TR－A－0147

RHYTHM TYPE AND ARTICULATORY DYNAMICS IN ENGLISH, FRENCH, AND JAPANESE

Eric Vatikiotis-Bateson   J.A.Scott Kelso

1992. 7. 10

RHYTHM TYPE AND ARTICULATORY DYNAMICS IN ENGLISH, FRENCH, AND JAPANESE[1]

Eric Vatikiotis-Bateson[2]

J.A. Scott Kelso[3]

Abstract

In the study reported here, movement data were analyzed for archetypes of the three most widely recognized temporal organization categories: English for stress timing, Japanese for mora timing, and French for syllable timing. Reiterant speech productions from the three languages were elicited and analyzed as commensurately as possible, using the experimental methodology employed originally by Kelso, Vatikiotis-Bateson, Saltzman, & Kay (1985) for two speakers of English. The primary aim was to show the extent to which simple kinematic analysis of a primary articulator (the lower lip-jaw complex) can reveal universal and language-specific aspects of temporal organization and prosody. For the most part, kinematic results were like those of Kelso et al. (1985): Most of the spatiotemporal variability of the movement behavior could be accounted for in the highly linear covariation of peak velocity and displacement. Moreover, there were clear condition-specific correlates of stress in English and French, of accent-related tone and mora complexity in Japanese, and speaking rate in all three languages on displacement and on the slope of the linear relation between peak velocity and displacement. These results are interpreted in terms of an abstract, yet simple, second-order system such as a linear spring-mass. By setting a small number of underlying parameters, such a system can characterize the overall spatiotemporal behavior of the lip-jaw system, as well as most of the specific linguistic and performance distinctions in stress, speaking rate, and the like.

1 Introduction

In an earlier study, Kelso, Vatikiotis-Bateson, Saltzman, and Kay (1985) examined the articulatory kinematics of lower lip-jaw motion during reiterant sentence productions by two English speakers. They showed that gestural displacement, duration, and peak velocity, while highly variable, do not vary independently of one another. They found, as had many before them, that the individual kinematics were consistently correlated with differences in "stress" (based on word and phrase prominence) and speaking rate. Stressed movement gestures were larger in displacement, longer in duration, and higher in peak velocity than unstressed gestures produced at the same speaking rate. Similarly, the kinematics of gestures produced at a conversational speaking rate were generally larger than those produced at a faster rate. Further, while there was a marked tendency for lip-jaw movements to take longer to go farther, resulting in a somewhat linear relation between displacement and duration, the relation was quite noisy. On the other hand, the relation between peak velocity and displacement had a strong linear component that accounted for most of the overall variance in the movement behavior.

Such linear covariation between peak velocity and displacement had been observed before in

speech production (e.g., Kozhevnikov & Chistovich, 1965; Ohala, Hiki, Hubler, & Harshman, 1968; Mermelstein, 1973; Sussman, MacNeilage, & Hanson, 1973, Kuehn & Moll, 1976), but it was not until the roughly contemporary studies by Ostry, Keller, & Parush (1983) and Kelso et al. (1985) that the relation was investigated in any detail for running speech. In addition to the highly linear covariation between peak velocity and displacement throughout the data range, both studies showed that variations in stress and, to lesser extent, speaking rate are characterized by systematic differences in the slope of the condition specific peak velocity-displacement relations. Also, both studies recognized that the relation could be modeled as the behavior of a second order dynamical system such as a linear mass-spring. Kelso et al. (1985) went on to show in detail how the spatiotemporal behavior of the system could be simulated by adjusting the settings of just two underlying model parameters — equilibrium position and stiffness — which can be approximated from the mean gestural displacement and the slope of the relation between peak velocity and displacement, respectively.

Approximating the observed, nearly sinusoidal motion of the speech articulators to the motion of a linear mass-spring has several advantages. First, the behavior of such systems is well-defined. A simple second-order equation of motion, $m\ddot{x} + b\dot{x} + k(x - x_0) = 0$, with only a few parameters corresponding to mass (m), viscosity (b), stiffness (k), and equilibrium position (x_0), can generate movements from an infinite number of initial conditions that will attain the same target or end point. Thus, different movement trajectories do not necessarily require different parametrization of the movement equation. Second, experimental data can be used to estimate the underlying dynamic parameters from the observable kinematic variables of acceleration ($\ddot{x}$), velocity ($\dot{x}$), and displacement (x) and their interrelation (e.g., Smith, Browman, McGowan, & Kay, 1991). Third, the temporal as well as the spatial characteristics of motion are fully determined in such a system. Movement duration need not be independently controlled, since it is inversely proportional to and, therefore, recoverable from spring stiffness, k, which can be estimated from the relation between a movement's peak velocity and displacement (see Vatikiotis-Bateson, 1988, Appendix A). Fourth, the same basic system can account for both cyclical and discrete movement behavior (Kay, Kelso, Saltzman, & Schöner, 1987; Saltzman & Kelso, 1987). If, as appears likely, other biological movement behaviors besides speech, such as (discrete) reaching and (cyclical) locomotion, can be successfully described as analogous second order systems, then this analysis may provide a way to identify what is common to all biological movement behaviors (Kelso & Tuller, 1984).

Thus, this approach offers a promising framework for characterizing potentially complex movement behaviors in terms of the function-specific settings of a few underlying parameters. In particular, the success of the simple model proposed by Kelso et al. (1985) in characterizing spatiotemporal behavior across changes in stress and speaking rate suggests that, with appropriate tuning, such a model might be applied universally across languages. The present study is a preliminary test of this suggestion. We analyzed comparable reiterant speech data using two reiterant syllables, *ba* and *ma*, from at least three speakers each of English, French, and Japanese. These languages are generally accepted to differ substantially in their temporal organization and prosody (Pike, 1943; Abercrombie, 1967; Bloch, 1950). Kinematic variables associated with lower lip-jaw motion were analyzed within and across two prosodic conditions and two instructed speaking rates for each language. The detailed results of this study are reported elsewhere (Vatikiotis-Bateson, 1988). Here, the results of that study are quickly summarized, and are used to evaluate hypotheses concerning the existence and implementation of universal and language-specific constraints on supralaryngeal movement behavior.

To this end, lip-jaw movement data are considered in two ways. First, the gestural kinematics are examined in order to assess the overall spatiotemporal character of different languages whose prosodic structures and perceived temporal organization are quite different. We demonstrate that differences observed among the language-specific data are commensurate with temporal differences we have come to expect through perceptual and other empirical observations. At the same time, we show that most of the observed differences in temporal organization and instructed differences in speaking rate can be ascribed to the scaling of stiffness, inferred from the slope of the relation between peak velocity and displacement. From this scaling we hypothesize that opening and closing movements of the lower lip-jaw complex may adhere to the constraints of an underlying, dynamical second-order system, though not one quite so

simple as an undamped linear springmass. The applicability of such a model to three languages so different in their temporal organization is, we argue, indicative of a universal constraint on speech movement behavior. Second, language-specific parametrization of these potentially universal constraints is demonstrated by showing how language-specific prosodic distinctions may be realized similarly within the hypothesized dynamical system through co-modulation of stiffness and equilibrium position, inferred from mean articulator displacement.

2 Methods and procedures

Since much of the experimental and analytic methodology used in this study has been described in detail elsewhere (Kelso et al., 1985; Kay, Munhall, Vatikiotis-Bateson, & Kelso, 1985; Vatikiotis-Bateson, 1988), only the experimental aspects specific to this study are described in detail.

2.1 *Subjects*

Five native speakers of English, four speakers of French, and five speakers of Tokyo Japanese took part in the study. All but one speaker of Japanese (NK), who served as a pilot subject, were naive to the purposes of the experiment, had never been exposed to the reiterant speech task, and were paid for their participation. Speaker NK, on the other hand, selected the Japanese stimuli and practiced them reiterantly before her experimental run.

2.2 *Reiterant Speech Task*

Reiterant, or mimicked, speech is a substitution task in which each syllable of a target phrase or sentence is replaced by a test syllable such as *ba* or *ma*, while trying to maintain the rhythmic and prosodic character of the original. The task was used extensively in early acoustic studies of metrics (Scripture 1899a,b; Wallin, 1901; Stetson, 1905). More recently, it has been used for both acoustic (e.g., Lindblom and Rapp, 1973; Liberman and Streeter, 1978; Larkey, 1983) and articulatory (Kelso et al., 1985) studies.

2.3 *Training Procedure*

The reiterant speech task was explained to English and French speakers as one involving syllable substitution and demonstrated using short phrases and sentences, such as "Mary had a little lamb", spoken at a conversational rate. Due to the controversy over whether the "unit" of timing in Tokyo dialect is the syllable or the mora (e.g., McCawley, 1978; Higurashi, 1984), no attempt was made to force Japanese speakers to use a unit that might be unnatural to them, especially in an already difficult task. Therefore, the reiterant speech task was simply demonstrated by speaker NK, using phrases containing only light, single mora syllables. Speakers practiced producing these phrases reiterantly using *ba* and *ma*, first at a comfortable rate and then as fast as possible, until they could do so fluently with proper intonation and the right number of syllables at two distinct (experimenter determined) rates. Finally, subjects were instructed to memorize the two sentences to be used as experimental stimuli and told not to practice them reiterantly.

2.4 *Stimuli*

2.4.1 *English*

In order to maintain commensurability with the Kelso et al. (1985) study, the same two sentences from the Rainbow Passage (Fairbanks, 1960), the same stress assignment, and the same exclusion of sentence initial and phrase final syllables were used here. The one notable exception to this was that speakers in this study occasionally paused (shown below by vertical hash-marks) between the fourth and fifth syllables of sentence 1 (...sunlight ‖ strikes...). Therefore, the fourth syllable, which in these cases becomes phrase-final, was excluded from analysis. In the sentences below, excluded syllables are within parentheses. Syllables marked for stress are underlined; all other syllables are treated as unstressed.

1. (When) the sun(light) ‖ strikes raindrops in the (air), ‖ they act like a pris(m) ‖ and form a rain(bow).
2. (There is), ‖ according to leg(end), ‖ a boiling pot of (gold), ‖ at one (end).
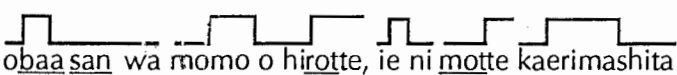
2.4.2 *French*

Two sentences were chosen from the "Maximes" of LaRochefoucauld, 'written' in the mid to late seventeenth century. Although not all speakers were acquainted with these particular maxims, they were familiar with the genre. The two sentences were chosen because of their length and structure. The second sentence in particular, with its embedded relative clause, is grossly similar to English Sentence 2. Speakers accepted both sentences as non-archaic.

1. La ferocité naturelle ‖ fait moins de cruel ‖ que l'amour propre.
2. L'interêt, ‖ qui aveugle les uns, ‖ fait la lumière des autres.

Per convention (e.g., Delattre, 1966; Selkirk, 1978; Anderson, 1982), stress was assigned to all non-schwa (mute "e") word-final (in trisyllablic words) and phrase-final syllables. Generally, the final syllables of *aveugle* (elided with *les*), *propre* (unreleased), and *autres* (unreleased) were not mimicked in the reiterant productions. Finally, following the lead of Vaissiere (1983), who cites an increasing tendency for speakers to place stress phrase-initially due to the influence of the French telecommunications media, all phrase-initial syllables were treated as stressed. In the two sentences above, stressed syllables are underlined and phrase boundaries are marked by double vertical strokes.

2.4.3 *Japanese*

The two sentences used here were taken from a Japanese folk tale — the "Momotaroo" — well-known to all five speakers. They were chosen to be roughly of the same length (syllable count) as the two English sentences and to include several types of multimora syllable (underlined). The sentences are presented below as transcribed and marked (superimposed lines) for high and low tone by speaker NK.

1. obaa san wa kawa ni sentaku ni dekake mashita

2. obaa san wa momo o hirotte, ie ni motte kaerimashita

In the Tokyo dialect, accented morae such as the /wa/ of *kawa* are denoted by a High-Low tone fall, occurring once within the accentual phrase (delimited by breaks in the tone level marking). In this study, tone level rather than accent is treated as the relevant binary variable roughly analogous to the stress distinctions used for French and English.

2.5 *Experiment Protocol*

To be sure that talkers had memorized the two sentences and that they could produce them at two distinct rates, the experimental session began with five normal recitations of each sentence at each speaking rate. They were then instructed that for the remainder of the experiment they would produce normal-reiterant utterance pairs — that is, a normal recitation immediately followed by its reiterant rendition — for the specified sentence (1 or 2), speaking rate (conversational or fast), and syllable identity (*ba* or *ma*). A balanced design was used to elicit ten normal-reiterant utterance pairs for each condition. This resulted in 80 reiterant utterances (= 10 repetitions x 2 sentences x 2 speaking rates x 2 syllable types) for later analysis. Including errors and technical adjustments, experimental sessions lasted approximately 45 minutes.

Of the 15 subjects in this study, ten (3 English, 3 French, 4 Japanese) succeeded in producing prosodically intact reiterant renditions of the target sentences. Although use of untrained speakers (14 of 15) resulted in fewer usable data than trained speakers would have provided, it demonstrated the "all or nothing" character of speakers' ability to produce reiterant speech. No practice effect was observed other than the commonly observed tendency for speakers to produce utterances faster as the experiment progressed.

One of the Japanese speakers (SM) often had trouble producing all phrases of an utterance correctly, especially when *ma* was the reiterant syllable. His data were analyzed only when he successfully mimicked two or more phrases within a given sentence rendition. A fifth

Japanese speaker's data (MY) are included in the analysis of overall kinematic patterning, but not in the condition-specific analysis of accentual tone and speaking rate. Although this speaker did not produce the correct number of reiterant syllables to match the original utterances (by either mora or syllable count; for details, see Vatikiotis-Bateson, 1988), her productions show the same language-specific, overall spatiotemporal characteristics as those of the other Japanese speakers.

2.6 Signal Recording And Conditioning

Vertical and horizontal movement of the lips and jaw were tracked midsagittally using a modified Selspot (Huntspot) system and recorded simultaneously with the acoustic output onto FM tape. A small infrared LED was placed on the vermilion border of each lip and on the chin. A fourth LED was placed on the bridge of the nose as a reference for head movement. The analog movement and audio signals were digitized at 200 Hz and 10 KHz, respectively. Then, the movement signals were numerically corrected for vertical head movement, low-pass filtered at 40 Hz, and differentiated to obtain instantaneous velocity. [A detailed description of the processing sequence and hardware is given in Kay et al., 1985].

2.7 Kinematic Analysis

Kinematic measures were made from the vertical movements of the lower lip LED, which contains the jaw's contribution to lower lip movement as well as that of the lip alone. Figure 1 shows the vertical change of lower lip position over time (middle trace), the instantaneous velocity (top trace), and the audio waveform, for a portion of a reiterant production using *ba*.

The continuous motion of the lower lip-jaw complex was divided into successive opening and closing gestures defined as lowering from peak consonant closure position to peak vowel opening and raising from peak vowel opening position to peak consonant closure, respectively. Measures of displacement, duration, and peak velocity of motion were obtained for each gesture (approximately 3000 per speaker) by means of an automated procedure that marks the position and time of waveform peaks and valleys. For position, the peaks and valleys correspond to the points of maximum consonant closure and vowel opening, respectively.

From these the displacement and duration of opening (peak-to-valley) and closing (valley-to-peak) gestures were calculated. In the velocity trace, valleys (labeled Vp_o) denote the
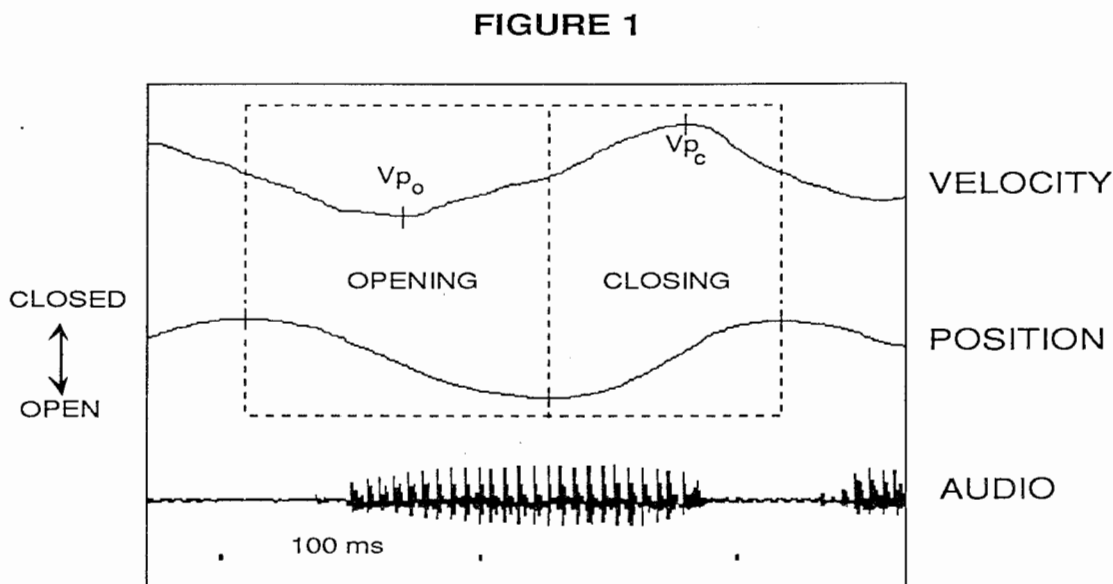
FIGURE 1

Figure 1. Time series representation of position, instantaneous velocity, and audio. Movements are divided into opening and closing gestures at peaks and valleys of position trace. Displacement and duration are computed between successive peaks and valleys. Each gesture has an associated peak velocity (Vp_o or Vp_c).

maximum, or peak, instantaneous velocity achieved during the opening gesture, and peaks (labeled Vp_c) the peak velocity for closing. A gesture was excluded when multiple peaks in the velocity profile made it difficult to define peak velocity for that gesture (less than 1% of the nonfinal gestures).

3 Results and discussion

In what follows, the kinematic variables associated with the data for speakers of the three languages are analyzed with respect to the cyclicity and continuity of motion, two relations among the three kinematic variables, and the kinematic correlates of language-specific prosodic variables. For each speaker, Analysis of Variance (ANOVA) was used to identify the correlations between individual kinematic variables and syllable identity (*ba vs. ma*), speaking rate (conversational *vs.* fast), gesture type (opening *vs.* closing), and the language-specific prosodic variable (stress for French and English and accent-related tone for Japanese). Linear regression was used to assess the relations between displacement and duration and between peak velocity and displacement. Other statistical analyses are described as they arise in the text.

3.1 *Cyclicity of motion*

Figure 2 shows times series representations of lower lip-jaw vertical position, instantaneous velocity, and associated speech acoustics for a representative reiterant production for a speaker of each language. In general, use of reiterant *ba* and *ma* resulted in rhythmic and continuous motion of the lower lip-jaw within a phrase. Opening movement gestures were usually longer in duration than closing ones. Thus, the movement cycle, defined here as the peak-to-peak interval containing a non-final opening gesture and the following closing gesture, was not durationally symmetrical. Differences between *ba* and *ma* productions were primarily ones of magnitude rather than kinematic patterning. Within a movement cycle, there was no break or flattening of the valley between an opening (lowering) gesture and the closing (raising) gesture immediately following it. This was also true of the peaks of motion, with one exception in Japanese.

3.1.1 *Moraic consonants in Japanese: An exception to cyclicity*

Speaker NK, shown in Figure 2, and two other Japanese speakers produced monosyllabic reiterant copies of multimora gesture sequences containing a geminate stop (e.g., /tt/ of *motte*) or heterosyllabic nasal+consonant cluster (e.g., /nw/ of *obaasan wa*, /nt/ of *sentaku*) with a geminate /b/ or /m/ in which lip closure was held for some period of time. Here, there was a break in the otherwise continuous alternation of opening and closing gestures. Because there was a sustained period during which there was no motion, such a plateau constitutes a portion of the speech production that cannot be accounted for within the simple dynamical scheme discussed here. It is an exception to the claim made originally by Kelso et al. (1985) that time need not be a controlled variable, since duration cannot be recovered from the relation between peak velocity and displacement when there is no motion.

There may be further problems for this scheme in that the stable period of silence during closure for the consonant cluster can be achieved via different coordinative configurations among the laryngeal and supralaryngeal articulators (see Vatikiotis-Bateson, 1988, Section 3.3.5.2). However, these types are the lone exception in this corpus. Gestures for all other mora types in the Japanese data and for all syllable types in the other two languages showed continuously cyclic opening and closing with no holds either at vowel minima or consonant maxima. This is true despite the widely different prosodic types they exemplified: the different types of stress in English and French, the phonemically long-voweled syllables in Japanese, etc. It is these other, regular gestures that are examined here. The exceptional moraic consonant gestures and their implications for the modeling discussed are left for a detailed follow-up study of glottal-oral coordination in Japanese (Vatikiotis-Bateson, in preparation).

3.1.2 *Differences between opening and closing gestures*

While motion was largely continuous, its cyclicity was not symmetrical; closing gestures were generally shorter in duration than opening ones. As shown in Table I, the results for English reproduce those reported by Kelso et al. (1985), who found temporal differences

FIGURE 2
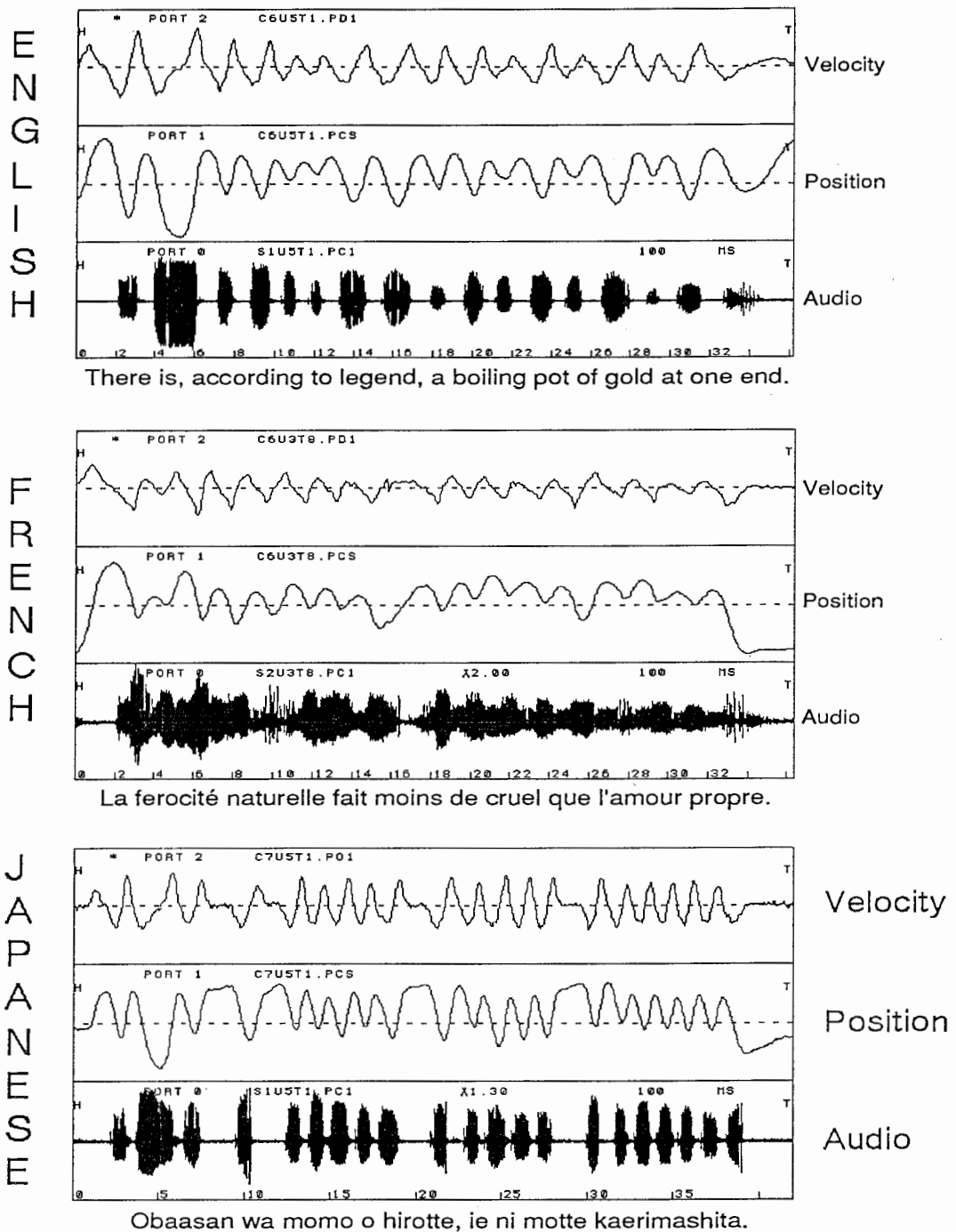
Figure 2. Position, instantaneous velocity, and audio traces for a representative reiterant sentence production by a speaker of each language — English: Sentence 2, Speaker RH, *ba*; French: Sentence 1, Speaker CG, *ma*; Japanese: Sentence 2, Speaker NK, *ba*.

between opening and closing gestures. Durations were consistently shorter for closing than opening gestures; accordingly, closing peak velocities generally were observed to be higher than opening ones. The one exception to this was speaker JK's *ma* productions in which there was no durational difference between opening and closing gestures overall due to a stress interaction. That is, her stressed opening gestures were shorter in duration and had smaller displacements than stressed closing gestures and, for that matter, stressed opening *ba* gestures (see Vatikiotis-Bateson, 1988, Table 2-2c).

Table I also shows that there is a similar durational asymmetry for the French and Japanese data. Furthermore, for all three languages, *ba* productions were consistently more durationally asymmetrical than *ma* productions. Within a language, no tendency was observed for speakers having slower absolute rates to produce relatively larger temporal asymmetries. Nor did within-speaker changes of speaking rate have systematic effects on the gestural asymmetry, as shown by the absence of interactions between speaking rate and gesture type for speakers of English and French, and by the inconsistency of the interaction among speakers of Japanese (Vatikiotis-Bateson, 1988). Thus, durational asymmetry was roughly the same for the three languages in which opening gestures were consistently longer than closing gestures.

3.1.3 *Effect of initial consonant*

As discussed in the previous section, the syllable-internal duration pattern tended to differ in that opening and closing gesture durations were slightly more symmetrical for *ma* than *ba*. There was also a small, but consistent, magnitude difference between *ba* and *ma* for all speakers in most conditions; all three kinematic measures were slightly larger for *ma* productions. There were no other systematic interactions between syllable identity and either the overall patterning of the kinematics or the condition-specific kinematic correlates of prosodic and speaking rate distinctions. However, if analyzed with greater scrutiny, a correlation might be found between the observed differences in durational symmetry and kinematic magnitude and possible differences in overall patterning (see Section 3.2.2, below). Therefore, it is with caution that, in the current presentation, the results for the two syllable types are usually treated interchangeably.

3.1.4 *Kinematic correlates of stress, tone, and speaking rate*

As shown in Table II, all speakers displayed a tendency for mean values of the three kinematic variables to covary across changes of speaking rate and the linguistic variable. Kinematic means for movement gestures produced at faster speaking rates were smaller than those produced at the slower, conversational rates. Stressed gestures in English and French were larger in displacement, duration, and peak velocity than unstressed ones. For the four Japanese speakers whose data could be analyzed for the prosodic variable, there were consistent effects of accent-related tone differences on the mean kinematics: high-tone gestures had smaller mean kinematics than low tone gestures. Note that the reiterant copies of the multimora /baa/ of *obaasan* produced by speakers NK and FE as a single long syllable are not included in this analysis, because there is a high-low fall in tone within the syllable.

We return to these condition-specific effects and their interrelations after first examining the overall relations among the three kinematic parameters in the three languages.

3.2 *Overall patterning of gestural kinematics*

To assess the overall kinematic patterning, the same data set was used as given in Table I. That is, data were analyzed for all five Japanese speakers and the analysis included the larger gestures corresponding to the underlying /baa/ of *obaasan* produced by speakers NK and FE.

3.2.1 *The relation between displacement and duration*

A limitation of examining measures of displacement (a spatial measure) and duration (a temporal measure) individually — as in Table I and in previous production studies (e.g., Lindblom, 1963; Sussman et al., 1973) — is that the inevitable variance in each measure remains an unexplained residue. Examination of the relations between the spatial and temporal components of an articulatory event might provide an account for the variance found in the individual kinematics.

The distance-time (d-t) relation provides the basic space-time view of the gestural data and a measure of the average speed ($V_{av} = d/t$) of articulator movements. A possible relation between displacement and duration is that they covary in a positive, linear fashion such that V_{av} is conserved; that is, as duration increases so too does displacement. This is more or less what has been observed for English (e.g., Nelson, 1983; Kelso et al., 1985; cf. Gay, 1981). However, it is possible that languages such as French and Japanese, whose temporal organizations are perceived to be much more regular, might show less tendency for displacement and duration to covary because of their relatively small (compared to English)

FIGURE 3

Figure 3. Scatterplots show the overall regression of gestural displacement (ordinate) on duration (abscissa) for the opening (left) and closing (right) gestures associated with *ma* productions for one speaker of each language.

durational variability. Both of these possibilities were examined in the present study.

Figure 3 contains scatterplots of opening gestures (left) and closing gestures (right) for the *ma* productions for representative speakers of English (top), French (middle), and Japanese (bottom). Table III shows coefficients and slopes of the linear regression of displacement on duration for these plots and all other like groups of data. Three observations can be made from the figure and table. First, there was a positive covariation between displacement and duration, whose linear component accounted for anywhere between 3% and 81% of the variability (French speaker DP's *ma* closing gestures and Japanese speaker Nk's *ba* opening gestures, respectively). In general, the linear regression accounted for more of the variance for opening than closing gestures regardless of language. Second, the linear component of the regression

was, on average, substantially higher for English than for French or Japanese speakers. Indeed, among the French and Japanese speakers, the linear component accounted for more than 50% of the variability for the data of only one speaker (Japanese NK, discussed below); in most of the other cases, it accounted for less than 25%. Third, even though the covariation was positive and fast speaking rate gestures often had steeper d-t slopes (implying higher average velocities) than durationally longer gestures produced at conversational rates, the slope values did not correspond well to the mean kinematics. When the fit of the linear regression was poor, the d-t slopes tended to underestimate average velocity. On the other hand, in cases where the linear component accounted for at least half of the variability, such as Japanese speaker NKs data and opening gestures for English, the slope values tended to overestimate observed average velocity (see Vatikiotis-Bateson, 1988, pp. 25-26 for further discussion).

The covariation between displacement and duration fits the expectation that movements take longer to go farther, but, because of the substantial variation about the regression line, the relation is highly variable and usually accounts for less than half of the overall variance. Indeed, for French and Japanese closing gestures, where durational variance was the smallest, the regression of displacement on duration, though reliable, typically accounts for a small percentage of the variability. It is unlikely, then, that conservation of average velocity played a role in producing the movement gestures considered in this study.

3.2.2 *The relation between displacement and peak velocity*

The relation between peak velocity and displacement (Vp-d) is of interest because it allows us to model observable movement behavior using second-order dynamics. If we restrict such modeling to the simple undamped linear mass-spring, then spring stiffness, **k**, is proportional to the slope of the Vp-d relation, whose units are temporal. Specifically, in such a system, **k** is proportional to ω_o^2 and $Vp = \omega_o A$, where ω_o is angular frequency and A is half the peak-to-valley displacement. Therefore, within this framework, movement duration may be recovered from the slope of the Vp-d relation.

In order for such modeling to be applied to speech production several predictions should hold: First and foremost, the linear component of the covariation between displacement and peak velocity should account for a substantial portion of the overall variability. Next, since stiffness varies inversely with duration, differences in mean duration should correspond to differences in the slope of the Vp-d relation (from which stiffness is inferred). Finally, since a perfectly linear covariation between peak velocity and displacement would imply isochronous movement behavior, some correspondence should be observed between temporal variability and the linear covariation of peak velocity and displacement. The extent to which these criteria are met by the data is illustrated below.

Figure 4 contains scatterplots of the covariation between peak velocity and displacement for the opening and closing *ma* gestures for one speaker of each language. Comparison of Figures 3 and 4 and Tables III and IV shows that, for nine of eleven speakers, the linear covariation of displacement and peak velocity accounted for substantially more of the overall kinematic variability (as much as 90%) than did the covariation of displacement and movement duration. This trend is observed for opening and closing gestures regardless of differences in syllable identity or language-specific differences in temporal organization and absolute speaking rate. The strength and universality of the linear covariation between displacement and peak velocity adequately satisfies the first criterion stated above.

One of the two exceptions to this was English speaker JK. Although her regression coefficients for the Vp-d relation tended to be higher than for the d-t relation, they were somewhat lower than average while the d-t regressions were somewhat higher than average. Her gestural durations were substantially longer than all other speakers and highly variable, which supports the prediction that the degree of linear covariation reflects temporal variability. The durational difference between opening and closing gestures was small for *ba* productions and nonexistent for *ma*. Closer inspection of her mean kinematics showed that stressed *ma* closing gestures were actually longer in duration (see Table VII) and lower in peak velocity (see Vatikiotis-Bateson, 1988, Table 2-2c) than opening stressed gestures. This anomaly accounts for both the lack of duration difference between opening and closing gestures and the weak regression coefficient for closing gestures.
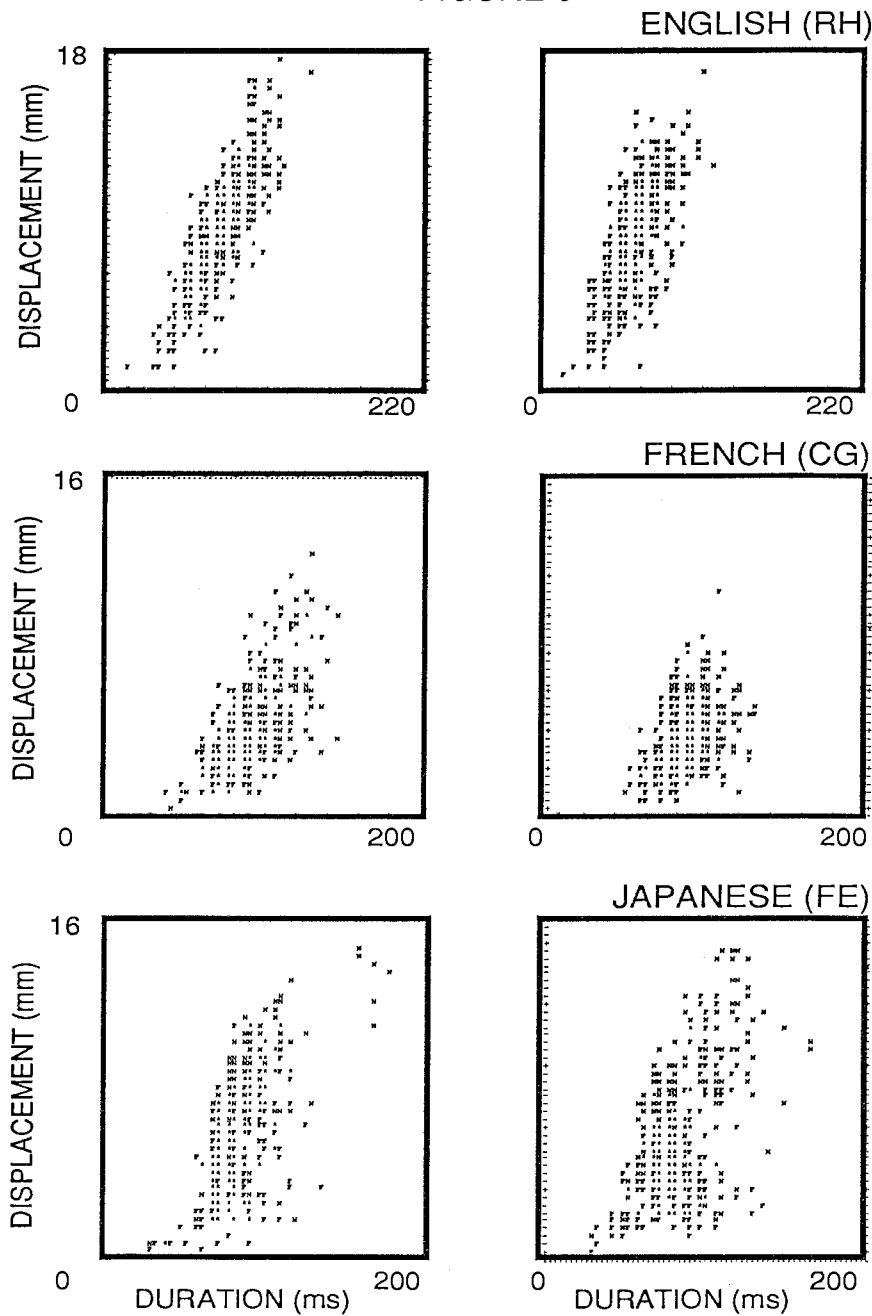
FIGURE 4
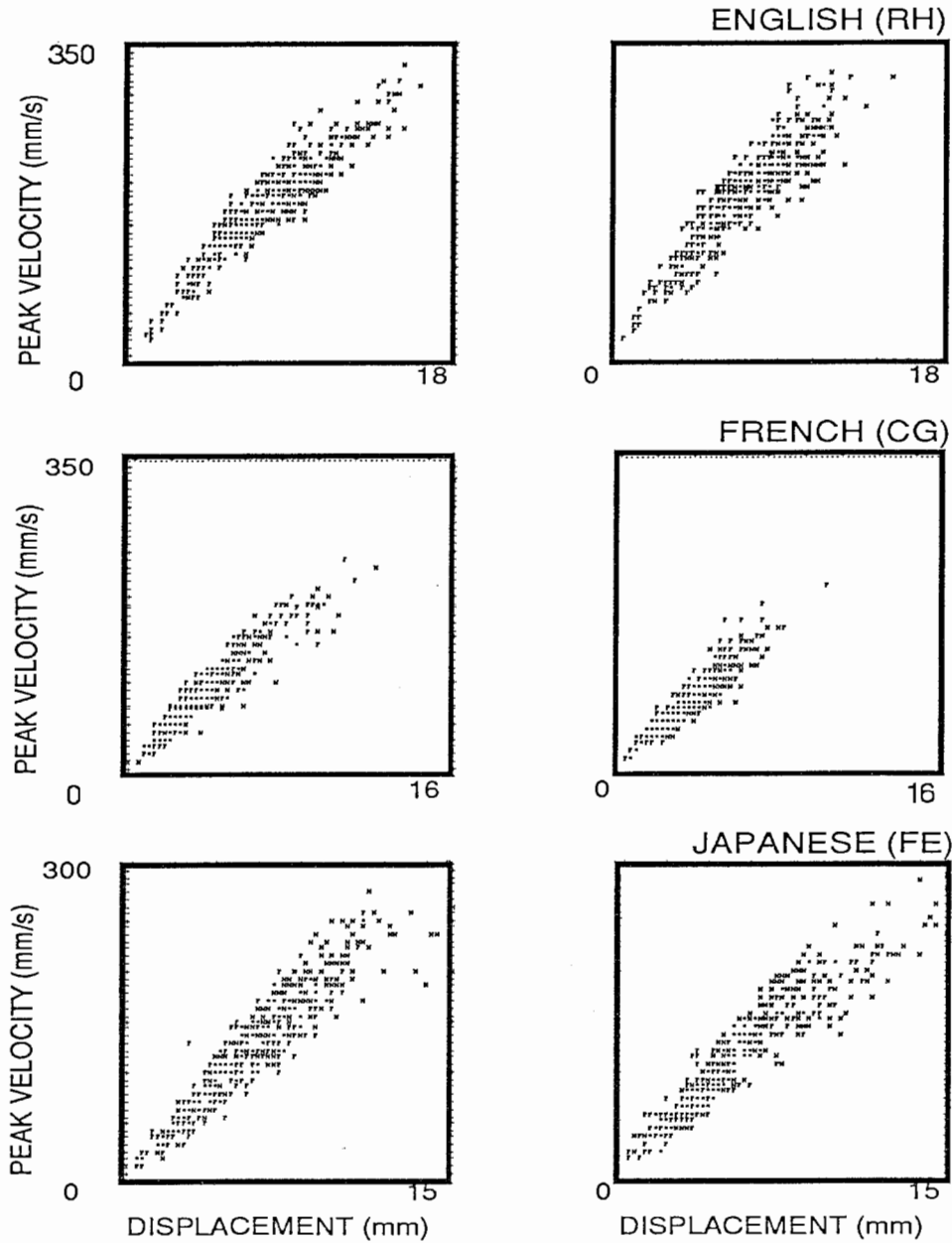
Figure 4. Scatterplots show the overall regression of gestural peak velocity (ordinate) on displacement (abscissa) for the opening (left) and closing (right) gestures associated with *ma* productions for one speaker of each language.

The other exceptional speaker was Japanese NK. Her d-t regression coefficients for opening gestures, which accounted for the most variability (80%) of any speaker, were actually higher than her Vp-d coefficients (67% of variability). However, closing gesture coefficients were about the same for *ma* productions and Vp-d coefficients were higher for *ba*. We cannot explain why this speaker's data patterned so differently from those of the others, but her training in speech and voice as well as her prior experience with the experimental task may have been factors. What is particularly interesting about her data is that, unlike speaker JK's highly variable and perhaps anomalous data, they may indicate a more elaborate production strategy combining adherence to a stiffness constraint and stylized control of amplitude-specific duration.

Table IV also shows that mean duration for opening gestures was longer than for closing in all but one case (English JK *ma*). In five other instances, the durational difference was small

(3 ms or less) and statistically unreliable for three of them (French DP *ba* and CG *ma* and Japanese FE *ma*). By the prediction that the slope of the Vp-d relation will vary inversely with mean duration, the slopes for closing gestures should be the same as or steeper than those for opening gestures depending on whether or not there is a reliable difference in duration. Slope differences were tested and showed this to be true for most speakers' (nine of eleven) *ba* productions.[4]

However, the prediction was not met for six speakers' *ma* productions. Because the difference in duration between opening and closing gestures was often smaller for *ma* than *ba* productions, we expected fewer reliable slope differences. Instead, there was a tendency for the slopes of opening and closing gestures to be the same when the durational difference was small and for the opening slope to be steeper than closing when there was no durational difference. This was true for four of the six violations for *ma* (and French speaker CG's *ba* productions). For example, Japanese speaker SM showed no slope effect and a weak durational difference; for Japanese speaker FE, there was no durational difference but the slope was steeper for opening gestures. There were no examples of the reverse tendency; namely, for closing slopes to be steeper when there was no durational difference.

It is not clear to what extent the difference in results for *ba* and *ma* constitute a difference in patterning. Although it is beyond the scope of the current study to provide an answer, the difference appears to be language-independent and deserves closer examination in the future. One possibility is that the results reflect speaker-specific differences in performing the reiterant speech task with *ma*. Several speakers (e.g., English speaker JK) reported that production of reiterant *ma* was more difficult. Another possibility is that there really are differences in the production of *ma* and *ba*, which might be revealed by a more thorough examination of the kinematics and underlying physiological events.

It is also possible that the observed difference has nothing to do with syllable identity; rather,the simple comparison of opening and closing gestures done in this study may run afoul of anatomical and physiological differences associated with their production. That is, much more force can be generated during closing movements than during opening. Appropriate to chewing, the jaw closing muscles (e.g., masseter and pteragoid) are massive compared to those of opening (e.g., anterior belly of the digastric). While the covariation of peak velocity and displacement was highly linear for both types of gesture, the basic motion equation may be parametrized differently to accommodate the structural difference of opening and closing gestures. The effect of this difference could be to change the relation between Vp-d slope and mean duration so that the Vp-d slope is slightly steeper for opening gestures than for that of closing gestures having the same duration. This difference would not be visible when durational differences are large, as in the case of *ba*.

Whatever the true cause of the difference between *ba* and *ma* productions, we can at least remove the possible effect of structural differences between opening and closing gestures. As shown in Table V, Vp-d slope consistently scaled inversely with mean duration when speaking rate conditions were compared for each gesture type. Speaking rate differences were clearly distinguished in the mean kinematics for 10 of 11 speakers. One Japanese speaker, SM, produced all of his gestures at a relatively fast rate, although listeners heard a clear distinction between the two instructed speaking rates. Appropriately there was no Vp-d slope difference for

---

[4]Slopes of linear regressions may be compared using a test for parallelism, which results in t-values, conservatively adjusted for different N's (Cohen & Cohen, 1975). We are indebted to Randy Flanagan and David Ostry for passing on the following formula.

$$t = \frac{m_2 - m_1}{\sqrt{\left(\frac{(N_1 - 2)(RMS_1) + (N_2 - 2)(RMS_2)}{(N_1 + N_2 - 4)}\right)\left(\left(\frac{1}{(N_1 - 1)(VAR_1)}\right) + \left(\frac{1}{(N_2 - 1)(VAR_2)}\right)\right)}}$$

This test is used for slope comparisons throughout this study. Since specific predictions are made concerning the direction of the difference, a one-tailed test is appropriate (Ferguson, 1981). Although one-tailed tests make it easier to obtain reliable differences in favor of the claims being made, reliable differences contrary to those claims are also more easily obtained. Given the large N's involved, difference are considered reliable at the 5% level if $t = 1.645$.

his productions, while Vp-d slope was reliably steeper for the fast rate conditions of all other speakers.

There is an important difference between the speaking rate comparison and the gesture type comparison. For the gesture type comparison, there was little or no difference in mean displacement between opening and closing gestures. However, faster rate gestures were substantially smaller in displacement than slower, conversational rate gestures. Thus, there was an inverse scaling between mean displacement and Vp-d slope. This should not be surprising given the positive covariation of displacement and duration and is discussed in more detail in Section 3.3.3.

The generality of the inverse scaling between Vp-d slope and duration can be seen in Figure 5, where slopes are plotted against means for the four conditions of gesture type and syllable identity for all speakers. The figure shows quite clearly the tendency for the productions of English speakers to have smaller Vp-d slope values and longer mean gesture duration than French and Japanese, and suggests that the data of all speakers may be described by a single, probably nonlinear function.
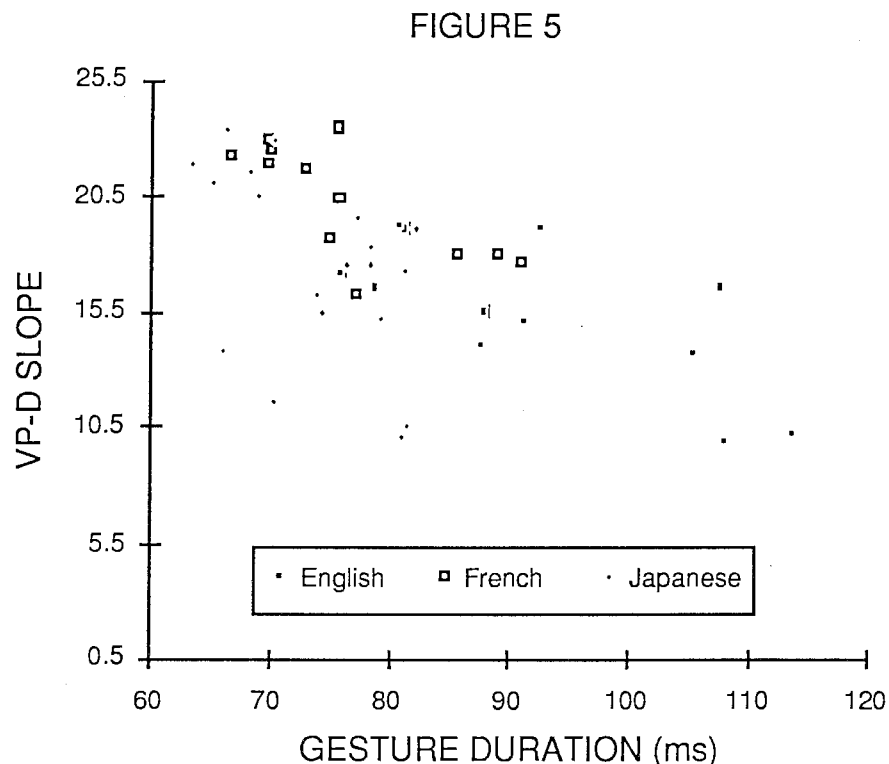
FIGURE 5

Figure 5. Slope (m) of the Vp-d relation is plotted against mean gestural duration (in ms) as a function of gesture type (opening vs. closing) and syllable identity (ba vs. ma) for all speakers' data. The best-fit line (second-order polynomial) shows the cross-language inverse scaling of Vp-d slope and mean duration.

Finally, we consider the third prediction that there is an inverse relation between the duration variability and the linear covariation of peak velocity and displacement. Unfortunately, the within-speaker results gave no indication that this prediction was met. However, a casual test of the prediction was made by again looking across the data for all speakers and the two conditions of gesture type and syllable identity (Figure 6). The data for French and English occupied the two extremes of the distribution, with the Japanese data more broadly distributed, but closer to the French than to the English. While the majority of r-values were greater than .90, there was an overall tendency for r-values to decrease as durational variability increased and,contrary to what might be inferred from the figure, this was not due solely to the three extreme values in the English set.

Thus, we conclude that both the spatial and temporal aspects of the overall movement behavior can be adequately characterized by the relation between peak velocity and displacement interpreted in terms of a simple second-order dynamical system. In addition to

the model's universal applicability to the data of all three languages, there are appropriate language-specific differences in speaking rate and temporal variability which could be controlled by setting the underlying stiffness parameter.

FIGURE 6
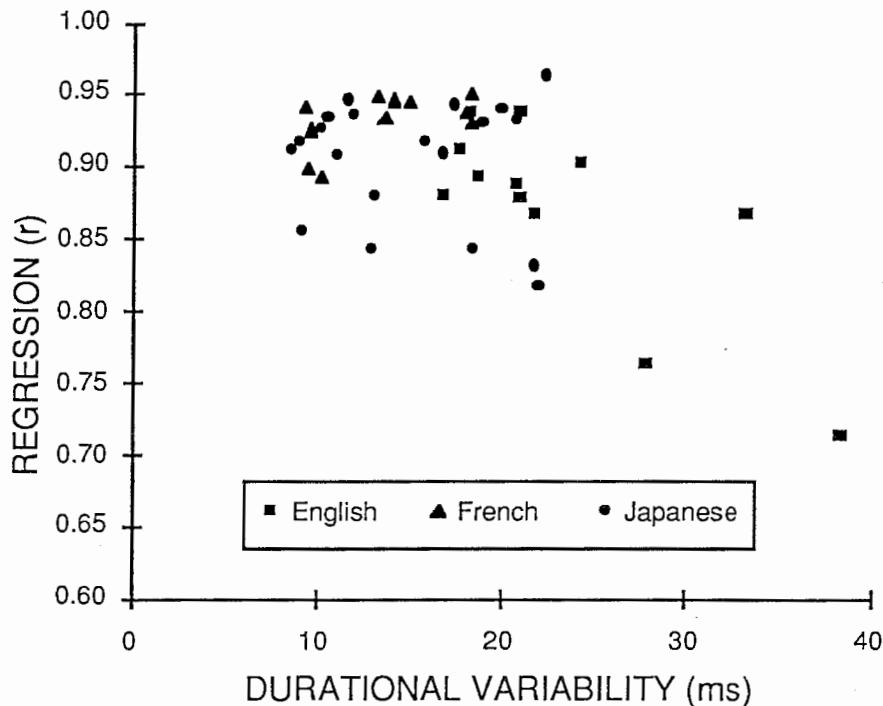
Figure 6. Linear regression coefficients (r) are plotted against durational variability (in ms) as a function of gesture type (opening *vs.* closing) and syllable identity (*ba vs. ma*) for all speakers' data. The inverse scaling denoted by the best-fit line (first-order polynomial) is not due to the extreme values for the English data (lower right).

3.3 *Kinematic analysis of language-specific variables*

Having found the relation between peak velocity and displacement to be useful in characterizing and distinguishing the overall movement behavior of three languages, we examine now whether the Vp-d relation can be used to characterize language-specific distinctions, such as stress in English and French and accent-related tone and mora complexity in Japanese.

3.3.1 *Stress in English and French*

Before examining the kinematic correlates of stress distinctions in French and English, it must be emphasized that we use the term "stress" rather loosely. In assigning stress, our aim was to introduce a prosodic dichotomy, following reasonable conventions for each language (Section 2.4). The resulting distinction more closely corresponds to prominence than stress, which is etiologically and functionally quite different for the two languages (De Groot, 1923; Grammont, 1933; Wenk & Wioland, 1982).

As noted earlier and shown in Table II, stress distinctions in both languages showed consistent correlates in the lip-jaw kinematics. Stressed gestures were larger in displacement, longer in duration (except French speaker BA's closing gestures), and higher in peak velocity than unstressed gestures. The biggest difference in behavior of the individual kinematics between the two languages was in the absence of stress effect on the duration of closing gestures in French (see Table VII); the duration of closing gestures in French was the same regardless of stress and its effect on displacement and peak velocity.

The question we consider now is whether the Vp-d relation can be used to uncover dynamic correlates of condition-specific distinctions in stress. Linear regressions of peak velocity against displacement were computed for each stress condition as a function of syllable identity, gesture type, and speaking rate. These are shown in Table VI. Similar to the findings of

—14—

Kelso et al. (1985) for English, the condition-specific regressions had a very strong linear component for both stressed and unstressed gestures, accounting for most of the condition-specific spatiotemporal variability. As shown in Table VI, the regression coefficients were even larger for the French data.

In Figure 7, best-fit regression lines for each stress-rate (and tone-rate) condition are plotted as a function of gesture type and syllable identity for a representative speaker of each language. Focussing on the English and French data, two things can be seen in the figure. First, while there was quite a bit of overlap in the distribution of the data, stressed and unstressed gestures occupied distinct regions of the overall distribution. Second, the Vp-d relation tended to be steeper for smaller unstressed gestures than larger stressed gestures. Although speakers did not show such dramatic results for every gesture-syllable condition, they consistently differentiated Vp-d slopes for at least one gesture type and usually in the direction predicted by the hypothe-sized relation between Vp-d slope and gestural duration. That is, Vp-d slopes tended to be steeper for that condition having the shorter mean duration (see Table VII).

Thus, two languages that differ in temporal organization and in the etiology and function of a prosodic distinction, which we are calling stress, showed the same patterns of behavior in and among the kinematic variables of the lip-jaw complex. In particular, the highly linear condition-specific covariation of peak velocity and displacement accounted for the bulk of the within-condition spatiotemporal variability and showed that, despite variability in the absolute kinematic magnitudes, the relation between variables is stable within stress conditions. The potential universality of this phenomenon is further demonstrated in the next section for the data from Japanese, in which the observed prosodic distinction is not perceived as even re-motely similar to stress in French or English.
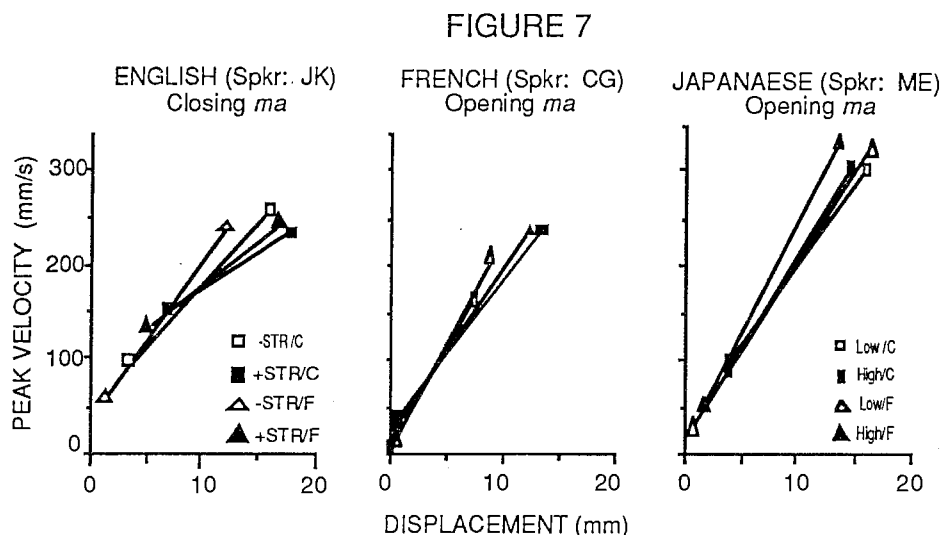
FIGURE 7

Figure 7. Representative data are shown for a speaker of each language. Regressions of peak velocity (ordinate) on displacement (abscissa) for the four stress-rate or tone-rate conditions are depicted by best-fit lines whose lengths denote the range of variation on displacement.

3.3.2 Accent related tone in Japanese

In this section, we consider the kinematic correlates of the high-low accent related tone distinction for the four Japanese speakers, whose data could be prosodically analyzed. The analyses reported here should not be confused with previous production studies of Japanese accent distinctions, which have focused primarily on measuring the temporal acoustics of minimal pairs differing only in accent — i.e., the presence or absence of a tone fall (e.g., Mitsuya and Sugito, 1978; Dalby and Port, 1981; Beckman, 1982). Generally, if there was a durational difference, the accented syllable was longer. In this study, accented syllables such as the /baa/ of obaasan were excluded from this analysis because of their tonal complexity.

As shown in Table II, high tone gestures were generally associated with smaller lip-jaw kinematics than low tone gestures, despite differences in speaking rate, gesture type, and syllable identity. The magnitude of the kinematic difference was not as pronounced for the

Japanese tone distinction as for French and English stress (Table II). In particular, the difference in mean displacement was smaller for the different tone conditions than for the different stress conditions of English and French. This resulted in a greater overlap in the distribution as shown in the rightmost panel of Figure 7. Despite this, the consistent patterning across speakers suggests that the phenomenon is real.

The fact that the individual kinematics revealed a supralaryngeal instantiation of a tone distinction previously assumed to be strictly laryngeal may not be surprising. When distinct pitch registers are produced, it is often observed that, in addition to greater tension of the vocal folds, high tones are produced with a concomitant raising of the larynx. This effectively shortens the vocal tract and raises formant frequencies. Since formant height is proportional to jaw position in open tube vowels such as /a/ — the lower the jaw the higher the formant values — it is possible that speakers may try to counteract the raising of formant values. Japanese speakers, then, could reduce or perhaps eliminate the difference in formant frequencies caused by the tone level distinction by not lowering the jaw so far during production of high tone vowels.[5]

Table VIII lists the results of the linear regressions of peak velocity on displacement for each tone-rate condition. As the table shows, r-values for the condition-specific regressions were generally large. The major exception is speaker NK whose closing *ma* gestures uniformly accounted for only 40-45% of the variability.

Table IX gives slope comparisons for the different tone level conditions as a function of speaker, speaking rate, gesture type, and syllable identity. As with the data for French and English, slope differences were not found for every comparison in which we might expect them; but, where slope differences were reliable, they tended to be steeper for the condition having shorter mean duration (Figure 7, bottom).

3.3.3 *CVV productions in Japanese*

The data for two Japanese speakers, NK and FE, provide yet another opportunity to test the ubiquity of the patterning among kinematic variables. Unlike the other speakers, these two produced long, monosyllabic reiterant copies of the two-mora /baa/ in *obaasan*. These gestures were substantially longer in duration, larger in displacement, and higher in peak velocity than one-mora gestures, as shown in Table X. Thus, they occupied a distinct region of the spatiotemporal distribution. Furthermore, the linear regression coefficients for the Vp-d relation were quite large, especially for speaker FE, despite the restricted range of the distribution and the small number of CVV gestures (one per utterance). Finally, the slope of the Vp-d relation was appropriately steeper for one-mora gestures which were shorter in duration than two-mora gestures.

This subset of the data provides an extreme demonstration of the inverse scaling between Vp-d slope and mean displacement that was observed in the comparisons of speaking rates and language-specific prosodic conditions. That is, condition-specific increases in Vp-d slope were associated with decreases in mean displacement as well as duration. In terms of the second-order system considered here, changes in displacement could correspond to changes in equilibrium position, x_0. Concomitant changes in stiffness and equilibrium position could account for the observed tendency for movement gestures to take longer to go farther.

4. General discussion

We have examined the kinematics associated with motion of a primary articulator complex for reiterant productions by speakers of three languages differing in their temporal organization. The same basic results were observed regardless of language-specific differences in absolute speaking rate and prosodic contrast, or instructed differences in syllable identity and speaking rate. This cross-language similarity suggests that linguistically relevant movement behavior, for all its apparent diversity, is realized within fairly narrow limits.

The main goal of this study was to determine whether these limits could be characterized in terms of a system having the properties of a simple oscillator, which is controlled by specification of stiffness and equilibrium position. To this end, the most important results are those concerning the positive covariation among kinematic variables. Movement gestures that were consistently larger in displacement were almost always found to be longer in duration and

---

[5]We are grateful to Arthur Abramson for helpful discussion of this issue.

higher in peak velocity. In addition to the covariation of kinematic means for speaking rate and stress or tone-level conditions, kinematic measures covaried on a gesture-by-gesture basis. Positive correlation among the kinematic variables was demonstrated across the full range of movement gestures through observation of both the highly linear covariation of peak velocity and displacement and a tendency for movement duration and displacement to covary.

4.1 *The relation between peak velocity and displacement*
The relation between peak velocity and displacement demonstrates the stability of spatiotemporal patterning within and across the three languages. Moreover, changes in the slope of the relation correspond to differences in movement duration and are consistent with the idea that orofacial motions are produced in part by specifying stiffness. The kinematic data meet three criteria for inferring stiffness: First, the linear component of the regression of peak velocity and displacement accounts for most of the variability and obtains at all levels of observation. Second, regression slopes, indicative of average stiffness, vary inversely with mean duration. This is particularly clear in capturing global differences in speaking rate and, to a lesser degree, the difference between opening and closing gestures. Also, the correspondence between duration and Vp-d slope is observed for language-specific distinctions — e.g, the distinction between one- and two- mora, open syllables in Japanese. Third, there is an overall tendency across speakers for regression coefficients to vary inversely with durational variability; as regression coefficients approach 1.0, indicative of uniform average stiffness, temporal variability decreases making movements more isochronous.

The finding that absolute speaking rate and temporal variability are correlated within and across languages suggests that movement behavior in all three languages may be constrained by a single continuous function. As commonly observed, mean duration and standard deviation are positively correlated and display fairly constant coefficients of variation. This phenomenon has been attributed to the statistical nature of Poisson distributions — non-normal distributions that characterize a wide range of durational measures in speech production (Crystal & House, 1986; also, see Vatikiotis-Bateson, 1988). According to this view, as speakers increase speaking rate, temporal variability will decrease along a continuum simply as a function of the shape of the temporal distribution of the data. This could be achieved through scaling of underlying stiffness.

In addition, the data suggest that different languages may occupy different regions of that continuum and coincide with our expectations about language-specific differences in temporal organization. When measured as the number of opening-closing gesture pairs per second, English speakers have the slowest absolute speaking rates, while French and Japanese speakers have the fastest. This dichotomy is consistent with the often proposed dichotomy in temporal organization that lumps syllable- and mora- timing together and distinguishes them from stress-timing (e.g., Pike, 1967). Furthermore, the durational values of this study are very similar to the acoustic duration measures reported by Dauer (1983) for these languages. Thus, differences in temporal organization could entail quite distinct settings of the stiffness parameter.

4.2 *The relation between displacement and duration*
The outputs of the simple oscillator model, which we have been considering, depend upon the specification of stiffness and equilibrium position. We have suggested that differences in stiffness and equilibrium position can be inferred from changes to the slope of the Vp-d relation and articulator displacement, respectively. We have found that the correspondence between duration and Vp-d slope extends to language-specific prosodic distinctions in stress or tone level, suggesting that such distinctions may be related to control of stiffness. However, prosodic distinctions are more consistently correlated with differences in mean displacement and, therefore, better correspond to the setting of a second inferred parameter, equilibrium position. For English, the relatively strong linear covariation between displacement and duration reflects a prosodic structure that is both temporal (i.e., longer stressed vs. shorter unstressed gesture durations) and spatial. Thus, English stress distinctions may be correlated with both stiffness and equilibrium position. For French, however, the covariation between displacement and duration is generally much weaker, especially for closing gestures in which stress effects are observed for displacement, but not necessarily for duration. In fact, unstressed closing gestures compensate somewhat for the durational effect of stress in opening

gestures for two of the three speakers, contributing to the overall stability of movement cycle duration. Thus, for French, equilibrium position may be the only parameter that consistently varies with stress. It is possible, then, that equilibrium position is the primary underlying parameter governing stress distinctions in both languages.

Depending on the contrast being compared, the Japanese data show parallels to both English and French. As observed for stress in English, heavy (two-mora) and light (one-mora) syllables are clearly distinguished by differences in both mean displacement and duration (consistent with differences in Vp-d slope). Similar to the French stress contrast, tone level is differentiated spatially regardless of durational differences, which tend to be small. Differential settings of equilibrium position could account for both distinctions.

It is possible, then, that prosodic distinctions and differences in speaking rate or gesture type are specified by different underlying parameters. For example, stiffness could be set for speaking rate once for each utterance while equilibrium position is modulated prosodically for each syllable. However, the positive covariation of displacement and duration and the inverse variation of mean displacement and slope of the Vp-d relation indicates that the two parameters are not independent in their kinematic realization.

This co-dependence of parameters could enhance the perceptual effectiveness of intentional variations in speaking rate or prosody across the range of different language structures, while adhering to fairly narrow, universal constraints on speech production. Although speculative, we think it is likely that perceptual constraints also condition how prosodic distinctions are modulated within the durational range appropriate to a particular temporal organization. It is not yet clear how small a temporal discrepancy can be heard or produced in sequence, since the phenomenon has been studied sporadically in a variety of contexts (e.g., Lehiste, 1977; Fujisaki, Nakamura, & Imoto, 1975; Morton, Marcus, & Frankish, 1976; Espinoza-Varas & Watson, 1986). Nor is it clear to what extent perception is conditioned or predisposed to durational contrasts — e.g., the demonstration by Bolton (1894) that English speakers will hear a stress-like duration alternation in isochronous sequences, provided the sequence is less than 4-5 Hz. However, it appears that temporal discrepancies must be at least 30-40 ms to be useful in rhythmic contrasts. Therefore, the duration differences between one- and two- mora syllables in Japanese and between stressed and unstressed syllables in English, which are about the same (35-40 ms), are large enough to be easily perceived. On the other hand, the overall temporal variability for one-mora gestures in Japanese is only 15-20 ms and, thus, unlikely to support a bimodal temporal distinction — e.g., between high and low tone syllables. This is consistent with the finding that non phrase-final stress contrasts in French are due primarily to differences in pitch and/or amplitude, rather than duration (Delattre 1966).

4.3 Determining the model system

The observed co-dependence between stiffness and equilibrium position has consequences for determining the second-order system that best describes the data. Clearly, it cannot be an undamped linear mass-spring in which movement time is independent of amplitude. In their analysis of English reiterant speech, Kelso et al. (1985; also, see Appendix A.1) tested the possibility that the system should be modeled as a single non-linear function, which happens to be largely linear for the range of values observed. They concluded that the data were better described as a composite of condition-specific linear functions than as part of a single nonlinear function. That is, linear stiffness would be set independently of equilibrium position according to the stress, speaking rate, and type (opening or closing) of each gesture. In view of the striking similarity in patterning observed across the wider range provided by data from three languages , this result seemed somewhat inelegant. Therefore, we tested the data of this study for nonlinearity using the method described in Appendix A.1. However, we were unable to find a stable nonlinear solution that accounts for as much variability in the data as the linear component of the overall covariation of peak velocity and displacement.

Another possible account for the inverse dependency observed between gestural displacement and Vp-d slope is to add a linear damping term to the linear mass-spring (see Appendix A.2). In particular, the system behaves as though it might be slightly underdamped, especially for the production of larger movement gestures — e.g., stressed in English and multimora productions in Japanese. Compared to the undamped system's behavior, the effect of damping would be to reduce peak velocity for a given displacement and to increase observed duration. Provided that damping increased with increases in movement extent, such a system could gen-

erate data consistent with the results of this study — i.e., slope of the Vp-d relation decreases and observed duration increases at larger displacements.

Modeling the data as a linearly damped linear system has the advantages of preserving the basic finding of this and the Kelso et al. study that stiffness is linear while capturing the overall tendency for movements to take longer to go farther. Also, this account is consistent with probable physiological factors, such as the increased viscosity of the mandibular joint and increased resistance to stretch in the muscles of the face and lip region, for larger displacements of the jaw and lips (e.g., McDevitt, 1989).

Although damping has been used previously to model speech kinematics, e.g., the examination of sequences of nonsense syllables (Masaki, Shirai, Imagawa, & Kiritani, 1985; Imagawa, Kiritani, Masaki, & Shirai, 1985; Flanagan, Ostry, & Feldman, 1990), the number of studies is small and the results are not always promising (Smith et al., 1991). It is possible that the apparent damping in these data is largely an artifact of transducing only the vertical component of lip-jaw motion. The combination of translational and rotational components of jaw motion (see Edwards, 1985; Edwards & Harris, 1990) could be biomechanically quite different at larger than at smaller displacements, such that the contribution of the rotational component increases with displacement (cf. Ostry, Flanagan, Feldman, & Munhall, 1992). Even though measured displacements are not very large (relative to possible non-speech displacements), a larger rotational component as displacement increases would cause the extent, but not necessarily the duration, of the true trajectory to be underestimated.

4.4 *Future directions*

There is always concern about the difference between reiterant speech and "natural" speech. The necessity of confining primary articulation to the lips and jaw certainly introduces a high degree of intra-articulator rhythmicity, not seen under normal adult conditions, and it probably slows the time course of the behavior (Liberman & Streeter, 1978). The durational asymmetry between opening and closing gestures is also probably a task-specific consequence. Its similarity to asymmetries observed for other repetitive activities and structures, e.g., chewing (Hiiemae and Crompton, 1985) and foot-tapping (Stetson, 1905), suggests inherent physiological and dispositional constraints on sequential behavior rather than anything peculiar to speech. So, it is not surprising to see similarities in the reiterant speech produced by speakers of three very different languages.

Use of reiterant speech makes it possible to identify and compare language-specific contrasts through simple kinematic analysis of a single position signal and the relation to its derivative. Yet, the reiterant productions meet language-specific expectations derived from experience with "natural" speech. The same articulatory correlates to stress and speaking rate distinctions have been observed for "more natural" productions. We doubt that the appropriate patterning of these contrasts is due to a special constraint on reiterant speech.

It may be a consequence of using reiterant speech that all the data fall so neatly on the same second-order function, within which most of the spatiotemporal distinctiveness in the data can attributed to the scaling of just two inferred dynamic parameters. Reiterant speech may oversimplify the picture and, for that reason, we are not overly concerned with the exact form of the model function. However, in our view, it is extremely unlikely that the apparently universal constraints on lip-jaw kinematics revealed here do not also apply to real speech. Indeed, stiffness and equilibrium position have been usefully applied to the modeling of real speech gestures (e.g., Browman & Goldstein, 1990).

Ultimately, the connections we have attempted to establish between intentional distinctions and dynamic parameters inferred from the kinematics must give way to modeling more directly the dynamics and biomechanics of the neuromotor and musculoskeletal systems. We can now collect high quality data for muscle activity and multidimensional movements and subject them to powerful correlation techniques such as neural networks. These techniques allow us to model the inherently dynamical relation between the action of muscles and their kinematic consequences. Parameters such as stiffness, damping, and equilibrium position can then be assessed as has recently been done for reiterant speech (Vatikiots-Bateson, Hirayama, & Kawato, 1991; Hirayama, Vatikiotis-Bateson, Kawato, & Jordan, in press) and is now being done for real speech (Hirayama, Vatikiotis-Bateson, Honda, & Kawato, in press).

5 Conclusion

In the foregoing, we have shown that much can be learned about the spatiotemporal organization of speech articulation from simple kinematic analysis of unidimensional articulator motion during reiterant speech production. For the speakers of each language, it was seen that the highly linear relation between gestural displacement and peak velocity accounts for most of the overall spatiotemporal variance of the system. Similarly, when specific conditions of a linguistic variable and speaking rate were considered, the condition-specific distributions demonstrated the same highly linear relation between peak velocity and displacement and occupied overlapping but usually distinct regions of the overall distribution. It was further shown that slope of the Vp-d relation, whether for a specific condition or for the entire data set, reflected measured mean gestural duration. Finally, we noted that there is an inverse relation between condition-specific mean displacement and slope of the Vp-d function. This is in keeping with the observed tendency for movement gestures to take longer to go farther. From these facts, we conclude that the motion of the system can be characterized in terms of an abstract second order dynamical system, whose underlying parameters, stiffness and equilibrium position, can be quantitatively inferred from the observed discrete kinematic measures (duration, displacement, and peak velocity) and their interrelations — specifically, the Vp-d and displacement-duration functions.

By comparing the French, Japanese, and English data, it was shown that the results for all three languages are qualitatively the same, yet quantitatively differ in accordance with independently demonstrated differences in temporal organization and prosody. It was suggested further that the temporal organization differences observed among languages may be based primarily on the severely constrained interaction of absolute speaking rate, production constraints on syllable structure, and inherent constraints on the perception of temporal distinctions. Finally, although further analysis is required to more adequately characterize the slight but consistently observed nonlinearity of the system, we conclude from these results that articulatory motion can be modeled in terms of a small number of universal underlying dynamic parameters whose values can be appropriately set to meet language-specific criteria.

References
Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine.

Anderson, S. R. (1982). The analysis of French shwa: Or how to get something for nothing. *Language, 58*, 534-573.

Beckman, M. (1982). Segment duration and the mora in Japanese. *Phonetica, 39*, 113-135.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language, 26*, 86-125.

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology, 6*, 145-238.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*.

Cohen, J. & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Crystal, T, H., & House, A. S. (1986). Variation of timing control: Maturational or statistical? *Journal of the Acoustical Society of America, 79*, Suppl. 1, S54.

Dalby, J., & Port, R. (1981). Temporal structure of Japanese: Segment, mora and word. *Research in Phonetics* (Indiana University Phonetics Laboratory), *2*, 149-172.

De Groot, A.W. (1926). Le syllabe. *Bulletin de Societé de Linguistique de Paris, 27*, 1-42.

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

Delattre, P. C. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics, 4*, 183-198.

Edwards, J. (1985). *Mandibular rotation and translation during speech*. Unpublished Doctoral Dissertation, CUNY.

Edwards, J., & Harris, K. (1990). Rotation and translation of the jaw during speech. *Journal of Speech and Hearing Research, 33*, 550-562.

Espinoza-Varas, B., & Watson, C. S. (1986). Temporal discrimination for single components of non-speech auditory patterns. *Journal of the Acoustical Society of America, 80*, 1685-1694.

Fairbanks, G. (1960). *Voice and articulation drillbook.* New York: Harper and Row.

Ferguson, G. A. (1981). *Statistical analysis in psychology and education.* New York: McGraw-Hill.

Flanagan, J.R., Ostry, D.J., Feldman, A.G. (1990). Control of human jaw and multi-joint arm movements. In G.E. Hammond (Ed), *Cerebral control of speech and limb movements.* Elsevier Science Publishers (North-Holland).

Fujisaki, H., Nakamura, K., & Imoto, T. (1975). Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech* (Vol. , pp. 197-220). New York: Academic Press.

Gay, T. J. (1981). Mechanisms in the control of speech rate. *Phonetica, 38*, 148-158.

Grammont, M. (1933). *Traite de phonetique.* Paris: Libraire Delgrave.

Hiiemae, K. M., & Crompton, A. W. (1985). Mastication, food transport, and swallowing. In M. Hillebrand, D. Bramble, D. Kiem, & D. Wake (eds) *Functional vertebrate morphology.* Cambridge, MA: Belknap.

Higurashi, Y. (1984). *The accent of extended word structure in Tokyo standard Japanese.* Tokyo: Educa.

Hirayama, M., Vatikiotis-Bateson, E., Kawato, M., & Honda, K. (in press). Neural network modeling of speech motor control. *Proceedings of the International Conference on Spoken Language Processing-1992.*

Hirayama, M., Vatikiotis-Bateson, E., Kawato, M., & Jordan, M. (in press). Forward dynamics modeling of speech motor control using physiological data. In Lippmann, R.P., Moody, J.E., & Touretzky, D.S.(Eds), *Advances in neural information processing systems 4.* San Mateo, CA: Morgan Kaufmann Publishers.

Imagawa, H., Kiritani, S., Masaki, S., & Shirai, K. (1985). Contextual variation in the jaw position for the vowels in /CVC/ utterances. *Annual Bulletin (Research Institute of Logopedics and Phoniatrics, Tokyo), 19*, 7-19.

Jordan, D.W., & Smith, P. (1977). *Nonlinear ordinary differential equations.* Oxford: Oxford University Press.

Kay, B.A., Kelso, J.A.S., Saltzman, E.L., & Schöner, G. (1987). The space-time behavior of single and bimanual rhythmical movements. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 178-192.

Kay, B. A., Munhall, K. G., V.-Bateson, E., & Kelso, J. A. S. (1985). Processing movement data at Haskins: Sampling, filtering, and differentiation. *Haskins Laboratories Status Report on Speech Research, SR-81*, 291-303.

Kelso, J. A. S., & Tuller, B. (1984). A dynamical basis for action systems. In Gazzaniga, M.S. (Ed), *Handbook of cognitive neuroscience.* New York: Plenum.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America, 77*, 266-280.

Kozhevnikov, V. A., & Chistovich, L. A. (1966). *Rech, Artikulyatsiya, i vospriyatiye, [Speech: Articulation and perception] (30*, p. 543, originally published 1965). Washington, DC: Joint Publications Res. Service.

Kuehn, D. P., & Moll, K. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics, 4*, 303-320.

Larkey, L. S. (1983). Reiterant speech: An acoustic and perceptual evaluation. *Journal of the Acoustical Society of America, 73*, 1337-1345.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5*, 253-264.

Liberman, M. Y., & Streeter, L. A. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America, 63*, 231-233.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*, 1773-1781.

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

Masaki, S., Shirai, K., Imagawa, H., & Kiritani, S. (1985). Differences in jaw opening for vowels due to speaking rate and word-internal position in the production of vowel sequence words. *Annual Bulletin (Research Institute of Logopedics and Phoniatrics, Tokyo), 19,* 29-46.

McCawley, J. (1978). What is a tone language? In V. A. Fromkin (ed.) *Tone: A linguistic survey.* New York: Academic Press.

McDevitt, W.E. (1989). *Functional anatomy of the masticatory system.* London: Wright.

Mermelstein, P. (1973). Articulatory model for the study of speech production. *Journal of the Acoustical Society of America, 53,* 1070-1082.

Mitsuya, F., & Sugito, M. (1978). A study of the accentual effect of on segmental and morais duration in Japanese. *Annual Bull. of the Res. Inst. Logopedics and Phoniatrics, 12,* 97-112.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review, 83,* 405-408.

Nelson, W. L. (1983). Physical principles of economies of skilled movements. *Biological Cybernetics, 46,* 135-147.

Ohala, J. J. Hiki, S., Hubler, S., & Harshman, R. (1968). Photoelectric methods of transducing lip and jaw movements in speech. *UCLA Working Papers in Phonetics, 10,* 135-144.

Ostry, D.J., Flanagan, J.R., Feldman, A.G., & Munhall, K.G. (in press). Jaw movement kinematics and control. In G.E. Stelmach and J. Requin, (Eds.), Tutorials in Motor Behavior II. North-Holland Press.

Ostry, D.J., Keller, E., & Parush, A. (1983). Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 622-636.

Pike, K. L. (1943). Phonetics. *Language and literature* (Vol. 21). Ann Arbor, MI: University of Michigan Press.

Pike, K.L. (1967). *Language in relation to a unified theory of the structure of human behavior.* The Hague: Mouton.

Saltzman, E. L., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review, 94,* 84-106.

Scripture, E. W. (1899a). Researches in experimental phonetics. *Yale Psychological Studies, 7,* 1-101.

Scripture, E. W. (1899b). Observations on rhythmic action. *Yale Psychological Studies, 7,* 102-108.

Selkirk, E. 0. (1978). The French foot: On the status of 'mute' e. *Journal of French Linguistics, 1,* 141-150.

Smith, C.L., Browman, C.P., McGowan, R.S., & Kay, B.L. (1991). Extracting dynamic parameters from speech movement data. *Haskins Laboratories Status Report on Speech Research, SR-105-106,* 107-140.

Stetson, R. H. (1905). A motor theory of rhythm and discrete succession. II. *Psychological Review, 12,* 293-350.

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

Vaissiere, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (eds), *Prosody: Models and measurement.* New York: Springer-Verlag.

Vatikiotis-Bateson, E. (1988). *Linguistic structure and articulatory dynamics.* Bloomington, IN: Indiana University Linguistics Club.

Vatikiotis-Bateson, E. (in preparation). Remote and local correlates of Japanese lexical accent in supralaryngeal articulation.

Vatikiotis-Bateson, E., Hirayama, M., & Kawato, M. (1991). Neural network modeling of speech motor control using physiological data. *Perilus, XIV,* 63-67.

Wallin, J. E. W. (1901). Researches on the rhythm of speech. *Yale Psychological Studies, 9,* 1-142.

Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10,* 193-216.

Appendices
A.1 *Nonlinear stiffness*

Because stiffness, inferred from the slope of the Vp-d relation, tends to decrease as mean displacement increases, the system's behavior resembles that of a nonlinear "soft" spring, . Furthermore, this "softening" of spring stiffness appears to vary among the languages in a lawful way. The effect of the nonlinearity is greater at the slower production rates of English than those of French or Japanese as shown by the fact that the overall linear covariation of peak velocity and displacement accounts for a smaller percentage of the spatiotemporal variance in the English data than that of the other two languages. It is possible that the relative strengths of the linear and nonlinear stiffness components for the different languages vary according to changes in the linear component alone. That is, the observed correlation between absolute speaking rates and the degrees to which the overall distributions adhere to a linear stiffness function might reflect a language-specific variation of linear stiffness, but a language-independent nonlinear stiffness component.

In their analysis of English reiterant speech, Kelso et al. (1985) also observed this relation between gestural duration and displacement. However, when the relation between acceleration (second temporal derivative of position) and displacement was examined around the spatial midpoint (inferred equilibrium position) of the movement gestures, condition-specific differences in slope of the $\ddot{x}/x$ relation were observed. They concluded from this that the overall distribution of the data was composed of different, condition-specific linear spring functions, whose stiffness was actively modulated according to displacement.

A very different approach was used here to assess the apparent overall nonlinearity of system stiffness. This method does not require measurement of the much noisier second time derivative of position, acceleration. Nonlinear regressions were computed for both individual speaker data and pooled data for each language. A Newton-Gaussian algorithm was used to fit the nonlinear function, $Vp^2 = \omega_0^2 A^2 - \delta A^4$, in which A is half the peak-to-valley displacement and $-\delta A^4$ defines the contribution of stiffness nonlinearity to the overall relation between stiffness and amplitude. This Vp-d function is derived from a nonlinear spring function of the form $F_s = -\omega_0^2 \Delta x + \delta \Delta x^3$; where F_s = spring force, ω_0^2 = (angular frequency)^2 = k = the mass-normalized, linear stiffness coefficient, and δ = the nonlinear stiffness coefficient. This Vp-d relation was derived for the mass-normalized nonlinear spring system using the harmonic balance method (Jordan & Smith, 1977). Note that when $\delta = 0$, this reduces to the familiar linear case of $Vp = \omega_0 A$. The regression analysis estimated w_0 and δ, which resulted in reasonable estimates of ω_0. However, in every case, the asymptotic correlation of ω_0 and δ was extremely high, indicating that the solution was unstable for these two parameters. Given that the linear Vp-d function, which estimates ω_0, already accounts for the bulk of the system's spatiotemporal variance, we conclude that the instability of the nonlinear function is due to the coupling of ω_0 and δ (i.e., active variations in one are rigidly linked to variations in the other).

A.2 *Linear damping*

In an undamped linear system, the relation between peak velocity and displacement is simply $Vp/A = \omega_0$. In the underdamped linear system, we assume the ratio of exponential growth to natural frequency, e, to be small and derive the following expressions: a) $\omega \approx \omega_0 (1 - \varepsilon)$, where ω is the observed (damped natural) frequency; and b) $Vp/A \approx \omega_0 [1 + (\pi/2 - 1) \varepsilon] + O(\varepsilon^2)$, where the final term denotes the second-order error term. Therefore, for a given displacement, the peak velocity and, hence, the slope of the Vp-d relation will be less than in the undamped system. Because the system is linear, the condition-specific regressions of peak velocity on displacement will still be linear.

Table I

Means and standard deviations of gestural displacement (DISP in mm), duration (DUR in ms), and peak velocity (PKV in mm/s). Data are grouped by language and speaker, gesture type (opening *vs.* closing), and syllable identity (*ba vs. ma*).

		OPENING			CLOSING		
		DISP	DUR	PKV	DISP	DUR	PKV
ENGLISH (N = 7241)							
RH	*ba*	7.4 (3.4)	88 (21)	137 (51)	7.3 (3.0)	79 (19)	163 (56)
	ma	8.4 (3.3)	88 (18)	172 (54)	8.1 (2.9)	76 (17)	187 (57)
MP	*ba*	3.6 (1.5)	91 (21)	69 (26)	3.5 (1.4)	81 (21)	87 (31)
	ma	4.5 (1.8)	93 (18)	92 (37)	4.3 (1.6)	81 (22)	102 (36)
JK	*ba*	9.1 (2.6)	114 (28)	149 (35)	9.0 (3.0)	105 (33)	178 (47)
	ma	9.4 (2.7)	108 (24)	179 (49)	9.3 (2.9)	108 (38)	167 (40)
FRENCH (N = 5842)							
BA	*ba*	6.9 (2.1)	75 (15)	157 (41)	6.7 (1.8)	67 (9)	183 (45)
	ma	7.2 (2.2)	76 (13)	163 (48)	6.9 (1.9)	70 (10)	176 (48)
DP	*ba*	5.0 (2.0)	73 (14)	123 (47)	5.0 (1.6)	70 (10)	121 (38)
	ma	5.5 (2.0)	76 (14)	138 (51)	5.3 (1.7)	70 (9)	126 (42)
CG	*ba*	3.3 (2.0)	91 (18)	66 (38)	2.7 (1.6)	77 (18)	58 (27)
	ma	4.1 (2.2)	89 (18)	90 (42)	3.5 (1.6)	86 (14)	69 (31)
JAPANESE (N = 11707)							
NK	*ba*	6.3 (2.2)	81 (22)	117 (27)	6.0 (2.2)	66 (17)	140 (34)
	ma	6.9 (2.2)	81 (22)	145 (27)	6.6 (2.3)	70 (19)	146 (32)
FE	*ba*	4.2 (3.3)	79 (21)	81 (54)	3.9 (3.4)	76 (22)	86 (62)
	ma	5.7 (2.9)	82 (19)	121 (59)	5.5 (3.0)	81 (20)	114 (56)
ME	*ba*	9.3 (2.6)	78 (11)	188 (50)	8.6 (2.5)	74 (13)	187 (45)
	ma	10.0 (2.5)	78 (11)	209 (51)	9.3 (2.3)	74 (13)	202 (43)
SM	*ba*	7.7 (2.1)	69 (10)	175 (47)	7.4 (1.8)	63 (9)	186 (43)
	ma	8.0 (1.9)	68 (9)	187 (44)	7.7 (1.6)	65 (9)	185 (40)
MY	*ba*	4.5 (2.4)	78 (18)	96 (44)	4.2 (1.8)	66 (12)	108 (45)
	ma	6.3 (2.4)	77 (16)	143 (51)	6.1 (2.2)	70 (12)	149 (54)

Table II

Means and standard deviations of gestural displacement (DISP in mm), duration (DUR in ms), and peak velocity (PKV in mm/s). Data are grouped by language and speaker, speaking rate condition (conversational *vs.* fast), and value of the prosodic variable (±Str for English and French; high *vs.* low tone for Japanese).

| | | CONVERSATIONAL | | | FAST | | |
		DISP	DUR	PKV	DISP	DUR	PKV
ENGLISH (N = 7241)							
RH	-Str	8.1 (3.0)	86 (20)	169 (50)	5.5 (3.1)	69 (16)	130 (60)
	+Str	10.2 (2.3)	97 (16)	196 (46)	8.0 (2.4)	81 (15)	176 (51)
MP	-Str	3.8 (1.5)	88 (18)	81 (31)	3.1 (1.3)	76 (13)	73 (30)
	+Str	5.2 (1.6)	101 (27)	107 (35)	4.2 (1.5)	84 (18)	97 (35)
JK	-Str	8.8 (2.7)	108 (33)	159 (43)	7.4 (2.3)	88 (20)	150 (42)
	+Str	11.6 (2.1)	136 (27)	193 (38)	10.5 (2.0)	115 (24)	187 (41)
FRENCH (N = 5812)							
BA	-Str	6.8 (1.4)	74 (8)	164 (37)	5.8 (1.2)	66 (8)	150 (37)
	+Str	9.3 (2.5)	84 (20)	215 (53)	6.9 (1.8)	67 (10)	174 (45)
DP	-Str	4.8 (1.5)	72 (11)	117 (38)	4.3 (1.5)	68 (10)	110 (39)
	+Str	7.0 (1.9)	79 (14)	165 (42)	6.3 (1.8)	76 (14)	153 (42)
CG	-Str	3.0 (1.4)	87 (17)	63 (29)	2.5 (1.3)	78 (16)	58 (28)
	+Str	4.8 (2.3)	96 (19)	94 (40)	4.4 (2.3)	88 (17)	92 (43)
JAPANESE (N = 7531)							
NK	High	7.6 (1.4)	82 (15)	153 (25)	5.4 (1.5)	67 (15)	124 (25)
	Low	7.5 (2.3)	85 (26)	147 (30)	5.4 (2.4)	66 (19)	124 (36)
FE	High	6.0 (2.7)	82 (20)	126 (53)	4.5 (2.5)	79 (19)	94 (47)
	Low	4.4 (2.6)	77 (14)	100 (58)	3.0 (1.9)	74 (15)	66 (38)
ME	High	10.4 (2.2)	80 (13)	213 (40)	8.8 (2.6)	74 (12)	187 (50)
	Low	10.1 (2.0)	79 (12)	211 (40)	7.7 (2.3)	71 (9)	171 (51)
SM	High	8.8 (2.0)	70 (11)	205 (46)	7.6 (1.9)	66 (10)	182 (41)
	Low	7.6 (1.6)	67 (9)	180 (40)	7.0 (1.7)	64 (9)	169 (42)

Table III

Number of observations (n), linear regression coefficient (r), and slope (m) for the linear regressions of gestural displacement on duration. Data for the regressions are grouped by language and speaker, gesture type (opening *vs.* closing), and syllable identity (*ba vs. ma*).

		OPENING			CLOSING		
		n	r	m	n	r	m
ENGLISH							
RH	*ba*	603	0.86	139.6	604	0.73	118.0
	ma	600	0.85	151.1	601	0.68	118.7
MP	*ba*	617	0.63	45.7	616	0.52	35.4
	ma	600	0.64	64.4	602	0.46	33.7
JK	*ba*	600	0.71	66.9	599	0.73	65.9
	ma	603	0.61	67.1	596	0.73	55.9
FRENCH							
BA	*ba*	470	0.69	95.4	471	0.39	75.5
	ma	469	0.68	114.1	468	0.28	51.3
DP	*ba*	526	0.65	93.5	526	0.35	58.4
	ma	511	0.50	72.2	501	0.18	34.5
CG	*ba*	456	0.50	55.4	486	0.59	50.0
	ma	474	0.61	73.0	484	0.40	48.3
JAPANESE							
NK	*ba*	603	0.89	88.0	603	0.85	111.6
	ma	569	0.90	89.7	566	0.82	100.8
FE	*ba*	511	0.77	122.5	508	0.53	79.8
	ma	528	0.62	93.4	526	0.49	74.6
ME	*ba*	684	0.56	136.0	645	0.63	118.5
	ma	671	0.56	128.0	632	0.65	116.7
SM	*ba*	340	0.46	95.1	340	0.38	79.5
	ma	360	0.36	75.4	355	0.26	46.1
MY	*ba*	816	0.72	96.8	816	0.42	66.7
	ma	817	0.62	93.3	818	0.46	85.4

Table IV

Mean gestural duration (DUR in ms), linear regression coefficient (r), and linear slope (m) for the regression of peak velocity on displacement (Vp-d). Data are grouped by language and speaker, gesture type (opening *vs.* closing), and syllable identity (*ba vs. ma*). Pairs of duration values in boldface indicate statistical equivalence. T-values (one-tailed) are given for Vp-d slope comparisons between opening and closing gestures. Values within ±1.65 are nonsignificant; values in boldface are contrary to the model prediction (see text for details).

		OPENING			CLOSING			t-value
		DUR	r	m	DUR	r	m	
ENGLISH								
RH	*ba*	88	.94	14.1	79	.89	16.5	15.88
	ma	88	.94	15.5	76	.88	17.2	8.85
MP	*ba*	91	.89	15.1	81	.88	19.1	14.58
	ma	93	.91	19.2	81	.87	19.2	**0.27**
JK	*ba*	114	.76	10.2	105	.87	13.6	14.82
	ma	**108**	.90	16.6	**108**	.71	9.9	**-24.90**
FRENCH								
BA	*ba*	75	.95	18.7	67	.90	22.3	10.90
	ma	76	.95	20.4	70	.89	23.0	6.91
DP	*ba*	**73**	.95	21.7	**70**	.93	21.9	0.76
	ma	76	.94	23.4	70	.94	22.5	**-3.51**
CG	*ba*	91	.95	17.6	77	.93	16.3	**-8.30**
	ma	**89**	.94	18.0	**86**	.93	18.0	0.06
JAPANESE								
NK	*ba*	81	.83	10.6	66	.91	13.8	21.43
	ma	81	.82	10.1	70	.84	11.7	8.20
FE	*ba*	79	.93	15.3	76	.96	17.6	19.49
	ma	**82**	.93	19.1	**81**	.94	17.3	**-9.65**
ME	*ba*	78	.94	18.2	74	.88	16.3	**-10.30**
	ma	78	.91	18.3	74	.84	15.4	**-10.80**
SM	*ba*	69	.93	20.5	63	.91	22.0	2.98
	ma	68	.92	21.6	65	.86	21.1	-0.74
MY	*ba*	78	.94	17.6	66	.95	23.4	45.56
	ma	77	.92	19.6	70	.94	22.9	17.90

Table V

Mean gestural duration (DUR in ms), linear regression coefficient (r), and linear slope (m) for the regression of peak velocity on displacement (Vp-d). Data are grouped by language and speaker, speaking rate (conversational *vs*. fast), and gesture type (Op *vs*. Cl). Pairs of duration values in boldface indicate statistical equivalence. T-values (one-tailed) are given for Vp-d slope comparisons between conversational and fast rate gestures. Values less than 1.65 are nonsignificant.

		CONVERSATIONAL			FAST			t-value
		DUR	r	m	DUR	r	m	
ENGLISH								
RH	Op	96	.90	14.4	79	.94	16.7	16.43
	Cl	85	.84	15.1	69	.93	20.5	19.83
MP	Op	100	.90	17.6	84	.92	19.5	7.87
	Cl	88	.87	18.4	74	.90	21.0	6.88
JK	Op	121	.79	13.6	100	.84	14.9	3.77
	Cl	116	.76	11.2	97	.81	12.8	5.63
FRENCH								
BA	Op	81	.95	19.6	70	.94	21.3	6.99
	Cl	73	.91	22.4	63	.87	26.7	6.31
DP	Op	76	.95	22.5	72	.94	23.1	2.61
	Cl	72	.93	22.0	68	.94	22.8	2.81
CG	Op	94	.94	17.8	86	.95	18.9	5.90
	Cl	86	.94	16.2	77	.94	18.2	11.57
JAPANESE								
NK	Op	91	.72	10.9	72	.78	11.5	1.65
	Cl	75	.79	10.4	61	.89	14.0	18.10
FE	Op	84	.90	16.6	77	.95	18.2	7.03
	Cl	81	.94	16.8	77	.97	18.2	10.89
ME	Op	82	.87	17.7	75	.94	19.4	7.76
	Cl	78	.75	13.3	70	.92	18.5	16.74
SM	Op	**70**	.93	21.3	**68**	.91	21.0	-0.59
	Cl	**66**	.89	22.0	**63**	.87	21.1	-1.12
MY	Op	85	.93	19.8	70	.94	21.1	8.40
	Cl	74	.94	22.3	62	.96	26.2	22.53

Table VI

Number of observations (n), linear regression coefficient (r), and slope (m) for the condition-specific regressions of peak velocity on displacement. Data for the regressions are grouped by language and speaker, syllable identity (*ba vs. ma*), gesture type (opening *vs.* closing) and stress-rate condition.

		ba								*ma*						
		Opening			Closing				Opening			Closing		
		r	m		r	m			r	m		r	m
ENGLISH													
RH	-str/C	.95	15.2		.89	18.6			.94	14.9		.90	19.9
	+str/C	.86	13.3		.79	14.8			.90	15.7		.79	17.2
	-str/F	.97	16.6		.96	22.6			.96	17.6		.94	23.0
	+str/F	.89	14.1		.89	19.6			.89	17.1		.89	20.6
MP	-str/C	.91	15.2		.91	20.1			.88	17.2		.89	21.4
	+str/C	.88	16.0		.81	17.2			.88	19.9		.69	15.4
	-str/F	.91	17.6		.88	21.3			.94	19.8		.91	24.1
	+str/F	.89	17.1		.87	18.9			.91	21.5		.85	20.2
JK	-str/C	.78	11.1		.86	14.0			.92	18.1		.86	13.4
	+str/C	.74	12.0		.64	9.8			.84	17.0		.56	7.2
	-str/F	.88	14.7		.87	16.6			.89	17.2		.86	16.7
	+str/F	.83	14.0		.81	14.1			.88	18.7		.56	9.3
FRENCH													
BA	-str/C	.93	21.5		.87	22.5			.92	21.3		.85	23.5
	+str/C	.94	18.1		.94	21.7			.94	20.9		.93	21.2
	-str/F	.91	21.7		.90	29.2			.91	23.6		.85	26.2
	+str/F	.94	21.1		.91	26.1			.94	21.0		.91	27.7
DP	-str/C	.93	24.6		.89	20.9			.95	27.3		.93	22.1
	+str/C	.92	20.0		.90	20.1			.95	21.3		.94	21.5
	-str/F	.94	26.3		.94	23.1			.91	26.4		.92	22.1
	+str/F	.94	21.3		.92	21.7			.92	23.5		.95	22.2
CG	-str/C	.94	20.3		.91	16.2			.94	21.1		.93	16.2
	+str/C	.93	15.4		.94	15.6			.91	15.6		.93	17.3
	-str/F	.89	19.2		.90	16.9			.93	22.4		.94	18.2
	+str/F	.95	17.2		.93	16.7			.94	17.4		.95	21.1

Table VII

Mean gestural duration (in ms) for each stress condition. Data are grouped by language and speaker, gesture type (opening *vs.* closing), syllable identity (*ba vs. ma*), and speaking rate (Conv *vs.* Fast). Pairs of duration values in boldface indicate statistical equivalence. Vp-d slopes were tested (one-tailed t) for stress differences. Asterisks indicate probabilities at the .05 (*), .01 (**), and .001 (***) levels. Reliable slope differences counter to the model prediction are given in boldface.

		Rate	Opening Duration		Slope Test	Closing Duration		Slope Test
			+str	-str		+str	-str	
ENGLISH								
RH	*ba*	Conv	104	88	**	91	79	***
		Fast	88	69	***	76	64	***
	ma	Conv	103	88	ns	88	76	**
		Fast	86	72	ns	73	64	**
MP	*ba*	Conv	109	88	ns	94	83	**
		Fast	88	75	ns	75	71	*
	ma	Conv	105	90	**	93	79	***
		Fast	89	79	ns	78	71	***
JK	*ba*	Conv	143	110	ns	132	99	***
		Fast	119	91	ns	115	83	**
	ma	Conv	132	105	ns	138	102	***
		Fast	107	90	ns	119	81	***
FRENCH								
BA	*ba*	Conv	96	73	***	**73**	**72**	ns
		Fast	73	67	ns	**61**	**61**	*
	ma	Conv	94	75	ns	**74**	**74**	*
		Fast	72	69	**	**64**	**66**	ns
DP	*ba*	Conv	88	69	***	**71**	**73**	ns
		Fast	82	66	***	**68**	**67**	ns
	ma	Conv	87	72	***	**70**	**73**	ns
		Fast	86	69	*	**67**	**68**	ns
CG	*ba*	Conv	107	90	***	84	80	ns
		Fast	96	82	*	77	72	ns
	ma	Conv	105	88	***	**90**	**90**	ns
		Fast	98	79	***	**82**	**80**	***

Table VIII

Number of observations (n), linear regression coefficient (r), and slope (m) for the condition-specific regressions of peak velocity on displacement. Data for the regressions are grouped by Japanese speaker, syllable identity (*ba vs. ma*), gesture type (opening *vs.* closing) and tone-rate condition.

		ba Opening		Closing		*ma* Opening		Closing	
JAPANESE		**r**	**m**	**r**	**m**	**r**	**m**	**r**	**m**
NK									
	Low/C	.80	14.9	.71	11.4	.85	14.2	.55	8.2
	High/C	.71	12.7	.72	14.3	.77	15.5	.58	10.9
	Low/F	.89	15.1	.83	13.0	.81	14.0	.59	8.6
	High/F	.92	16.6	.87	18.0	.82	15.6	.68	14.6
FE									
	Low/C	.94	18.2	.93	16.4	.92	20.8	.85	14.6
	High/C	.98	18.9	.97	25.6	.97	21.9	.96	23.8
	Low/F	.94	16.6	.96	18.0	.95	20.8	.93	16.6
	High/F	.98	20.0	.96	25.6	.92	17.9	.95	21.3
ME									
	Low/C	.89	17.5	.80	16.0	.84	17.0	.69	12.8
	High/C	.88	19.3	.73	13.9	.91	19.3	.74	12.6
	Low/F	.95	18.4	.92	19.6	.93	19.1	.87	18.0
	High/F	.96	21.0	.96	22.7	.95	23.1	.91	17.8
SM									
	Low/C	.90	20.2	.91	22.7	.94	22.1	.91	24.0
	High/C	.89	20.4	.86	20.4	.95	29.1	.83	23.3
	Low/F	.91	19.7	.93	21.7	.87	19.8	.82	20.0
	High/F	.92	22.1	.94	23.8	.93	23.1	.86	21.2

Table IX

Mean gestural duration (in ms) for each tone condition. Data are grouped by language and speaker, gesture type (opening *vs.* closing), syllable identity (*ba vs. ma*), and speaking rate (Conv *vs.* Fast). Pairs of duration values in boldface indicate statistical equivalence. Vp-d slopes were tested (one-tailed t) for tone differences. Asterisks indicate probabilities at the .05 (*), .01 (**), and .001 (***) levels. Reliable slope differences counter to the model prediction are given in boldface.

		Rate	Opening Duration		Slope Test	Closing Duration		Slope Test
			Low	High		Low	High	
NK								
	ba	Conv	89	83	ns	73	68	*
		Fast	72	64	ns	59	53	***
	ma	Conv	90	81	ns	77	72	ns
		Fast	72	64	ns	65	58	***
FE								
	ba	Conv	84	77	ns	77	72	***
		Fast	81	67	***	72	68	***
	ma	Conv	86	79	ns	80	78	***
		Fast	85	76	***	76	79	***
ME								
	ba	Conv	84	75	*	**80**	**81**	ns
		Fast	78	69	***	**71**	**70**	***
	ma	Conv	84	77	**	**78**	**79**	ns
		Fast	79	70	***	**71**	**71**	ns
SM								
	ba	Conv	73	68	ns	**65**	**66**	ns
		Fast	70	65	ns	**60**	**61**	ns
	ma	Conv	75	65	***	**66**	**68**	ns
		Fast	**60**	**61**	**	**62**	**65**	ns

Table X

Kinematic means (DUR in ms, DISP in mm, PKV in mm/s) and linear regression results (coefficients, r, and slopes, m) for renditions of underlying single mora (CV) and two-mora (CVV) open syllables. Data are grouped by speaker, gesture type (opening *vs.* closing), syllable type (*ba vs. ma*), and mora complexity (CV *vs.* CVV).

		OPENING					CLOSING				
		r	m	DUR	DISP	PKV	r	m	DUR	DISP	PKV
Spkr: NK											
ba	CV	.88	13.3	77	6.0	115	.88	14.8	64	5.6	135
	CVV	.76	10.0	145	10.9	141	.57	10.8	101	11.6	212
ma	CV	.86	13.1	77	6.5	143	.72	10.3	68	6.2	140
	CVV	.57	8.2	140	12.0	175	.46	6.7	102	12.6	225
Spkr: FE											
ba	CV	.97	18.6	76	3.7	77	.95	18.8	75	3.3	75
	CVV	.94	8.2	120	9.5	138	.94	12.0	97	11.1	212
ma	CV	.95	21.2	79	5.3	117	.93	18.6	79	5.0	106
	CVV	.88	14.2	118	9.8	172	.92	15.3	106	11.5	206