TEMPORAL RELATIONS WITHIN SPEECH UNITS
Summary of Moderator's Introduction

Ilse Lehiste, Department of Linguistics, Ohio State University

The title of the symposium leaves open the question of the type and size of the speech units. The contributors to the symposium have indeed chosen to address themselves to units of quite different types and sizes. Likewise, they have approached the problems connected with the temporal structure of speech units both from the perspective of speech production and from that of speech perception. The contributions include highly theoretical papers, papers presenting detailed results of experiments, and papers falling between these two poles. Some systematization appears to be in order. I would like to present herewith a framework within which I believe the issues can be profitably formulated for the discussions which I hope will follow.

The framework involves three dimensions. One of them concerns the relationship between timing control in production and the role of timing in perception. The second dimension deals with the direction of determination in the temporal organization of spoken language; specifically, with the question whether the timing of an utterance is determined by its syntax, or whether there exist rhythmic principles in production and perception that are at least partly independent of syntax. The third dimension follows directly from the previous two and relates to the type and size of speech units. What is the nature of those units, and are they to be established on the basis of a morphosyntactic analysis of the sentence, or on some kinds of independent phonetic criteria?

Clearly both production and perception are involved in oral communication by spoken language, and it would seem unnecessary to elaborate the point. However, I have had occasion to argue-- against considerable weight of opinion--that durational differences in production, be they ever so significant statistically, cannot play a linguistically significant role if they are so small as to be below the perceptual threshold. It would be wise, I think, to remind oneself periodically of "the evident fact that we speak in order to be heard in order to be understood" (Jakobson et al. 1952). I hope, therefore, that in our discussion of temporal relations within speech units, models of production and models of perception will be related to each other.

The second and third questions concern the direction of determination: does phonology follow syntax, or are we dealing with interacting, but parallel hierarchies? Some researchers have developed programs for generating the temporal structure of a sentence on the basis of segments and syntactic structure, without paying any attention to rhythm. This is, I believe, due to a particular theoretical orientation. Generative phonology operates with segmental features; even suprasegmental features are attached to segments. And in a generative grammar, phonetic output is the last step in the generation of a sentence. An independent rhythm component simply has no place in the theory. For those scholars, then, the speech units are segments, phrases, clauses, and sentences. (And it is quite interesting to see them struggle with units not foreseen in the theory, like syllables and phonetic words.) Researchers who are not fully committed to this theoretical viewpoint operate with certain other units, such as speech measures or metric feet. Again, the reality of both kinds of units can be studied from the point of view of production as well as from that of perception.

Practically all the issues I have outlined are treated in the papers contributed to this symposium. Production is the main concern of the papers of Allen, Bannert, Klatt, and Öhman et al.; perception is the focus in the papers of Carlson et al., Donovan and Darwin, Fujisaki and Higuchi, Huggins, and Nooteboom.

Among the papers dealing with production, Bannert considers the effect of sentence accent on the duration of VC sequences, employing a rather complex concept--vowel-to-sequence ratio V/(V+C). The relationship between the VC-unit and its two parts represents a measure of the temporal structure of quantity of complementary length. Bannert shows that this unit is useful in describing the effect of the addition of sentence accent to quantity in Stockholm Swedish; it remains to be demonstrated whether the unit is as significant for perception as it is for production.

The paper by Klatt presents a detailed scheme for the synthesis by rule of segmental durations in English sentences. It is an almost pure example of the approach that starts from an abstract linguistic description and ends up as a sequence of segments, whose durations are conditioned by other segments and by syntactic con-

straints. Interestingly, a companion paper by Carlson et al. testing the output of Klatt's synthesis algorithm arrives at the conclusion that certain aspects of the durational pattern have greater perceptual importance than others. Vowel duration is more important than consonant duration; the durations between stressed vowel onsets seem to constitute a particularly important aspect of sentence structure.

The papers by Öhman et al. and by Allen concern themselves with production models in general. Öhman's et al. paper argues for a gesture theory of speech production. Their examples deal primarily with the assignment of fundamental frequency and are thus somewhat outside of the current topic. Allen's paper draws a useful distinction between descriptive models and theoretical models of speech timing, and makes the intriguing prediction that theoretical models may be about to undergo substantial modification, primarily due to the emergence of an "action theory" of speech production. According to that theory, neural activity is hierarchically organized into successively higher levels of coordination, until the highest level of all can only be described in terms of the overall goal of the action.

Among the papers devoted primarily to perception, Nooteboom presents a decision strategy for the disambiguation of vowel length in Dutch. The strategy is complex, but listeners are fully capable of applying it in ongoing perception. Fujisaki and Higuchi present an analysis of the temporal organization of segmental features in Japanese disyllables consisting only of vowels, and find that although the onsets of the transition for the second vowel are distributed over a relatively wide range, a perceptual analysis of the onset of the second vowel shows relatively little temporal variation. It thus seems that the apparent diversity of the onset of transition in various disyllables is introduced to maintain the uniformity of perceived duration of segments. Fujisaki and Higuchi consider their results supportive of a model in which the motor commands and the articulatory/acoustic realizations of successive segments are programmed in such a way that the perceptual onsets of successive segments are isochronous.

The last two papers are likewise concerned with speech rhythm. Huggins finds that a correct rhythmic pattern, which is basically isochronous, enhances intelligibility, while a badly distorted

timing pattern impairs it seriously, even though all phonemes are identifiable. Donovan and Darwin deal with the perceived rhythm of speech, and give special consideration to the problem of isochrony. Their paper tests, among others, a hypothesis that I had formulated in 1973 and discussed in more detail in 1977. My observation was that listeners tend to hear utterances as more isochronous than they really are, and that listeners perform better in perceiving actual durational differences in non-speech as compared to speech. I concluded from this that isochrony is largely a perceptual phenomenon. Donovan and Darwin have confirmed these results. They make two points in addition: first, that isochrony is a perceptual phenomenon which is not independent of intonation, and second, that it is a perceptual phenomenon confined to language--reflecting underlying processes in speech production. Donovan and Darwin question the value of seeking direct links between syntax and segmental durations rather than indirect ones by way of an overall rhythmic structure.

I should like to propose a few direct questions for starting the discussion. What is the relationship between rhythm and syntax? How should rhythm be integrated into models of speech production and perception? What are the physiological constraints within which the production and perception of temporal structure must take place? What, indeed, is the nature of the temporal relations within speech units?

References

Jakobson, R., C.G.M. Fant, and M. Halle (1952): Preliminaries to Speech Analysis, Cambridge, Massachusetts: MIT Press, tenth printing, 1972.

Lehiste, I. (1973): "Rhythmic units and syntactic units in production and perception", JASA 54, 1228-1234.

Lehiste, I. (1977): "Isochrony reconsidered", JPh 5, 253-263.