IMPLICATIONS OF TEMPORAL COMPENSATION FOR SPEECH PRODUCTION MODELS

LINDA SHOCKEY AND RICHARD GREGORSKI

Our most recent research has been concerned with the temporal aspects of speech and especially with temporal compensation. Our efforts were aimed at establishing the maximum domain of temporal compensation with a view towards delimiting the size of the unit of temporal programming. At present we are concerned with the implications of our results for a model of speech production.

Shockey, Gregorski, and Lehiste (1971) have suggested that utterances are temporally programmed in units larger than the syllable and tentatively propose that the domain of temporal organization may extend to an entire short utterance. We propose that a model of speech production allow for this extent of programming.

The current status of speech production models is rather primitive, especially in the area of timing relationships, which are usually represented by a rhythm generator component. Our efforts here are limited to comments on the function of the timing unit, assuredly a small contribution, but one which will eventually have to be taken into account.

It appears that present models of speech production consider the syllable as the basic unit of articulatory programming. Kozhevnikov and Chistovich (1965), in their model for the generation of a word, suggest that the syllable elements are "consecutively connected to each other through coincidence circuits", and propose a rhythm generator as part of the mechanism for syllable excitation. Since our results show a high degree of temporal interaction between both syllables and larger units, there must be postulated a temporal integration component which adjusts the articulatory program to approximate a temporal template. This mechanism should perhaps be introduced before Kozhevnikov and Chistovich's rhythm generator and syllable coincidence circuits, to allow for the execution of the adjusted articulatory program.

Another more elaborate model is Tatham's (1970) design for implementing articulatory speech synthesis by rule. His model includes a process which generates a target program for each segment in the syllable as dictated by the linguistic structure of the language and which at the same time considers the nature of the other segments to be included in the syllable. The composition of the syllable has been previously deter-

mined in this model using cohesion factors deduced from EMG data and compensation factors deduced from exploratory studies in temporal compensation. Next his model places limits on coarticulation factors which restrict the freedom of range of articulatory variables (presumably regarding duration as well as segmental quality) but which do not alter the established target program.

We suggest two minor modifications of this model: (1) reconsider the use of temporal compensation as a determiner of English syllable structure, and (2) incorporate in a near-final stage of the model a temporal program including timing information for an entire short utterance or phrase.

Our first suggestion is based on the information derived from our most recent study (1971) that some degree of negative correlation exists between most adjacent segments with no clear patterns emerging which would give evidence for programming in terms of either VC or CV syllables. Thus, we feel it is unwarranted to base syllabification on temporal compensation data. Of course, our suggestion in no way precludes the use of physiological data in establishing syllabic structure.

Our second suggestion, also derived from our compensation study, is based on the fact that, even when a rate normalization technique is not followed, the sum of the segmental variances for a short utterance is far greater than the variance of the utterance taken as a whole, implying that a super-program for timing exists at some level. Tatham's limits on articulatory variables do not take into consideration the fact that durational variation is not arbitrary, i.e. that each segment varies non-independently in time. We propose that a low-level 'optimal duration of utterance' program, based on rate, be considered, such that temporal freedom is controlled and durations are complementary.

In summary, we suggest that speech production model builders provide for a temporal adjustment component which allows for the utterance-size temporal integration manifested in our compensation research.

Department of Linguistics
Ohio State University
Columbus, Ohio

REFERENCES

Kozhevnikov, V.A. and L.A. Chistovich
  1965 *Speech: Articulation and Perception*, translated by J.P.R.S. (Washington, D.C., JPRS, Moscow-Leningrad).
Tatham, M.A.A.
  1970 "A Speech Production Model for Synthesis-by-Rule", *Ohio State University Working Papers in Linguistics* 6 (September).
Shockey, L., R. Gregorski, and I. Lehiste
  1971 "Word-Unit Temporal Compensation", *Ohio State University Working Papers in linguistics* 9.

DISCUSSION

GODFREY (Dayton, Ohio)
Do you know of any studies on these timing effects in other languages with different syllabic structures and/or stress rules than English?

SHOCKEY
Professor Ilse Lehiste, our colleague at Ohio State, is reporting on a similar study in Estonian at this Congress. Estonian has both different stress rules and different syllabic patterns from those found in English. [Lehiste's article may be found on pp. 929-939 of this work.]

PURCELL (Los Angeles)
In answer to Mr. Godfrey's question whether timing phenomena had been investigated for languages other than English, I would like to say that I have recently been involved in an investigation of timing and the modifications of segmental duration in Serbo-Croatian. Although the study is in a preliminary stage, it seems possible to state that factors such as those described by Shockey and Gregorski operate in Serbo-Croatian, although the situation is more complex because of phonemic length distinctions which appear in Serbo-Croatian.

HARMS (Honolulu)
Have you — or do you plan to — extended this analysis into an *A B A B...* dyadic speech chain?

SHOCKEY
No, we have no plans at this time to investigate discourse, although we agree that it is an intriguing idea.

NOOTEBOOM (Eindhoven)
The two major arguments in your paper were based on the negative correlation between adjacent segments and the relation between the standard deviation for the segments and that for the whole. I will comment on both. If you assume that there is some uncertainty in the mechanical movement of the articulators, plus a feedback from the moment of arrival of a neural command instead of a feedback from the moment of achievement of the gesture, one would expect to find a negative correlation between adjacent segment durations, and still could accept the improbable chain model. If the durations of the segments in a word were independent of one another, one would expect that the standard deviation of the word duration was about the square root of the sum of the standard deviations of the segments. Which relation did you find?

SHOCKEY

To your first point, since we also find strong negative correlations between portions of the utterance comprised of two or more segments, we fail to see in what way our results would provide evidence for either the chain or comb model. Secondly, as we mentioned in the body of the paper, we feel that our results point to the conclusion that segments are not produced independently.

LASS (Morgantown, W.Va.)

What would you anticipate happening to the temporal compensation hypothesis if one purposefully increased or decreased the rate of his utterances?

SHOCKEY

Kozhevnikov and Chistovich's (1965) data indicate that quite different sorts of correlations occur between adjacent segments in Russian at different speeds. We do not know whether this will hold true for English. Our research has been concerned only with normal speech rate.