Journal of Experimental Psychology:
Human Perception and Performance
1983, Vol. 9, No. 5, 829-833

430

Converging Evidence for the Role of Relative Timing in Speech

Betty Tuller
Cornell University Medical College
and Haskins Laboratories

J. A. Scott Kelso
Haskins Laboratories and University of Connecticut

Katherine S. Harris
City University of New York, Graduate Center, and Haskins Laboratories

In an earlier article (Tuller, Kelso, & Harris, 1982a) we suggested that the timing of consonant-related muscle activity was constrained relative to the period between onsets of muscle activity for successive vowels. Here, we reexamine those data based on reservations posed by Barry (1983). Next, we present a kinematic study of articulation that extends and strongly supports our original observations. Finally, we very briefly survey some converging lines of evidence for a functionally significant vowel-to-vowel period in speech and how this may relate to the role of temporal invariance in motor skills in general.

In his review, Barry (1983) made some well-reasoned comments, which have given us further insight into our previously presented data and have encouraged us to look at the results of a newly completed study within a similiar perspective. Barry's first point is that our results may be, in some sense, a statistical artifact. Just as most of the durational stretching and shrinking across rate and stress changes occurs in the vowel portion of the acoustic signal, the vowel-related electromyographic (EMG) activity is also the most elastic part of production. Changes in duration of consonant-related activity are smaller, though systematic (cf. Tuller, Harris, & Kelso, 1982). This alone, according to Barry, might account for the fact that the correlations we computed of the interval between the onsets of muscle activity specific to production of successive vowels and the timing of muscle activity for the intervening consonant (Barry, 1983, Figure 1, section a) are higher than the correlations between the onsets of muscle activity for successive consonants and the timing of activity for the intervening vowel (Barry, 1983, Figure 1, section b). To explore this possibility, we followed Barry's suggestion and correlated the period between successive consonant onsets with the vowel-onset-to-consonant-onset interval. In all cases, this resulted in a lower correlation than our original measure. The shape of the histogram of correlations based on Barry's suggested analysis, presented in Figure 1, section A, is significantly different (Kolmogorov-Smirnov test for $r > .8$, $p < .001$) from the distribution arising from our procedure, that is, the one obtained by correlating the period between vowel onsets with the interval from vowel onset to consonant onset (see Figure 1, section B).

Although this analysis shows that the correlation measure we used will give higher correlations than the one Barry (1983) suggested as a substitute, the results do not address a crucial point that underlies our argument and is obliquely addressed by Barry. We believe that we obtain our correlation results because the small changes in duration of consonant-related activity are correlated with the relatively larger changes in duration of vowel-related activity over the averaged effects of stress and speaking rate on an ensemble of tokens. If this is true on the average across stress and rate conditions, the same relations should hold for individual tokens within stress and rate conditions. As we pointed out in a previous article (Tuller, Kelso, & Harris, 1982) there is no need to assume that changes in vowel- and consonant-related activity are ratiomorphic, and, indeed, neither we nor Barry believes they usually are. However, we cannot examine this point in detail using EMG data because it is not always possible to define onsets and offsets in individual repetition tokens of an utterance (see Baer, Bell-Berti, & Tuller, 1979, for a discussion of temporal measures of individual vs. averaged EMG records). For this reason, we describe a more recent experiment in which articulator movement trajectories were measured, which can, of course, be analyzed on a token-by-token basis.

Since the publication of our previous article in this journal (Tuller, Kelso, & Harris, 1982a), we

This work was supported by National Institute of Neurologic and Communicative Disorders and Stroke (NINCDS) Grant NS-17778 to Cornell University Medical Center, NINCDS Grant NS-13617, U.S. Office of Naval Research Contract No. N00014-83-C-00833, and Biomedical Research Support Grant RR-05596 to Haskins Laboratories.

Requests for reprints should be sent to Betty Tuller, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511.

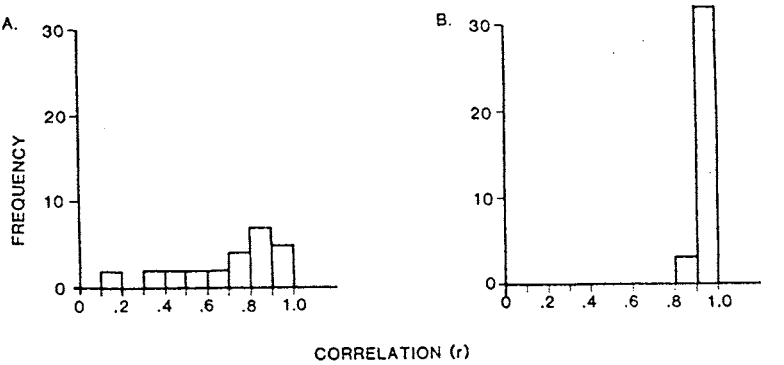

Figure 1. (A) Distribution of correlations for the period between onsets of muscle activity for successive consonants and the latency of onset of vowel-related muscle activity. (B) Distribution of correlations for the period between onsets of muscle activity for successive vowels and the latency of onset of consonant-related muscle activity.

have extended our observations to the kinematics of the jaw and lips during speech (Tuller, Kelso, & Harris, 1982b). Briefly, subjects produced utterances of the form /bVCab/, where V was either /a/ or /æ/ and C was from the set /p, b, v, w/. Each utterance was spoken with two stress patterns and at two self-selected speaking rates, conversational and relatively fast. In essence, the experimental design incorporated and extended the earlier design of our EMG study. Ten to 12 repetitions were produced of each utterance type. Articulatory movements in the up–down direction were monitored by an optoelectronic device that tracked the movement of lightweight, infrared, light-emitting diodes attached to the subjects lips and jaw. (Details of data collection and processing may be found in Tuller, Kelso, & Harris, 1982b.)

To examine more closely whether the high correlations obtained in the EMG experiment are a function of using means in the analyses or perhaps are solely due to the effect of variations in vowel duration, we performed three analyses of /bapab/ (the one utterance common to both experiments) produced by the only subject who participated in both studies. First, we asked the original question about stress and rate variations: Does the interval from vowel onset to consonant onset change systematically as a function of a vowel-to-vowel period? To this end, correlations were computed between the period from the onset of jaw lowering for the first vowel to the onset of jaw lowering for the second vowel and the interval between the onset of jaw lowering for the first vowel and the onset of consonant-specific movement (i.e., a close movement analogue of our earlier EMG measure; Figure 2, section A). In separate analyses, the onset of movement for the medial labial consonant was defined by either the onset of upper lip lowering or the onset of lower lip raising (independent of simultaneous jaw movements). Each correlation was based on 35 data points. The Pearson's product-moment correlations were .97 and .96 for the lower and upper lip, respectively (Figure 2, sections B and C). These kinematic results, obtained from measures of individual repetitions of each utterance type, essentially mirror our earlier EMG findings, which were based on utterance-ensemble averages.

In a second analysis, we examined the movement analogue of Barry's suggested analysis by correlating the interval between onsets of upper lip lowering (or lower lip raising) for successive consonants with the interval between vowel onset (as indexed by the onset of jaw lowering) and the following consonant onset. These correlations were significantly lower (using Fisher's *r*-to-*z* transform) than those obtained by our original definition of period and latency (when consonant production is indexed by upper lip movement, $r = .70$ vs. .96, $t[32] = 3.704$, $p < .001$; when consonant production is indexed by lower lip movement, $r = .76$ vs. .97, $t[32] = 4.384$, $p < .001$). Again, the variations in vowel duration alone cannot account for the systematic relationship between the timing of consonant articulation and the period between successive vowels.

A final analysis was undertaken to examine specifically whether the high correlations obtained are simply a function of the change in vowel duration contributing to both variables or whether they reflect some organizational attribute of each repetition's internal structure. To this end, we explored whether the small changes in duration of consonant-related articulatory movements were correlated with corresponding changes in vowel-related gestures (i.e., Barry's, 1983, suggested correlation of "period" and "period minus latency"). For all repetitions of /bapab/ at both stress and rate levels, we determined the duration of vowel-related movements, defined as the interval from the onset of jaw lowering for

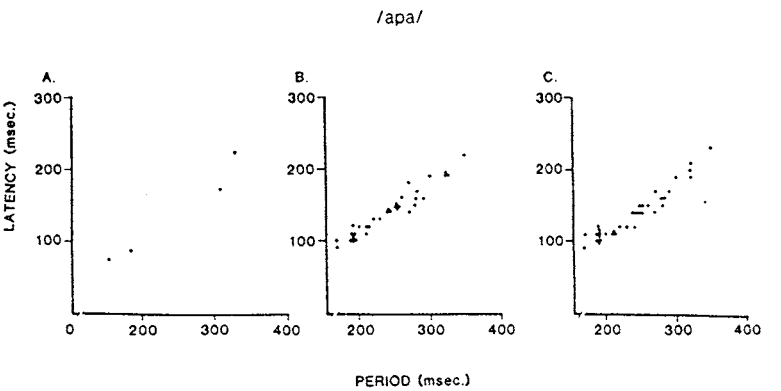

Figure 2. Timing of consonant articulation ("latency") as a function of the vowel-to-vowel period. Each graph contains data from two stress patterns and two rates produced by the same speaker. (Section A: The onset of consonant-related activity in orbicularis oris is graphed relative to the interval between epochs of activity in anterior belly of digastric, $y = .89x + 107$, $r = .89$. Each point represents the mean of EMG data for 12 repetitions of "pa-pap." Section B: Timing of lower lip raising as a function of the vowel-to-vowel period indexed by jaw lowering movements. Each point represents one token of the utterance "ba-pab," $y = .66x - 18$, $r = .97$. Section C: Same as section B but with consonant articulation indexed by the onset of upper lip lowering, $y = .7x - 28$, $r = .96$.)

the first vowel to the onset of lip movement for the /p/, and the duration of movement specific to the consonant, defined as the interval from the onset of lip movement for the /p/ to the onset of jaw lowering for the second vowel. We then calculated the correlation between members of the pairs. If these correlations are significantly greater than zero, then the temporal relations between a vowel and its following consonant are not random and, although vowel duration does contribute to the high correlations, it is not the only significant factor. The duration of vowel and consonant movements were positively correlated; when consonant production was indexed by upper lip movement, $r = .74$, $t(32) = 5.37$, $p < .001$; when consonant production was indexed by lower lip movement, $r = .72$, $t(32) = 5.14$, $p < .001$. In conclusion, we feel that our results cannot be accounted for by vowel variation alone but indicate that the timing of consonant articulation is constrained relative to the timing of articulation for the flanking vowels.

To unpack Barry's third point, we must return to consideration of the EMG data. Barry speculated on the interpretation of results reported in our 1982 article relative to our own earlier findings that the temporal overlap of muscle activity for certain vowels and consonants altered little over marked changes in syllable duration (Tuller, Harris, & Kelso, 1981). Consider the schematic in Figure 3. The interval AC represents the duration of muscle activity specific to the first vowel; the interval BE represents the duration of activity in a different muscle for production of the consonant; and DF is the duration of muscle activity for the second vowel. The "overlap intervals" we have referred to are the time from the onset of consonant-related activity to the offset of activity specific to the preceding vowel (BC in Figure 3), and the time between the onset of activity for the second vowel and the offset of activity for the preceding consonant (DE). In our earlier work, we examined the duration of overlapping activity in a lip muscle (orbicularis oris), acting for production of the consonants "p" and "b," and a tongue muscle (genioglossus), acting for production of the long vowels "ee" and "ay" in utterances such as "pee-peep" and "pay-payp". The overlap intervals (BC and DE in Figure 3) remained remarkably constant across two stress patterns and two speaking rates. In a companion article (Tuller, Kelso, & Harris, 1981), we extended these observations to the activity of various other articulator muscles; in fact, these were the same recordings analyzed for our 1982 article. Although the relatively constant temporal overlap of activity in orbicularis oris and genioglossus again resulted, other muscle comparisons showed different patterns. For example, for the production of "pa-pap" the temporal relationship between activity in a jaw-lowering muscle (the anterior belly of digastric) and activity in a lip muscle (or-

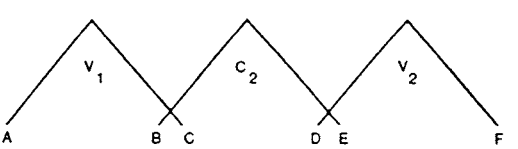

Figure 3. Schematic electromyographic (EMG) activity for a vowel-consonant-vowel (VCV) triad.

bicularis oris) changed systematically as speaking rate increased. Our conclusion was that the temporal overlap of muscle activity in vowel–consonant and consonant–vowel pairs does not, as a rule, remain fixed over metrical variations in speaking rate and syllable stress.

Following from this conclusion, we wish to point out that our thoughts have altered somewhat as to why the overlap interval between orbicularis oris and genioglossus remained unaltered in both experiments (see also Raphael, 1975). It may be that our assumption that the tongue is completely free to assume any position during production of /p/ is in fact incorrect (see also Alfonso & Baer, 1982; Bell-Berti, 1980; Harris & Bell-Berti, in press; Houde, 1967). Rather than conceiving different articulators as being either crucially involved or uninvolved in producing a given sound, we might do better to consider the entire vocal tract as involved in producing all sounds with only the relative importance of individual articulators shifting as the phonetic structure changes. Thus, the constant overlap of orbicularis oris and genioglossus may reflect the articulatory organization that in some way maximizes conditions for production of the bilabial stop consonant and does not reflect feedback-dependent (or for that matter feedback-independent) control of the timing of successive segments.

In Barry's final comment, he expressed surprise that we find that vowel-to-consonant timing is stable relative to the interval between successive vowels even though the vowel and consonant are separated by a syllable boundary. He suggested that the subject was performing an articulatory syllabification different from that which we have represented orthographically. Thus, perhaps the subject was saying something like "peep-eep" rather than "pee-peep." Apart from the fact that such a production strategy seems counterintuitive, we should remark that the intervocalic /p/ was usually aspirated, thus conforming to the conventional description of a syllable-initial form.

Leaving aside the question of articulatory strategies, an issue we have not addressed in any detail, we should remark that temporal and spatial coarticulatory effects are very well documented in the literature. These indicate that syllable boundaries do not necessarily disrupt acoustic or articulatory interactions between segments and, perhaps more to the point, that transsyllabic interactions may be stronger than intrasyllabic ones. For example, the measured acoustic duration of a vowel is strongly affected by the number of transsyllabic consonants that immediately follow it (Lindblom & Rapp, 1973). An effect on acoustic vowel duration of preceding, intrasyllabic consonants has not always been found (for review see Elert, 1964; also see Lindblom & Rapp, 1973). In addition, the acoustic duration of a vowel before a voiceless stop consonant (such as /p/) has long been known to be shorter than the same vowel occurring before a voiced stop consonant (such as /b/), both within ("rip" vs. "rib") and across ("rapid" vs. "rabid") syllables (House, 1961; Klatt, 1973; Petersen & Lehiste, 1960). Transsyllabic articulatory effects have also been documented. As a recent example, Harris and Bell-Berti (in press) reported that in sequences such as [iʔi] and [uʔu] the glottal stop [ʔ] does not cause relaxation of the tongue for [i] sequences or the lips for [u] sequences. In other words, the syllable boundary between the first vowel and the stop does not seem to be articulatorily marked. More generally, there may not be any isomorphism between articulatory syllabification and syllabification as defined by linguists (i.e., if linguists could agree on the rules for syllabification; cf. Bell, 1978).

In his comments, Barry agreed with us that it is at least "plausible" that vowel-to-vowel timing is important for rhythmic structuring. In fact, there are many pieces of evidence in the literature (in addition to the two articles Barry cited) that suggest a functionally significant vowel-to-vowel period (and perhaps, by extension, that commonalities among segments are exploited in production; cf. Fowler, 1977). First, the description of English as being "stress-timed" is based on the perception of stressed vowels as occurring at approximately equal intervals. Although there is little support for a strict stress-timing hypothesis, there is evidence that speakers maintain at least a tendency toward stress-timing that may be more closely associated with the timing of the stressed vowels than with the accompanying consonants (for review see Fowler, in press).

A second source of evidence that a vowel-to-vowel articulatory period may be functionally significant is the literature on compensatory shortening and coarticulation. We have already mentioned that intervocalic consonants shorten the measured acoustic duration of the surrounding vowels. This may mean that all aspects of the articulation of vowels are produced in shorter time periods when consonants follow them. Alternatively, it may mean that the consonants and vowels are produced in concert, with the trailing edges of the vowels progressively "overlaid," as it were, by the consonants. In other words, consonants and consonant clusters might be produced on a background of continuous vowel articulation. An articulatory organization of this sort was first proposed by Öhman (1966) to explain the changes in formant transitions for intervocalic consonants as a function of the flanking vowels. More recent articulatory evidence that the influence of both preceding and following vowels is apparent throughout the intervocalic consonant (Barry &

Kuenzel, 1975; Butcher & Weiher, 1976; Gay, 1977; Harris & Bell-Berti, in press; Sussman, MacNeilage, & Hanson, 1973) might also be interpreted as indicating a significant vowel-to-vowel articulatory period.

In conclusion, let us reiterate our previous conviction that these data are compatible with a style of motor organization in which the relative timing among individual EMG or kinematic events is preserved in the face of scalar changes in, for example, absolute duration and amplitude of EMG activity or articulator displacement and velocity (for reviews see Kelso, 1981; Kelso, Tuller, & Harris, 1983). In fact we believe, with Bernstein (1967), that the cooperativity observed among muscles and joints during coordinated activity is best described by a partitioning of variables into two classes; those that can effect scalar changes in a behavior and those that preserve its internal temporal "topology." Temporal invariance across scalar variation may be a design feature of all motor systems and may constitute one of Nature's solutions to the problems of coordinating complex systems, like speech, that possess many degrees of freedom.

References

Alfonso, P. J., & Baer, T. Dynamics of vowel articulation. *Language and Speech*, 1982, *25*, 151–173.

Baer, T., Bell-Berti, F., & Tuller, B. On determining EMG onset time. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication preprints, 97th meeting of the Acoustical Society of America, Cambridge, Massachusetts, 1979*. New York: Acoustical Society of America, 1979.

Barry, W. J. Some problems of interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 1983, *9*, 826–828.

Barry, W., & Kuenzel, H. Co-articulatory airflow characteristics of intervocalic voiceless plosives. *Journal of Phonetics*, 1975, *3*, 263–282.

Bell, A. Segment organization phenomena and their explanations. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments*, New York: North-Holland, 1978.

Bell-Berti, F. Velopharyngeal function: A spatial-temporal model. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 4). New York: Academic Press, 1980.

Bernstein, N. A. *The coordination and regulation of movements*. London: Pergamon Press, 1967.

Butcher, A., & Weiher, E. An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, 1976, *4*, 59–74.

Elert, C. C. *Phonological studies of quantity in Swedish*. Uppsala, Sweden: Almqvist & Wiksell, 1964.

Fowler, C. *Timing control in speech production*. Bloomington: Indiana University Linguistics Club, 1977.

Fowler, C. A. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General*, in press.

Gay, T. Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America*, 1977, *62*, 183–193.

Harris, K. S., & Bell-Berti, F. On consonants and syllable boundaries. In L. Raphael, C. Raphael, & M. Valdovinos (Eds.), *Language and cognition: Essays in honor of Arthur J. Bronstein*. New York: Plenum Press, in press.

Houde, R. A. *A study of tongue body motion during selected speech sounds*. Unpublished doctoral dissertation, University of Michigan, 1967.

House, A. On vowel duration in English. *Journal of the Acoustical Society of America*, 1961, *33*, 1174–1178.

Kelso, J. A. S. Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, N.J.: Erlbaum, 1981.

Kelso, J. A. S., Tuller, B., & Harris, K. S. A "dynamic pattern" perspective on the control and coordination of articulation. In P. MacNeilage (Ed.), *The production of speech*. New York: Springer-Verlag, 1983.

Klatt, D. H. Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America*, 1973, *54*, 1102–1104.

Lindblom, B., & Rapp, K. Some temporal regularities of spoken Swedish. *Papers from the Institute of Linguistics* (University of Stockholm), 1973, *21*, 1–59.

Öhman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 1966, *39*, 151–168.

Peterson, G., & Lehiste, I. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 1960, *32*, 693–703.

Raphael, L. J. The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics*, 1975, *3*, 25–33.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research*, 1973, *16*, 397–420.

Tuller, B., Harris, K. S., & Kelso, J. A. S. Articulatory motor events as a function of speaking rate and stress. *Haskins Laboratories Status Report on Speech Research*, 1981, SR-65, 33–62.

Tuller, B., Harris, K., & Kelso, J. A. S. Stress and rate: Differential transformations of articulation. *Journal of the Acoustical Society of America*, 1982, *71*, 1534–1543.

Tuller, B., Kelso, J. A. S., & Harris, K. Phase relationships among articulatory muscles as a function of speaking rate and stress. *Haskins Laboratories Status Reports on Speech Research*, 1981, SR-65, 63–90.

Tuller, B., Kelso, J. A. S., & Harris, K. Interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, *8*, 460–472. (a)

Tuller, B., Kelso, J. A. S., & Harris, K. On the kinematics of articulatory control as a function of stress and rate. *Haskins Laboratories Status Reports on Speech Research*, 1982, SR-71/72, 81–88. (b)

Received April 18, 1983 ■