COMMENTS ON "SOME OBSERVATIONS ON THE ORGANISATION AND RHYTHM OF SPEECH"

Fredericka Bell-Berti

Haskins Laboratories, New Haven, CT and St.John's University, Jamaica, NY, USA

ABSTRACT

The proposal that the word is the basic organizing unit of speech production is satisfying as well as being a proposal that can be supported by a substantial body of data. These comments review some of the supportive data and also raise questions about the origins of utterance-level speech-timing effects.

1. INTRODUCTION

The notion of the word as the basic organizing unit of speech production is both intriguing and intuitively satisfying. A good measure of this satisfaction derives, I think, from the idea that the primary organizational unit of speech may be the same for all languages, regardless of their status as "syllable-timed" or "stress-timed," or unspecified on that dimension. Or, in another light, since speech tempo is usually specified as a measure of syllables per unit of time, if "stress-timing" assumes that the constant duration intervals occur between stressed syllables, then, for some languages at least, "the durations of utterances are determined by syllable count, but not all syllables count" [26].

However, before embracing a model having the word as the basic unit of speech production, it would be useful to be able to answer what, on the surface, appears to be a very simple question: "What is a word?" The most widely offered definition of "word" is that it is a string of characters set off by spaces, and the orthographic conventions for representing words are language-specific. What is not considered in this definition (or attempt at definition) is the nature of a "word" for an illiterate speaker, for a listener to a foreign language, or for a child first learning to speak. For these groups of speech users and hearers, the "words" may be quite different entities than they are for accomplished literate speaker/hearers of a language. (To make this point explicit: consider the many anecdotes of the sort in which a string of segments intended to convey one set of "words" is heard as another: e.g., "pullet surprises" for "Pulitzer Prizes.") And although lack of a satisfactory definition of "word" seriously limits our ability to confirm or reject this model immediately, the study necessary to answer questions abut the nature the word ought to provide evidence useful in evaluating this model. That is, identifying the nature of the differences (if any) among the "words" of these groups of speakers should provide the linguistic framework necessary for evaluating models of speech production in which "words" play a critical organizing role.

2. SOME RELEVANT STUDIES

Dauer's [11] evidence that their is no more isochrony in English and Thai (both stress-timed languages) than in Spanish (a syllable-timed language) or Italian and Greek (both unclassified on this dimension) offers compelling support for rejection of the stress group as the temporal organizing unit of speech. It will be necessary, though, to reconcile Umeda's [28] data with the word as the organizing unit. Her data do not provide evidence that the number of syllables in a word influences the durations of the vowels in the word; rather, she found that vowel durations could be predicted from a number of phonological conditions, including whether a vowel occurred pre-pausally, whether it occurred in a stressed syllable, and whether it occurred word-finally. Thus, Umeda's data suggest that, other things being equal, syllable number is not very important--instead, one syllable is much like another. This result contrasts with that of Lehiste [22], who found a reduction in syllable duration as suffixes were added to monosyllabic stems. (We must, of course, consider the possible effects of the different tasks on the experimental outcomes: in Umeda's study, *vowel* durations were measured in extended, continuous--i.e., 10-20 minute--speech samples, whereas in Lehiste's study, *syllable* durations were measured in much shorter speech samples.) It is possible to interpret Lehiste's data in terms of the word as the organizing unit: within the word, durations are determined by phonological conditions including the number of syllables. This leads, of course, to questions of the syllable as the organizing unit of speech. However, questions of speech timing and speech rhythm necessarily require units larger than the syllable, since questions of *relative* segment duration mandate comparisons among parts of larger pieces of speech.

Returning, then to the proposal that the word is the temporal organizing unit of speech, it may be instructive to examine some physiological and acoustic data collected for other purposes, to see if and how they may support this model, or how the model must be modified in order to account for these data. To begin, such a model must account for observed differences in the organization of speech gestures in utterances produced at different speaking rates, including differences in the relative magnitude and timing relations of the articulatory gestures for successive segments. There are studies [13, 29] whose data are in conflict with the target-undershoot model [23], suggesting that although increasing speech tempo results in vowels of shorter duration, it does not result in spectrally reduced vowels. This result is taken as support for the word as the temporal organizing unit, and suggests that there is a reorganization of the word's articulatory patterns to achieve the same spectral targets as occur in slower speech. These results, however, also suggest a homogeneity among speakers in the way that they achieve changes in speech tempo, a homogeneity that is contradicted by Harris' [16] data, in which one of her three subjects showed overshoot and the other two showed undershoot of vowels in speech produced with increased tempo. In addition, the genioglossus muscle electromyographic (EMG) data accompanying Harris' acoustic data support the notion of reorganization of the magnitude of the articulatory gestures, and not simply their timing (or overlap, [23]). It is also worth noting, I think, that the relation between acoustic overshoot and undershoot and the underlying muscle potentials is not a simple one. That is, one subject's peak EMG activity was reduced for increased tempo utterances, one subject's activity was substantially increased for increased tempo utterances, and the third subject's activity was only modestly increased for increased tempo utterances.

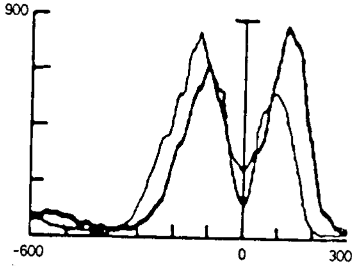

Figure 1. Ensemble-averaged EMG potentials (in.μv) from *genioglossus m.* for 16-20 tokens each of /əpípipə/ (thin line) and /əpipípə/ (heavy line). Zero, the reference point for signal alignment, represents the beginning of /p/ closure in the acoustic waveform.

A model of speech timing that takes the word, rather than the stress group, as its basic unit of organization can also be supported by the EMG and acoustic data of Bell-Berti and Harris [5]. That study reports on the effect of changing primary stress and speaking rate on the separation of lingual EMG activity associated with the production of two /i/ vowels separated by an intervening /p/ or /b/. Briefly, they reported a direct relation between the duration of the medial stop closure and the depth of the trough between the flanking vowel gestures (Fig. 1). The depth of this trough was taken to be a reflection of the change in relative timing (and hence, overlap) between the end of the pre-stop vowel and the beginning of the post-stop vowel. Similar data on the depth of the trough in labial EMG data for /u/ vowels separated by alveolar consonants have also been reported [8, 14, 15]. This view would have the durations of segmental articulatory gestures determined by the position of the segment within the word, the segmental composition of the word, the location of stressed syllables, and speaking rate, but would have the "edge" relations between the gestures for successive segments remain quite stable across substantial changes in all of these parameters (Fig. 2). That is, the timing of the beginning of a gesture for one segment will be relatively unchanged in relation to that segment's acoustic onset (or, viewed another way, the end of the gesture for the preceding segment). This view is supported by the data from a number of studies [5, 6, 8, 9, 15]. That is, these studies and the model they support [1, 7] may be compatible with a model of speech organization having the word as its primary temporal organizing unit.

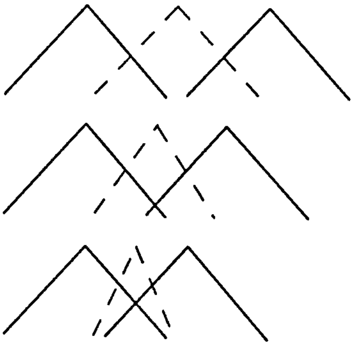

Figure 2. Schematic representation of gestures in a VCV utterance. As the medial (consonant) gesture shortens, the flanking vowel gestures move together, increasing temporal overlap. No changes in gesture magnitude are represented.

Even more directly to the point is Krakow's [20] study of the articulatory organization of syllables. She has shown that the position of a nasal consonant within a syllable determines the relative timing of maximum velum and lip displacements, with velum lowering movements generally being enhanced for syllable-final nasal consonants. In addition, however, Krakow has shown small, but stable, scaling effects on lip and velum movements for word-marginal (both syllable- and word-initial or final) nasals, compared with movements for word-medial (but syllable-initial or syllable-final) nasals. That is, although syllable position determines the basic patterns of articulator movements and interarticulator coordination, the position of a segment within a *word* does effect articulatory organization.

3. UTTERANCE-LEVEL PATTERNS

It is also obvious that there are utterance-level effects on segment durations in speech [18]. Thus, we know that segments occurring late in an utterance will be longer than those occurring earlier in an utterance--the final-lengthening effect [e.g., 24, 25]. So other questions we should address are: What are the sources of utterance-level timing patterns? and, How do 'words' fit into the larger units of language?

One possibility is that some aspects of speech timing are determined by the linguistic characteristics of an utterance (the inherent segment durations and the phonetic, semantic, and syntactic context in which a segment occurs), while other aspects of speech timing are determined by the neurological, muscular and mechanical components of the speech system. Bonnot [10] has proposed that there are two levels of speech-timing control, a motor planning level that results in Tatham's [27] "notional time," whose output timing pattern is the result of linguistic and motor programming interactions, and a motor execution level that results in Kent's [17] "clock time."

One utterance-level timing pattern that has been studied widely is the final lengthening effect [24, 25], whose importance in perceiving speech has been suggested by Klatt and Cooper [19], who have shown that listeners expect longer durations for words in phrase- and sentence-final positions. It has seemed reasonable, at least until this point, to assign final-lengthening to the motor-planning level [12]. However, some recent data [2, 3, 4] from studies of acoustic segment durations in French-speaking normal and cerebellar dysarthric subjects seem to shed a different light on the origin of the final-lengthening effect. In those studies, the final lengthening that was a characteristic of the speech of the normal adult speakers was absent in the speech of a group of ataxic dysarthric speakers whose speech was also marked by an overall slowing of speech tempo (measured as overall utterance duration). That is, the reduced speech rate of the ataxic speakers was not simply the result of a global slowing of speech; rather, the durational relations within an utterance were disrupted. These dysarthric speakers suffered from cerebellar disease, and the cerebellum is though to have a "setting" function for motor activity (possibly through muscle spindle biasing [21]). One possible interpretation of these data [2, 3,], then, is that final-lengthening originates at the motor-execution level. Alternatively, however, it may be that there is a limit on how much reduced speaking rate may be and still have final elements that are relatively slower than those occurring earlier in the utterance [4].

4. CONCLUSION

It seems, then, that substantial support can be found in the speech production literature for the word as the basic unit of speech organization. It is also reasonable to assume, though, that, in addition to basic units, we must also identify the larger linguistic units that affect speech timing (e.g., utterance-level effects), as well as the role that physiological systems may play in determining the temporal characteristics of speech. Furthermore, the development of a comprehensive model of speech timing requires that we explore the interactions between the linguistic and physiological systems involved in producing the timing patterns speech.

5. ACKNOWLEDGMENTS

The preparation of this paper was supported by NIH Grants DC-00121 and RR-05596 to the Haskins Laboratories.

6. REFERENCES

[1]BELL-BERTI, F. (1980), "A spatial-temporal model of velopharyngeal function", in N. J. Lass (Ed), *Speech and Language: Advances in basic research practice* (Vol. IV). New York: Academic Press.

[2]BELL-BERTI, F. & CHEVRIE-MULLER, C. (to appear), "Motor levels of speech timing: Evidence from studies of ataxia", in H. F. M. Peters & C. W. Starkweather (Eds.), *Speech Motor Control and Stuttering*. Amsterdam: Elsevier.

[3]BELL-BERTI, F., GELFER, C. E., BOYLE, M., & CHEVRIE-MULLER, C. (1990), "Neurological factors in speech timing", *Journal of the Acoustical Society of America, 88,* S128 (A).

[4]BELL-BERTI, F., GELFER, C. E., BOYLE, M., & CHEVRIE-MULLER, C. (1991), "Speech timing in ataxic dysarthria", *Proceedings of the XIIth International Congress of Phonetic Sciences*, Paper #614 Aix-en-Provence, France.

[5]BELL-BERTI, F., & HARRIS, K. S. (1974), "More on the motor organization of speech gestures", *Haskins Laboratories Status report on Speech Research, SR37/38,* 73-77.

[6]BELL-BERTI, F. & HARRIS, K. S. (1979), "Anticipatory coarticulation: Some implications from a study of lip rounding", *Journal of the Acoustical Society of America, 65,* 1268-1270.

[7]BELL-BERTI, F., & HARRIS, K. S. (1981), "A temporal model of speech production", *Phonetica, 38,* 9-20.

[8]Bell-Berti, F. & Harris, K. S. (1982), "Temporal patterns of coarticulation: Lip rounding", *Journal of the Acoustical Society of America, 71,* 449-454.

[9]BELL-BERTI, F. & KRAKOW, R. A. (submitted), "Anticipatory velar lowering: a coproduction account", *Journal of the Acoustical Society of America.*

[10]BONNOT, J.-F. (1989), "Timing intrinsèque et timing extrinsèque: le temps est-il une variable controlée?", *Journal d'Acoustique, 2,* 287-296.

[11]DAUER, R. M. (1983), "Stress-timing and syllable-timing reanalyzed", *Journal of Phonetics, 11,* 51-62.

[12]EDWARDS, J., BECKMAN, M. E., & FLETCHER, J. (1991). The articulatory kinematics of phrase-final lengthening. *Journal of the Acoustical Society of America, 89,* 369-382.

[13]GAY, T. (1978a), "Effect of speaking rate on vowel formant movements", *Journal of the Acoustical Society of America, 63,* 223-230.

[14]GAY, T. (1978b), "Articulatory units: Segments or Syllables?", in A. Bell and J. B. Hooper (Eds.) *Syllables and segments.* Amsterdam: North-Holland.

[15]GELFER, C. E., BELL-BERTI, F., & HARRIS, K. S. (1989), "Determining the extent of coarticulation: Effects of experimental design", *Journal of the Acoustical Society of America, 86,* 2443-2445 (L).

[16]HARRIS, K. S. (1978), "Vowel duration change and its underlying physiological mechanisms", *Language and Speech, 21,* 354-361.

[17]KENT, R. D. (1983), "The segmental organization of speech", in P. F. MacNeilage (Ed.), *The Production of Speech,* New York: Springer-Verlag, pp. 57-89.

[18]KLATT, D. (1976), "The linguistic uses of segment duration in English: Acoustic and perceptual evidence", *Journal of the Acoustical Society of America, 59,* 1208-1221.

[19]KLATT, D., & COOPER, W. E. (1975), "Perception of segment durations in sentence contexts", in A. Cohen and S. Nooteboom (Eds.), *Structure and Process in Speech Production.* Heidelberg: Springer Verlag.

[20]KRAKOW, R. A. (1989), *"The articulatory organization of syllables: A kinematic analysis of labial and velar gestures"*, Unpublished Ph. D. dissertation, Yale University, New Haven, CT.

[21]LARSON, C. R., & SUTTON, D. (1978), "Effects of cerebellar lesions on monkey jaw-force control: Implications for understanding ataxic dysarthria", *Journal of Speech and Hearing Research, 21,* 309-323.

[22]LEHISTE, I. (1972), "The timing of utterances and linguistic boundaries", *Journal of the Acoustical Society of America, 51,* 2018-2024.

[23]LINDBLOM, B. E. F. (1963), "Spectrographic study of vowel reduction", *Journal of the Acoustical Society of America, 35,* 1773-1781.

[24]LINDBLOM, B. E. F., LYBERG, B., & HOLMGREN, K. (1981), *"Durational patterns of Swedish phonology: Do they reflect short-term motor memory processes?"*, Bloomington, Indiana: Indiana University Linguistics Club.

[25]LINDBLOM, B. E. F., & RAPP, K. (1973), "Some temporal regularities in spoken Swedish", *Papers in Linguistics, University of Stockholm, 21,* Stockholm.

[26]LISKER, L. (1976), "Phonetic aspects of time and timing", *Haskins Laboratories Status report on Speech Research, SR47,* 113-120.

[27]TATHAM, M. A. A. (1970), "A speech production model for synthesis-by-rule", *Ohio State University Working Papers in Linguistics, 6.* (Cited in R. D. Kent, "The segmental organization of speech", in P. F. MacNeilage (Ed.), (1983), *The Production of Speech,* New York: Springer-Verlag, pp. 57-89.

[28]UMEDA, N. (1975), "Vowel duration in American English", *Journal of the Acoustical Society of America, 58,* 434-445.

[29]VAN SON, R. J. J. H., & POLS, L. C. W. (1989), "Comparing formant movements in fast and normal rate speech", in J. P. Tubach and J. J. Mariani (Eds.), *Eurospeech 89, the European Conference on Speech Communication and Technology,* CEP Consultants, Edinburgh, Volume 2, pp. 665-668. (Cited in Nooteboom, 1991).