The Journal of Neuroscience, November 1994, 14(11): 6585–6597

Speech Motor Coordination and Control: Evidence from Lip, Jaw, and Laryngeal Movements

Vincent L. Gracco and Anders Löfqvist

Haskins Laboratories, New Haven, Connecticut 06510

The movements of the lower lip, jaw, and larynx during speech were examined for two different speech actions involving oral closing for /p/ and oral constriction for /f/. The initial analysis focused on the manner in which the different speech articulators were coordinated to achieve sound production. It was found that the lip, jaw, and laryngeal movements were highly constrained in their relative timing apparently to facilitate their coordination. Differences were noted in the degree to which speech articulator timing covaried dependent on the functional characteristics of the action. Movements associated with coordinating multiple articulators for a single sound were more highly constrained in their relative timing than were movements associated with sequencing of individual sounds. The kinematic patterns for the different articulators were found to vary in a number of systematic ways depending on the identity of the sound being produced, the phonetic context surrounding the target sound, and whether one versus two consonants were produced in sequence. The results are consistent with an underlying organization reflecting the construct of the phoneme. It is suggested that vocal tract actions for the sounds of the language are stored in memory as motor programs and sequenced together into larger meaningful units during speaking. Speech articulator motion for the different vowel sounds was found to be influenced by the identity of the following consonant, suggesting that speech movements are modified in chunks larger than the individual phonetic segments. It appears that speech production is a hierarchical process with multiple levels of organization transforming cognitive intent into coherent and perceptually identifiable sound sequences.

[Key words: speech motor control, speech movement coordination, speech motor programs, lips, jaw, larynx]

As a highly developed skilled motor behavior, speech production provides a rich environment for observing the functional synergies and coordinative principles that underlie a uniquely human behavior. Like most motor behaviors, speaking requires the interaction of multiple effectors (speech articulators) into larger functional aggregates. These articulatory aggregates are the framework for speech motor control and their activation is associated with sound production. As such, the ultimate goal of speech movement coordination is generally known. One issue of interest in speech motor control as well as motor control in general is the manner in which the nervous system controls the multiple degrees of movement freedom (Bernstein, 1967). It is generally accepted that the nervous system employs simplifying strategies to reduce the potentially independent variables (motor units, muscles, joints) in most motor behaviors to a controllable number (Turvey, 1977; Lacquaniti and Soechting, 1982; Gracco, 1988; MacPherson, 1988a,b; Soechting and Lacquaniti, 1989; MacKenzie, 1992). Recently, analysis of the relative timing of the lips and jaw suggests that the multiple articulators are interdependently modulated such that timing variation in one articulator is accompanied by proportional changes in the timing of all the active articulators (Gracco and Abbs, 1986; Gracco, 1988, 1994). Rather than considering each articulator as independently controlled it has been suggested that speech articulators are functionally constrained. That is, rather than explicitly controlling the timing of the different neuromuscular elements involved in the production of a particular sound, the nervous system controls the coordinative requirements of all the active effectors as a unit (Gracco, 1990, 1991)

To date the most direct evidence for constraining speech movement timing has come from a relatively simple articulatory event, oral closing (Gracco and Abbs, 1986; Gracco, 1988, 1994). Oral closing for bilabial sounds such as /p/, /b/, or /m/ simply involves the approximation of the two lips momentarily. It is not clear how general such coordinative interactions are among different speech articulators and whether such interactions change for different speech sounds. As noted above speech production is dependent on the actions of articulators other than the lips and jaw. One speech articulator that is also involved in many of the sounds of English is the larynx. The larynx is a time-varying valve involved in the initiation and arrest of vocal fold vibration for various vowel and consonant sounds. For voiceless consonant sounds, such as /p/, /t/, /k/, /f/, /s/, /sh/, the larynx must open in conjunction with the raising of tongue or lips to create an occlusion (/p/, /t/, or /k/) or constriction (/f/, /s/, /sh/), generating the necessary aerodynamic conditions within the vocal tract. For vowels following a voiceless consonant, the vocal folds, in conjunction with the jaw and tongue, approximate to provide a vibrating sound source. For each of these situations, voiceless consonants and vowels, the laryngeal action must be integrated with the movements of other speech articulators. Examination of the timing relations among the component articulators should provide insight into the speech movement coordination process. One focus of the present investigation is to examine the manner and degree to which the lips and jaw are coupled in their timing to the larynx for the production of vowel and voiceless consonant sounds. By examining the rel-

Received Feb. 25, 1994; accepted Apr. 27, 1994.

This research was supported by NIH Grants DC-00121, DC-00594, and DC-00865 from the National Institute on Deafness and Other Communication Disorders, and by Esprit-BR Project 6975-Speech Maps.

Correspondence should be addressed to Vincent L. Gracco, Ph.D., Haskins Laboratories, 270 Crown Street, New Haven, CT 06510.

ative timing among the lips, jaw, and larynx it should be possible to evaluate the degree and character of the temporal coupling associated with the different sound categories (consonant versus vowels).

From previous investigations it is also clear that the principles of speech movement coordination are not rigidly specified and vary at least according to movement direction. For example, lip and jaw motion for oral closing is tightly coupled and the timing of each articulator demonstrates significant covariation (Gracco and Abbs, 1986; Gracco, 1988). For oral opening, however, these articulators do not display the same degree of temporal coupling (Gracco, 1988, 1994). Rather, for oral opening associated with a vowel sound, the timing constraint among the lips and jaw is apparently relaxed compared to their timing during oral closing. One possibility is that oral opening, generally associated with vowel production, and oral closing, generally associated with consonant production, are two distinct classes of speech motor actions with different principles underlying their coordination and control. Moving toward a vowel target, for example, may not require the same degree of temporal coupling among the contributing articulators as moving toward certain consonant targets (Gracco, 1994). It may not be surprising, then, that the lips and jaw are not as tightly coupled in their timing for oral opening as for oral closing. However, as with previous investigations, the context in which such observations have been made have been limited. The present investigation will focus on a larger phonetic context than has been previously examined.

A final focus of the present investigation of some theoretical importance for speech motor control is determining the characteristics of the underlying neural representation. While it is generally agreed that speech motor output is dependent on some underlying neural representations (production units) the form of such representations have yet to be determined. One possibility is that the units for speech are motor programs uniquely specified for the individual sounds (phonemes) of the language (Gracco, 1990, 1991). This conceptualization would require a finite number of motor programs (one per phoneme) that would be activated and sequenced into larger aggregates associated with syllables, words, phrases to allow meaningful communication. An alternative conceptualization involves a set of fundamental articulatory actions or features that are assembled and coordinated according to the phonetic context of the message (Kelso, 1986). One way to distinguish between these two alternatives is to examine the changes with context across articulators. Contextual variations influencing more than a single articulator might suggest that the entire vocal tract is being manipulated rather than the action of a single articulator. The difference between these two alternatives relates to the size of the fundamental units for speech production (phonetic segments vs articulatory gestures) and the level at which control is exerted (single or multiple articulators). Through a detailed examination of the movement differences associated with different sounds in sequence it will be possible to identify the specific kinematic adjustments that differentiate sounds and provide an objective method of characterizing speech articulator actions.

Materials and Methods

Three adult males (aged 40–48 years) served as subjects for the present investigation. Articulatory motions of the upper lip, lower lip, and jaw in the horizontal and vertical dimensions and changes in glottal area (or aperture) were obtained. Movements of the lips and jaw were transduced optoelectronically using small light-emitting diodes (LEDs) placed midsagittally on the vermilion border of the upper and lower lips. Changes in the positions of the LEDs were sensed by a planar diode located in the focal plane of a camera mounted on a tripod and placed approximately 25 inches from the subject. For jaw motion, a custom-fitted splint was constructed for each subject that fit snugly around the lower molars on one side. A piece of stainless steel wire was molded into the splint and bent to exit the corner of the mouth with minimal obstruction to the subjects articulation. The wire was bent to the midsagittal plane and an LED was placed on the extension of the jaw splint close to the chin allowing direct transduction of jaw motion. Glottal aperture was obtained using transillumination of the larynx. A flexible endoscope with a DC light source was passed through the nose and suspended in the oropharynx. The endoscope provided a light source that was registered at a sensor secured to the neck and placed external and inferior to the thyroid cartilage. The luminance registered at the sensor has been shown to vary as a function of changes in glottal area associated with opening and closing the glottis for voiceless sounds (Löfqvist and Yoshioka, 1980; Baer et al., 1983). Figure 1 is a schematic representing the experimental setup. Lip, jaw, and glottal signals were sampled at 500 Hz (12-bit resolution) and subsequently smoothed (42-point triangular window) and numerically differentiated (central difference) in software.

Subjects repeated one of seven words in the carrier phrase "It's a ______ again" at a comfortable speaking rate and loudness. The words used contained one of four vowels in combination with either the voiceless consonants /p/ and /f/ or the consonant sequence /ft/. The words included (1) sapapple, (2) supper, (3) suffer, (4) safe, (5) safety, (6) sipping, (7) sifting.

For "supper" and "suffer" the same vowel was used with a different following consonant: "safe" and "safety" differ in the presence of the consonant sequence (/ft/); "sipping" and "sifting" differ by the consonant sequence and the identity of the voiceless consonant (/p/ versus /f/). The words were repeated in blocks of 10 and each block was repeated four times. For subject ES, a number of repetitions were discarded because of poor transillumination signal quality due to the tongue obscuring part of the DC light source. The number of repetitions for each word per subject was, for S:VG, 40, 40, 39, 40, 39, 40, 39; S:AL, 40, 40, 40, 37, 40, 40, 40; S:ES, 32, 32, 37, 23, 39, 39, 39 for words 1–7, respectively.

Presented in Figure 2 are the signals recorded and the measurement points identified. To evaluate articulator coordination two temporal intervals were examined in detail. These include the temporal relationship between (1) jaw lowering and glottal closing for the vowel following the initial voiceless consonant /s/, and (2) glottal opening and lower lip raising for the occlusion (/p/) or constriction (/f/). In all cases, the relative timing of the articulatory events were based on the time of peak velocity and referenced to the peak glottal opening associated with the in "It's" in the carrier phrase. Because most of the motion of the lip and jaw was confined to the vertical plane (with respect to gravity), the kinematic measures will focus on this single dimension. Movements examined included the jaw lowering displacement and velocity for the different vowels, the lip raising displacement and velocity for the consonants, and the glottal aperture velocity for the opening and closing phases for the different consonants.

Results

Speech movement coordination—relative timing

In the present investigation the relative timing of the lip and jaw were examined with respect to the action of the larynx. All the words examined began with /s/, which requires a stable and high jaw position (relative to the maxilla) for the tongue articulation. In addition, /s/ is a voiceless consonant requiring larynx abduction or glottal opening. For the different sound sequences the jaw is then lowered from its relatively high position and the larynx is closed to allow phonation for the different vowels. In the present context this vowel-related action was then followed by lower lip raising and glottal opening to produce the subsequent voiceless consonants.

The initial comparison focused on the timing of the jaw lowering and the larynx closing action associated with opening the oral aperture for the different vowel sounds. As mentioned above,

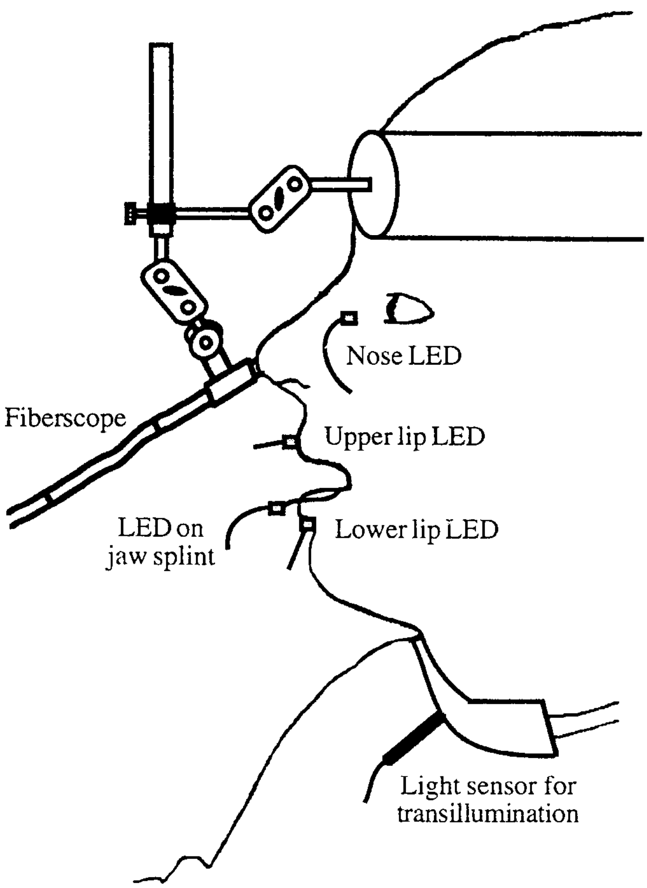

Figure 1. A line drawing of the experimental setup. A fiberscope, providing a DC light source, was passed through the nose and suspended in the pharynx. The light passed through the glottal opening in the larynx and the luminance was sensed from a sensor placed around the neck. The degree of luminance changed as a function of the glottal opening for the voiceless consonant sounds and was recorded as an analog voltage. Light emitting diodes (LEDs) were placed on the bridge of the nose, the upper lip, lower lip, and on a jaw splint that exited from the mouth and provided signals corresponding to the motion of the respective articulators in the horizontal and vertical dimensions (see text for details). The LEDs were pulsed and the light emitted was sensed at a planar diode located in the focal plane of a camera mounted on a tripod.

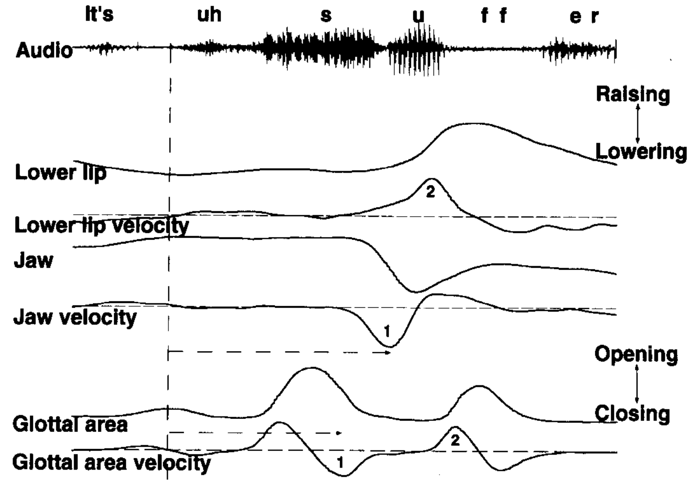

Figure 2. A schematic of the signals recorded for a single token of the phrase "It's a suffer" and the measurement points used in the present investigation. The signals from top to bottom include the acoustic signal recorded with a microphone, the vertical lower lip movement, the lower lip velocity, the vertical jaw movement, the jaw velocity, and the glottal area (aperture) and the change in glottal area (velocity). The *dashed line* indicates the midpoint of the glottal opening for "It's" and is used as the line-up point for all the timing measures (see text for details). The *horizontal dashed lines* illustrate one of the timing measures: the time of peak glottal closing (*1*) and the time of peak jaw lowering for the vowel sound (*2*). In addition to the timing measures, the jaw lowering displacement and associated peak velocity, the lower lip raising and associated peak velocity, and the peak glottal opening and closing velocities were also obtained.

all times are relative to the peak glottal opening for the "it's" in the carrier phrase and all timing measures reflect the occurrence of peak velocity associated with the respective lip, jaw, or laryngeal actions. The left portion of Figure 3 presents scatterplots of the time of the jaw lowering peak velocity and the time of the peak glottal closing velocity for the different vowels for the three subjects. The data have been grouped according to the different vowels following the /s/ sound. The vowel /U/ refers to "supper" and "suffer," /I/ refers to "sipping" and "sifting," /eI/ refers to "safe" and "safety," and /ae/ refers to "sapapple." As shown in the figure, there is a tendency for the timing of the jaw lowering to covary with the timing of the glottal closing. The correlation coefficients for the different vowels and subjects are presented in Table 1.

All correlations were significant ($p < 0.01$) although the magnitude of the relations varied quite a bit within and across the three subjects. For all subjects, the glottal closing peak velocity occurred in advance of the jaw lowering peak velocity. This can be seen in the mean interval between the glottal closing peak velocity and the jaw lowering peak velocity presented in the right side of Figure 3. The positive value for each vowel indicates that the glottal adjustment is initiated prior to the jaw adjustment associated with the tongue action. For two of the three subjects the same trend was noted with the largest interval associated with the vowel /ae/ and the smallest interval associated with the vowel /I/. Interestingly, for these two subjects the intervals were positively correlated with the magnitude of the jaw lowering peak velocity (see below).

In contrast to the opening action, the closing action of the lips and larynx for the different consonant sounds was found to be highly correlated in relative timing. Correlation coefficients for the timing of lip raising and glottal opening are presented in Table 2.

In comparison to the correlations presented in Table 1, the correlations for the closing action were higher for all subjects with coefficients ranging from $r = 0.93$ to $r = 0.99$. Presented in the left portion of Figure 4 are scatterplots of the time of lip raising and glottal opening peak velocity for the three subjects. With few exceptions, the timing relations are similar across contexts. Presented in the right portion of the figure are the

Table 1. Correlation of the time of the glottal closing velocity following /s/ with the time of peak velocity for the jaw lowering movement for the vowel for the three subjects

Subject	/ae/	/U/	/I/	/eI/
VG	0.930	0.699	0.899	0.930
AL	0.549	0.800	0.886	0.743
ES	0.705	0.893	0.818	0.864

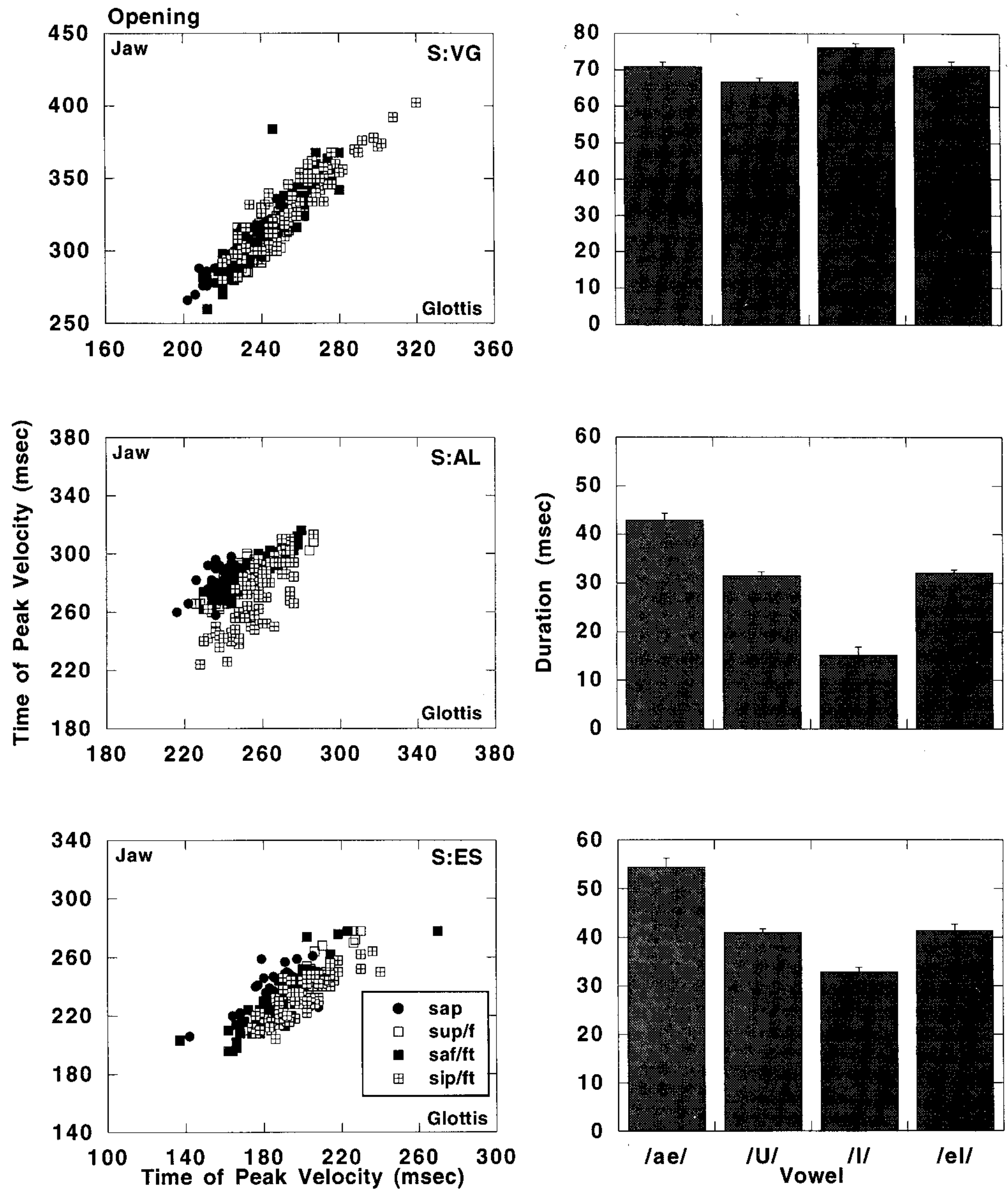

Figure 3. *Left*, Scatterplots of the time the jaw lowering (opening) peak closing velocity (in milliseconds) as a function of the glottal peak closing velocity for the four different vowels (/ae/, /U/, /I/, and /ai/) for the three subjects. *Right*, Mean differences between the time of the glottal closing velocity and jaw lowering velocity. The positive difference indicates that the time of glottal closing peak velocity always preceded the time of jaw lowering peak velocity. Error bars indicate 1 SE.

mean intervals between the time of lip raising velocity and the time of glottal opening velocity for the three different consonant contexts. Similar to the opening sequence, the lower lip peak velocity always preceded the glottal opening peak velocity. Similar to the oral opening results, the rank order of the intervals was not consistent across the different subjects. Since the relation between lower lip raising and peak glottal opening actions may be an important variable associated with the different consonants, it was also of interest to determine whether these two events demonstrated systematic consonant-related changes. To address this issue the difference between the time of peak glottal opening and the time of peak displacement for lip raising was obtained. The results for the three subjects are presented in Figure 5. The positive values indicate that the glottal peak opening occurred after the peak raising displacement of the lower lip while the negative value for S:ES for /f/ indicates that the order was reversed. While the differences across the consonant conditions were statistically different, there was no consistent trend across the three subjects. However, it appears that for two subjects the interval for /f/ is smaller than for /p/ and the interval for /ft/ in longer than either of the single consonants.

Table 2. Correlation of the time of peak lower lip raising velocity with the time of the glottal peak opening velocity for the consonant for the three subjects

Subject	/p/	/f/	/ft/
VG	0.956	0.941	0.986
AL	0.967	0.974	0.966
ES	0.970	0.982	0.929

Movement adjustments

Oral opening. The results suggest that the lip, jaw, and laryngeal movements are coupled in their timing and that the degree of coupling is greater for oral closing than for oral opening. In order to evaluate the manner in which these actions differ kin-

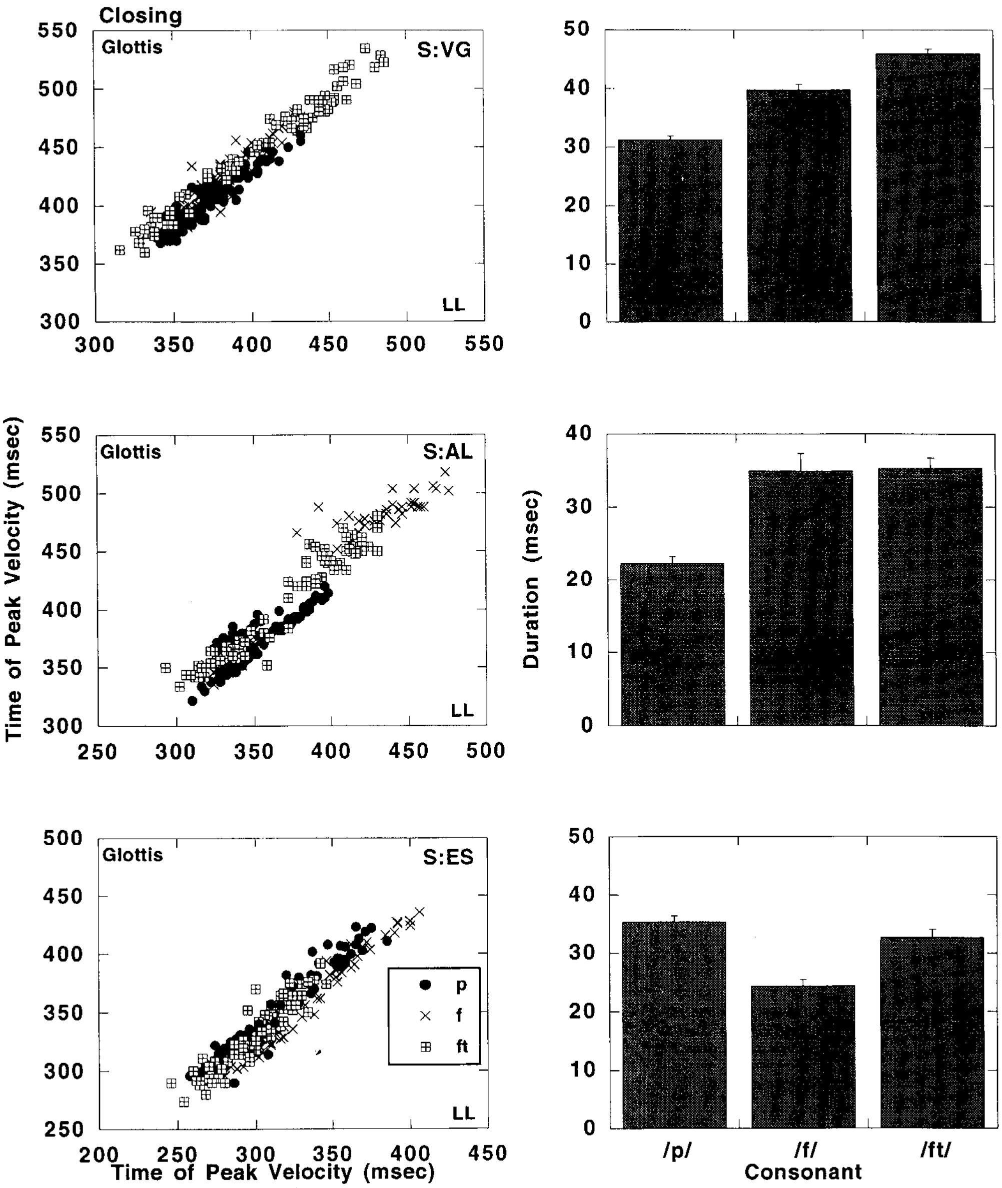

Figure 4. *Left*, Scatterplots of the time the lower lip raising (closing) peak velocity (in milliseconds) as a function of the glottal opening peak velocity for the different consonants (/p/, /f/, and /ft/) for the three subjects. *Right*, Mean differences between the time of the glottal opening peak velocity and lower lip raising peak velocity. The positive difference indicates that the time of lower lip raising peak velocity always preceded the time of glottal opening peak velocity. Error bars indicate 1 SE.

ematically, the movements of the jaw, lip, and larynx were examined in detail. As mentioned previously, all utterances examined were initiated from the same initial conditions. The different vowel sounds resulted in a range of jaw opening displacements and corresponding velocities. The first analysis focused on the relationship between jaw lowering displacement and velocity. As shown on the right side of Figure 6, the correlation of velocity and displacement is quite strong for all subjects. With the exception of the one cluster of data points for subject AL, each subject's velocity/displacement relationship can be described by a single function. The left side of the figure presents the average jaw lowering displacement for the different vowels. It can be seen that jaw displacement varied in a systematic way for the different vowel sounds (see also Macchi, 1988; Oshima and Gracco, 1992). Of the vowels used in the present study, the vowel /ae/ is produced with the lowest jaw position and consequently has the largest opening displacement while the vowel /I/ is produced with the highest jaw position and has the smallest displacement. The range of displacements for the three subjects varied considerably, however, the pattern across subjects was the same.

It was also found that the lowering motion of the jaw was dependent on the identity of the following consonant. For example, the words "supper" and "suffer" have the same vowel but different following consonants. There was a tendency for the jaw opening displacement for the same vowel to be reduced when followed by /f/ than /p/. This is illustrated in the average lip, jaw, and glottal signals presented in Figure 7. Shown are averages (n = 40) of the lower lip, jaw, and glottal signals for the utterances "It's a supper" and "It's a suffer" spoken by S:VG. The vertical jaw lowering displacement is reduced and the resulting jaw raising is of greater displacement and higher position when the consonant is /f/ compared to /p/. As summarized in the top portion of Figure 8, this trend was observed for S:AL but not S:ES. From the middle portion of Figure 8 it can be seen that for the words in which the vowel sound was the same but the consonant was /p/ compared to /ft/ (sipping vs sifting) a similar pattern was observed. In contrast, the bottom portion

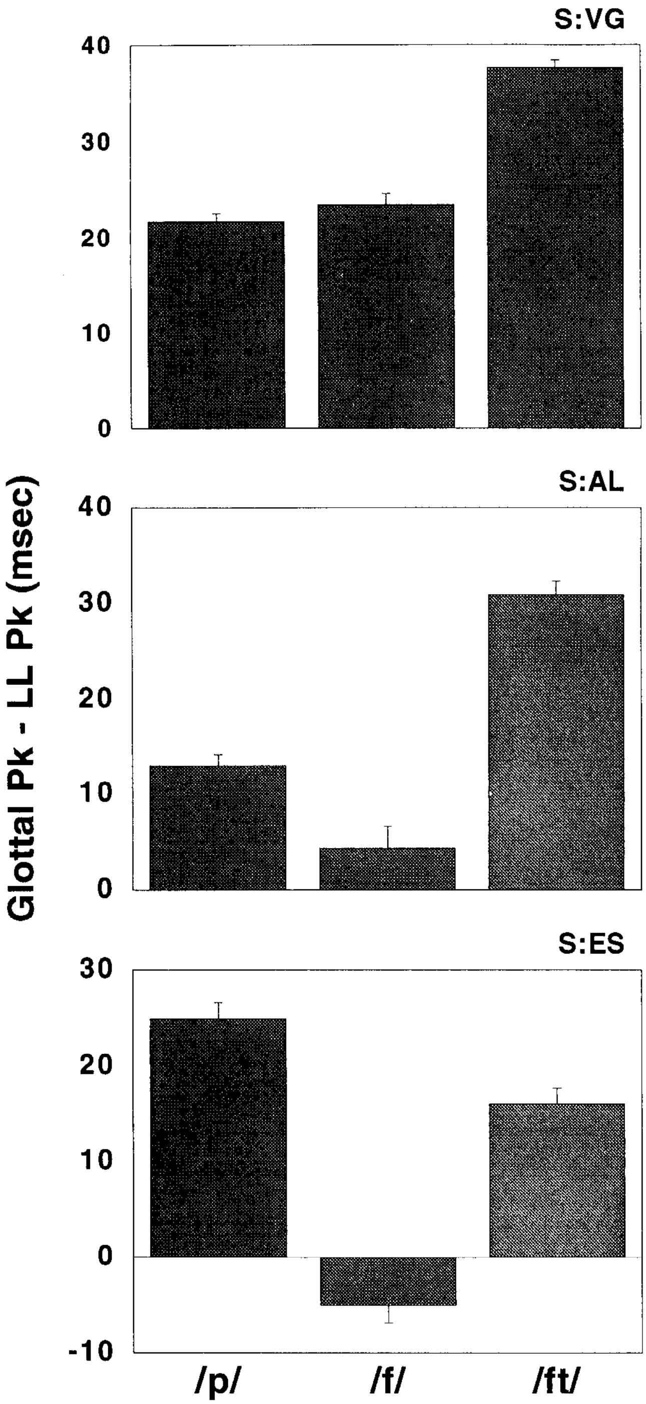

Figure 5. The time interval between the peak glottal opening and lower lip raising for the three subjects. Error bars indicate 1 SE. Similar to the results using the time of peak velocity, the positive values indicate that the maximum displacement for lower lip raising occurred before the maximum glottal opening. Only /f/ for S:ES showed a negative value indicating a reversal in the lip–glottal sequence.

of the figure illustrates that the jaw opening displacement for the same vowel did not differ when the following consonants was /f/ versus /ft/ (safe vs safety). It should also be noted that for the two subjects that showed a reduction in jaw lowering extent when the following sound was /f/ compared to /p/, a similar reduction was noted for the jaw lowering peak velocity. That is, the reduction in the jaw movement displacement was not due to the raising movement moving closer to the lowering movement and truncating the final displacement. Rather, the jaw lowering motion for a specific vowel was actively adjusted dependent on the identity of the subsequent consonant.

Oral closing. Further inspection of the average signals in Figure 7 also reflect some additional characteristics of the differences associated with the identity of the oral closing consonant. The extent of lower lip movement for /p/ and /f/ are similar although there appears to be differences in the velocity of the raising. Second, the glottal aperture is larger for /f/ than for /p/. Figure 9 presents the average displacement and peak lower lip raising velocity for the different contexts for the three subjects. The displacement for /p/ and /f/ were generally similar with the exception of the results for S:AL. In contrast, the closing peak velocity was significantly different with /p/ raising velocity higher than /f/ for all subjects. The movement durations were shorter for /p/ than /f/ for all subjects; the average durations were 136, 89, and 89 msec for /p/ and 175, 138, and 140 msec for /f/ for subjects VG, AL, and ES, respectively. It can also be seen that the consonant sequence /ft/ produces some changes in the kinematic characteristics of the lip raising action. In general the lip displacement is reduced and the velocity is lower for /ft/ compared to /f/ at least for two of the three subjects.

Another example of the effect of two consonants in a sequence can be seen in Figure 10. Shown are averages of the lower lip and jaw movements associated with the words "safe" and "safety." The jaw lowering movement for the vowel is similar for the two words. However, the lip and jaw raising movements are different in extent for /f/ compared to /ft/. Since the jaw is involved in elevating the tongue for the /t/ the jaw continues past the position for /f/. As a result the lip raising action is adjusted for the greater jaw contribution to the initial raising. A summary of the lip and jaw contribution to the oral closing is presented in Figure 11. Since it was shown previously that the jaw lowering movement extent for oral opening varied as a function of the vowel identity, it was necessary to normalize the lip and jaw raising to the extent of jaw lowering for each vowel. A gain was derived as the ratio of the jaw lowering displacement to the lip and jaw raising displacement. As can be seen there is a trend for the gain to be higher for /ft/ than for /p/ and /f/ and the gain for /p/ and /f/ are not significantly different. For the jaw, the gain increases from /p/ to /f/ to /ft/ for two of the three subjects (S:VG and S:AL).

All components of the glottal signal were found to differ according to the consonant sequence. The initial analysis focused on the characteristics of the glottal signal for the different consonants. Each glottal action for a voiceless consonant has two distinct phases; an abductory (opening) phase and an adductory (closing) phase. In order to determine whether each phase of the glottal action is an independent action or an interdependent action in which the phases are modulated as a unit, the peak opening and closing glottal velocities for each consonant as well as the initial /s/ were examined. Shown in Figure 12 are scatterplots of the opening and closing velocities for /s/ (left) and the /p/, /f/, and /ft/ (right) for the three subjects. The opening and closing velocities for both comparisons systematically covary. From the differences in the data ranges it can be seen that the peak glottal velocity for opening and closing for /s/ was always higher than for any of the other three consonants. To varying degrees it was also the case that the opening and closing velocity exhibited a hysteresis with the opening velocity higher

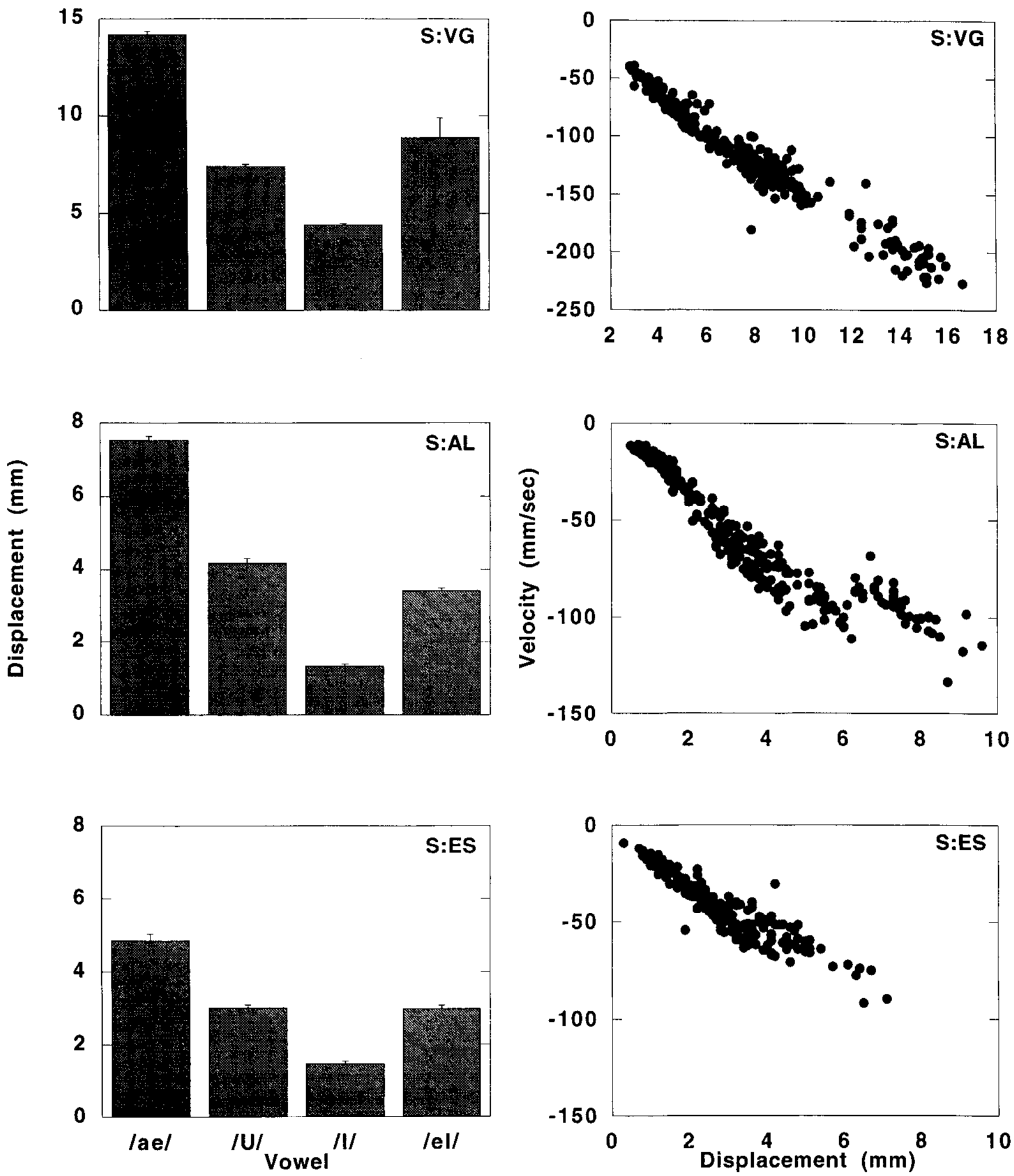

Figure 6. Left, Average jaw lowering displacement (in millimeters) for the four vowels for the three subjects. Error bars indicate 1 SE. *Right*, Scatterplots of the peak lowering velocity (in millimeters/second) as a function of lowering displacement for the four vowels.

than the closing velocity for /s/ and for the other consonants as a group ($p < 0.0001$ for all subjects).

In order to compare the glottal characteristics for the different consonants it was first necessary to amplitude and time normalize the glottal signal. It was reasoned that during the course of the experiment any deviations in the timing or amplitude of the glottal signal unrelated to the phonetic context, such as speaking rate variations or light source changes due to movement of the endoscope, would be evident in the signal for /s/ and normalizing to the /s/ kinematics would minimize any spurious changes. With the exception of the /s/ opening glottal velocity before /ft/ for S:VG there were no significant consonant-related differences for either /s/ opening or closing glottal velocity. As such, all the glottal signals for /p/, /f/, and /ft/ were normalized to the glottal signal for the /s/ in each target word. Shown in Figure 13 are the normalized mean glottal opening peak amplitude, duration, and opening and closing velocities for the three subjects. As shown, the glottal aperture is larger, and opens and closes faster for /f/ compared to /p/. Interestingly, the glottal opening movement duration is longer for /f/ consistent with the slower lip raising movement. Apparently the larger glottal opening is a functional adjustment associated with the aerodynamic or kinematic requirements for /f/ that are different than those for /p/. An additional comparison can be made from the figure. The consonant sequence /ft/ (two voiceless consonants) results in a consistent change in the glottal signal. The glottal aperture for the consonant sequence is larger and longer than for /f/ while the opening and closing velocities are lower. It appears that the glottal signal is some form of additive function of two voiceless phonetic segments.

Discussion

Speaking is a sensorimotor process in which cognitive/linguistic intent is transformed into conformational changes in the vocal tract generating the appropriate conditions for the acoustic structure characteristic of a language. This transformation is a time critical process in which multiple muscles and accompanying speech articulators must be coordinated in space and time to produce a variety of vocal tract adjustments. One purpose of the present investigation was to determine whether different

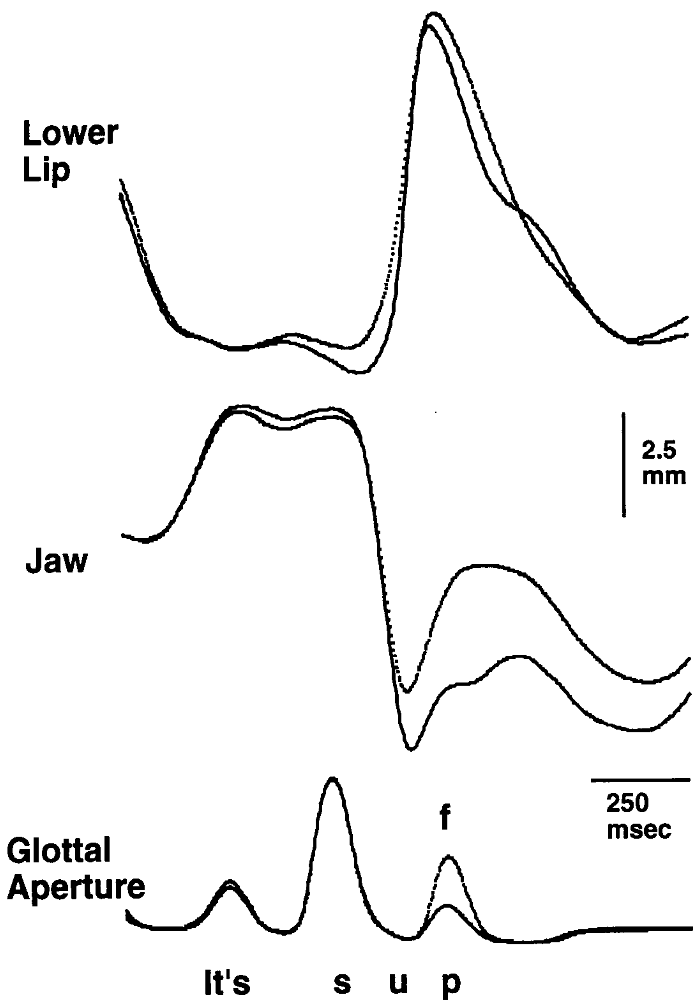

Figure 7. Averaged lower lip, jaw, and glottal signals ($n = 40$) for S:VG for the phrases "It's a supper" and "It's a suffer"; the *dotted line* indicates the /f/. Raising motion is up for the jaw and lower lip; increases in glottal aperture are also up. There are two points of interest: the jaw lowering movement is reduced for "u" when the following consonant is /f/ compared to /p/, and the glottal aperture is larger for /f/ than for /p/.

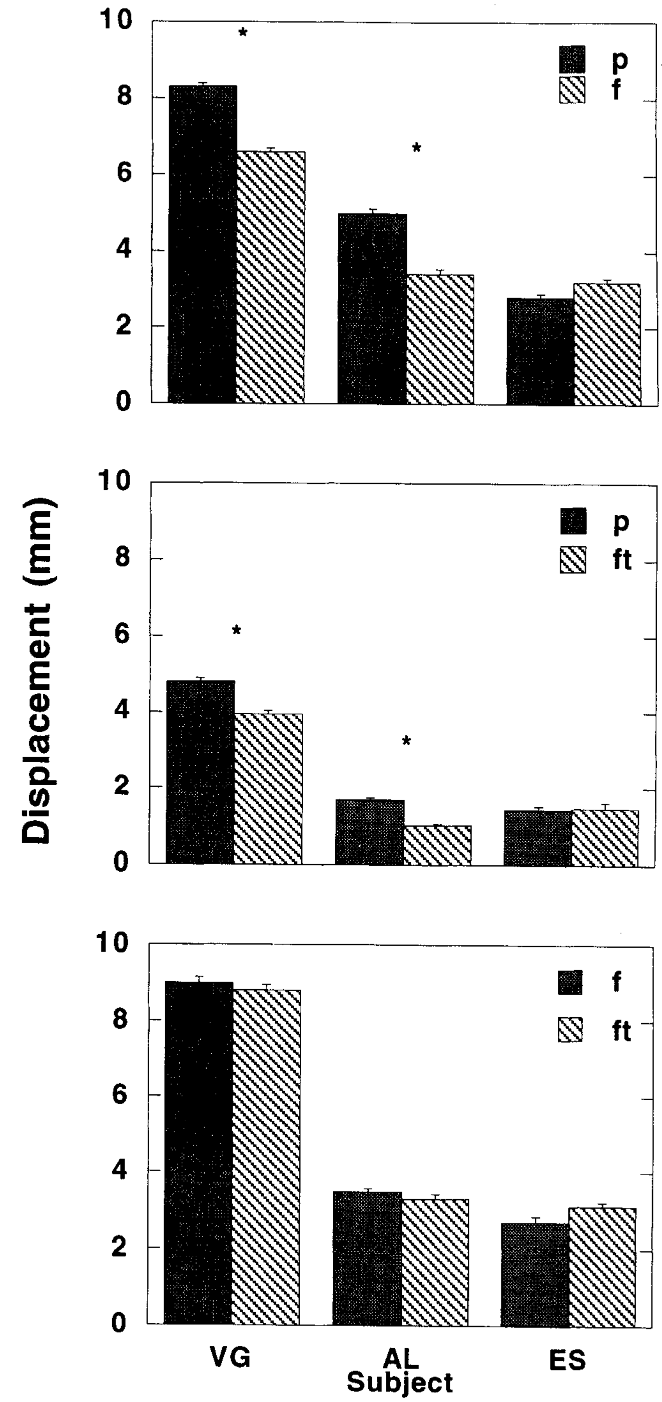

Figure 8. Average jaw lowering displacement (in millimeters) for the three subjects comparing the effects of the following consonant on the preceding jaw lowering movement for the same vowel. The *top panel* contrasts the jaw lowering displacement for the vowel /U/ in the words "supper" and "suffer"; the *middle panel* contrasts the vowel /I/ before /p/ and /ft/ in the words "sipping" and "sifting"; the *bottom panel* contrasts the vowel /eI/ before /f/ and /ft/ in the words "safe" and "safety." Error bars indicate 1 SE. The differences for S:VG and S:AL for the top and middle comparisons were statistically different ($p < 0.001$); for S:ES neither comparisons reached significance ($p > 0.1$). For the /f/, /ft/ comparisons (*bottom*), there were no significant differences ($p > 0.1$).

articulators cooperating to produce the same sound are coupled in their timing and thereby extend previous observations of speech movement coordination to include an important but relatively inaccessible articulator, the larynx. The results suggest that speech movement timing is a highly systematic and constrained process in which individual articulator actions are controlled as a unit rather than as individual degrees of freedom. A second purpose of the present investigation was to provide detail on the size and characteristics of the underlying units for speech production and to determine the manner in which speech movements are adjusted for phonetic context. Interarticulator timing and the different sound-specific articulator adjustments suggest that speech production units are organized at a level reflecting sound generating segments. Finally, phonetic context was found to produce systematic variations in the relative contribution of the different articulators to the overall movement patterns suggesting an important distinction between the units (speech motor programs) and the adjustments of the units (speech motor programming). These issues will be discussed in the following sections.

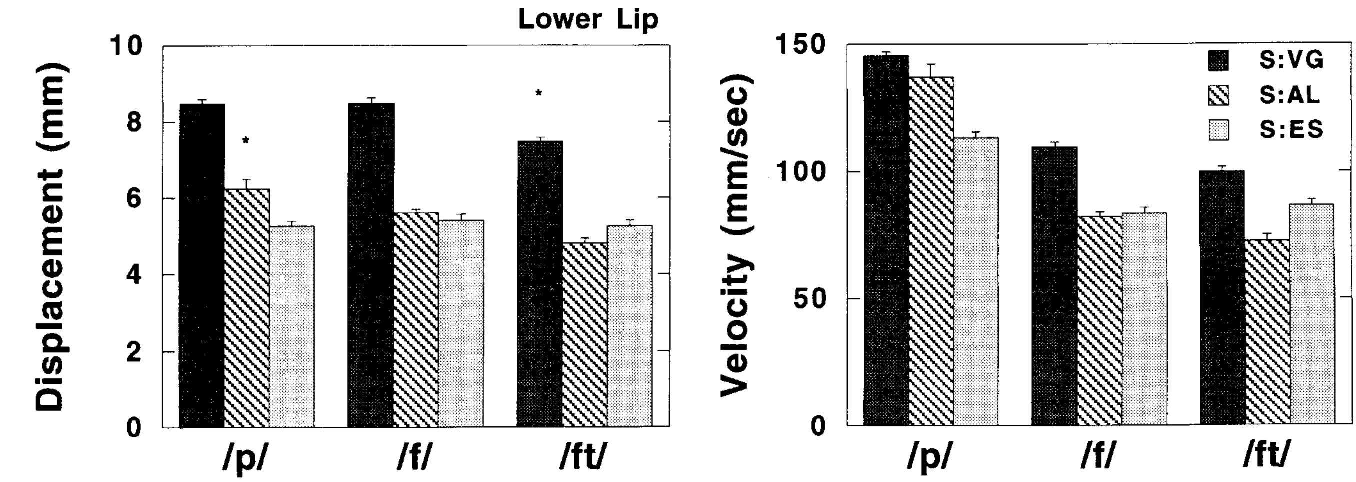

Figure 9. Lower lip raising displacement (*left*) and peak velocity (*right*) for the three subjects for /p/, /f/, and /ft/. There was a slight tendency for a reduction in lower lip displacement for /f/ and/or /ft/ compared to /p/ for two of the subjects (S:VG and S:AL). In contrast, the peak raising velocity demonstrated a robust reduction for /f/ compared to /p/ for all subjects with a smaller difference noted for /ft/ compared to /f/. Error bars indicate 1 SE.

Speech movement coordination

The present results extend previous observations on speech movement timing to include the temporal coordination of the lips, jaw, and larynx. These three articulators are critically involved in many of the sounds of English and the significant covariation in their timing reflect some properties of the speech production process. In previous studies it has been shown that the consistency of speech movement timing is partially dependent on the specific articulator action (opening/lowering or closing/raising) which is generally associated with different classes of speech sounds (e.g., vowels and consonants; Gracco, 1988, 1994). Speech motor actions associated with time critical closing adjustments for certain voiceless consonant sounds, such as /p/ and /f/ in the present study, appear to be highly constrained in their timing. This is apparently to assure that functionally related actions generate the necessary and sufficient aerodynamic conditions to produce perceptually acceptable acoustic products. In contrast to previous results (Gracco, 1988, 1994) which demonstrated a lack of robust relative timing among the lips and the jaw during oral opening, the present results suggest that similar constraints are operating for oral opening actions as well. The difference from previous studies is related to the articulators examined. In the previous studies, the lips and jaw were only examined and the apparent difference between the two general actions appears to be related to different articulators being involved in different linguistic-motor actions that overlap in time (see also Gracco, 1994). At the onset of oral opening the lips are still involved in the consonant sound while the jaw becomes functionally decoupled from the consonant and is directly involved in the following vowel sound. As shown in the present investigation for oral opening, the timing of the jaw and larynx are coupled in their relative timing as their actions are functionally related to the production of the vowel. The systematic and consistent timing covariation among the articulators for oral opening and closing indicate that timing constraints may be a fundamental property of speech movement coordination. However, it is the case that the relative strength of the coupling varied, with the oral closing actions more highly correlated than oral opening. One possible explanation is that these two actions, oral closing and oral opening, reflect two important but distinct characteristics of speech production. In the case of the oral closing, examination of the relative timing focused on a single speech motor action (the production of a specific phoneme) and the consistent timing of the movements represents the coordination of multiple speech articulators within a specific action unit. In the case of oral opening, examination of the relative timing focused on a transition region between two contiguous speech actions (the transition between a consonant and a vowel). As such, the present investigation examined multiarticulator coordination within a speech production unit and the sequencing of such units into larger aggregates. It is also the case that the oral closing actions were associated with rapid movements

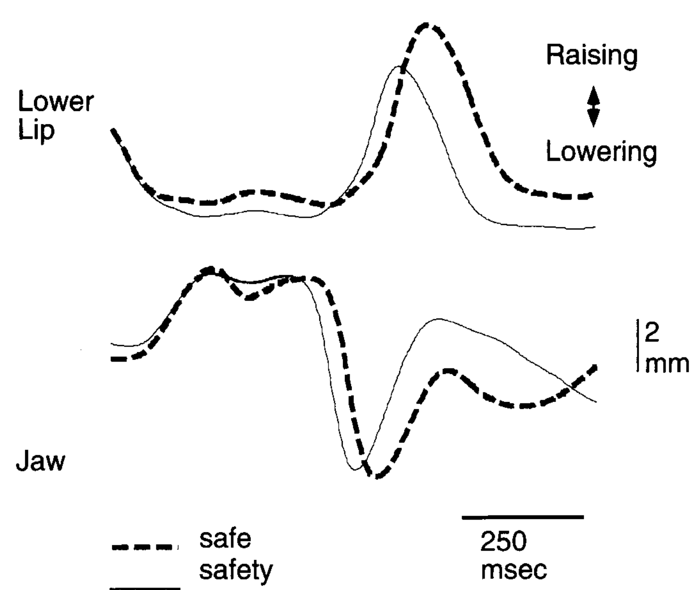

Figure 10. Average lower lip and jaw vertical movements ($n = 40$) for S:VG illustrating the displacement differences due to the consonant sequence /ft/ compared to /f/. The *dashed line* indicates the word "safe"; the *solid line* indicates the word "safety." The lower lip is reduced in raising displacement and increased in jaw displacement for the /ft/.

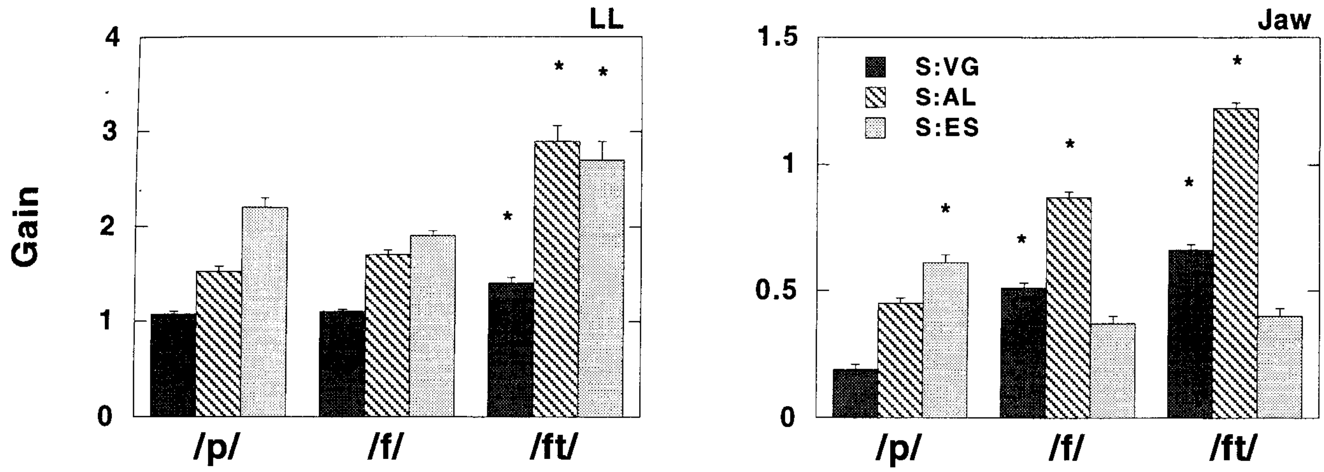

Figure 11. Lower lip (*LL*; *left*) and jaw gain (*right*) defined as the ratio of the raising displacement to the opening displacement for the preceding vowel. Lower lip gain is consistently higher for the consonant sequence /ft/ as is the jaw with the exception of S:ES. In addition, the jaw gain is higher for /f/ compared to /p/ for S:VG and S:AL. *Asterisks* indicate a significant difference ($p < 0.01$).

and high pressure consonants sounds while the oral opening actions were associated with slower movements and low pressure vowel sounds. The extent to which these factors influence the relative timing among speech articulators is open to empirical investigation. The next section will focus on some of the characteristics of the units for speech production followed by a discussion of the potential mechanisms for sequencing and adjusting the units for phonetic contexts.

Speech motor programs

Speech motor programs can be thought of either as high level goals or procedures for implementation of intent (Schaffer, 1992). A synthesis of these two views can been suggested in which speech motor programs are viewed as neuromuscular configurations that define the structure (intent) of the vocal tract for each unique element (sound) of the language (Gracco, 1990, 1991). Speech motor programs reflect a characteristic neuromuscular configuration that specifies the muscles to be activated and some general characteristics of that activation (Gracco, 1991, 1994). Similar to the concept of a motor plan (Evarts et al., 1971) a finite number of such programs would be established during speech motor development and modified periodically for changes in vocal tract shape due to growth. As noted here and elsewhere, speech motor actions appear to be organized at a functional (sound producing) level with control exerted over large regions of the vocal tract rather than over the action of individual articulators (Gracco and Abbs, 1986; Gracco, 1990, 1991). The present results are consistent with this conception in that for the three different consonant sounds examined (/s/, /p/, /f/) the articulatory configurations were unique and significantly different along a number of kinematic dimensions. Based on these and previous results demonstrating the consistent relative timing among functionally related articulator coordination it is further suggested that an important component of each speech motor program is the relative timing among the neuromuscular elements. Rather than explicitly controlling the timing among articulators such as the lips, jaw, and larynx, their coordination is an inherent component of the unit (program). These learned motor programs are stored in memory and provide the physiological framework for the sounds of the language reflecting the physiological instantiation of the phoneme.

Speech motor programming

One of the criticisms with the construct of motor programs underlying voluntary behavior is the lack of adaptability often cited as a limitation for such a metaphor (Kelso, 1986; Kugler and Turvey, 1987). The lack of adaptability is of some significance since it is well known that the phonetic context of a particular sound can substantially modify its peripheral (kinematic and acoustic) manifestations. The same sound produced at the beginning versus the end of a syllable and between different vowels will display different movement patterns. The widespread presence of contextual variation has even led to an extreme, though currently unpopular view, that all possible variations of the sounds of the language are stored in memory as part of the speech coding process (Wickelgren, 1969). Based on the results from the present investigation and results from investigations of the sensorimotor mechanisms of speech motor control, a more realistic perspective can be presented. A number of investigations have demonstrated that mechanical perturbations to the lips and jaw result in short-latency (within a reaction time) responses in all the articulators activated for the specific speech sounds (Folkins and Abbs, 1975; Abbs and Gracco, 1984; Kelso et al., 1984: Gracco and Abbs, 1985; Shaiman, 1989). It appears that somatic sensory receptors located within the vocal tract have the requisite properties to interact with the central motor commands to provide adaptive adjustments in the speech motor programs resulting from peripheral variations in phonetic context (Gracco, 1987; Gracco and Abbs, 1988). On-line sensorimotor mechanisms provide one means to adjust central commands to changes in peripheral conditions. The present results also suggest that an additional central mechanism is operating for contextual adjustments. For two of the three subjects it was shown that the jaw lowering extent for the vowel /U/ was affected by the identity of the following consonant. When the following consonant was /f/ the jaw lowering movement was reduced in amplitude compared to when the following consonant was /p/. This affect was not merely the result of the

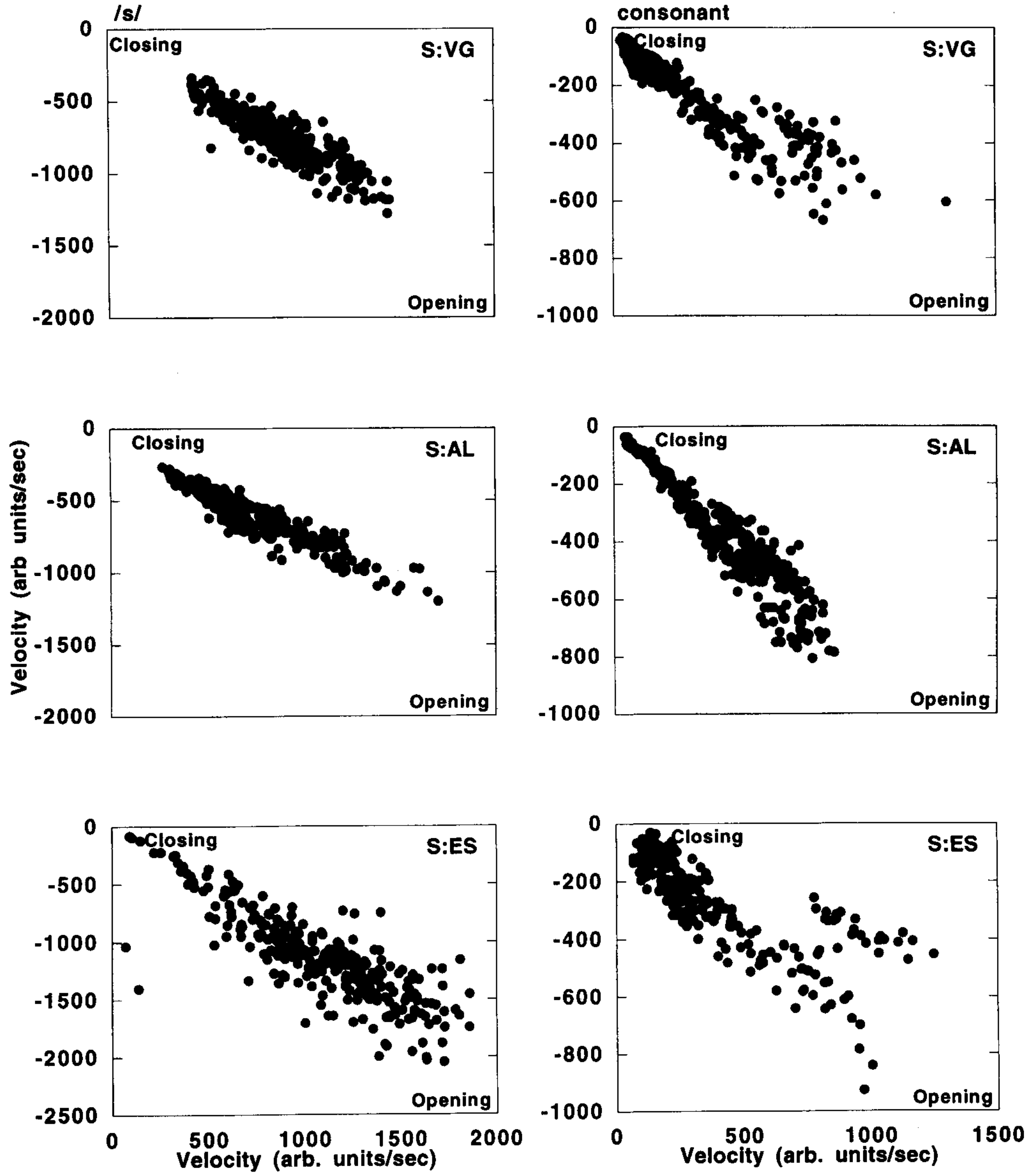

Figure 12. Scatterplots of the glottal opening and glottal closing peak velocity for the /s/ in the different words (*left*) and the consonants /p/, /f/, and /ft/ (*right*) for the three subjects. The opening and closing velocities covary strongly for both conditions. Moreover, the values are generally higher for /s/ than for any of the consonants and the opening velocity is generally higher than the closing velocity.

consonantal raising movement truncating the jaw lowering movement since the jaw lowering velocity was also reduced in the /f/ context. This phenomenon of coarticulation, or the anticipatory modification of speech output due to context, is of some neurophysiological significance. It suggests that what is to be said is planned in advance and the overall context can influence aspects of the central commands (see also Whalen, 1990). These two complementary processes operating on a framework of learned motor programs provide the flexibility characteristic of speech production. It suggests that a distinction can be made between speech motor programs, as goal directed phonetically based actions, and dynamic (programming) processes that provide adaptive and on-line adjustments to speech motor sequences. Moreover, speech motor adjustments associated with the consonant were distributed to the preceding vowel action suggesting that the speech motor programming operates over an interval on the order of a movement cycle involving two or more phonetic segments (see also Gracco, 1994).

Speech motor sequences

The present investigation allowed an examination of the kinematic effects of two consonants produced in sequence. The consonant sequence /ft/ involves two voiceless sounds that overlap. Functionally, the consonant sequence /ft/ requires the lip to contact the upper teeth followed by tongue tip contact with the roof of the mouth for the /t/. As shown in Figure 10 comparing /safe/ with /safety/ the jaw position is higher for /safety/. It can also be seen that the jaw moves continuously from a minimum for the vowel to some maximum value associated with the /t/. The jaw position for the /f/ in the /ft/ sequence was not unambiguously identifiable. This is a characteristic of many speech motor actions and is the basis for the difficulty in segmenting

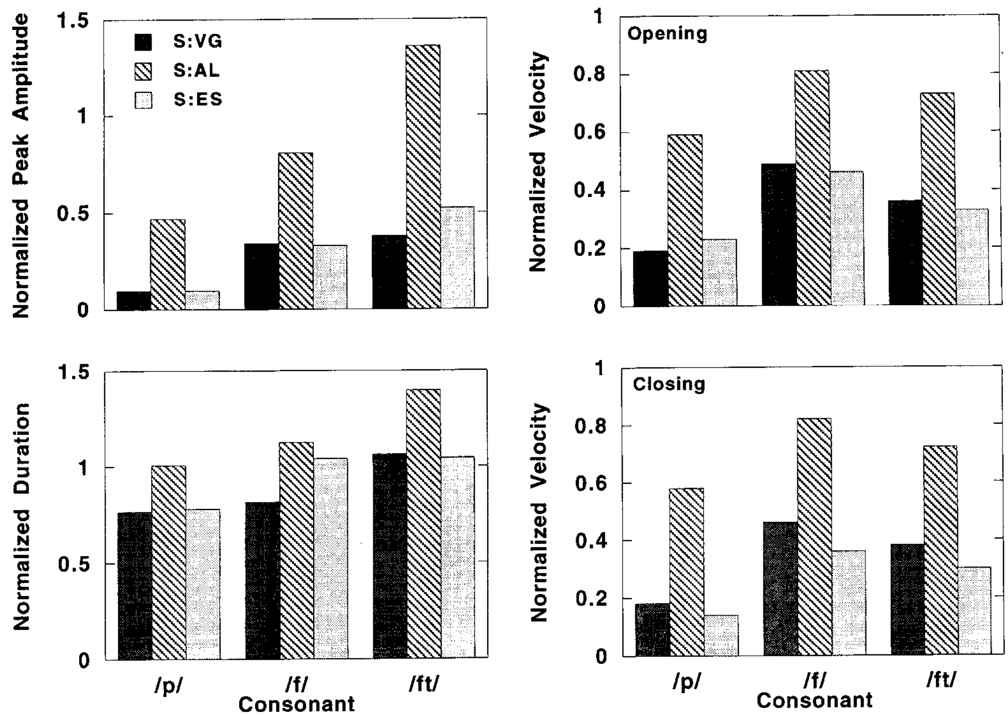

Figure 13. Normalized peak glottal opening, opening duration, and opening and closing peak velocity for the three consonants. The glottal opening is larger and longer in duration for /ft/ compared to /f/, which is larger and longer in duration for /f/ compared to /p/ for all subjects. The opening and closing velocity is greater for /f/ compared to /p/ with a consistent reduction in velocity for /ft/ for the three subjects.

continuous motion into the underlying discrete units. While the jaw passes through some spatial target for /f/ it does not (and need not) stop its motion. It is suggested from jaw movement considerations that two successive target positions are reflected in the single trajectory (see also Flanagan et al., 1993, for arm movements to displaced targets). For the larynx, the consonant sequence also resulted in an apparent blending of the two voiceless consonants with the resulting glottal amplitude and/or duration for /ft/ larger and/or longer than /f/ (see also Munhall and Löfqvist, 1992). As such, the consonant sequence produced a hybrid pattern adjusted to accommodate the longer duration voiceless segment.

Conclusions

The present investigation was initiated to evaluate the coordination and motor control for speech by examining the interactions of the lips, jaw, and larynx in different phonetic contexts. While limited in scope the present results suggest a number of general properties of speech production and its motor control. The timing among functionally related articulators suggests that speech movements are organized into aggregates larger than individual articulators. The different kinematic patterns associated with the different consonant and vowel sounds examined in the present investigation further suggest that different sounds have different neuromotor specifications. It appears that each sound in the language has associated with it a neuromotor representation reflecting the muscles to be activated and their specific spatiotemporal coordination. These fundamental units (speech motor programs, coordinative structures) provide the framework for speech production and appear to reflect the neurobiological equivalent of the phoneme. Additional modulatory processes exist to scale and sequence the phonetic units into larger sequences for communication (syllables, words, phrases, etc.) by adjusting the vocal tract characteristics over an interval larger than individual phonemes (phonetic segments). As suggested recently, the unit of programming is, minimally, on the order of a movement cycle (or syllable) and within this interval contextual variations are adjusted based on the immediate state of the vocal tract and the compatibility of the neighboring sounds (Saltzman and Munhall, 1989; Gracco, 1994). The dynamic nature of speech production results in blending of movements that modify the peripheral manifestation of the underlying units and obscures their identification. As such, the neural control specifications of the units must be sufficiently relaxed to allow for contextual variations. This is also reflected in the ability of the listener's perceptual system to handle the lack of invariance and maintain highly reliable information transfer.

References

Abbs JH, Gracco VL (1984) Control of complex motor gestures: orofacial muscle responses to load perturbations of the lip during speech. J Neurophysiol 51:705–723.

Baer T, Löfqvist A, McGarr NS (1983) Laryngeal vibrations: a comparison between high-speed filming and glottographic techniques. J Acoust Soc Am 73:1304–1308.

Bernstein N (1967) The co-ordination and regulation of movements. New York: Pergamon.

Evarts EV, Bizzi E, Burke R, DeLong M (1971) Central control of movement. Neurosci Res Prog Bull 6:1–70.

Flanagan JR, Ostry DJ, Feldman AG (1993) Control of trajectory modifications in target-directed reaching. J Mot Behav 25:140–152.

Folkins JW, Abbs JH (1975) Lip and jaw motor control during speech: responses to resistive loading of the jaw. J Speech Hearing Res 18: 207–220.

Gracco VL (1987) A multilevel control model for speech motor activity. In: Speech motor dynamics in stuttering (Peters IH, Hulstijn W, eds), pp 57–76. Wien: Springer.

Gracco VL (1988) Timing factors in the coordination of speech movements. J Neurosci 8:4628–4634.

Gracco VL (1990) Characteristics of speech as a motor control system. In Cerebral control of speech and limb movements (Hammond HG, ed), pp 3–28. Amsterdam: North Holland/Elsevier.

Gracco VL (1991) Sensorimotor mechanisms in speech motor control.

In: Speech motor control and stuttering (Peters H, Hultsijn W, Starkweather CW, eds), pp 53–78. Amsterdam: North Holland/Elsevier.
Gracco VL (1994) Some organizational characteristics of speech movement control. J Speech Hearing Res 37:4–27.
Gracco VL, Abbs JH (1985) Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. J Neurophysiol 54:418–432.
Gracco VL, Abbs JH (1986) Variant and invariant characteristics of speech movements. Exp Brain Res 65:156–166.
Gracco VL, Abbs JH (1987) Programming and execution processes of speech motor control. Potential neural correlates. In: Motor and sensory processes of language (Keller E, Gopnick M, eds), pp 163–202. Hillsdale, NJ: Erlbaum.
Gracco VL, Abbs JH (1988) Central patterning of speech movements. Exp Brain Res 71:515–526.
Kelso JAS (1986) Pattern formation in speech and limb movements involving many degrees of freedom. In: Generation and modulation of action patterns (Heuer H, Fromm C, eds), pp 105–128. Berlin: Springer.
Kelso JAS, Tuller B, V-Bateson E, Fowler C (1984) Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. J Exp Psychol [Hum Percept] 10:812–832.
Kugler PN, Turvey MT (1987) Information, natural law and the self-assembly of rhythmic movement. Hillsdale: Erlbaum.
Lacquaniti F, Soechting JF (1982) Coordination of arm and wrist motion during a reaching task. J Neurosci 2:399–408.
Löfqvist A, Yoshioka H (1980) Laryngeal activity in Swedish obstruent clusters. J Acoust Soc Am 68:792–801.
Macchi M (1988) Labial articulation patterns associated with segmental features and syllable structure in English. Phonetica 45:109–121.
MacKenzie CL (1992) Constraints, phases and sensorimotor processing in prehension. In: Tutorials in motor behavior (Stelmach GE, Requin J, eds), pp 181–194. Amsterdam: Elsevier.
MacPherson JM (1988a) Strategies that simplify the control of quadrupedal stance. I. Forces at the ground. J Neurophysiol 60:204–217.
MacPherson JM (1988b) Strategies that simplify the control of quadrupedal stance. II. Electromyographic activity. J Neurophysiol 60: 218–231.
Munhall K, Löfqvist A (1992) Gestural aggregation in speech: Laryngeal gestures. J Phonetics 20:111–126.
Oshima K, Gracco VL (1992) Mandibular contributions to speech production. In: Proceedings of the Conference on Spoken Language Processing, pp 775–778.
Saltzman E, Munhall K (1989) A dynamical approach to gestural patterning in speech productions. Ecol Psychol 1(4):333–382.
Schaffer LH (1992) Motor programming and control. In: Tutorials in motor behavior (Stelmach GE, Requin J, eds), pp 181–194. Amsterdam: Elsevier.
Shaiman S (1989) Kinematic and electromyographic responses to perturbation of the jaw. J Acoust Soc Am 86:78–87.
Soechting JF, Lacquaniti F (1989) An assessment of the existence of muscle synergies during load perturbations and intentional movements of the human arm. Exp Brain Res 74:535–548.
Turvey MT (1977) Preliminaries to a theory of action with reference to vision. In: Perceiving, acting and knowing: towards an ecological psychology (Shaw R, Bransford J, eds). Hillsdale, NJ: Erlbaum.
Whalen DH (1990) Coarticulation is largely planned. J Phonetics 18: 3–35.
Wickelgren (1969) Context-sensitive coding, associative memory, and serial order in (speech) behavior. Psychol Rev 76:1–15.