

A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling

J. A. S. Kelso

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511 and Departments of Psychology and Biobehavioral Sciences, The University of Connecticut, Storrs, Connecticut 06268

Eric Vatikiotis-Bateson

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511 and Department of Linguistics, Indiana University, Bloomington, Indiana 47405

Elliot L. Saltzman

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Bruce Kay

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511 and Department of Psychology, The University of Connecticut, Storrs, Connecticut 06268

(Received 4 May 1984; accepted for publication 7 September 1984)

The departure point of the present paper is our effort to characterize and understand the spatiotemporal structure of articulatory patterns in speech. To do so, we removed segmental variation as much as possible while retaining the spoken act's stress and prosodic structure. Subjects produced two sentences from the "rainbow passage" using reiterant speech in which normal syllables were replaced by /ba/ or /ma/. This task was performed at two self-selected rates, conversational and fast. Infrared LEDs were placed on the jaw and lips and monitored using a modified SELSPOT optical tracking system. As expected, when pauses marking major syntactic boundaries were removed, a high degree of rhythmicity within rate was observed, characterized by well-defined periodicities and small coefficients of variation. When articulatory gestures were examined geometrically on the phase plane, the trajectories revealed a scaling relation between a gesture's peak velocity and displacement. Further quantitative analysis of articulator movement as a function of stress and speaking rate was indicative of a language-modulated dynamical system with linear stiffness and equilibrium (or rest) position as key control parameters. Preliminary modeling was consonant with this dynamical perspective which, importantly, does not require that time *per se* be a controlled variable.

PACS numbers: 43.70.Aj, 43.70.Bk

INTRODUCTION

It has often been supposed that temporal organization in biological systems is ultimately governed by neural rhythm generators, biological clocks, metronomes, etc. Physiologists and psychologists, confronted with order in the time domain, have not hesitated to posit clocks whose "ticks" define when muscles will activate (e.g., Kozhevnikov and Chistovich, 1965; Rosenbaum and Patashnik, 1980). Our approach, however, has been directed towards identifying and understanding spatiotemporal pattern in articulatory events as a dynamic property of natural systems rather than as the result of the operation of some special neural or mental time-keeping device (cf. Kelso *et al.*, 1981). Once elaborated, we believe this dynamical perspective may afford a principled account of the ubiquity of temporal constraints in movement in general and in speech in particular. For example, the internal phasing relations among muscles and kinematic components in rhythmic activities such as locomotion, scratching, respiration, and mastication are preserved across scalar changes in force and rate (cf. Kelso, 1981; Grillner, 1982, for reviews). Similarly, in electromyographic and kinematic work on speech (Tuller *et al.*, 1982c, 1983; Tuller and Kelso, 1984), timing of consonant production relative to vowel production was found to be invariant over

substantial changes (induced by stress and rate) in the duration of the vocalic cycle. These data—along with other evidence (reviewed by Fowler, 1983)—suggest a vowel-to-vowel organization that places constraints on speech timing.

Although speech certainly involves many of the same body parts as chewing, its rhythmic basis is not clear, in spite of the fact that linguists and others have long claimed speech to be rhythmic, and people perceive it to be so (e.g., Lehiste, 1972; Lenneberg, 1967; Lisker, 1975; Pike, 1945). Yet experimenters have had enormous difficulty identifying rhythmicity in either the articulatory or the acoustic domain. One possible reason—as pointed out by Fowler (1983) with respect to acoustic studies—is that experimental measurements typically used may be inappropriate for capturing the natural, temporal structure of spoken sequences. Speaking is an inherently multidimensional process; during speech different articulators are involved to different degrees and the spatiotemporal overlap among movements is considerable. Confronted with so many simultaneous or nearly simultaneous events, there seems little chance of our identifying any basic temporal regularity, even though our perceptual impressions lead us to suppose that one exists.

Our approach in the present work was to strip away, as much as possible, the influence of segmental variation on

articulatory movement, by asking subjects to speak "reiterantly." That is, speakers substituted the syllable /ba/ or /ma/ for each real syllable in the utterance, while mimicking the utterance's normal prosodic structure. The benefit of the reiterant technique is that, by minimizing segmental variability while preserving the prosodic pattern (Liberman and Streeter, 1978; Nakatani, 1977), we are able to measure the movements of articulators (in this case the lips and jaw) that are consistently involved in the production of /ba/ and /ma/. In principle, this procedure affords an analysis of articulator patterns in a simple and accessible form.

We recognize that the relationship between real speech and reiterant speech is not always transparent. We should stress, however, that the main thrust of the present work is to use reiterant speech as a *tool* to examine articulator motions in a speechlike task. We do not claim any necessary generalization to real speech although one might exist (see also Larky, 1983). For instance, Liberman and Streeter (1978) show the pattern of acoustic syllable durations to be similar between real and skilled reiterant speech although the absolute durational values are very different. In terms of production, it seems unlikely to us that the control of the lip-jaw system for the production of a reiterant /ba/ is fundamentally different when the same syllable is produced during natural speech. Indeed, we shall quantitatively describe certain kinematic relationships (e.g., between an articulator's peak velocity and displacement) that have been observed in many other nonreiterant speech production studies.

In the present paper, we outline a geometric approach for characterizing the dynamic properties underlying articulatory movements during reiterant speech. We use the phase portrait to facilitate the analysis of relevant articulatory variables when speakers produce these simple sequences of syllables. To our knowledge, phase portrait techniques have rarely been employed in speech production studies, even though their role is to describe the *forms of motion* in complex, multidegree-of-freedom systems (cf. Abraham and Shaw, 1982). Were one to count the neurons, muscles, and joints that cooperate to produce even a simple utterance, literally thousands of such elements would be involved. Yet normal speech is usually coherent and organized: A low dimensional pattern emerges from a system of high dimensionality that can be controlled with relatively few dynamic parameters.¹ Thus our approach is one in which we attempt to characterize regularities of articulator pattern in terms of a relatively abstract *functional* organization (cf. Kelso and Tuller, 1984a). We do not attempt to model peripheral biomechanics or neurophysiological mechanisms. Rather we use the phase portrait as a way of uncovering qualitatively the system's control structure and as a preface to a quantitative treatment of articulatory trajectories. In doing so we observe both invariant and systematically varying features of motion when stress and speaking rate are changed. Perhaps most important, our results, analyzed geometrically and interpreted from a dynamic perspective, do not require the assumption that time itself is a controlled variable. Instead, the form of articulator trajectories over time is seen as a consequence of a control structure whose dynamic parameters are functionally equivalent to those of a mechanical

mass-spring system, namely: equilibrium (or rest) position, which is the position at which the net force on the mass is zero; and linear stiffness, which is the reactive force per unit displacement.

I. METHODS AND PROCEDURES

Two adult speakers [one male (SK, the first author and a native speaker of an Ulster dialect of English), and one female (DW, a speaker of a New Jersey dialect of American English)] recited the first and last sentences of the "rainbow passage": (1) "When the sunlight strikes raindrops in the air, they act like a prism and form a rainbow," and (2) "There is, according to legend, a boiling pot of gold at one end." After reciting each sentence, speakers mimicked the prosodic pattern 2-4 times, substituting only /ba/ or only /ma/ for each syllable. So, for example, "When the sunlight strikes raindrops in the air" would be mimicked as "*ba ba ba ba ba ba ba ba ba*" (where italics indicates a hypothetical stress pattern for the syllables). Upon completion of the task at a normal, conversational rate, it was then repeated at a faster rate. One of the speakers (SK) repeated this procedure at a later date. In all, 392 syllables at each rate were analyzed. We also obtained measures of each speaker's preferred frequency of jaw movement over an extended period of time, by asking the subject to "wag" the jaw at a comfortable amplitude and frequency "as if you were going to do it all day." "Wagging" movements were then sampled over a 30-s interval.

For speech and nonspeech tasks, vertical displacements of the lips and jaw were tracked using a device similar in principle to the commercially available SELSPOT system, which employs infrared LEDs that can be placed midsagittally on the nose, lips, and point of the chin. Modulated light from the diodes is captured by a camera equipped with a Schottky planar diode located in its focal plane. The output of the photodiode is fed to associated electronics that decode the signals and compute pairs of x and y coordinates. Up to eight channels of coordinate potentials may be generated simultaneously, each with a bandwidth of 0-500 Hz. These potentials are then fed to first-stage dc offset preamplifiers which center the signals about the 0-dc level. Following the offset adjustment, the coordinate values are transmitted via dc coupled amplifiers, checked by means of a monitoring oscilloscope, and recorded. Once the subject was seated with the LEDs in place, calibration was achieved by raising the camera a known distance (2 cm) and recording the output of the lower lip LED. Simultaneous acoustic recordings were also made. The movement data were recorded on FM tape and sampled at 200 Hz in later computer analysis. This included numerical smoothing (using a 25-ms triangular window) and differentiation (using a two-point central difference algorithm; James *et al.*, 1977) for obtaining the derivatives of motion (velocity, acceleration).

Figure 1 shows an example of the position and velocity of the lower lip and jaw (i.e., the LEDs attached to lower lip and jaw) for the first part of sentence 1, "When the sunlight strikes raindrops in the air," where /ba/ is the reiterated syllable. In the movement traces, peaks and valleys denote the high and low vertical positions achieved by the indicated

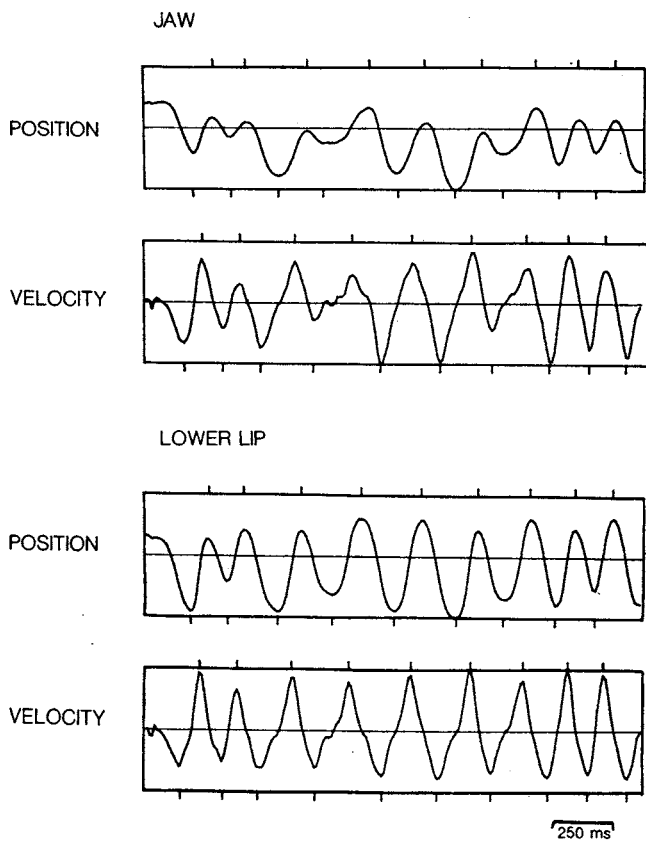


FIG. 1. Position and velocity over time of lower lip and jaw LEDs for the reiterant production of "When the sunlight strikes raindrops in the air." /ba/ is the reiterant syllable.

articulators. Thus peaks occur during lip closure for the bilabial stop and valleys occur during production of the low vowel /a/. In the velocity traces, peaks and valleys are the maximum velocities attained going into and out of a closure, respectively. The peaks and valleys were determined by a computer program which also calculated means (M) and standard deviations (s.d.) for peak-to-peak cycle duration and displacement of opening (peak-to-valley) and closing (valley-to-peak) gestures.

II. RESULTS AND DISCUSSION

Each of the following sections is designed to be self-contained in that a discussion accompanies each set of empirical findings. First we present data pertaining to the global temporal regularity of articulator movement that was observed in the experiments. Second, a qualitative dynamic analysis of articulatory motion is presented using the phase portrait to describe the forms of motion that are produced. Following is a quantitative kinematic analysis of motion and its derivatives that details effects of the local changes induced by stress and speaking rate transformations. We try to maintain continuity of presentation in this quantitative section by proceeding from lower-order to higher-order kinematic relations. Finally we present some of our preliminary efforts to model the present articulatory findings using an approach based in dynamical systems theory and supported by recent results in the field of physiological motor control.

A. Global temporal regularity

First we show separately for the two rates and two reiterant syllables the mean duration between successive peaks and the associated standard deviations. The values shown in Table I are averaged across subjects and sentences for both jaw and lower lip motions (i.e., motions of the jaw and lower lip LEDs). In order to study articulatory motions *per se*, we have removed intervals that span major syntactic breaks and the first and last syllables of the sentence, i.e., where startup, pauses, and lengthening effects predominate.

The durational data show quite low variability regardless of rate, with coefficients of variation in the 10%–20% range. The two speakers are also very similar in their durational behavior as revealed in the small between-subject standard deviation of the means. Mean cycle durations for the three experimental sessions were 211 ms (approximately 5 Hz) for the normal rate and 167 ms (approximately 6 Hz) for the fast rate. In this case, the jaw exhibits a periodicity similar to that of the lower lip. Not surprisingly, the data contrast with those of Ohala's (1975) earlier study in which 10 000 consecutive jaw opening gestures were obtained during a 1.5-h reading period. Ohala (1975) found large durational variance (presumably because of the presence of pauses and segmental factors) accompanied by a dominant, but weakly defined, periodicity of about 250 ms (4 Hz). Ohala and others (e.g., Lindblom, 1983) have suggested that this periodicity may correspond to the "preferred frequency of the mandible." However, the preferred wagging frequencies of jaw movement for our two speakers (0.81 and 2.04 Hz, s.d.s = 0.06 and 0.21 Hz, respectively) are much slower than the frequencies found either by us for reiterant speech or by Ohala for read speech. It is clear then, that neither the sharply defined periodicity observed by us in reiterant speech nor the weakly defined cycling found by Ohala in read speech is the same as the preferred frequency of the mandible in our nonspeech task (see also Nelson *et al.*, 1984, for differences between preferred frequencies of mandible movement in speechlike and nonspeech tasks). We also found that the pe-

TABLE I. Means and standard deviations of peak-to-peak duration in ms and frequency (f) in Hz for jaw (J) and lower lip (LL) during reiterant speech at two rates. Between-subject standard deviations are in parentheses.

	/ba/		/ma/		Overall	
	J	LL	J	LL	J	LL
Normal						
m	213 (5)	212 (8)	212 (3)	211 (3)	212 (4)	211 (6)
s.d.	42 (6)	40 (6)	41 (1)	37 (1)	42 (4)	39 (5)
f	4.70 (0.11)	4.72 (0.19)	4.72 (0.06)	4.73 (0.06)	4.72 (0.08)	4.73 (0.14)
Fast						
m	168 (5)	168 (5)	166 (3)	165 (4)	167 (4)	167 (4)
s.d.	33 (9)	29 (8)	29 (3)	30 (5)	31 (7)	30 (6)
f	5.95 (0.17)	5.95 (0.18)	6.03 (0.11)	6.06 (0.15)	5.98 (0.14)	6.00 (0.16)
n	512	512	272	272	784	784

riodicity was unaffected by the syllable that was used to mimic real speech. The largest mean durational difference regardless of rate condition between /ba/ and /ma/ for any articulator was 3 ms (see Table I). In short, when segmental variation is minimized, it is possible to identify relatively stable articulatory periodicity. The periodicity is not *perfectly* isochronous because there are systematic variations concomitant with stress and rate (see Sec. II C).

B. A geometric (qualitative dynamic) analysis²

In the following geometric analysis, phase plane trajectories are generated by continuously plotting the relationship between, in this case, articulator position x and its derivative, velocity \dot{x} . As an example, consider the idealized case shown in Fig. 2. The upper trace is a computer generated sine wave of 5 Hz with a peak-to-valley displacement defined to be 20 mm. The peak position corresponds to the consonant closure, and the valley position to the maximum opening for the vowel. Points of maximum downward (opening) and upward (closing) velocity fall at the midpoints of the position trace. To create a phase plane trajectory shown on the lower part of Fig. 2, we plot successive position points and their corresponding velocities as coordinates on a plane whose vertical axis denotes position and whose horizontal axis denotes velocity.³ The arrowheads on the circle denote the direction of motion on the plane. Thus one cycle or orbit corresponds to the interval between successive closures, with the opening gesture on the left half and the closing gesture on the right. *Note that time itself is not an explicit variable in this description.*

Figure 3 shows phase plane trajectories for the jaw and lower lip LEDs of "When the sunlight strikes raindrops in the air," using reiterant /ba/ spoken at a normal rate. Qualitatively, the shapes of the trajectories are quite similar across the ten syllables plotted. There is a strong tendency, for example, for displacement and peak velocity to covary directly (see Sec. II C). Normal and fast reiterant produc-

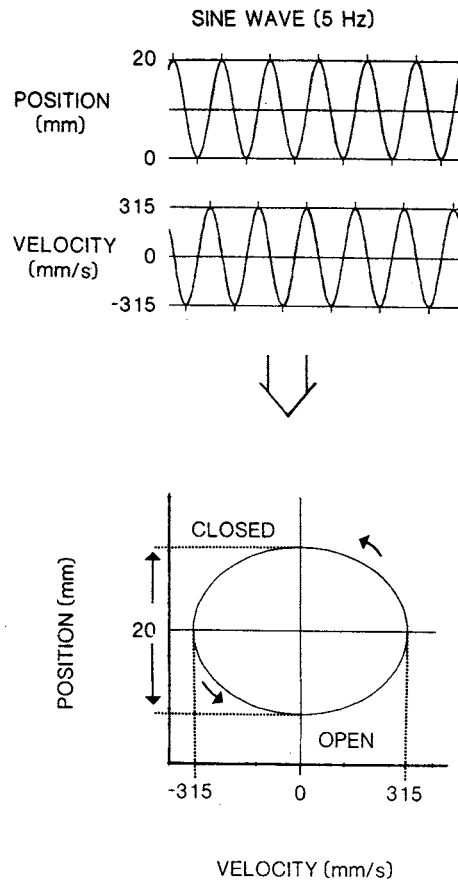


FIG. 2. Top: idealized position and velocity over time of articulator movement. Bottom: corresponding phase plane trajectories. Abscissa is velocity, ordinate is position (see text for details).

tions for subjects SK and DW of the second part of the first sentence, "they act like a prism and form a rainbow," are shown in Figs. 4 and 5. The mutual relationship between the kinematic variables of position and velocity is accentuated by the rate manipulation, particularly for subject SK. Once

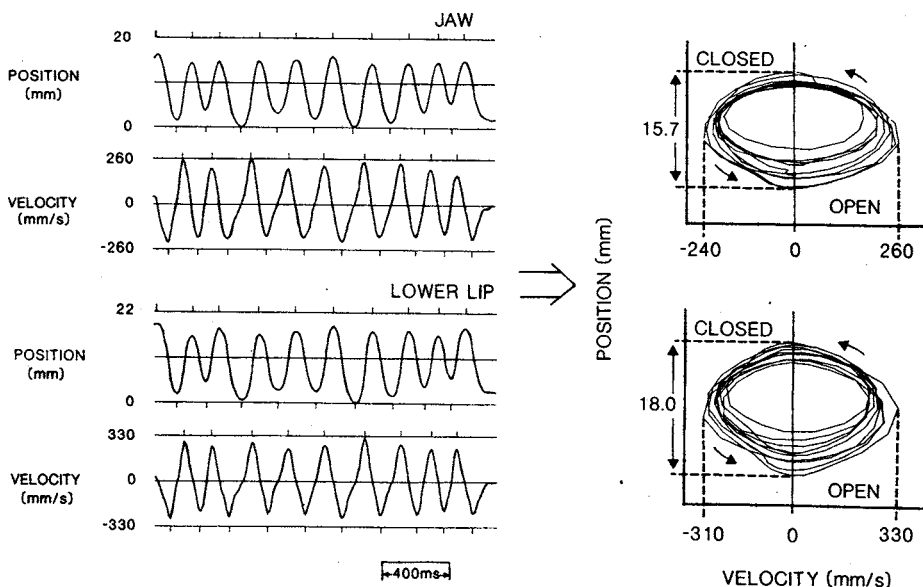


FIG. 3. Left: position and velocity over time of jaw and lower lip LEDs for sentence produced with reiterant /ba/ at a normal rate. Right: corresponding phase plane trajectories.

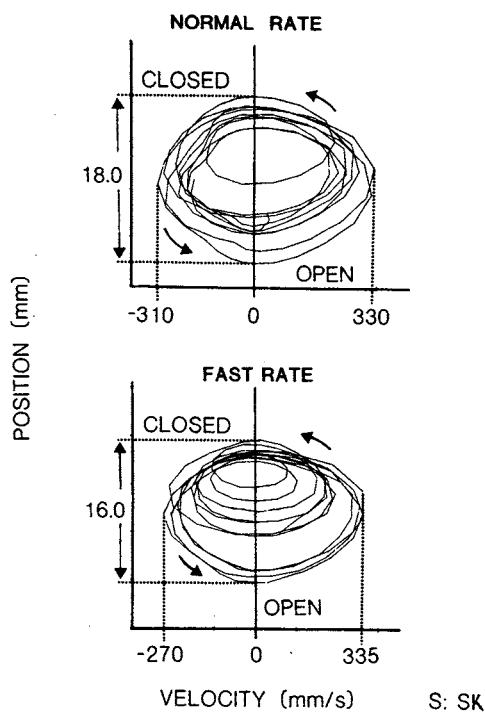


FIG. 4. Phase plane trajectories of lower lip motions for the second part of sentence 1, "They act like a prism and form a rainbow" produced at normal and fast speaking rates with /ba/ as the reiterant syllable; subject is SK.

again, even when there is a clear distinction between the trajectories corresponding to stressed and unstressed syllables, their orbital shapes are generally similar. The unstressed (sometimes reduced) syllables are characterized by

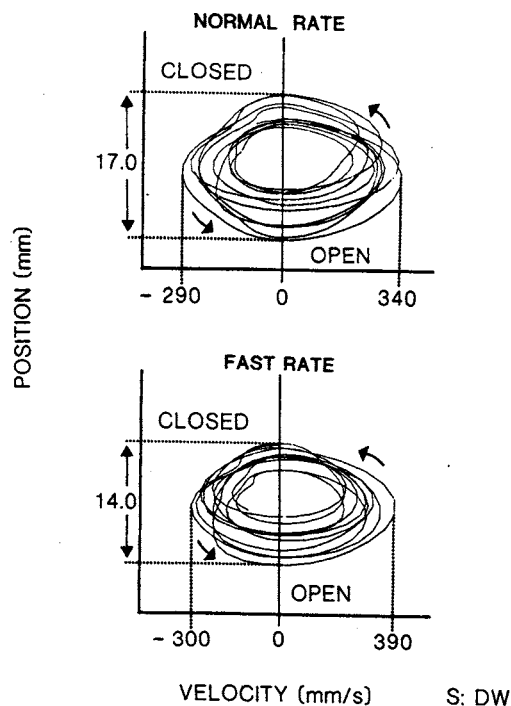


FIG. 5. Phase plane trajectories of lower lip motions for the second part of sentence 1, "They act like a prism and form a rainbow" produced at normal and fast speaking rates with /ba/ as the reiterant syllable; subject is DW.

smaller displacements and peak velocities than the stressed syllables, thus maintaining a global similarity of (elliptical) trajectory shape across unstressed and stressed gestures. Also observed, however, are subtle differences between trajectory shapes associated with different gestural displacements. For example, the orbits appear to be slightly more compressed horizontally for larger displacement gestures relative to shorter displacement gestures. In Sec. II C, we will quantify both the global similarities and subtle differences among gestural trajectory shapes.

C. Quantitative kinematic analysis

In Sec. II C we proceed to quantify specific effects of speaking rate and stress on articulatory movements in an effort to answer the following questions. First, what kinematic variables or relations among variables might inform us about the *control* of speech gestures? Second, what kind of regularity, if any, exists in the motions of speech articulators across changes in stress and speaking rate, and how might such regularity be rationalized? Although we appreciate that there are many idiosyncratic differences among speakers, dialects, and languages, our emphasis here is on identifying what is common across such diversity. In short, can we begin to define a "deep structure" for speech motor control that can be recognized in the face of much surface variability, and, if so, on what principle(s) is it based?

We begin with an analysis of the space-time characteristics of articulator movement and its derivatives, with the emphasis now on the gesture (opening and closing) rather than the cycle. Because of the enormous amount of kinematic data involved, we restrict our concerns (unless otherwise indicated) to (a) the motions of the jaw and lower lip complex for the syllable /ba/ during reiterant speech, and (b) the single experimental session for each speaker, i.e., omitting the repeated session. This amounts to 232 gestures for speaker DW (116 opening and 116 closing) and 464 gestures for speaker SK (232 opening and 232 closing).

The general statistical analysis of the kinematic variables takes the form of a gesture (opening, closing) \times stress (stressed, unstressed) \times rate (normal, fast) analysis of variance for each dependent variable, followed by correlational analysis between variables (e.g., displacement versus time) where appropriate. In order to facilitate communication of the results we report the degrees of freedom for the statistical main effects and interactions only once. For subject DW the numerator and denominator degrees of freedom are 1 and 224; for subject SK they are 1 and 456.

1. Displacement, movement time, and their relation

Tables II and III provide the mean displacement and mean movement times of the opening and closing gestures for the syllable /ba/, as a function of speaking rate and stress. The mean data order systematically for both kinematic variables in both subjects, although the magnitude of change across rate and stress is idiosyncratic. Similar results have been reported by others (e.g., Kuehn and Moll, 1976; Tuller *et al.*, 1982b). For displacement, since the lips always return to closure, the main effect of gesture type (opening versus closing) was not significant in either subject's data;

TABLE II. Kinematic values of displacement, time, and peak velocity across rate and stress variations (opening gestures, /ba/).

		Stressed			Unstressed		
		<i>d</i> ^a	<i>t</i>	<i>V_p</i>	<i>d</i>	<i>t</i>	<i>V_p</i>
Normal	DW m	14.58	123.9	229.2	11.80	112.4	204.0
	s.d.	3.68	20.6	74.2	3.30	22.4	65.9
	n	24			34		
	SK m	16.02	140.4	262.5	12.63	114.7	238.6
	s.d.	1.40	19.0	24.9	3.35	30.1	55.8
	n	48			68		
Fast	DW m	13.41	103.3	241.0	10.38	85.3	216.3
	s.d.	2.83	12.7	48.8	3.27	11.3	63.6
	n	24			34		
	SK m	14.85	120.0	241.1	8.27	81.3	170.2
	s.d.	1.46	17.9	32.9	3.85	20.6	64.4
	n	48			68		

^a Displacement (*d*) in mm, time (*t*) in ms, and peak velocity (*V_p*) in mm/s.

*F*s = 0.10 and 0.55, *p*s > 0.05 for DW and SK, respectively. Nor were there two- or three-way interactions with gesture type. Stressed gestures had larger displacements than unstressed gestures; *F*s = 39.19 (DW) and 415.44 (SK), *p*s < 0.0001. Rate had a generally similar effect: Normal rate gestures were produced with larger displacements than fast gestures, *F*s = 11.26 (DW) and 136.18 (SK), *p*s < 0.001. Unlike DW, subject SK revealed a stress × rate interaction on the displacement measure, *F* = 35.44, *p* < 0.0001. A simple main effects analysis of this interaction was entirely consonant with the main effects, however: The difference in displacement as a function of rate was more apparent in unstressed gestures, *F* = 162.92, *p* < 0.0001 than stressed gestures, *F* = 8.70, *p* < 0.004. Similarly, differences in displacement as a function of stress were manifest particularly at a fast speaking rate, *F* = 346.77, *p* < 0.0001, although they were highly significant at the normal rate as well, *F* = 104.11, *p* < 0.0001 (see Tables II and III). No other interactions were significant for either subject.

TABLE III. Kinematic values of displacement, time, and peak velocity across rate and stress transformations (closing gestures, /ba/).

		Stressed			Unstressed		
		<i>d</i> ^a	<i>t</i>	<i>V_p</i>	<i>d</i>	<i>t</i>	<i>V_p</i>
Normal	DW m	13.88	91.8	297.0	12.14	82.5	265.3
	s.d.	2.94	17.8	73.6	2.91	20.1	74.7
	n	24			34		
	SK m	15.99	106.4	321.7	12.09	84.8	272.2
	s.d.	1.26	15.1	19.2	2.77	14.9	52.8
	n	48			68		
Fast	DW m	13.07	66.3	348.4	10.33	64.7	280.5
	s.d.	2.39	14.8	65.6	3.05	11.1	96.5
	n	24			34		
	SK m	14.88	86.5	323.0	8.17	66.5	202.3
	s.d.	1.41	12.1	27.7	3.11	10.6	68.1
	n	48			68		

^a Displacement (*d*) in mm, time (*t*) in ms, and peak velocity (*V_p*) in mm/s.

For movement time, opening gestures as a class took longer than closing gestures. All the movement time values for similar conditions reported in Table II (opening) are greater than those reported in Table III (closing), a finding substantiated by a significant gesture main effect for DW, *F* = 171.43, *p* < 0.0001 and SK, *F* = 240.57, *p* < 0.0001. Stressed gestures take longer than unstressed gestures, *F*s = 20.21 and 223.62, *p*s < 0.0001 for DW and SK, respectively. For subject DW a simple main effects analysis of the significant gesture × stress interaction, *F* = 4.31, *p* < 0.04 revealed that the stress effect was greatest for opening gestures, *F* = 21.58, *p* < 0.001 (compare Tables II and III). For subject SK, the gesture × stress interaction was also significant, *F* = 10.34, *p* < 0.002: The difference in movement time between stressed and unstressed conditions was greater for opening gestures, *F* = 165.08, *p* < 0.0001 than closing gestures, *F* = 68.89, *p* < 0.0001.

Speaking rate had a systematic effect on movement time. Gestures produced at a normal rate took longer than those at a faster rate, *F* = 104.50 (DW) and *F* = 181.84 (SK), *p*s < 0.0001. For subject SK, there was also a gesture × rate interaction, *F* = 6.60, *p* < 0.02. Again, the rate effect between gestures was a matter of degree; movement time differences between rates were more apparent in opening gestures, *F* = 128.86, *p* < 0.001 than closing gestures, *F* = 59.98, *p* < 0.0001, although clearly the effect was highly significant in both gesture types.

In summary, in both subjects, the main effects of stress and rate predominate for both displacement and movement time as dependent measures, although these effects tend to be greater in opening gestures than closing gestures. Generally speaking, stressed gestures display greater articulatory displacement and longer duration than unstressed gestures. Rate has similar effects. Gestures produced at faster speaking rates are accomplished with smaller displacements and in shorter movement times than those at a normal conversational pace.

Viewed from an overall perspective based on the mean data of each subject, we can make a rather simple statement regarding the displacement–time relation independent of movement phase (opening versus closing), rate, or stress. Namely, on the average, *displacement covaries directly with duration*. Smaller (larger) displacements tend to be observed at fast (normal) rates and in unstressed (stressed) environments; duration of motion adjusts in a corresponding fashion.

These overall effects, therefore, suggest a systematic and apparently quite linear relationship between spatial and temporal dependent measures. However, examination of the scatter plots for each subject in Fig. 6 (opening phase) and Fig. 7 (closing phase) reveals a somewhat more complicated picture. For subject SK the data follow the general picture outlined above; amplitude and duration vary in a quite linear way. The overall correlation for opening gestures is *r* = 0.82 and for closing gestures *r* = 0.76 (*p*s < 0.01). Moreover, the displacement–time correlations for individual conditions, shown in Table IV, are, with a single exception, significant (*p*s < 0.05). The picture is rather different for subject DW, however. In the data for DW the individual displacement–

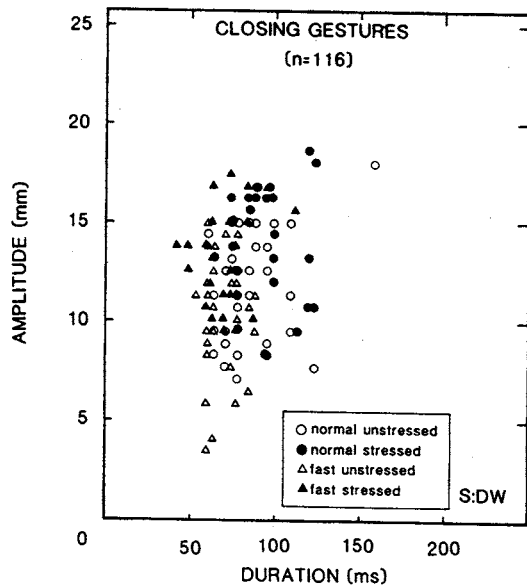
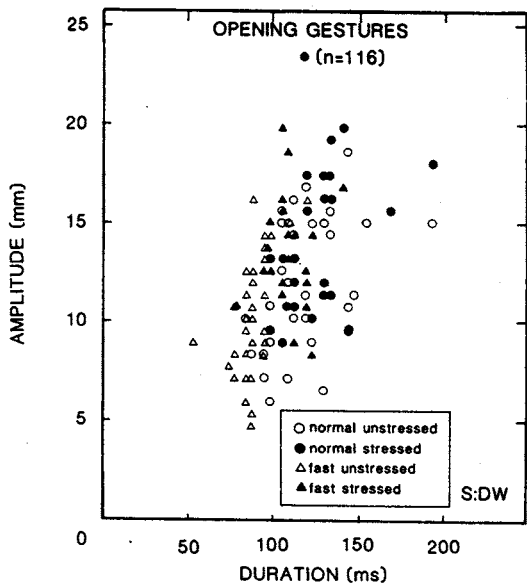
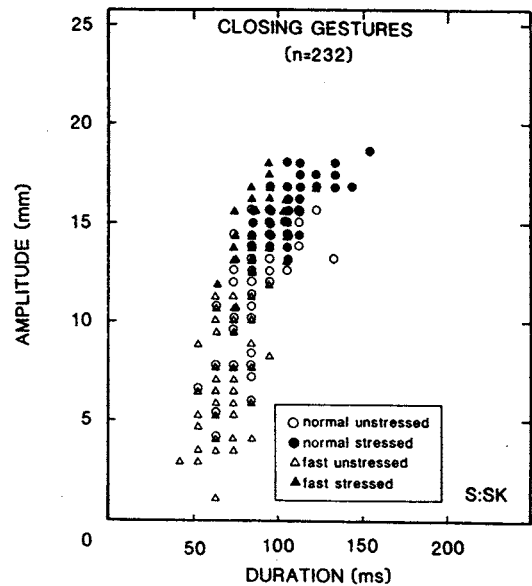
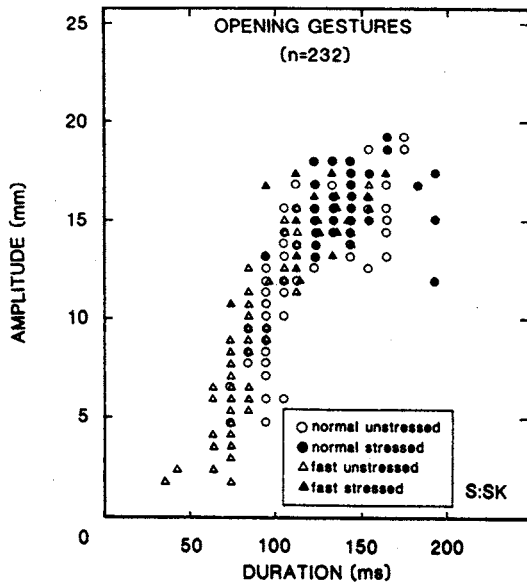


FIG. 6. Scatter plot of amplitude and duration of each subject's lower lip motions for opening gestures associated with the consonant-vowel (CV) portion of the syllable. Points are differentiated by rate and stress, as shown in the legend.

FIG. 7. Scatter plot of amplitude and duration of each subject's lower lip motions for closing gestures associated with the vowel-consonant portion (VC) of the syllable. Points are differentiated by rate and stress, as shown in the legend.

time pairs are widely distributed and in only one out of a possible eight conditions (unstressed opening gestures produced at a normal rate) is there a significant correlation (see Table IV). When opening and closing gestures are analyzed as a group for DW, significant correlations are obtained, $r_s = 0.46$ and 0.26 ($ps < 0.05$), respectively, although the proportion of variance accounted for is small.

To summarize, the coupling between displacement and time is quite different for the two subjects. One subject (SK) reveals a rather orderly relation between these variables across rate, stress, and movement phase (opening versus closing). The other subject (DW) shows a high degree of

overlap among conditions and a much more homogeneous distribution of displacement-time data pairs. Indeed, the proportion of variance accounted for by this relationship is so small as to suggest that, for DW, displacement and time are essentially independent.

How might these apparent discrepancies between subjects in the displacement-time performance space be interpreted? One account that merits mention is that the speech motor system adheres to a minimum cost function such as "least effort," which might give rise to tradeoffs in articulatory displacement and duration. This notion of movement costs is elaborated in some detail in a recent paper by Nelson

TABLE IV. Linear correlations (r) and regression slopes (m) of displacement–time relationship across rate and stress transformations ($/ba/$).

		Opening gestures			
		Stressed		Unstressed	
		m	r	m	r
Normal	DW	0.06	0.33	0.07	0.46 ^a
	SK	0.02	0.24	0.08	0.72 ^a
Fast	DW	0.01	0.05	0.10	0.35
	SK	0.05	0.55 ^a	0.15	0.82 ^a

		Closing gestures			
		Stressed		Unstressed	
		m	r	m	r
Normal	DW	0.02	-0.11	0.05	0.31
	SK	0.05	0.60 ^a	0.12	0.65 ^a
Fast	DW	0.02	0.23	0.02	0.08
	SK	0.06	0.50 ^a	0.15	0.52 ^a

^a $p < 0.05$.

(1983) and has been applied to an analysis of jaw movements in repetitive speech and nonspeech gestures (Nelson *et al.*, 1984). The key idea is that articulatory movements during speech are accomplishing system “goals” in the physically most economical fashion, i.e., according to some “ease of movement” criteria (see also Lindblom, 1983), which in turn imposes boundary constraints on speech motor programming (Nelson, 1983). Such criteria may be met by minimizing a number of possible articulatory cost indices such as “effort” (proportional to peak velocity, which bears a direct relation to the impulse or integral of the force–time curve for a given movement) or “jerk” (the first derivative of acceleration). Nelson (1983) shows that although a wide variety of “movement ease” cost functions may be minimized, the displacement–duration relation remains roughly the same. Thus a common feature of all such functions is that “cost” increases (on whatever dimension) are associated with moving a given distance in less time or moving a greater distance within a given time. To do either requires an increase in peak velocity, acceleration, jerk, etc. (see also Hogan, 1984).

In the displacement–time space a relationship, such as that displayed by subject SK in Figs. 6 and 7 is suggestive of a fairly constant articulatory cost (cf. Nelson, 1983, Fig. 5). Thus it could be argued that gestures of short amplitude and duration (e.g., fast unstressed gestures) do not necessarily cost the system any more than larger amplitude movements of greater duration (corresponding, say, to normal stressed gestures). Distance and time mutually adapt to the linguistic requirements of the activity in such a way as to preserve a relatively constant cost.

A problem, however, with this analysis of “economy of effort” in speech is that it appears to pertain, at best, to only one of our subjects and to only one of the three subjects in the Nelson *et al.* (1984) study. Several possibilities could account for such a state of affairs. One is that it could reflect differences in the skill level of producing reiterant speech. That is, the less constrained, more variable relation between dis-

placement and time in subject DW suggests that her mode of motor control is not following a strategy of minimum cost. DW may, in fact, have to *discover* exactly what that strategy is. It is well appreciated in the literature (e.g., Larkey, 1983) that reiterant speech is itself a skill, and it was certainly our impression that subject DW was not as skilled at “converting” real speech into reiterant speech as was subject SK. How cost functions change with increasing skill is a topic open to much further research.

Given that the displacement–time relation is not consistent between subjects in the present study or in the literature in general (see Nelson *et al.*, 1984; Parush *et al.*, 1983; Tuller *et al.*, 1982b), the question is: Are there other observables that might afford insight into the *similarity* among subjects in this task? Are subjects really as different in performing reiterant speech as the displacement–time distributions suggest? As we shall see, examination of the higher derivatives of motion not only affords a window into the nature of the system’s underlying dynamic organization, but also suggests that the differences between subjects might be due to the surface nature of the displacement–time description.

2. Peak velocity and the peak velocity–displacement relation

The phase plane data discussed in Sec. II B reveal at least two interesting features about a given gesture’s velocity pattern that merit further quantification. First, the patterns are largely unimodal (see Figs. 3–5) in that both opening and closing gestures possess single velocity peaks. Related to this, peak velocity (V_p) bears a direct relationship to total impulse (i.e., the integral of the force magnitude as a function of time), and thus can usefully be used to index the “effort” underlying the movement (e.g., Nelson, 1983; Schmidt *et al.*, 1979). Since variables like stress have been associated with articulatory effort (e.g., Ohman’s, 1967, stress pulse theory) it is of interest to quantitatively assess if and how peak velocity changes with gesture type, rate, and stress conditions. Second, and perhaps more important, is the apparent regularity—evident on the phase plane—in the covariation between a gesture’s peak velocity (V_p) and its displacement (d). We consider first the statistical effects on peak velocity itself; then we evaluate and interpret the relationship between peak velocity and displacement.

A cursory look at Tables II and III indicates that V_p , like displacement and movement time, varies systematically with stress and rate, although in somewhat idiosyncratic ways. The gesture type main effect is significant for both subjects, $F_s = 59.08$ and 111.01 , $ps < 0.0001$, for DW and SK, respectively. For similar conditions, all the V_p values in Table III (closing) are greater than Table II (opening). Stress had predictable effects on peak velocity regardless of gesture type. As in the recent results of Stone (1981) on jaw movement, and Ostry *et al.* (1983) on tongue dorsum movement, stressed gestures are produced with higher peak velocities than unstressed gestures, $F = 15.03$, $p < 0.002$ and $F = 201.48$, $p < 0.0001$, for DW and SK, respectively.

As others have found, however, the effect of speaking rate on the V_p measure was not so consistent across subjects (e.g., Abbs, 1973; Kuehn and Moll, 1976; Ostry *et al.*, 1983;

Tuller *et al.*, 1982b). For subject DW, peak velocity was greater for the faster speaking rate, $F = 4.94$, $p < 0.03$. For SK, the opposite occurred (see Tables II and III), $F = 94.41$, $p < 0.0001$. Moreover, there was a stress \times rate interaction for SK, $F = 40.06$, $p < 0.0001$ but not DW, $F = 0.86$, $p > 0.05$. For SK, although stressed gestures are always produced more rapidly than unstressed gestures in both speaking rates, $F = 30.93$, $p < 0.0001$ (normal) and $F = 210.61$, $p < 0.0001$ (fast), only unstressed gestures differentiate between normal and fast speaking rates. For subject SK's unstressed gestures, normal speaking rates have higher V_p than fast speaking rates, $F = 132.50$, $p < 0.0001$. For stressed gestures, no significant differences in V_p occur between speaking rates (see Tables II and III), $F = 1.97$, $p > 0.05$.

Because stress has very systematic effects on a variety of variables (including not only the kinematics reported here, but EMG as well, e.g., Tuller *et al.*, 1982a) and the effects of rate are less systematic across subjects (particularly for V_p), it can be argued that stress and rate are qualitatively different kinds of articulatory transformations (see Tuller *et al.*, 1982a, for review). However, the differences observed between stress and rate remain puzzling at least when viewed on single dimensions (e.g., EMG amplitude, duration, and articulator velocity), and further work is necessary to establish the validity of this claim. One potential issue—yet to be fully explored—is that the subject is usually free to vary the elected rate whereas stress constraints are more clearly defined. Systematic control of speaking rate may prove useful and enlightening.

The linkage between peak velocity and displacement, however, is less ambiguous. This finding in itself is not new; it has been reported before in other studies of articulation, often as an incidental result (e.g., Kent and Moll, 1972; Kozhevnikov and Chistovich, 1966; Kiritani *et al.*, 1982; Kuehn and Moll, 1976; Ohala *et al.*, 1968; MacNeilage, 1970; Perkell, 1969; Sussman *et al.*, 1973). The particular articulator involved does not appear to be a factor; the relationship exists in movements at the supralaryngeal level including the tongue dorsum, tongue tip, lips, and jaw. More recently it has been observed in both abduction and adduction of the vocal folds (Munhall and Ostry, 1984).

Quantitative analysis of the present data reveals that V_p and d are highly correlated for opening and closing gestures in both subjects. For subject SK the correlations, collapsed across stress and rate, are 0.87 for the opening phase and 0.94 for the closing phase. For subject DW the correlations are 0.84 and 0.76 for opening and closing gestures, respectively ($ps < 0.01$). Compared to the displacement–time relationship, which was very different between subjects (cf. Figs. 6 and 7), the scatter plots displayed in Figs. 8 and 9 for opening and closing gestures, respectively, show a much greater degree of overall similarity between subjects in both phases of motion.

The high linearity, of course, is a reflection of the overall temporal stability present in the opening and closing phases of the articulatory movements across rate and stress transformations. Since the slope of the V_p – d relationship for a given gesture type has the units of frequency, a perfect

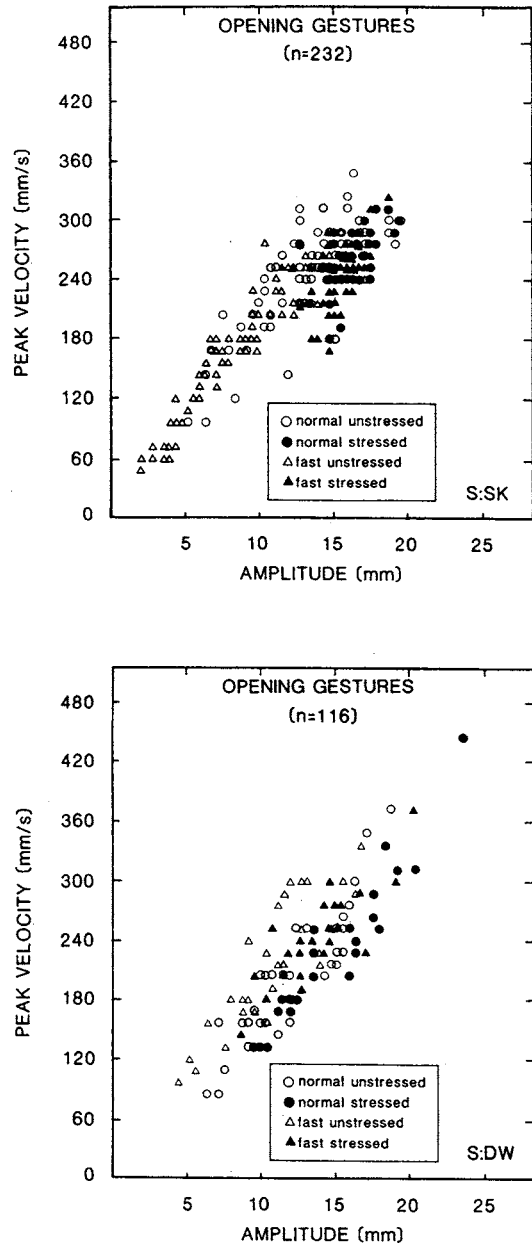


FIG. 8. Scatter plot of peak velocity versus (peak-to-valley) amplitude (lower lip) of each subject's opening gestures associated with the consonant-vowel portion of the syllable. Legend specifies conditions.

correlation between the two variables would indicate that the opening or closing gestures were of the same frequency, i.e., were perfectly isochronous. There are, however, local effects of stress and rate when the data are partitioned into subcategories, as can be seen from the absolute values of displacement, peak velocity, and duration given in Table II for opening gestures and Table III for closing gestures. In Table V we present the linear regression slopes and correlations of the peak velocity–displacement relationship for opening and closing gestures as a function of stress and rate. Overall, although the correlations are generally high and significantly different from zero, the slopes of the relationship between peak velocity and displacement are quite vari-

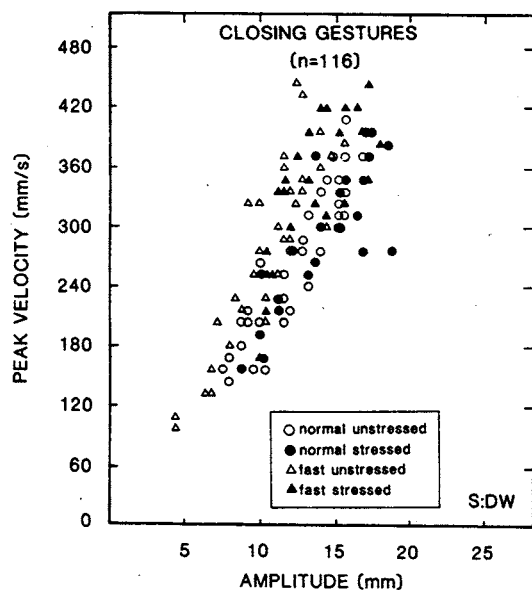
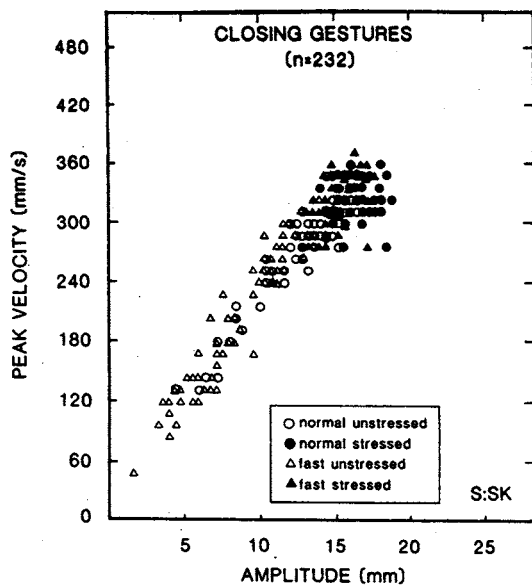


FIG. 9. Scatter plot of peak velocity versus (valley-to-peak) amplitude (lower lip) of each subject's closing gestures associated with the vowel-consonant portion of the syllable. Legend specifies conditions.

able across subcategories. How might the slopes of the kinematic relation between V_p and d be interpreted with respect to the control processes underlying the reiterant speech task? First we address the significance of the overall V_p - d relation; then we consider the specific effects of rate and stress.

Recent theoretical considerations and empirical findings in the motor control field support an account of the V_p - d relation that is based on a movement's dynamics, not its kinematics. Relations among kinematic variables are useful to describe the space-time behavior of articulators, but it is dynamics that cause such motions. That is, it is important to realize that changes in displacement and its time derivatives (velocity and acceleration) are consequences of dynamical

TABLE V. Linear correlations (r) and regression slopes (m) of peak velocity-displacement relationship across rate and stress transformations ($/ba/$). All r 's except those marked by an asterisk are significant at $p < 0.01$ or greater.

		Opening gestures			
		Stressed		Unstressed	
		m	r	m	r
Normal	DW	19.02	0.94	18.02	0.90
	SK	9.29	0.53	13.19	0.79
Fast	DW	14.18	0.82	17.02	0.87
	SK	10.78	0.48	15.68	0.94
		Closing gestures			
		Stressed		Unstressed	
		m	r	m	r
Normal	DW	20.44	0.82	20.19	0.79
	SK	3.04	0.25*	18.00	0.95
Fast	DW	20.25	0.74	27.39	0.87
	SK	13.20	0.68	20.92	0.95

systems with parameters such as mass, stiffness, and damping. It is possible, however, to infer the structure of the underlying dynamics from the kinematics of articulator motions during either discrete or rhythmic tasks.

It is now generally recognized that many features of single-dimensional movements in discrete targeting tasks can be generated by second-order, linear models whose parameters include damping, stiffness, and rest angle (cf. Bizzi, 1980; Cooke, 1980; Fel'dman and Latash, 1982; Kelso and Holt, 1980 for reviews). In short, the limb exhibits behavior qualitatively similar to a damped mass-spring system for these tasks (Fel'dman, 1966). Such systems are intrinsically self-equilibrating in the sense that the "endpoints" or "movement targets" are achieved regardless of initial conditions. In normal and deafferented animals, for example, it has been shown that desired head (Bizzi *et al.*, 1976) and limb positions (Polit and Bizzi, 1978) are attainable without starting position information even when the limb is perturbed on its path to the goal. Similarly, Kelso (1977) demonstrated that finger localization ability is not seriously impaired in functionally deafferented humans, or individuals with the metacarpophalangeal joint capsule surgically removed, in spite of changes in initial conditions or unexpected perturbations (Kelso and Holt, 1980; see also Kelso and Tuller, 1983, and Tye *et al.*, 1983 for evidence in speech). Closed-loop notions that rely on peripheral feedback break down in the face of such evidence. Further, kinematic variables need not be controlled explicitly. In a dynamic system like a damped mass-spring (or *point attractor*, Abraham and Shaw, 1982), kinematic variations in displacement, velocity, and acceleration occur as a result of the specified parameter values, and sensory "feedback" in its conventional form is not required. *Nor, importantly, is duration a controlled variable* (see Sec. II B).

For sustained, stable cyclic movements of dissipative systems the appropriate dynamic regime is a limit cycle (or

periodic attractor, Abraham and Shaw, 1982). In such systems, the same orbit is achieved regardless of initial conditions or temporary perturbations. In the absence of imposed perturbations, such systems can display near-sinusoidal steady-state motions that may be treated as if they were generated by simpler nondissipative mass-spring dynamics. As mentioned earlier, a constant slope in the relationship between each gesture's peak velocity and displacement for a given set of gestures indicates that the gestures are perfectly isochronous. With regard to an hypothesized underlying linear (harmonic) or nonlinear (anharmonic) mass-spring model, the $Vp-d$ slope is indicative of the stiffness over the range of gestural displacements examined. Roughly speaking, a constant $Vp-d$ slope for a given gestural subset implies that the average mass-normalized stiffness (K_{av}^*) of the spring functions underlying the gestures is the same across the observed range.⁴ Recently, Ostry *et al.* (1983) have shown in a study of tongue dorsum movement that the slope of the $Vp-d$ relation varies systematically with stress, but less so as a function of rate. In their data, particularly for opening gestures, the slope of the relationship was greater for unstressed than stressed gestures, suggesting to them that the tongue muscle system was actually *stiffer* in the unstressed environment (see also Laferriere, 1982, for evidence leading to the same conclusion). More recently, observations of tongue dorsum kinematics as a function of rate, vowel (/u/, /o/, and /a/), and consonant (/k/ and /g/) have been interpreted as indicative of an underlying mass-spring control regime with *constant* linear stiffness for a given gesture (Ostry and Munhall, 1984; see also Munhall and Ostry, 1984).

Our data also suggest that unstressed gestures are characterized by greater stiffness (K_{av}^*) values (as revealed in Table V by the slopes of the $Vp-d$ relations and the phase portraits) than stressed ones. This is apparent in three out of four cases for both opening and closing gestures (Table V). Interestingly, we show also that the $Vp-d$ slopes for closing gestures (again with a single exception) are greater than those for opening gestures, particularly for unstressed syllables. Like the Ostry *et al.* (1983) tongue data, the rate effects on the slope of the $Vp-d$ relationship are less clear cut. In only five of eight possible cases, slope increases as a function of rate. With one notable exception, however, in which a fourfold increase in slope occurs, slope changes between fast and normally produced gestures are fairly small.

Although the data in general suggest that stiffness (K_{av}^*) is a key system parameter, a comparison of the $Vp-d$ slope data (which indexes K_{av}^*) and the displacement data shown in Tables II and III reveals that stiffness is not constant for movements of different displacements within a given stress condition (see also Ostry and Munhall, 1984). In fact, stiffness changes invariably with displacement both within and across stress categories.

3. The acceleration-displacement function

There are at least two possibilities that can account for the observed change in stiffness (K_{av}^*) as a function of displacement. One is that a linear spring function holds, for which spring force equals $-kx$ and for which *different* values of linear stiffness k are elected for, say, shorter ampli-

tude, unstressed gestures than larger amplitude, stressed gestures. An alternative notion is that during reiterant speech the jaw-lip system behaves like a soft, nonlinear mass-spring system where, for example, spring force equals $-kx + ex^3$, with k and e denoting linear and cubic stiffnesses, respectively (cf. Jordan and Smith, 1977; Kelso *et al.*, 1983). For such springs, the net stiffness decreases nonlinearly with deviation from the equilibrium position. Hence, shorter amplitude gestures, involving relatively small deviations from equilibrium, are characterized by higher average stiffnesses over the course of the movement than larger amplitude gestures (see also footnote 4). This second hypothesis is presaged on the assumption that all the motions we have observed arise from a single underlying nonlinear spring function with constant linear and cubic stiffness coefficients. Since a gesture's linear stiffness coefficient is indexed by the slope of the acceleration-displacement function near the gesture's midpoint (corresponding roughly to its equilibrium position), we can distinguish between these foregoing alternatives.

The acceleration data of the lower lip-jaw combination were obtained by velocity differentiation and smoothed over a 25-ms interval (see Sec. I). Linear instantaneous, mass-normalized stiffness, K^* , was estimated using a computer routine that first found the midpoint of a given opening or closing gesture and then obtained the position (x) and acceleration (X) coordinates of the data sample to each side of the midpoint. This procedure allowed us to compute the slope around the hypothesized equilibrium position. If K^* is unchanged across conditions the slopes should be statistically equal. Thus if the data lie on a single spring function (linear or nonlinear) K^* should be identical close to the movement's midpoint. Different slopes of the x, X function, however, would suggest separate spring functions with distinct linear stiffness components.⁵

Figure 10 (inset) shows how K^* was estimated and also

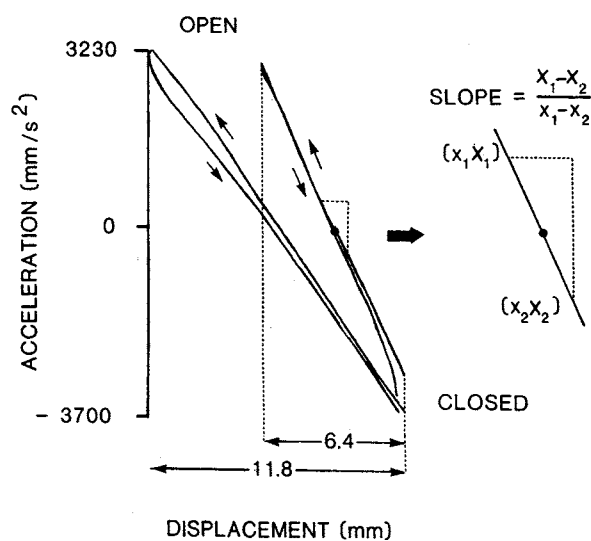


FIG. 10. Acceleration versus displacement from rest position for the opening and closing gestures associated with a stressed and unstressed syllable (see text). The smaller displacements and steeper slopes correspond to the unstressed gestures. The opening gestures start at the bottom right; closing gestures start at top left.

an example of the acceleration–displacement differences between the opening and closing gestures of a stressed versus an unstressed syllable, the fourth and fifth syllables (italicized) of the reiterant versions of “There is *ac-cor-ding* to legend ...” (normal rate, SK). Differences in slope are apparent, with the shorter amplitude, unstressed gestures displaying greater K^* values than the longer amplitude, stressed gestures.

Statistical analysis bears this picture out. The mean estimated K^* and its standard deviation are provided for each subject and each gesture type as a function of conditions in Table VI. Stressed gestures as a class have lower K^* values than unstressed gestures, $F_s = 9.38$ and 192.13 , $p_s < 0.01$ for DW and SK, respectively. Subject DW displays a gesture type main effect, $F = 19.16$, $p < 0.0001$ with K^* significantly greater for closing than opening gestures. Additionally, for SK there is a gesture type \times stress interaction, $F = 20.39$, $p < 0.0001$, and also a gesture \times stress \times rate interaction, $F = 4.70$, $p < 0.04$. A simple main effects analysis of these interactions revealed that for SK: (a) K^* is greater for unstressed gestures than stressed gestures for both opening, $F = 168.85$, $p < 0.0001$, and closing gestures, $F = 43.67$, $p < 0.0001$; (b) K^* is greater for stressed gestures in the closing phase than the opening phase, $F = 8.80$, $p < 0.004$; and (c) K^* is greater for unstressed gestures in the opening phase, $F = 12.17$, $p < 0.0006$, particularly at the fast speaking rate (see Table VI).

The effect of rate is highly significant for both subjects. In all the cells of similar conditions, K^* is greater at the faster speaking rate than it is in syllables produced at a conversational pace, $F = 69.43$, $p < 0.0001$ (DW) and $F = 90.80$, $p < 0.0001$ (SK). Subject SK also reveals a stress \times rate interaction, $F = 41.78$, $p < 0.0001$, although DW does not, $F = 1.22$, $p > 0.05$. For subject SK: (a) at both rates, K^* is greater in unstressed than stressed gestures, $F = 27.39$, $p < 0.0001$ (normal) and $F = 206.55$, $p < 0.0001$ (fast); and (b) only in unstressed gestures and in stressed closing gestures, however, is K^* greater for fast than for normal speaking rates (see Table VI).

TABLE VI. Estimated stiffness (K^*) in units of acceleration per unit displacement across rate and stress transformations. Standard deviation is in parentheses.

		Opening gestures	
		Stressed	Unstressed
Normal	DW	1443 (405)	1703 (561)
	SK	1781 (342)	2413 (703)
Fast	DW	1931 (836)	2336 (787)
	SK	1803 (427)	3555 (1211)
		Closing gestures	
		Stressed	Unstressed
Normal	DW	1854 (364)	1889 (496)
	SK	1981 (338)	2308 (344)
Fast	DW	2409 (378)	2633 (452)
	SK	2193 (337)	3078 (880)

These data correspond rather well to the peak velocity–displacement findings discussed in Sec. II C 3. The present acceleration–displacement results, however, afford an additional conclusion, namely, that linear mass-normalized stiffness (K^* , estimated around the equilibrium point of the motion) is *not* the same for short amplitude, unstressed gestures as it is for large amplitude, stressed gestures. In short, different stress categories are characterized by different K^* values. A similar conclusion applies to rate changes. In all the cells from comparable conditions shown in Table VI, faster speaking rates are accompanied by higher estimated K^* values, and, as we reported in Sec. II C 1, smaller displacements. Thus although a constant linear stiffness model is a reasonable first approximation it does not handle all of the kinematic variations in our data that are induced by stress and rate. For whatever reasons, no doubt in part linguistic, linear stiffness is modulated according to the stress (or amplitude) of the gesture. Increasing stiffness for unstressed (shorter amplitude) gestures may be a way for the English language, conventionally classified as stress timed, to differentiate its stress categories. Interestingly, recent theorizing in speech perception argues for a perceptual metric that is closely tied to articulatory dynamics (e.g., Summerfield, 1979; Studdert-Kennedy, 1983). The notion, based in part on studies of visual perception (e.g., Runeson and Frykholm, 1981) is that perception of events is not simply based on surface kinematics, but on the underlying relations among dynamic parameters that govern such events. The present findings, in showing a clear relation between stress and linear stiffness values, provides an initial grounding for these speculations. The data also show that faster speaking rates are associated with higher estimated linear stiffnesses, though, like the Ostry *et al.* (1983) tongue data, the rate effects are not quite as consistent.

D. Summary and preliminary dynamic modeling

To summarize, the present data offer insights into both the similarities and differences in our subjects' articulatory behavior. The movements of both subjects can be assumed to emerge from the same underlying dynamic organization. That is, a periodic attractor (limit cycle) control regime can capture the forms of motion produced by both subjects. The slopes of the peak velocity–displacement and the acceleration–displacement functions point to linear mass-normalized stiffness K^* as a key dynamic parameter. The subjects differ, however, in the degree to which estimated K^* and overall gestural displacement are coupled *across* movement conditions. Subject SK shows an inverse relation between stiffness and displacement for opening ($r = -0.77$) and closing ($r = -0.73$) gestures. Thus larger (smaller) amplitude motions that accompany stressed (unstressed) gestures and normal (fast) rates are associated with lower (higher) stiffness. For DW, however, the correlation between estimated stiffness and displacement (like DW's displacement–movement time relation) is low (-0.18 for opening and -0.25 for closing gestures), perhaps because of the reasons discussed in Sec. II C 1. In short, the “strength” of the constraint between K^* and displacement may be a useful way to conceptualize between-subject differences.

LOWER LIP AND JAW

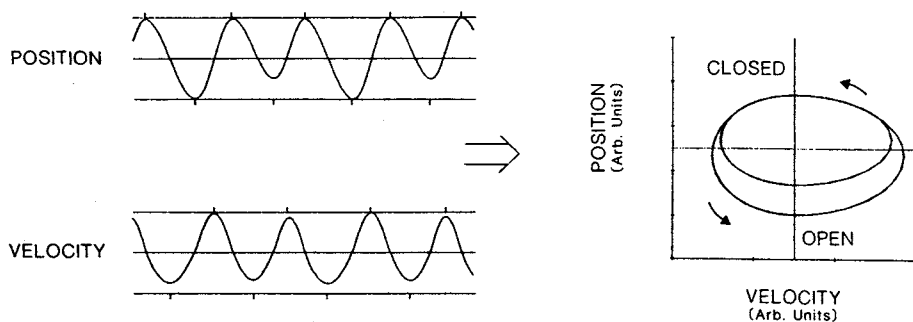


FIG. 11. Computer simulation of resultant lower lip position and velocity time series (left) and corresponding phase portraits for a pattern of reiterant, alternating stress syllables.

The present findings can be couched conveniently within a recent dynamic modeling and computer simulation framework developed for multiarticulator systems by Saltzman and Kelso (1983). Briefly, the unique feature of this approach is that invariant dynamical equations of motion are established functionally (i.e., at an abstract *task level* of movement description) for the particular end effectors directly involved in the task's accomplishment. For example, Saltzman and Kelso (1983) demonstrate that a constant set of dynamic parameters defined for a given task, e.g., a hand reaching for a target, can be used to specify context- (task and posture) dependent patterns of change in the articulator-level dynamic parameters (e.g., joint stiffness, damping, and equilibrium points of shoulder, elbow, and wrist). Among other advantages the approach allows for task-specific trajectory shaping (e.g., Bizzi *et al.*, 1982) and the compensatory behaviors typically involved in speech (e.g., Folkins and Abbs, 1975; Kelso *et al.*, 1984).

At a recent conference, Browman *et al.* (1984) reported that the task-dynamic approach can be fruitfully applied to understanding speech organization. For example, we have used the average values of amplitude and duration from the present data (for stressed and unstressed gestures at a particular rate) to estimate the dynamic parameters (equilibrium positions and stiffnesses) in a functional mass-spring model for the control of lip aperture defined by the vertical distance between the upper and lower lip. These lip aperture parameters remain invariant throughout a given lip opening or closing gesture, and during each gesture are transformed into contextually varying patterns of dynamic parameters at the articulatory level (upper lip, lower lip, and jaw degrees of freedom as defined in the Haskins Laboratories' software articulatory synthesizer; Rubin *et al.*, 1981). Thus inserting our empirically estimated dynamic parameters for lip aperture into the task-dynamic model, we can generate sets of simulated articulator trajectories associated with lip opening and closing. Figure 11 illustrates simulated time series and phase plane trajectories for the resultant vertical motion of the lower lip during a reiterant bilabial task with simple alternating stress.

In these simulations, the equilibrium position for a given cycle (closure-to-closure) is specified at the onset of the opening gesture and is located halfway between the maximum opening position and the (relatively fixed) closure position. However, because closing gestures are faster than opening gestures (compare Tables II and III) stiffness is specified

twice during the cycle: once at the start of the opening gesture and once at the start of the closing gesture. Although the present example simply shows an alternating stress pattern, clearly this procedure can be executed on a syllable-by-syllable basis. Although the model is presently undergoing refinement (e.g., to incorporate fully limit cycle dynamics), Browman *et al.* (1984) have used the displacement-time data shown in Fig. 11 as input parameters to an articulatory synthesizer with promising acoustic and perceptual results. The point here, however, is that the simulation illustrates how articulatory trajectories can be generated from a simple specification of dynamic parameters without explicit or detailed trajectory planning.

III. CONCLUSIONS

The phase portrait methodology introduced in Sec. II B, along with a detailed analysis of articulatory kinematics, allow us a window into the hypothesized dynamic structure underlying the production of simple, reiterant syllables. It is popular to propose "time control" as the basis of temporal organization in speech, as if the system somehow had to program and/or keep continuous track of time (e.g., Lindblom, 1963; Lindblom *et al.*, 1984). Different time control schemes, according to this notion, correspond to stress and rate, while other kinematic variables, such as velocity, are computationally derived (cf. Kuehn and Moll, 1976; Laferriere, 1982). In an alternative view, which we have applied here, spatiotemporal pattern arises as a consequence of a dynamic regime in which—at worst—only two articulatory parameters, stiffness and rest position, are specified according to stress and rate requirements. Similar arguments have been proposed for the space-time structure of multidegree of freedom limb movements (Kelso *et al.*, 1979, 1983). The dynamic description captures the forms of articulatory motion observed in our phase portraits across rate and stress conditions. It recognizes in full that articulatory motions *evolve* in time but it undercuts the necessity to explicitly *regulate* time as a controlled variable. Dynamics can provide a grounding for, and a principled analysis of so-called intrinsic timing theories of speech production (Fowler *et al.*, 1980). According to the present findings and supplemented by preliminary modeling, movement time results from an underlying dynamic organization that is specified according to linguistic requirements and that remains invariant throughout the production of a given speech gesture.

ACKNOWLEDGMENTS

This research was supported by NIH grant NS-13617, Biomedical Research Support grant RR-05596, and contract no. N00014-83-C-0083 from the U.S. Office of Naval Research. Part of the work was reported by J. A. S. Kelso and E. V. Bateson at the 105th Meeting of the Acoustical Society of America in Cincinnati, OH in a paper entitled "On the cyclical basis of speech production" [J. Acoust. Soc. Am. Suppl. 1 73, S67 (1983)]. We thank Betty Tuller and David Ostry for helpful comments on an earlier version of the paper and J. S. Perkell and W. L. Nelson for detailed and constructive reviews.

¹For many examples of complex, multicomponent systems in physics, chemistry, and biology, whose cooperative behavior can be described by a small set of dynamic or "order" parameters, see Haken (1975, 1977). For examples in speech and other biological motions, see Kelso and Tuller (1984b).

²The field of qualitative dynamics has a rich history dating back to Poincaré (1899) (see Abraham and Shaw, 1982; Garfinkel, 1983). In this vein, we combine geometry and dynamics to reflect our concern with the forms of articulator motion (indicated by patterns of displacement, velocity, and acceleration over time) that are created by a functionally defined dynamical organization (e.g., point attractor or periodic attractor dynamics).

³Note that the plotting convention here is not the one typically used in dynamics, which plots position on the horizontal axis and velocity on the vertical axis. Since the displacements measured here are vertical, not horizontal, we have simply switched the axes to conform to the behavior of the lip-jaw system and to facilitate visualization of the process.

⁴Both linear and nonlinear mass-spring systems can display near-sinusoidal cyclic motions whose observed peak-to-peak period $T = 2\pi/\Omega_c$ where Ω_c denotes the observed angular frequency for the cycle. For systems with constant parameters, the peak-to-valley duration ($D_p = \pi/\Omega_p$) and the valley-to-peak duration ($D_v = \pi/\Omega_v$) are equal and, consequently, $T = D_p + D_v = 2D = 2D_p$, and $\Omega_c = \Omega_v = \Omega_p$. More generally, in cases where motion during each half-cycle is near sinusoidal but of different duration we have $\Omega_c = 2[\Omega_v\Omega_p/(\Omega_v + \Omega_p)]$. A linear undamped mass-spring system (harmonic oscillator) may be characterized by the following equation of motion with constant parameters: $m\ddot{X} + k\Delta x = 0$, where m = mass, k = linear stiffness, $\Delta x = (x - x_0)$ with x_0 = rest position, and x and \dot{X} represent position and acceleration, respectively. Such systems display cyclic motions with period $T = 2\pi/\Omega_c$, where $\Omega_c = \omega_0 = (k/m)^{1/2}$, and ω_0^2 (denoted K^* in Sec. II C 3) defines the mass-normalized linear stiffness of the system. Due to system linearity, the instantaneous system stiffness is independent of displacement and, hence, both the instantaneous and the "average" stiffness of the system for motion cycles of different amplitudes are simply equal to k . Normalizing with respect to mass, we see that the average mass-normalized stiffness described in the text, (K^*) is simply $k/m (= \omega_0^2 = \Omega_c^2)$ for linear mass-spring systems. Additionally, the peak velocity (V_p) for harmonic oscillators is $\omega_0 A$, where A denotes the maximum displacement from the rest position during cyclic motion. Consequently a plot of V_p vs A for different amplitude cycles of a given linear oscillator shows a straight line whose slope equals ω_0 . Thus for a given linear mass-spring system the V_p - A slope is equal to $\omega_0 = \Omega_c = (K^*)^{1/2}$ and is constant across the entire displacement range. For undamped mass-spring systems with nonlinear stiffness functions (anharmonic oscillators), however, the average mass-normalized stiffness (K^*) for motion cycles of different amplitude is not so simply related to the system's instantaneous stiffness. For example, for a soft nonlinear spring (cf. Jordan and Smith, 1977) the equation of motion is $m\ddot{X} + k\Delta x - e\Delta x^3 = 0$, where e = cubic stiffness, and all other terms are as in the linear case. Here, the system's instantaneous stiffness does not equal k but is a nonlinearly decreasing function of the amplitude of motion. Thus the system's (K^*) will vary for cycles of different amplitude with (K^*) decreasing for increasing amplitudes. Additionally, the plot of V_p vs A for different cycles will have a slope that is a decreasing function of amplitude, unlike the linear systems described above. Yet, like these linear systems, the V_p - A slope is still proportional to $(K^*)^{1/2}$.

⁵We had two concerns about the derivation of acceleration. First, we wanted to ascertain how the elected smoothing window changed the values of the central portion of the trajectories where the slope of the acceleration-displacement function was calculated in this study. Second, and related, we wished to ascertain how our derived (smoothed and differentiated) acceleration compared with actual accelerometric data. Two independent methods were used to examine the effects of numerical smoothing and differentiation on the acceleration data; in both cases, the effects were small. (1) To assess the effects of smoothing on the reported data itself, the x, \dot{x} slope was calculated for a subset of subject SK's reiterant productions (16 opening gestures representative of the overall stress/rate distribution) at four degrees of smoothing: 15 ms, 25 ms, twice at 25 ms (the condition used in the text), and once at 25 ms and again at 45 ms. Increased smoothing reduced mean slope values, $F(3,60) = 3.48, p < 0.05$, but did not change the pattern relative to overall gesture displacement, $F(6,56) = 0.37, p > 0.1$. (2) The influence of the double differentiation (i.e., acceleration derived from position) and concomitant smoothing procedures was tested at a movement frequency (5 Hz) comparable to that used by speakers in the present study (see Table I), by comparing the output of an Entran accelerometer (model EGC-240-10D) to the second derivative of position output smoothed twice at 25 ms. Taking into account the gain reduction induced by smoothing (see above), we found the average, absolute difference between transducer (unsmoothed) and numerical (smoothed twice at 25 ms) acceleration to be less than 5% of the range of measured movements. The cross correlation between the raw, unsmoothed and the smoothed, derived signal was $r = 0.98$. Note: Not all the x, \dot{x} functions approximated straight lines as closely as those illustrated in Fig. 10. Some were S shaped ("hooked" at displacement extrema). However, our smoothing procedures did not remove the "hooks." More important, our estimates of K^* were not affected by the presence of such "hooks."

- Abbs, J. H. (1973). "The influence of the gamma motor system on jaw movements during speech: A theoretical framework and some preliminary observations," J. Speech Hear. Res. 16, 175-200.
- Abraham, R. H., and Shaw, C. D. (1982). *Dynamics—The Geometry of Behavior* (Aerial, Santa Cruz, CA).
- Bizzi, E. (1980). "Central and peripheral mechanisms in motor control," in *Tutorials in Motor Behavior*, edited by G. E. Stelmach and J. Requin (North-Holland, Amsterdam).
- Bizzi, E., Accornero, N., Chapple, W., and Hogan, N. (1982). "Arm trajectory formation in monkeys," Exp. Brain Res. 46, 139-143.
- Bizzi, E., Polit, A., and Morasso, P. (1976). "Mechanisms underlying achievement of final head position," J. Neurophysiol. 39, 435-444.
- Browman, C., Goldstein, L., Kelso, J. A. S., Rubin, P. E., and Saltzman, E. L. (1984). "Articulatory synthesis from underlying dynamics," J. Acoust. Soc. Am. Suppl. 1 75, S22.
- Cooke, J. D. (1980). "The organization of simple, skilled movements," in *Tutorials in Motor Behavior*, edited by G. E. Stelmach and J. Requin (North-Holland, Amsterdam).
- Fel'dman, A. G. (1966). "Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Mechanographic analysis of execution by man of the simplest motor tasks," Biophysics 11, 766-775.
- Fel'dman, A. G., and Latash, M. L. (1982). "Afferent and efferent components of joint position sense: Interpretation of kinesthetic illusions," Biol. Cybernet. 42, 205-214.
- Folkens, J. W., and Abbs, J. H. (1975). "Lip and jaw motor control during speech: Responses to resistive loading of the jaw," J. Speech Hear. Res. 18, 207-220.
- Fowler, C. (1983). "Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet," J. Exp. Psychol.: Gen. 112, 386-412.
- Fowler, C. A., Rubin, P., Remez, R. E., 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).
- Garfinkel, A. (1983). "A mathematics for physiology," Am. J. Physiol. Reg. Int. Comp. Phys. 245, R455-R466.
- Grillner, S. (1982). "Possible analogies in the control of innate motor acts and the production of sound in speech," in *Speech Motor Control*, edited by S. Grillner, B. Lindblom, J. Lubker, and A. Persson (Pergamon, Oxford).
- Haken, H. (1975). "Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems," Rev. Mod. Phys. 47, 67-121.

- Haken, H. (1977). *Synergetics: An Introduction* (Springer-Verlag, Heidelberg).
- Hogan, N. (1984). "An organizing principle for a class of voluntary movements," *J. Neurosci.*, in press.
- James, M. L., Smith, G. M., and Wolford, J. C. (1977). *Applied Numerical Methods for Digital Computation* (Harper and Row, New York), 2nd ed.
- Jordan, D. W., and Smith, P. (1977). *Nonlinear Ordinary Differential Equations* (Clarendon, Oxford).
- Kelso, J. A. S. (1977). "Motor control mechanisms underlying human movement reproduction," *J. Exp. Psychol. Human Percept. Perform.* 3, 529-543.
- Kelso, J. A. S. (1981). "Contrasting perspectives on order and regulation in movement," in *Attention and Performance*, edited by J. Long and A. Baddeley (Erlbaum, Hillsdale, NJ), Vol. IX.
- Kelso, J. A. S., and Holt, K. G. (1980). "Exploring a vibratory systems analysis of human movement production," *J. Neurophys.* 43, 1183-1196.
- Kelso, J. A. S., Holt, K. G., Rubin, P., and Kugler, P. N. (1981). "Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data," *J. Motor Behav.* 13, 226-261.
- Kelso, J. A. S., Putnam, C. A., and Goodman, D. (1983). "On the space-time structure of human interlimb coordination," *Q. J. Exp. Psychol.* 35A, 347-376.
- Kelso, J. A. S., Southard, D. L., and Goodman, D. (1979). "On the nature of human interlimb coordination," *Science* 203, 1029-1031.
- Kelso, J. A. S., and Tuller, B. (1983). "'Compensatory articulation' under conditions of reduced afferent information: A dynamic formulation," *J. Speech Hear. Res.* 26, 217-224.
- Kelso, J. A. S., and Tuller, B. (1984a). "Converging evidence in support of common dynamical principles for speech and movement coordination," *Am. J. Physiol.* 246, R928-R935.
- Kelso, J. A. S., and Tuller, B. (1984b). "A dynamical basis for action systems," in *Handbook of Cognitive Neuroscience*, edited by M. S. Gazzaniga (Plenum, New York).
- Kelso, J. A. S., Tuller, B., V. Bateson, E., and Fowler, C. A. (1984). "Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures," *J. Exp. Psychol. Human Percept. Perform.*, in press.
- Kent, R. D., and Moll, K. (1972). "Cinefluorographic analyses of selected lingual consonants," *J. Speech Hear. Res.* 15, 453-473.
- Kiritani, S., Imagawa, H., Takahashi, T., Masaki, S., and Shirai, K. (1982). "Temporal characteristics of the jaw movements in the production of connected vowels," *RILP* 16, 1-10.
- Kozhevnikov, V. A., and Chistovich, L. A. (1966). *Rech , Artikulyatsiya, i Vospriyatiye (Speech: Articulation and Perception)* (Joint Publ. Res. Serv., Washington, DC) (Orig. pub. 1965).
- Kuehn, D., and Moll, K. (1976). "A cineradiographic study of VC and CV articulatory velocities," *J. Phonet.* 4, 303-320.
- LaFerriere, M. (1982). "Stress and tongue blade movement in alveolar VC gestures," *J. Acoust. Soc. Am. Suppl.* 1 72, S104.
- Larkey, L. S. (1983). "Reiterant speech: An acoustic and perceptual evaluation," *J. Acoust. Soc. Am.* 73, 1337-1345.
- Lehiste, I. (1972). "The timing of utterances and linguistic boundaries," *J. Acoust. Soc. Am.* 51, 2018-2024.
- Lenneberg, E. H. (1967). *Biological Foundations of Language* (Wiley, New York).
- Liberman, M., and Streeter, L. A. (1978). "Use of nonsense-syllable mimicry in the study of prosodic phenomena," *J. Acoust. Soc. Am.* 63, 231-233.
- Lindblom, B. (1963). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* 35, 1773-1781.
- Lindblom, B. (1983). "Economy of speech gestures," in *The Production of Speech*, edited by P. F. MacNeilage (Springer-Verlag, New York).
- Lindblom, B., Lubker, J., Gay, T., Lyberg, B., Branderud, P., and Holmgren, K. (1984). "The concept of target and speech timing," unpublished manuscript (Dept. Phonetics, Stockholm University, Sweden).
- Lisker, L. (1975). "Phonetic aspects of time and timing," invited paper to 100th Meeting of the American Speech and Hearing Association, Washington, D.C.; also Haskins Lab. Stat. Rep. Speech Res. SR-47, 113-120.
- MacNeilage, P. F. (1970). "Motor control of serial ordering of speech," *Psychol. Res.* 77, 182-196.
- Munhall, K., and Ostry, D. (1984). "Ultrasonic measurement of laryngeal kinematics," in *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver, CO), in press.
- Nakatani, L. H. (1977). "Computer-aided signal handling for speech research," *J. Acoust. Soc. Am.* 61, 1056-1062.
- Nelson, W. L. (1983). "Physical principles of economies of skilled movements," *Biol. Cybernet.* 46, 135-147.
- Nelson, W. L., Perkell, J., and Westbury, J. (1984). "Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations," *J. Acoust. Soc. Am.* 75, 945-951.
- Ohala, J. J. (1975). "The temporal regulation of speech," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, London).
- Ohala, J. J., Hiki, S., Hubler, S., and Harshman, R. (1968). "Photoelectric methods of transducing lip and jaw movements in speech," *UCLA Work. Pap. Phonet.* 10, 135-144.
- Ohman, S. E. G. (1967). "Numerical model of coarticulation," *J. Acoust. Soc. Am.* 41, 310-320.
- Ostry, D. J., and Munhall, K. (1985). "Control of rate and duration in speech," *J. Acoust. Soc. Am.*, to be published.
- Ostry, D. J., Keller, E., and Parush, A. (1983). "Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement in speech," *J. Exp. Psychol. Human Percept. Perform.* 9, 622-636.
- Parush A., Ostry, D. J., and Munhall, K. G. (1983). "A kinematic study of lingual coarticulation in VCV sequences," *J. Acoust. Soc. Am.* 74, 1115-1125.
- Perkell, J. S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT, Cambridge, MA).
- Pike, K. (1945). *Intonation of American English* (Univ. Michigan P., Ann Arbor, MI).
- Poincaré, H. (1899). *Les methods nouvelles de la mecanique celeste*, Vol. III.
- Polit, A., and Bizzi, E. (1978). "Processes controlling arm movements in monkeys," *Science* 201, 1235-1237.
- Rosenbaum, D. A., and Patashnik, O. (1980). "Time to time in the human motor system," in *Attention and Performance VIII*, edited by R. S. Nickerson (Erlbaum, Hillsdale, NJ).
- Rubin, P., Baer, T., and Mermelstein, P. (1981). "An articulatory synthesizer for perceptual research," *J. Acoust. Soc. Am.* 70, 321-328.
- Runeson, S., and Frykholm, G. (1981). "Visual perception of lifted weight," *J. Exp. Psychol. Human Percept. Perform.* 7, 733-740.
- Saltzman, E. L., and Kelso, J. A. S. (1983). "Skilled actions: A task dynamic approach," *Haskins Lab. Stat. Rep. Speech Res. SR-76*, 3-50.
- Schmidt, R. A., Zelaznik, H. N., Hawkins, B., Frank, J. S., and Quinn, J. T., Jr. (1979). "Motor-output variability: A theory for the accuracy of rapid motor acts," *Psychol. Rev.* 86, 415-451.
- Stone, M. (1981). "Evidence for a rhythm pattern in speech production: Observations of jaw movement," *J. Phonet.* 9, 109-120.
- Studdert-Kennedy, M. (1983). "On learning to speak," *Human Neurobiol.* 2, 191-195.
- Summerfield, Q. (1979). "Use of visual information for phonetic perception," *Phonetica* 36, 314-331.
- 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., and Kelso, J. A. S. (1984). "The relative timing of articulatory gestures: Evidence for relational invariants," *J. Acoust. Soc. Am.* 76, 1030-1036.
- Tuller, B., Harris, K., and Kelso, J. A. S. (1982a). "Stress and rate: Differential transformations of articulation," *J. Acoust. Soc. Am.* 71, 1534-1543.
- Tuller, B., Harris, K., and Kelso, J. A. S. (1982b). "Articulatory control as a function of stress and rate," *Haskins Lab. Stat. Rep. Speech Res. SR-71/72*, 81-88.
- Tuller, B., Kelso, J. A. S., and Harris, K. (1982c). "Interarticular phasing as an index of temporal regularity in speech," *J. Exp. Psych. Human Percept. Perform.* 8, 460-472.
- Tuller, B., Kelso, J. A. S., and Harris, K. S. (1983). "Further evidence for the role of relative timing in speech," *J. Exp. Psych. Human Percept. Perform.* 9, 829-833.
- Tye, N., Zimmerman, G., and Kelso, J. A. S. (1983). "Compensatory articulation in hearing-impaired speakers: A cinefluorographic study," *J. Phonet.* 11, 101-115.