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

The rhythmic structure

By rhythmic structure we refer to the temporal structuring of syllables and syllables in the speech signal. This means that the rhythmic structure is a result of both the articulatory planning system and the stress-timing system, as effects of both will be mixed in the speech signal. However, within the conceptual frame as given above, together with an analysis of 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

In the following, we will present some data from Swedish, Finnish, and Spanish. The data come from different functional linguistic environments. The temporal aspects of stress-timing seem to occur also in languages assumed to be syllable-timed. Assumptions that the temporal adjustments are related to articulatory planning strategies.

The data are more thoroughly accounted for in [2], p. 117-156.

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 temporal adjustments 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.

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

The message structure, with its high degree of flexibility, is important in order to convey the intended message to the listener. Especially important characteristics of the listener are served in order to convey the intended message. The message structure diff

3. Temporal adjustments to regularity will only occur in languages of functional importance.

The planning mechanism is sensitive to the specifications in the message structure. Therefore, general characteristics as well as specific cues of the input structure and the planning mechanism are important in order to achieve the appropriate output of the planning mechanism. Obviously there is a complex interplay between the input structure and the planning mechanism. This interplay seems to be conditioned by the specific linguistic environment.

The data support the predictions. The next section elaborates the interplay between the input structure and the planning mechanism.
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.

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

/6/ Fischer-Jörgensen, 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, 137-139.