648

Patterns of interarticulator phasing and their relation to linguistic structure

Susan Nittrouer,[a] Kevin Munhall,[b] J. A. Scott Kelso,[c] Betty Tuller,[d] and Katherine S. Harris[e]
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

(Received 25 April 1988; accepted for publication 5 August 1988)

Work by Tuller and Kelso [J. Acoust. Soc. Am. **76**, 1030–1036 (1984)] and Kelso *et al.* [J. Phon. **14**, 29–59 (1986)] has demonstrated stable relations between jaw and lip movements in /bV#CVb/ utterances across rate and stress conditions. Specifically, the onset of lip movement toward the intervocalic consonant was found to be constant with respect to the vowel-to-vowel jaw cycle in both time and relative phasing. An attempt was made to replicate and extend this work by investigating interarticulator phase relations for utterances having a broader range of linguistic organization: In addition to rate and stress, syllable structure (open versus closed syllables) and identity of the intervocalic consonant (/p/ vs /m/) were manipulated. Results showed that the upper lip's lowering onset varied systematically with respect to the jaw vowel cycle as a function of both rate and stress. In addition, syllable structure and consonant identity influenced the relation of lip and jaw gestures. There was a general tendency for any condition that shortened the first vowel to produce earlier onsets of the upper lip relative to the jaw. However, the within-condition jaw cycle duration variability did not correlate with the within-condition variability in phase. Thus it seems that stable interarticulator phase relations maintain not only the integrity of phonological structure, as suggested by Kelso *et al.*, but structural integrity at other levels of linguistic organization as well.

PACS numbers: 43.70.Bk, 43.70.Aj, 43.70.Fq

INTRODUCTION

A central goal of speech research has been to relate the traditional units of linguistic theory (features, phonemes, and syllables) to the physical structures of speech (acoustic and articulatory patterns). Originally, attempts were focused on the acoustic signal so that the objective was to compile a code book of acoustic entries with the corresponding linguistic units (Cooper *et al.*, 1952). However, attempts to relate linguistic segments to articulatory events soon followed (e.g., MacNeilage, 1963). Both lines of work failed in their attempts to relate linguistic and physical structures isomorphically, and so translation accounts (Fowler *et al.*, 1980) of both speech production and perception proliferated to fill the gap (e.g., Liberman *et al.*, 1967; Daniloff and Hammarberg, 1973; Hammarberg, 1976; MacNeilage and Ladefoged, 1976; Studdert-Kennedy, 1976; Wickelgren, 1976). All such accounts require elaborate assumptions about the processes that speakers and listeners go through to translate linguistic entities into observable, physical units, and back again.

More recently, Browman and Goldstein (1986), following the work of Fowler *et al.* (1980) and Saltzman and Kelso (1987), have developed an approach that abandons the concept of linguistic units as abstract, mental entities. Instead, the articulatory structure of speech is examined for lawful regularities from the perspective that such regularities could provide new descriptions of linguistic organization using observable parameters. An essential correlate of this approach is the notion that listeners directly apprehend these lawful articulatory regularities (Fowler, 1986), and so derive linguistic structure.[1] With this approach, speech production and perception are no longer cumbersome or indirect because speakers and listeners produce and perceive the actual matter of speech.

While a fully satisfactory description of these articulatory regularities has not yet been developed, there has been a long tradition of attempts to describe stability in speech production. For example, Sussman *et al.* (1973) and Gracco and Abbs (1986) reported stability in movement sequencing of the articulators in repetitions of the same utterance. Kent and Netsell (1971) and Löfqvist and Yoshioka (1984) reported stability in sequencing over transformations of rate and stress. These studies parallel attempts to describe stability over rate changes among component movements of other sequenced activities, such as typing or walking. It is obvious in all these activities that the simplest possible rule, that components of the sequence expand and contract propor-

---

[a] Present address: Boys Town National Institute for Communication Disorders in Children, 555 N. 30th St., Omaha, NE 68131.

[b] Present address: Department of Communicative Disorders, Elborn College, University of Western Ontario London, Ontario N6G 1H1, Canada.

[c] Present address: Center for Complex Systems, Florida Atlantic University, P. O. Box 3091, Boca Raton, FL 33431.

[d] Present address: Department of Psychology, Florida Atlantic University, P. O. Box 3091, Boca Raton, FL 33431.

[e] Also with: Ph.D. Program in Speech and Hearing Sciences, The City University of New York Graduate Center, 33 W. 42 St., New York, NY 10036.

© 1988 Acoustical Society of America

tionately with rate changes, does not apply (Gentner, 1987). Nevertheless, some lawful timing relations can and have been described among system components involved in many activities.

Some success in discovering lawful regularities among component movements in articulation has been demonstrated. Tuller and Kelso (1984) used a very simple paradigm to explore articulator timing relations over suprasegmental changes. Subjects were asked to produce disyllables of the form /bV#CVb/, with the stress on either the first or the second syllable, and variable speaking rate. Figure 1 displays movements of the jaw, upper lip, and lower lip corrected for jaw movement and the acoustic signal for one token of /ba#'pab/. The jaw vowel period in this analysis was defined as the time between onset of jaw lowering for the first vowel and onset of jaw lowering for the second vowel (A to B), while consonant latency was defined as the time between onset of jaw lowering for the first vowel and onset of movement of either lip for the medial consonant (A to either C or E).[2] Figure 2 more clearly illustrates these articulatory intervals for the jaw and the upper lip. Tuller and Kelso reported Pearson product-moment correlation coefficients ranging in value from 0.84–0.97 between the jaw vowel period and the latency of either lip.

There are several shortcomings in this approach. First, the statistical analysis relied on correlation of a part, the latency, with the whole, the period, of which it is a part. This procedure results in high correlation values because of statistical artifact (Barry, 1983; Munhall, 1985; Benoit, 1986; Sock and Jah, 1986). Second, the analyses of disyllable production have examined only a limited syllabic structure, so it is difficult to know whether the described regularity, if not artifactual, is independent of syllable structure. Third, only the onsets of consonant and vowel events are represented in the analysis, while the trajectories of the movements are ignored.

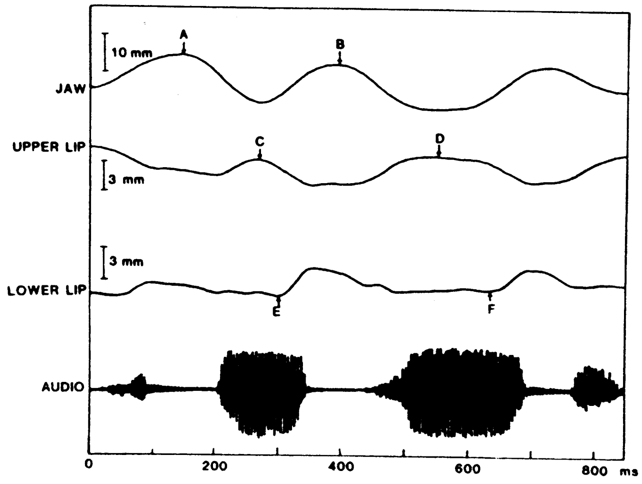

FIG. 1. Movements of the jaw, upper lip, and lower lip corrected for jaw movement and the acoustic signal for one token of /ba#'pab/. Articulator displacement (y axis) is shown as a function of time (x axis). Onsets of jaw lowering for the first and second vowels are represented by points A and B, respectively. Onsets of upper and lower lip movements for /p/ are represented by points C and E, respectively (from Tuller and Kelso, 1984).

Kelso *et al.* (1986; see, also, Kelso and Tuller, 1987) suggested an alternate approach to the analysis of consonant–vowel relations which eliminates two of these problems: the statistical problem of the correlational method and the failure to use whole trajectories. In this approach, the movement of the jaw associated with the vowel-to-vowel cycle is represented in the phase plane (i.e., displacement versus velocity). Thus space as well as time is incorporated into the metric. Because the opening and closing movements are not usually symmetrical, jaw position and velocity are both normalized to the same numerical interval, -1 to $+1$. Normalization of jaw position is computed on the complete cycle, while velocity is normalized on the half-cycle in which the lip movement onset occurs [see Kelso *et al.* (1986) and Kelso and Tuller, 1987, for further discussions of normalization]. Lip closure onsets then may be given as angles on the jaw position–velocity phase plane, as illustrated in Fig. 3. Applying this analysis method to the data of Tuller and Kelso (1984), Kelso *et al.* reported stable phase relations between jaw movement cycles and upper lip movement onsets across variations in speaking rate and stress. Based on this result, it was suggested that phonological units might be reconceptualized as characteristic phase relations among component gestures. Again, the notion of phonological units as abstract, mental entities is abandoned, specifically in favor of the idea that phonological structure is realized, independent of other aspects of linguistic structure, as stable spatiotemporal relations among articulatory gestures. The purpose of the present experiment was to test this idea by investigating the patterns of interarticulator phase relations using utterances demonstrating a broader range of linguistic structure than those used by Kelso *et al.*

I. METHOD

A. Speakers

Four adult speakers (two male and two female) served as subjects for this experiment. All were native speakers of American English, and none had a history of speech or hearing problems. All demonstrated perceptible upper lip movement during informal observation of their conversational speech.

B. Materials and procedures

Four utterances were chosen to vary syllable structure and the identity of the intervocalic consonant:

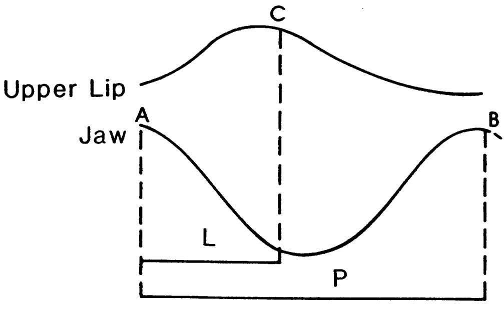

FIG. 2. Upper lip and jaw trajectories for the first syllable in a disyllabic utterance such as the /bV#CVb/ used by Tuller and Kelso (1984). $L=$ consonant latency; $P=$ vowel period (from Kelso and Tuller, 1987).

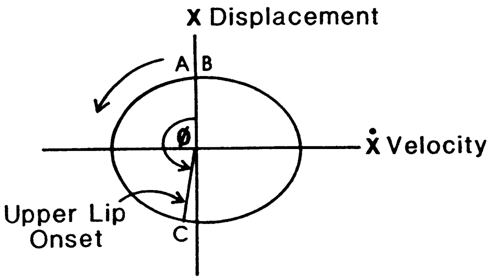

FIG. 3. Onset of upper lip movement toward an intervocalic consonant represented as an angle on the jaw position–velocity phase plane (from Kelso and Tuller, 1987).

(1) /pa#map/, (2) /pam#ap/, (3) /pa#pap/, and (4) /pap#ap/. Speakers were instructed to produce tokens at either a normal or a self-selected faster rate and to place stress on either the first or second syllable. Thus there were 16 conditions in all: two rates (fast or slow) × two stress patterns (first syllable stressed or unstressed) × two syllable structures (first syllable open or closed) × two intervocalic consonants (/m/ or /p/).

Twenty tokens in each condition were obtained by speakers reading the stimuli in the carrier phrase "It's a ____ again." Stimuli were presented in lists of eight, randomized on stress, syllable structure, and identity of the intervocalic consonant. Speaking rate was blocked across lists by instructing speakers to alternate rate from list to list.

C. Data recording

Light-emitting diodes (LEDs) were attached to the vermilion borders of the upper and lower lips and to the jaw and nose at the points that seemed least affected by lip movements. A multichannel optoelectronic device (Tuller and Kelso, 1984) recorded both the vertical and horizontal movements of the LEDs on FM tape. Acoustic recordings were made simultaneously on a separate channel of the same FM tape.

The acoustic and vertical movement channels were digitized on a VAX 780 computer using a 10-kHz sampling rate for the acoustic data and a 200-Hz sampling rate for the movement data. Signals were low-pass filtered with a high-frequency cutoff appropriate to prevent aliasing. The channel on which nose movement was recorded was subtracted from both the upper lip and jaw channels in order to remove any components of the signals associated with general head movement. Only the upper lip was used in these analyses, as in Kelso *et al.* (1986), because it is more independent of jaw movement than is the lower lip. The acoustic channel was used to locate those portions of the waveforms associated with the test stimuli. Velocity records for upper lip and jaw movements (with the nose subtracted out) were obtained by software computation of the first derivative. These displacement and velocity records were then used for identifying vowel periods and consonant latencies in each token, according to the definitions provided by Tuller and Kelso (1984); that is, vowel period was defined as the interval between the onset of jaw lowering for the first vowel and the onset of jaw lowering for the second vowel, and consonant latency was defined as the interval between the onset of jaw lowering for the first vowel and the onset of upper lip lowering for the medial consonant (see Fig. 2). Movement onsets were indexed by points of zero velocity.

II. RESULTS

A. General description

Table I provides mean jaw vowel cycle durations and phase angles for each condition for each speaker. The duration of the jaw cycle varied from 141 ms to 369 ms, with fast utterances and utterances in which the first syllable was unstressed generally demonstrating shorter values than slow utterances and utterances in which the first syllable was stressed. When the first syllable was closed or the intervocalic consonant was the unvoiced /p/, jaw cycle durations tended to be shorter than when the first syllable was open or the intervocalic consonant was the voiced /m/. Within-condition standard errors of the mean for jaw cycle durations ranged from 1.75 ms to 8.39 ms for all four speakers, with a mean of 4.05 ms across speakers.

Phase angles also were affected by linguistic and nonlinguistic manipulations, and the patterns of results were the same as those for jaw cycle durations. Speaking faster, destressing or closing the first syllable, or placing an unvoiced consonant in the intervocalic position tended to decrease phase angle. Within-condition standard errors of the mean for phase ranged from 0.85 to 9.85 deg with an across-subject mean of 2.38 deg. Thus both jaw cycle durations and phase angles were highly consistent within individual conditions.

B. Systematic patterns of upper lip phase angle results

Kelso *et al.* (1986) reported that the angle on the jaw phase plane at which the upper lip began its downward trajectory remained quite stable across manipulations in rate and syllabic stress. Based on the patterns of results described above, it seems that in the present experiment phase angles varied systematically with both rate and stress, as well as with syllable structure and identity of the intervocalic consonant. In general, any condition that served to shorten the jaw cycle duration seemed to decrease the angle within that cycle at which the upper lip began its downward movement. Four specific predictions could be made based on this generality:

(1) Fast tokens would demonstrate smaller phase angles than slow tokens.

(2) Tokens with the first syllable unstressed would have smaller phase angles than tokens with the first syllable stressed.

(3) Tokens with the syllable break occurring after the intervocalic consonant (i.e., closed first syllables) would show smaller phase angles than tokens with the break before the intervocalic consonant (i.e., open first syllables).

(4) Tokens with an intervocalic /p/ would demonstrate smaller phase angles than tokens with an intervocalic /m/.

To investigate whether or not the data supported these hypotheses, an analysis of variance was done for each subject individually, with the results shown in Table II. Only those *F* ratios statistically significant at the 0.01 level are listed, with

TABLE I. Mean jaw vowel cycle (jvc) durations in milliseconds and upper lip phase angle (pha) in degrees for each speaker in each condition. Each value represents the mean of 20 tokens, with the rate × stress (RS) means representing 80 tokens each (SS = slow/stressed, SU = slow/unstressed, FS = fast/stressed, and FU = fast/unstressed).

Rate/stress	SS		SU		FS		FU	
	jvc	pha	jvc	pha	jvc	pha	jvc	pha
Speaker AC								
(1) pa#map	254	185	210	170	193	171	165	161
(2) pam#ap	245	181	206	171	185	166	163	159
(3) pa#pap	246	177	196	171	198	172	164	153
(4) pap#ap	236	177	197	165	192	167	167	153
RS $\bar{X}$ =	245	180	202	169	192	169	165	157
Speaker SN								
(1) pa#map	369	225	220	205	288	209	167	194
(2) pam#ap	353	215	242	199	261	196	170	194
(3) pa#pap	349	213	210	198	274	202	152	194
(4) pap#ap	331	203	235	198	246	192	170	188
RS $\bar{X}$ =	351	214	227	200	267	200	165	193
Speaker CS								
(1) pa#map	337	196	161	182	268	192	141	172
(2) pam#ap	334	197	163	179	275	195	143	174
(3) pa#pap	342	195	163	175	274	198	144	177
(4) pap#ap	322	197	164	187	258	191	150	178
RS $\bar{X}$ =	334	196	163	181	269	194	145	175
Speaker ES								
(1) pa#map	312	193	230	177	253	188	183	167
(2) pam#ap	326	196	239	184	277	193	191	177
(3) pa#pap	293	187	204	164	244	185	172	115
(4) pap#ap	286	186	223	177	249	183	193	165
RS $\bar{X}$ =	304	191	224	176	256	187	185	156

the one exception noted in the table. As can be seen, we obtained statistically significant main effects of rate and stress for all four subjects. Results of manipulating the position of syllable break and for changing the identity of the medial consonant were significant also for three of the four speakers.

To determine whether these statistically significant differences were in the directions predicted, condition means

TABLE II. The F ratios for analyses of variance of phase angles done on tokens from each speaker. Only ratios significant at the 0.01 level are reported (except for the main effect of rate for CS, which was significant at the 0.05 level). Main effects were rate (fast or slow), stress (first syllable stressed or unstressed), syllable structure (syllable break before or after medial consonant), and consonant identity (/p/ or /m/).

Speaker	AC	SN	CS	ES
Main effects				
rate (r)	155.40	66.62	$(4.09)^{0.05}$	92.39
stress (s)	159.82	65.25	82.23	379.53
syllable break (b)	7.36	26.84		81.45
consonant identity (c)	16.67	20.14		140.31
Two-way interactions				
rs		7.05		49.47
rb				18.24
rc				18.19
sb		8.09		59.09
sc				36.17
bc				12.86
Three-way interactions				
rsb				18.73
rsc	10.32			22.94
rbc				10.69
sbc				35.98
Four-way interaction				
$rsbc$				15.79

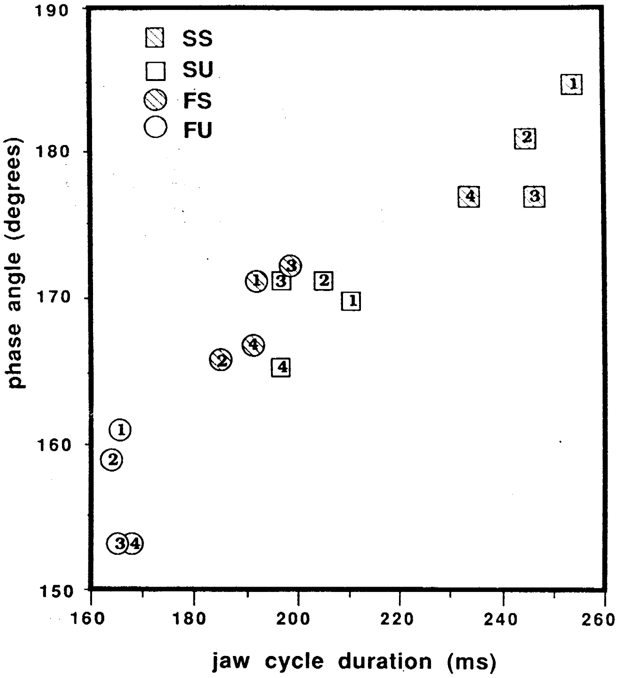

FIG. 4. Plot of condition–mean phase angles (degrees) as functions of condition–mean jaw cycle durations (ms) for subject AC (striped squares = slow/first syllable stressed, open squares = slow/first syllable unstressed, striped circles = fast/first syllable stressed, open circles = fast/first syllable unstressed, 1 = pa#map, 2 = pam#ap, 3 = pa#pap, and 4 = pap#ap).

for phase angle were plotted as functions of condition means for jaw cycle duration for each speaker. Figures 4–7 display these plots.

In Fig. 4, it can be seen that mean phase angle for subject AC varied by roughly 40 deg across conditions. Mean jaw cycle duration varied by about 100 ms, which is one-third the range demonstrated by the other three speakers. From the way the symbols for each rate/stress condition cluster, it is

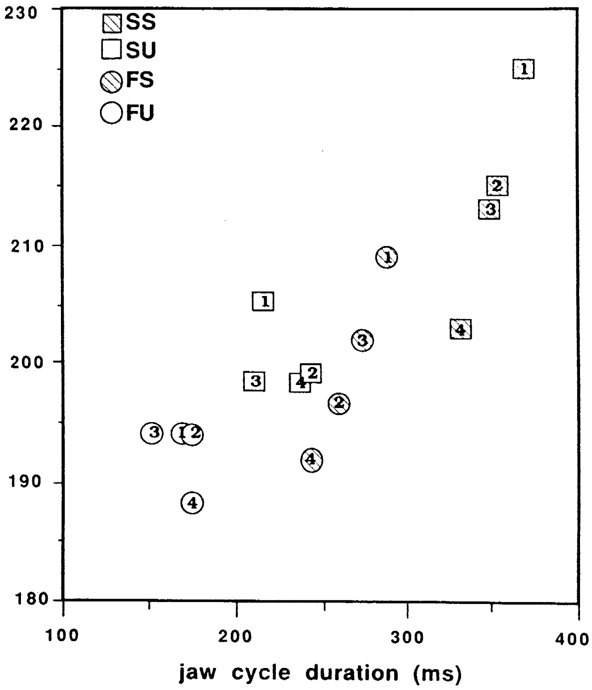

FIG. 5. Plot of condition–mean phase angles (degrees) as functions of condition–mean jaw cycle durations (ms) for subject SN (see the caption for Fig. 4 for details).

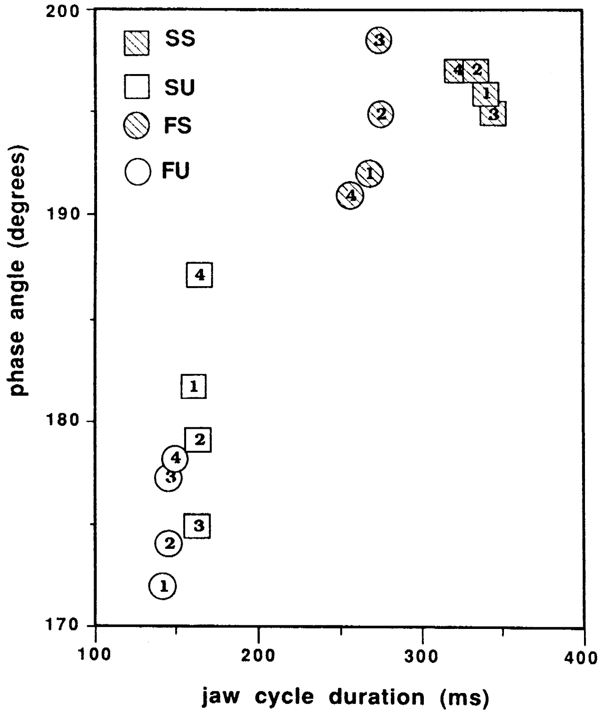

FIG. 6. Plot of condition–mean phase angles (degrees) as functions of condition–mean jaw cycle durations (ms) for subject CS (see the caption for Fig. 4 for details).

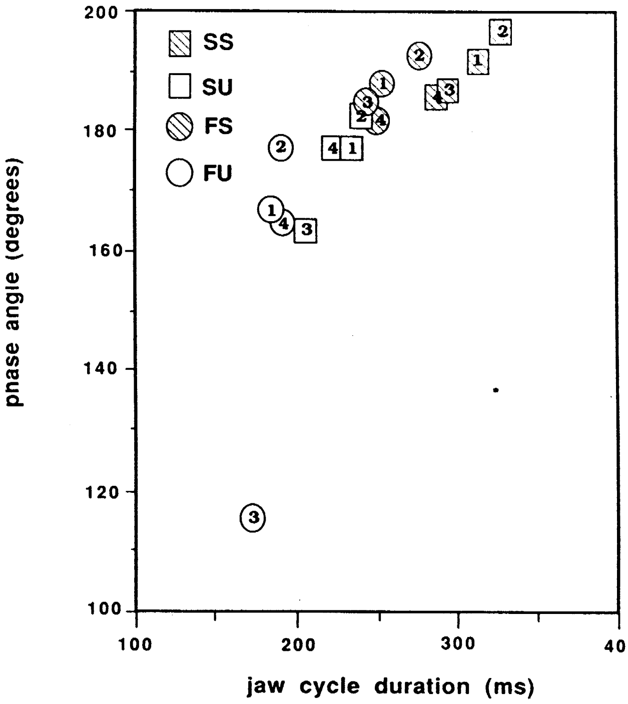

FIG. 7. Plot of condition–mean phase angles (degrees) as functions of condition–mean jaw cycle durations (ms) for subject ES (see the caption for Fig. 4 for details).

clear that rate and stress had relatively larger effects than either the position of the syllable break or the identity of the medial consonant. However, it is also clear that position of syllable break and consonant identity did have some effect, as indicated by the significant main effects on phase angle of these factors. In the slow, stressed condition, the order of utterance type is close to what would be predicted, which would be numbers 1–4, decreasing both in jaw cycle duration and in phase angle (going from the upper, right-hand corner to the lower, left-hand corner). In the fast, stressed condition, closing the first syllable had a relatively larger effect than placing an unvoiced consonant in the intervocalic position (numbers 1 and 3 above 2 and 4), while in the fast, unstressed condition, it was the voicing of the intervocalic consonant that had a relatively greater effect on phase angle (numbers 1 and 2 above 3 and 4). Also, it is apparent from the clustering of the open squares and the striped circles in the middle of Fig. 4 that changing stress to the second syllable or increasing rate had approximately the same effect on both phase angle and jaw cycle duration.

Figure 5 shows that, for subject SN (the first author), mean jaw cycle duration varied by roughly 300 ms, and phase angle varied over a range of 50 degrees. From the scatter of the points within the two stressed conditions (striped symbols), it is apparent that position of syllable break and consonant identity had stronger effects for this speaker than for speaker AC, at least in the stressed conditions (numbers 1–4 decrease in both jaw cycle duration and phase angle). This finding is also reflected in the larger *F* ratios for position of syllable break and for consonant identity in the analysis of phase angle for this speaker than for speaker AC.

While the range of jaw cycle duration for speaker CS (Fig. 6) was the same as that for speaker SN, the range of phase angles was reduced (30 deg as opposed to 50 deg). From the way the striped and the open symbols cluster for CS, it appears that stress had a much greater effect than rate on both jaw cycle duration and phase angle, a finding that is reflected in the larger phase angle *F* ratio for stress than for rate. Figure 6 also shows that syllable structure and consonant identity did not affect jaw cycle duration, and the direction of these effects for phase angle did not always match the predictions, as indicated by the lack of statistical significance for these effects.

For 15 of the 16 conditions, the range of phase angles for subject ES (Fig. 7) is approximately 40 deg, which is similar to the other speakers. However, for one fast/unstressed utterance (/pa#pap/), the angle on the jaw phase plane at which the upper lip began its downward trajectory was much earlier than would be expected (115 deg). This one, large difference may account to a great extent for the significant interactions found for this speaker, both because this is not the utterance that we would have predicted to show the earliest onset and because of the sheer magnitude of this effect.

C. Continuous scaling or main effects

While the preceding analysis demonstrates a relation between jaw cycle duration and phase angle, it does not indicate whether this relation represented a continuous or a discrete effect. To help answer this question, we looked at the patterns of Pearson product-moment correlation coefficients. If the relation between these two variables represented a continuous scaling, then we would expect high correlations within conditions as well as across conditions. If, on the other hand, the relation between these two variables represented discrete changes among conditions (that is, differences in main effects only), then we would expect low within-conditions correlations even when across-conditions correlations are high. We first computed the within-conditions correlations. Only 11 of the 64 correlations were significant, and seven of these were for speaker ES. Consequently, it was concluded that the relation between jaw cycle duration and upper lip phase angle was not continuous. We next tested for a discrete effect by computing correlation coefficients between these two variables using the condition means for each speaker. This reduced the degrees of freedom from 319 to 15. All correlation coefficients computed in this way were highly significant. Thus the relation between jaw cycle duration and upper lip phase angle can best be described as one of main effects only, i.e., utterance type.

As one more way of investigating this relation, we computed correlation ratios (η^2) between phase angle and utterance type and jaw cycle duration and utterance type. Unlike the Pearson product-moment correlation coefficient, the correlation ratio η^2 may be used when the relation between two variables is not necessarily linear, as is the case when one variable is nominal and the other is continuous (Ferguson, 1981). The values of η^2 are given by the ratio of a between-groups sum of squares to a total sum of squares and thus directly estimates the proportion of the variance in one variable (here, phase angle or jaw cycle duration) directly attributable to the other variable (here, utterance type). A separate analysis was done for each subject and the results are shown in Table III. For the sake of comparison, the squared values of Pearson product-moment correlation coefficients (r^2) between phase angle and jaw cycle duration, computed using all individual tokens, are also given for each subject. These values indicate the proportion of variance in phase angle associated with jaw cycle duration *per se*. As can be seen, a large proportion of the variation in jaw cycle duration can be explained by utterance type. Although not as much variation in phase angle is explained by utterance type, in all cases, utterance type explains more of the variation

TABLE III. The η^2 values for each speaker indicating the proportion of variance in jaw vowel cycle (jvc) duration and phase angle (pha) associated with utterance type (utt). Also shown are r^2 values indicating the proportion of variance in phase angle associated with jaw cycle duration.

	η^2(jvc)	η^2(pha)	r^2(pha×jvc)
AC	0.826	0.544	0.429
SN	0.917	0.410	0.361
CS	0.939	0.239	0.209
ES	0.836	0.765	0.462

than jaw cycle duration. Thus jaw cycle duration and upper lip phase angle are strongly associated with the linguistic and nonlinguistic structure of the utterance being produced by the speaker.

D. Comparison of present results with earlier results

As mentioned previously, this study was designed to extend the earlier work of Kelso *et al.* (1986), and the utterances from the two studies form a partially overlapping set: One utterance, /pa#pap/, from this investigation was almost identical to the /ba#pab/ of the earlier work. Table IV shows the comparable phase angle data from the present experiment and from Kelso *et al.* An inspection of the table shows that phase angles vary far more substantially in the present study, leading to the conclusion that there were no significant differences in phase angle over rate and stress conditions in the previous work but were significant differences here. There is no firm explanation for the discrepancy obvious to us. The number of observations in this experiment was a little larger than in the previous one (20 vs 10 or 12 observations for each condition), but were the standard errors tended to be smaller here, but this does not explain the large difference in range of phase angles.[3] One possibility we considered was that speakers in the present experiment may have used a broader range of syllable durations overall, with a resulting expansion of phase angles. When we examined the subset of the earlier data presently available, it was found that the subjects of that study used speaking rates that ranged between about 150–440 ms per syllable. The range for subjects in the present experiment was from about 140–370 ms per syllable. Thus the range of syllable durations is a little smaller and, consequently, we would expect a compressed, rather than an expanded range of phase angles. The slight difference in phonetic frame (/p/ vs /b/) should be irrelevant to the difference in outcome, as far as we know. We note that Lubker (1986) reports a finding of great variability in lip–jaw phase relations in a series of words, but his data are presented only as anecdotal comments on the Kelso *et al.* paper. It is possible that differences in randomization of stimuli within lists could have influenced the difference in outcome: Both stress and speaking rate were blocked across lists in Kelso *et al.*, whereas stress was randomized within a list for this study. However, this difference should only have increased within-condition variability for this work, making it less likely that a statistically significant difference among conditions would be found.

We also looked at Pearson product-moment correlation coefficients between upper lip latency and jaw period, as Tuller and Kelso (1984) had done. Following the suggestion of Munhall (1985) concerning this procedure, we computed correlation coefficients within each of the conditions for each subject and compared these values to those that would be expected based on the part–whole relationship of upper lip latency to jaw period. The obtained coefficients ranged from +0.28– +0.99, with a mean of +0.80 (s.d. = +0.20). None of the 64 obtained values exceeded the expected part–whole correlations, and hence, the results were statistically insignificant (Benoit, 1986; Sock and Jah, 1986). This finding was not unexpected in light of the analysis of similar data by Munhall. Given that we believe, as do others, that there are many pitfalls inherent in a part–whole correlational analysis, we did not pursue the matter further.

III. DISCUSSION

In contrast to the findings of Kelso *et al.* (1986), we found no support for the notion that the relative phasing of jaw vowel gestures and upper lip consonant gestures are stable across manipulations in linguistic and nonlinguistic factors. In fact, the evidence from the present experiment suggests that the intersegmental organization of gestures is a function of the utterance being produced. In other words, the phase relations between articulatory gestures used in the production of adjacent segments varies systematically based on linguistic and nonlinguistic structure, which includes speaking rate, stress pattern, syllable structure, and consonant identity.

This suggestion is not meant to refute the idea that phonological structure within a word may be instantiated by interarticulator phase relations. Instead, what the present data suggest is that the spatiotemporal relations among ar-

TABLE IV. Mean upper lip phase angle (pha) and standard error (SE) relative to vowel-to-vowel jaw trajectory for /Ca#paC/ utterances for speakers from the present experiment and from Kelso *et al.* (1986) (SS = slow/stressed, SU = slow/unstressed, FS = fast/stressed, and FU = fast/unstressed).

	pha	SE	pha	SE	pha	SE	pha	SE
Present experiment								
Speaker	AC		SN		CS		ES	
SS	177	1.48	213	2.96	195	0.86	187	1.07
SU	171	1.56	198	2.48	175	3.47	164	5.26
FS	172	1.36	202	3.01	198	4.56	185	0.90
FU	153	2.85	194	2.18	177	3.01	115	4.81
Kelso *et al.* (1986)								
Speaker	JE		NM		BT		CH	
SS	184	6.41	178	2.35	163	2.93	188	6.45
SU	183	2.07	175	3.21	174	3.50	184	3.01
FS	177	3.82	176	1.73	166	2.68	183	3.94
FU	179	3.81	172	2.50	167	3.96	177	4.08

ticulatory gestures vary systematically with the structure of the utterance at all levels, not just the phonological level. It seems that articulatory gestures can overlap to varying degrees and do so as a function of the phonological composition and linguistic organization of the utterance being produced. In fact, one of the most fundamental aspects of language, that it is composed of segmentable units that can be combined in a great variety of ways while still respecting the principles of linguistic organization (MacNeilage *et al.*, 1985), probably depends on the ability of speakers to overlap articulatory gestures. A related notion must be that listeners perceive the articulatory gestures, as well as the patterns of overlap among these gestures, and use both kinds of information to recover linguistic structure at all levels of organization concurrently.

While these accounts of the present findings are reasonable and in accordance with current theories of gestural organization (e.g., Browman and Goldstein, 1986), there is still a great deal of work to be done before we have a complete understanding of intergestural organization. For example, we looked only at the relative phase of articulatory movements associated with adjacent segments. Here, the jaw trajectory represented a vocalic gesture, while the onset of lip movement was a consonantal gesture. It may be the case that the gestures associated with adjacent segments can overlap by varying amounts, but the relative timing of articulatory movements associated with any one segment is stable. Thus, for example, we may have found that the relative phasing of upper lip lowering and lower lip raising were more stable across conditions.[4]

In spite of this need for further study, we feel that the present data provide valuable information concerning interarticulator phase relations. These data suggest that the spatiotemporal organization of intersegmental gestures is not stable across linguistic and nonlinguistic manipulations, but instead varies systematically and discretely as a function of the exact utterance being produced. Thus linguistic structure at all levels of organization may be inextricably related in its representation in these spatiotemporal patterns.

ACKNOWLEDGMENTS

This work was supported by NIH Grants NS-13617 and NS-07237 to Haskins Laboratories. We are grateful to Steven Barlow, Catherine Browman, Carol Fowler, Ron Netsell, Elliot Saltzman, and three anonymous reviewers for their comments on an earlier draft of this article, and to Eric Vatikiotis-Bateson for his reanalysis of a subset of the data from the present study and that of Kelso *et al.* (1986).

---

[1] Fowler (1986), as well as Browman and Goldstein (1986), deals mainly with the phonological level of linguistic structure.

[2] Period and latency intervals were defined in this way because work by Tuller *et al.* (1982) demonstrated a stronger relation between these articulatory intervals than between any other possible pair.

[3] One of our colleagues at Haskins Laboratories reanalyzed the data from the raw waveforms of the utterance /ba#pab/ for one subject of Kelso *et al.* (1985) (CH) and for 12 tokens of the utterance /pa#pap/ for one of our subjects (AC). In both cases, the results demonstrated statistically significant effects of rate and stress on phase angle, but the magnitude of these effects was slightly reduced for AC when only 12 tokens were analyzed instead of the 20 used in the present analysis. Furthermore, standard errors of the mean were similar in the two reanalyses (a mean standard error of 1.77 for CH and of 1.71 for AC). These values are similar to those found for AC in the present study but reduced from what was reported for CH by Kelso *et al.*

[4] While the results of Löfqvist and Yoshioka (1984) suggest this possibility, their results are inconclusive because of their use of part–whole correlational analysis similar to that of Tuller and Kelso (1984).

Barry, W. J. (**1983**). "Some problems of interarticulator phasing as an index of temporal regularity in speech," J. Exp. Psychol.: Human Percept. Perform. **9**, 826–828.

Benoit, C. (**1986**). "Note on the use of correlations in speech timing," J. Acoust. Soc. Am. **80**, 1846–1849.

Browman, C. P., and Goldstein, L. M. (**1986**). "Towards an articulatory phonology," Phonol. Yearbook **3**, 219–252.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (**1952**). "Some experiments on the perception of synthetic speech sounds," J. Acoust. Soc. Am. **24**, 597–606.

Daniloff, R. G., and Hammarberg, R. E. (**1973**). "On defining coarticulation," J. Phon. **1**, 239–248.

Ferguson, G. A. (**1981**). *Statistical Analysis in Psychology and Education* (McGraw-Hill, New York).

Fowler, C. A. (**1986**). "An event approach to the study of speech perception from a direct-realist perspective," J. Phon. **14**, 3–28.

Fowler, C. A., Rubin, P., Remez, R., and Turvey, M. T. (**1980**). "Implications for speech production of a general theory of action," in *Language Production*, edited by B. Butterworth (Academic, New York), Vol. 1, pp. 373–420.

Gentner, D. R. (**1987**). "Timing of skilled motor performance: Tests of the proportional duration model," Psychol. Rev. **94**, 255–276.

Gracco, V. L., and Abbs, J. H. (**1986**). "Variant and invariant characteristics of speech movements," Exp. Brain Res. **65**, 156–166.

Hammarberg, R. (**1976**). "The metaphysics of coarticulation," J. Phon. **4**, 353–363.

Kelso, J. A. S., and Tuller, B. (**1987**). "Intrinsic time in speech production: Theory, methodology, and preliminary observations," in *Motor and Sensory Processes of Language*, edited by E. Keller and M. Gopnik (Erlbaum, Hillsdale, NJ), pp. 203–222.

Kelso, J. A. S., Saltzman, E. L., and Tuller, B. (**1986**). "The dynamical perspective on speech production: Data and theory," J. Phon. **14**, 29–59.

Kent, R. D., and Netsell, R. (**1971**). "Effects of stress contrasts on certain articulatory parameters," Phonetica **24**, 23–44.

Liberman, A. M., Cooper, F. S., Shankweiler, D. S., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.

Löfqvist, A., and Yoshioka, H. (**1984**). "Intrasegmental timing: Laryngeal–oral coordination in voiceless consonant production," Speech Commun. **3**, 279–289.

Lubker, J. (**1986**). "Articulatory timing and the concept of phase," J. Phon. **14**, 133–137.

MacNeilage, P. F. (**1963**). "Electromyographic and acoustic study of certain final clusters," J. Acoust. Soc. Am. **35**, 461–463.

MacNeilage, P. F., and Ladefoged, P. (**1976**). "The production of speech and language," in *Handbook of Perception: Language and Speech*, edited by E. Carterette and M. Friedman (Academic, New York), Vol. 7, pp. 75–120.

MacNeilage, P. F., Studdert-Kennedy, M., and Lindblom, M. (**1985**). "Planning and production of speech: An overview," Asha Rep. **15**, 15–21.

Munhall, K. G. (**1985**). "An examination of intra-articulator relative timing," J. Acoust. Soc. Am. **78**, 1548–1553.

Saltzman, E. L., and Kelso, J. A. S. (**1987**). "Skilled actions: A task dynamic approach," Psychol. Rev. **94**, 84–106.

Sock, R., and Jah, O. (**1986**). "Old wine in new bottles, new wine in old bottles: Action theory and (para)metric phonology views in the domain of wolof," Bull. Inst. Phon. Grenoble **15**, 117–133.

Studdert-Kennedy, M. (**1976**). "Speech perception," in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass (Academic, New York), pp. 243–293.

Sussman, H. M., MacNeilage, P. F., and Hanson, R. J. (**1973**). "Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations," J. Speech Hear. Res. **16**, 397–420.

Tuller, B., Kelso, J. A. S., and Harris, K. S. (**1982**). "Interarticulator phasing as an index of temporal regularity in speech," J. Exp. Psychol: Human Percept. Perform. **8**, 460–472.

Tuller, B., and Kelso, J. A. S. (**1984**). "The timing of articulatory gestures: Evidence for relational invariants," J. Acoust. Soc. Am. **76**, 1030–1036.

Wickelgren, W. A. (**1976**). "Phonetic coding and serial order," in *Handbook of Perception: Language and Speech*, edited by E. Carterette and M. Friedman (Academic, New York), Vol. 7, pp. 227–264.