

Michael D. McClean · Stephen M. Tasko

Association of orofacial with laryngeal and respiratory motor output during speech

Received: 19 October 2001 / Accepted: 5 June 2002 / Published online: 6 September 2002
© Springer-Verlag 2002

Abstract Speech motor coordination most likely involves synaptic coupling among neural systems that innervate orofacial, laryngeal, and respiratory muscles. The nature and strength of coupling of the orofacial with the respiratory and laryngeal systems was studied indirectly by correlating orofacial speeds with fundamental frequency, vocal intensity, and inspiratory volume during speech. Fourteen adult subjects repeated a simple test utterance at varying rates and vocal intensities while recordings were obtained of the acoustic signal and movements of the upper lip, lower lip, tongue, jaw, rib cage, and abdomen. Across subjects and orofacial speed measures (14 subjects × 4 structures), significant correlations were obtained for fundamental frequency in 42 of 56 cases, for intensity in 35 of 56 cases, and for inspiratory volume in 14 of 56 cases. These results suggest that during speech production there is significant neural coupling of orofacial muscle systems with the laryngeal and respiratory systems as they are involved in vocalization. Comparisons across the four orofacial structures revealed higher correlations for the jaw relative to other orofacial structures. This suggests stronger connectivity between neural systems linking the jaw with the laryngeal and respiratory systems. This finding may be relevant to the frame/content theory of speech production, which suggests that the neural circuitry involved in jaw motor control for speech has evolved to form relatively strong linkages with systems involved in vocalization.

Keywords Speech motor coordination · Jaw · Lips · Tongue · Laryngeal · Respiratory · Human

Introduction

Speech production involves the integrated activity of neural systems that regulate orofacial, laryngeal, and respiratory movements. Areas of the brain involved in this process include sensorimotor cortex, cerebellum, basal ganglia, thalamus, auditory pathways, and the brain stem (Barlow and Farley 1989). It has been effectively argued that the neural circuitry and peripheral anatomy underlying speech motor control have evolved and specialized in order to meet its unique demands (Lieberman 1984; MacNeilage 1998). A critical aspect of such specialization most likely concerns the coordination of orofacial structures with those of the laryngeal and respiratory systems. The high degree of temporal-spatial precision involved in this process is illustrated by the rapid posturing of the vocal folds that occurs when voiceless consonant sounds are preceded and followed by vowels (see, e.g., Gracco and Löfqvist 1994).

In light of the importance of coordinating the lips, tongue, and jaw with laryngeal and respiratory movements for normal speech production, improved quantitative measures are needed that reflect the strength of neural coupling between these systems. The concept of neural coupling is used here in a broad sense to include both common efferent drive to the motoneurons of different subsystems (e.g., the lips and larynx) and mechanoreceptor input between subsystems. Evidence for this latter type of neural coupling is provided by studies involving mechanical stimulation of the peripheral speech structures. For example, Munhall et al. (1994) described rapid compensations in laryngeal movements during speech in response to unexpected mechanical perturbations of the lips. Also, it is known that perioral mechanical stimulation can evoke short-latency laryngeal reflex responses as reflected in voice fundamental frequency changes (Larson and Sapir 1995).

M.D. McClean (✉) · S.M. Tasko
Audiology and Speech Center, Walter Reed Army Medical Center,
Washington, DC 20307-5001, USA
e-mail: michael.mcclean@wmich.edu
Tel.: +1-616-3878049
Fax: +1-616-3878044

Present address:

M.D. McClean, Department of Speech Pathology and Audiology,
Western Michigan University, Kalamazoo, MI 49008-3825, USA

Present address:

S.M. Tasko, Department of Speech Pathology and Audiology,
Western Michigan University, Kalamazoo, MI 49008-3825, USA

Understanding the neural coupling among speech motor subsystems can be facilitated by the use of correlation and parametric analyses to relate the peripheral output of different muscle systems during natural speech. This approach is illustrated in a study of Gracco and Löfqvist (1994), who found that the timing of the glottal closing gesture during speech was highly correlated with the timing of lower-lip and jaw movements. Evidence for correlated output from speech motor subsystems also is provided by the work of Dromey and Ramig (1998a, 1998b), who reported increased lip velocities during speech at elevated lung volumes and vocal intensities. Additional data suggesting neural linkage between the orofacial and respiratory systems can be found in studies that address the maintenance of upper airway patency for vegetative breathing. For example, orofacial muscle activity is known to vary systematically across distinct phases of nonspeech breathing (Fregosi and Fuller 1997; Lansing et al. 1991; Sauerland and Mitchell 1975).

Improved quantitative measures of neural coupling of the orofacial with the laryngeal and respiratory neuromotor systems would contribute to improved understanding of group and individual characteristics of normal and disordered speech motor coordination. Such measures also would permit the study of as yet unexplored aspects of speech motor coordination. For example, it has been suggested that over the course of human evolution the neural systems regulating cyclic jaw movements for speech shifted from the lateral premotor cortex controlling mastication to areas of medial premotor cortex more closely linked to phylogenetically older regions involved in animal vocalization (MacNeilage 1998). Also, there is evidence that vocal tract modulation of the voice signal by infants during the babbling stage of speech development is achieved primarily with oscillatory jaw movement (Davis and MacNeilage 1995; Green et al. 2000). Thus, during speech acquisition relatively strong neural coupling is established between jaw movement and modulation in the amplitude of the acoustic speech signal. Based on these considerations, one might reasonably expect stronger neural linkage of the laryngeal and respiratory systems with the jaw compared to the lips and tongue.

The present study explores the strength and pattern of association of orofacial motor output with that of the laryngeal and respiratory systems during speech as reflected in kinematic and acoustic measures. Correlations between output parameters of different speech muscle systems could be due to biomechanical, acoustic, and neural interaction among the systems involved. In this study, output measures have been selected that can reasonably be assumed to be positively associated with the level of muscle activation. Significant correlations between output measures of different muscle systems would presumably then be due to common efferent drive to the relevant motoneuron pools and/or mechanoreceptor input from one system to another.

In the case of the orofacial system, measures of lip, tongue, and jaw peak speed or tangential velocity were employed as the primary output measure. Peak speed was selected in part because EMG and kinematic data indicate a positive association of lip muscle activation levels and velocity (Gay and Hirose 1973; McClean and Clay 1995). Also, orofacial mechanoreceptors are especially sensitive to the velocity of mechanical stretch (Larson et al. 1978), and are likely to respond well to tissue velocity changes during natural movements (Grill and Hallet 1995). While comparable analyses of orofacial movement have not been carried out, neuroimaging data indicate that activation levels within sensorimotor cortex, cerebellum, thalamus, and basal ganglia are positively associated with arm movement velocity (Turner et al. 1998).

Laryngeal motor output in the present study is inferred chiefly from measures of fundamental frequency. The focus on fundamental frequency derives from the results of numerous studies demonstrating positive associations between fundamental frequency and the level of intrinsic laryngeal muscle activation (Atkinson 1978; Finnegan et al. 2000; Sapir et al. 1984; Titze et al. 1989). Muscular drive to the respiratory system is inferred from the extent of inspiratory chest wall movements on individual speech breaths, considering evidence that inspiratory volume reflects level of inspiratory muscle activity (e.g., Gandevia et al. 1999). Vocal intensity is also studied, but it is assumed to be related to muscle activation within both the respiratory and laryngeal muscle systems (Baker et al. 2001; Finnegan et al. 2000; Hixon et al. 1973; Stathopoulos and Sapienza 1993).

Correlation analyses were carried out relating upper lip, lower lip, jaw, and tongue speed to vowel fundamental frequency, vowel intensity, and inspiratory volume during speech production. The principal goals were to determine the strength and frequency of occurrence of significant correlations among various output measures. It is assumed that the size and direction of resulting correlations indirectly reflect the strength of neural coupling between underlying muscle systems as they contribute to the speech motor process. In light of the suggestion that neural systems regulating jaw movement during speech have relatively strong connections with brain regions regulating vocalization (MacNeilage 1998), we evaluated the hypothesis that there is stronger neural coupling of the laryngeal and respiratory systems with the jaw compared to the lips and tongue.

Materials and methods

Data acquisition procedures

Data acquisition procedures were carried out on 15 adult male speakers of American English with normal speech and hearing. All subjects gave their informed consent prior to participation in the study. The study was approved by the Walter Reed Army Medical Center's Human Use Committee and therefore performed in accordance with the ethical standards of the 1964 Declaration of Helsinki. Data from one of the subjects were not included in the

analysis due to unusual kinematic patterns seen in lower-lip and jaw movement. Specifically, the speed-distance plots for this subject did not display uniform linear scatter and showed marked clustering across conditions, making it inappropriate to perform partial correlation analysis. The remaining 14 subjects, whose data were used, had a mean age of 22.4 years, ranging from 18 to 52 years.

The experimental procedure involved simultaneous kinematic and audio signal recording while subjects produced the utterance “a bad daba” at varying levels of vocal loudness and speech rate. This speech utterance was selected primarily because it involves large amplitude orofacial movements that are amenable to automated software measurement. The purpose of varying vocal loudness and rate was to obtain a wide output range in kinematic and acoustic measures of interest in order to facilitate correlation analysis. Subjects produced the test utterance in five 30-s blocks, with each block corresponding to one of five distinct speaking conditions. All subjects produced these conditions in the following order: normal speech (i.e., habitual loudness and rate), loud speech, soft speech, fast speech, and slow speech. There was an interval of approximately 2 min between conditions in which subjects received further instruction and practiced the upcoming condition. During this time, they repeated the test utterance a number of times at a normal loudness and rate. For the loud and soft conditions, subjects were briefly trained to produce the test utterance at a normal rate but at intensity levels that were approximately 6 dB above and 6 dB below their normal level. For the fast and slow rate conditions, they were instructed to produce the test utterance at a normal intensity but at rates they judged to be twice and half their normal speech rates. It should be noted that when speakers perform these tasks, speech rate and intensity tend to covary. That is, significant changes in speech rate occur with variations in intensity and vice versa. However, the pattern of these changes differs markedly across speakers. The primary intent here was not to vary speech rate and intensity in a parametric manner, but rather to have speakers produce a wide range of output within the parameter spaces of interest in order to enhance the sensitivity of the correlation analysis.

Recordings of the two-dimensional positions of the upper lip, lower lip, tongue blade, and jaw within the midsagittal plane were obtained with a Carstens AG100 electromagnetic movement analysis system (Tuller et al. 1990). Sensor coils (3×2×2 mm) were attached to the bridge of the nose, upper-lip vermillion, and lower-lip vermillion with biomedical tape, and to the tongue blade (1 cm from the tip) and base of the lower incisors with surgical adhesive (Isodent). An Ambulatory Monitoring respirace system was used to measure circumferential displacements of the rib cage and abdomen. Vital capacity and isovolume maneuvers were recorded prior to recording the speech sample. Subjects performed isovolume maneuvers by closing off their upper airway near their resting expiratory level and then moving the air volume in a cyclic manner between the rib cage and abdomen. The acoustic speech signal was transduced with a Shure M93 miniature condenser microphone positioned 7.5 cm from the mouth. The microphone-amplifier setup was calibrated to permit measurement of absolute sound pressure levels. This calibration was performed with a 1-kHz tone having a known absolute sound pressure level. The orofacial kinematic signals and the audio signal were digitized respectively at 250 Hz/channel and 16 kHz to one computer. The chest wall kinematic signals and audio signal were digitized at 2 kHz/channel to a second computer.

Signal processing and measurement

The orofacial and respiratory movement signals acquired with the two computers were synchronized offline. This was done by down sampling the 16 kHz audio signal to 2 kHz, then cross correlating the two audio signals, and adjusting for the phase lag of the peak of the correlation function. The upper lip, lower lip, jaw, and tongue movement signals were low-pass filtered at 8 Hz, and the chest wall and nose signals were low-pass filtered respectively at 5 Hz and 3 Hz. All filtering was performed with a fifth-order Butterworth

filter using a zero-phase distortion digital filter (Matlab 5). While head movements during recording were slight (<1.0 mm), nose sensor movements were subtracted from the upper lip, lower lip, tongue, and jaw movement signals in the *X* and *Y* dimensions in order to minimize any head movement contributions. The lower lip and tongue signals were then decoupled from the jaw using a method described by Westbury et al. (2002). This technique was developed for situations where only one point on the jaw is recorded, and it takes into account both jaw translation and rotation. In this method jaw rotation is estimated by assuming one half degree of rotation for every millimeter displacement away from the highest jaw position along its first principal component of movement. The half-degree value was empirically derived from a kinematic dataset of 42 speakers producing a single sentence with two landmarks recorded on the jaw along with landmarks on the lips and tongue. Westbury et al. compared lower lip and tongue movement signals obtained with the rotation estimate method with those obtained using two jaw landmarks and found that the error levels for displacement and speed were minimal. Specifically, for the lower lip and tongue blade, the median positional errors obtained with the rotation estimate method were 0.13 mm and 0.18 mm respectively. The interquartile ranges of the distribution of speed errors centered approximately on zero were 1.5 mm/s and 1.4 mm/s respectively for the lower lip and tongue blade.

Speed histories for the upper lip, lower lip, jaw, and tongue blade were derived from the *X* and *Y* velocities, as obtained with a three-point central method for differentiation. For the test utterance, these speed histories showed well-defined peaks associated with lip closing on /bæd/, lip opening and tongue closing in /bæd/, and tongue opening and lip closing in /dæb/. The various speed peaks are identified in Fig. 1 with capital letters and numbers. Measures of distance and peak speed were obtained on the 13 individual movements associated with the speed peaks labeled in Fig. 1. Interactive software was used to identify individual speed peaks and speed minima bounding each peak. These minima typically were associated with a change in the movement direction, and therefore the intervening period corresponded to a closing or opening movement. The five jaw speed peaks in Fig. 1 correspond respectively to vocal tract closing and opening movements on initial /b/, closing and opening on /d/, and closing on the final /b/. The lip and tongue speed peaks occurring in synchrony with these jaw peaks are associated with the lip and tongue closing and opening. The speed history for each structure was integrated over the interval between the minima, which provided measures of movement distance.

Peak speed served as the primary dependent measure of orofacial function, and movement distance was employed as an independent control variable. For each orofacial structure, peak speed and distance measures were based on the mean peak speed and distance across the different movements. Averaging across movements was judged appropriate, because the interest here was in global variations in level of muscle system activity associated with the entire test utterance. Also, analyses in this and an earlier study (McClellan 2000) have shown consistent positive correlations between the peak speeds within structures (e.g., U1, U2, U3) across repetitions of the test utterance employed here.

Vowel fundamental frequencies and intensities for each test utterance were measured by positioning a cursor at the approximate midpoints of waveform displays of the two /æ/ vowels. An autocorrelation method was used to estimate average fundamental frequencies over an 80-ms interval centered on these cursor placements. In a few cases involving fast or soft speech, this interval was set to a smaller value but was never less than 50 ms. Vowel intensity was calculated as the RMS level over the same mid-vowel intervals used to measure fundamental frequency and was expressed as absolute sound pressure level. The two horizontal lines under the acoustic waveform in Fig. 1 indicate the intervals where fundamental frequency and intensity were measured. The fundamental frequency and intensity of each test utterance were taken as the average across the two /æ/ vowels.

The gains of the rib cage and abdominal movement signals were equated using the isovolume maneuvers. The rib cage and

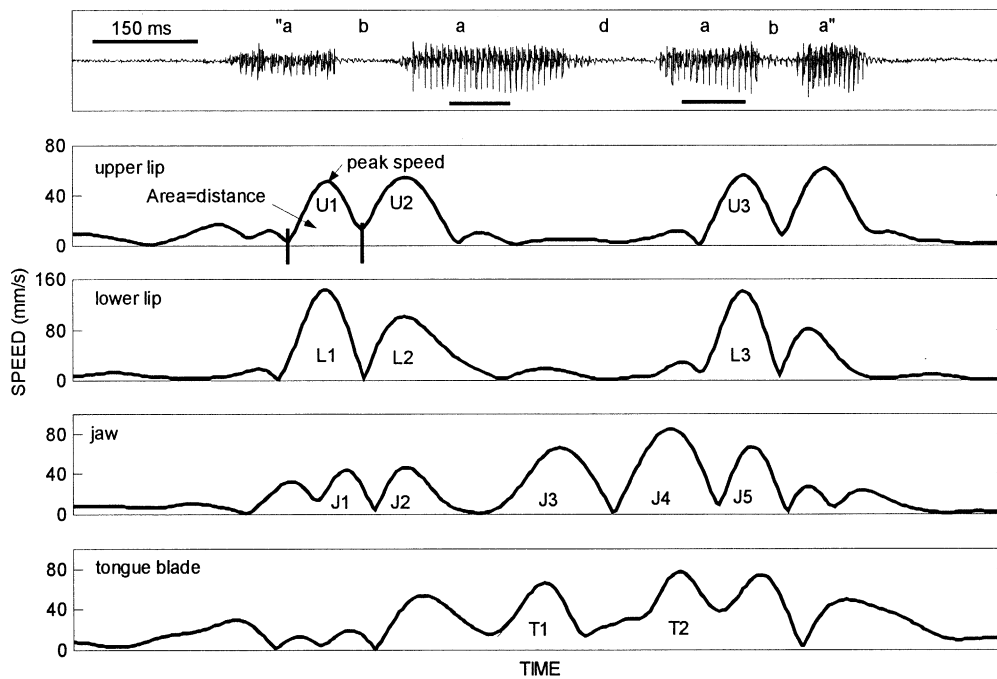


Fig. 1 Acoustic speech signal (*top trace*) and orofacial speed histories associated with a single production of the test utterance. *Horizontal lines* under the two stressed vowels indicate the 80-ms

time intervals over which fundamental frequency and sound pressure level were measured

abdominal signals were then summed and expressed as a percent of vital capacity. Measures of inspiratory volume of individual speech breaths were taken on the summed signal. This involved cursor placement at the approximate time point of peak inspiration, and automated extraction of expiratory minima at the end of each speech breath and maxima in the subsequent speech breath. In most cases, subjects produced only one speech token per breath group, but this varied with subject and speech condition. The correlation analyses as presented below, was performed both with and without trials involving more than one speech utterance per breath group. Equivalent results were obtained in the two cases and, therefore, results presented here involved analysis of the larger dataset that included speech breaths with multiple tokens. In other words, individual speech tokens sometimes shared the same inspiratory volume measure.

Correlation analysis

The associations of interest in this study were those relating upper lip, lower lip, tongue, and jaw speeds to vowel fundamental frequency, vowel intensity, and inspiratory volume of the speech breath. Statistical significance of individual correlations was set at $P=0.004$. This level was based on a 5% family error rate distributed across the 14 subjects within each type of correlation (e.g., upper-lip speed with fundamental frequency).

In correlating vocal intensity to orofacial speeds, it was necessary to control for their respective associations with movement distance. Previous studies have demonstrated strong associations between movement velocity and displacement (Ostry and Munhall 1985; Vatikiotis-Bateson and Kelso 1993). This finding was replicated here for all subjects when correlating average peak speed and average distance across speech tokens. In all 56 cases (14 subjects \times 4 structures), peak speed and distance showed significant positive correlations ($P<0.0005$), with a median correlation of 0.90. In cases where both peak speed and vowel intensity were correlated with displacement, positive correlations between speed and vowel intensity could be explained by their shared variance with distance.

This would argue against speed-intensity correlations being due to common neural drive. Instead, such relationships might be attributed to the association of orofacial displacement with the acoustic impedance and radiation characteristics of the vocal tract (Schulman 1989). Thus, partial correlation analysis was used to control for movement distance when correlating speed and vocal intensity. In this way significant correlations between speed and intensity might reasonably be attributed to neural coupling between the underlying neuromotor systems.

Results

Figure 2 provides examples from two subjects of scatter plots where significant correlations were seen relating orofacial speed to laryngeal and respiratory measures. The plot of vocal intensity and jaw speed involves the residuals from the regression equations relating the two variables to jaw distance; this reflects the method employed for partial correlation. Linear associations were generally noted in both types of plots shown in Fig. 2. In the few cases where nonlinearity was apparent, correlation levels were only slightly improved by data transformation. Thus, only linear correlation analysis was performed.

All 14 subjects showed significant correlations relating orofacial speed to fundamental frequency and intensity, while only six subjects showed significant correlations between orofacial speed and inspiratory volume. Figure 3 shows frequency histograms of correlations relating fundamental frequency, intensity, and inspiratory volume to orofacial speed. In these histograms, correlations were combined for the 14 subjects and 4 orofacial speed measures. With the degrees of freedom associated with

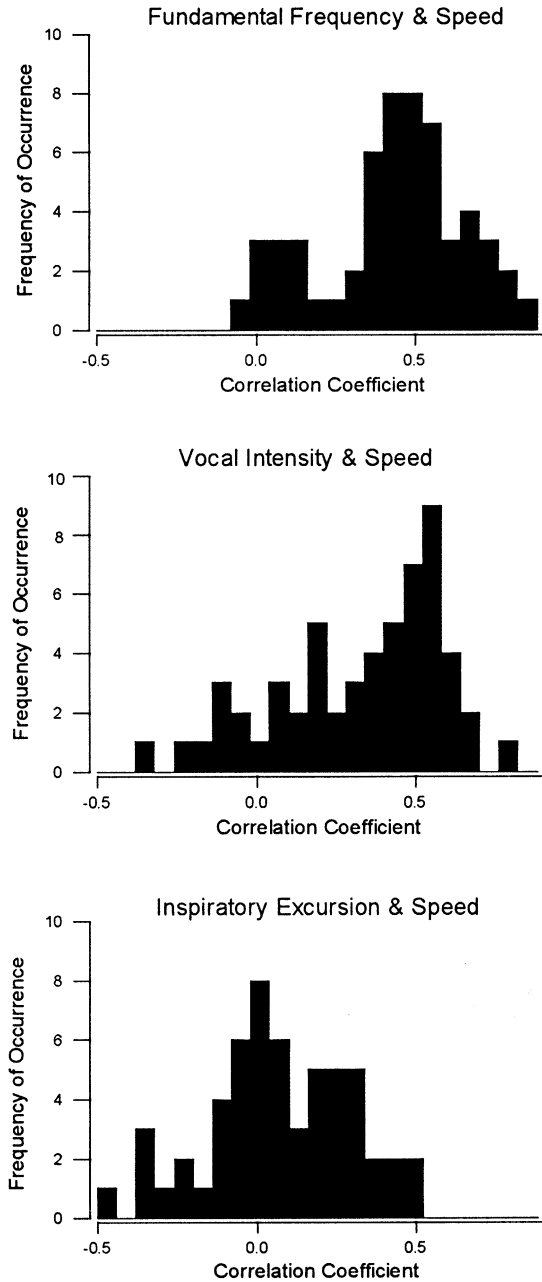
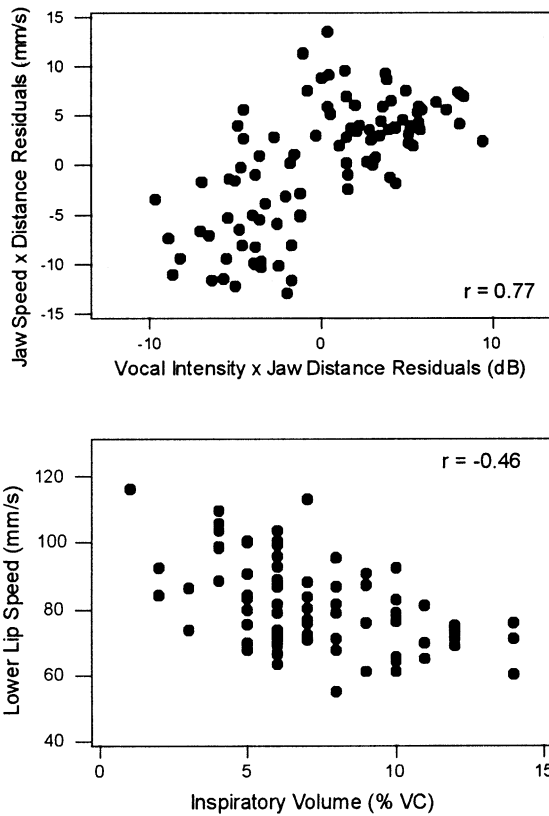


Fig. 2 Typical scatter plots on two subjects showing the type of data used to obtain correlation measures. The *top plot* shows residuals from regression equations relating jaw speed and vocal intensity to jaw distance

the various correlations, the cutoff for statistical significance for individual correlations was approximately 0.30. Significant correlations were obtained for fundamental frequency in 42 of 56 cases, for intensity in 35 of 56 cases, and for inspiratory volume in 14 of 56 cases. It can be seen in Fig. 3 that significant correlations were exclusively positive in the case of fundamental frequency. Positive correlations also predominated for intensity and inspiratory volume, but there were five instances of significant negative correlation, one involving vocal intensity and four involving inspiratory volume.

The frequencies of significant correlations relating the speeds of different orofacial structures to fundamental frequency, intensity, and inspiratory volume are summarized in Table 1. Examination of Table 1 indicates that the

Fig. 3 Histograms showing frequency of occurrence of correlation levels associated with fundamental frequency, intensity, and inspiratory lung volume. Data for all 14 subjects and the 4 orofacial speed measures are combined in each histogram

Table 1 Percentages of subjects out of 14 who showed statistically significant correlations relating orofacial speed measures to vowel fundamental frequency, vowel intensity, and speech breath inspi-

ratory volume. An alpha level of $P < 0.004$ was used for individual correlations based on a 5% family error rate, applying a Bonferroni correction within each cell for $n = 14$ subjects

Orofacial structure	Fundamental frequency	Vocal intensity	Inspiratory volume
Upper lip speed	79	50	21
Lower lip speed	86	71	36
Jaw speed	86	79	29
Tongue speed	50	50	14

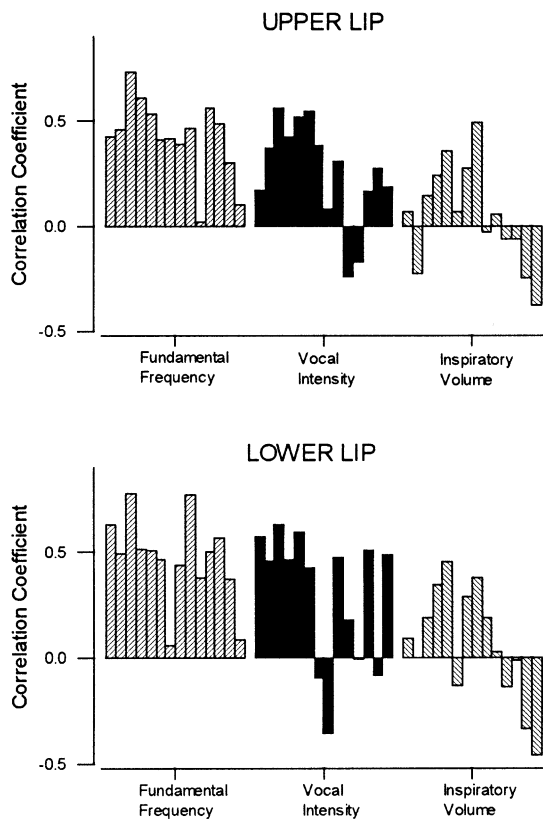


Fig. 4 Bar graphs on individual subjects showing correlation levels for upper lip speed (*top*) and lower lip speed (*bottom*) as related to fundamental frequency, vocal intensity, and inspiratory volume. The order of the 14 subjects is the same for each bar graph cluster

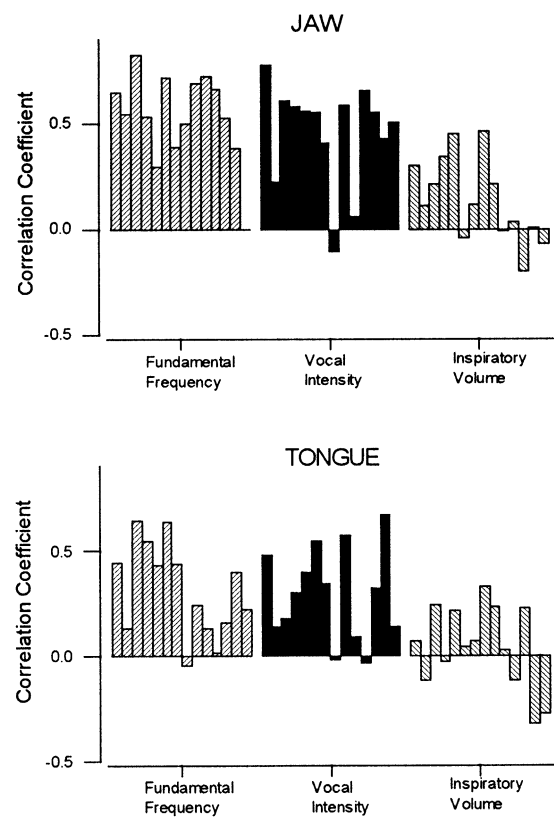


Fig. 5 Bar graphs on individual subjects showing correlation levels for jaw speed (*top*) and tongue speed (*bottom*) as related to fundamental frequency, vocal intensity, and inspiratory volume. The order of the 14 subjects is the same for each bar graph cluster

percentage of subjects who exhibited significant correlations tended to be reduced for inspiratory volume relative to fundamental frequency and intensity, and for the tongue relative to the lips and jaw.

The trends seen in Table 1 are further illustrated in Figs. 4 and 5, which show the correlation levels for individual subjects organized in relation to orofacial structures and associated voice acoustic and lung volume measures. Within the different bar-graph clusters (groups of 14 subjects), the correlation levels are arranged in a fixed order with respect to individual subjects. The data in Figs. 4 and 5 show a relatively high level of consistency across subjects in the correlation levels relating upper lip, lower lip, and jaw speed to fundamental frequency, and jaw speed to vocal intensity. Other types of correlation show considerably greater intersubject variability.

Comparison of the plots in Figs. 4 and 5 suggests that the strength of correlations relating orofacial speed to laryngeal and respiratory measures may have differed between orofacial structures. To statistically evaluate this possibility, sign tests were performed in which the correlations relating the speed of each orofacial structure to laryngeal and respiratory measures were compared with the correlations for the other three structures on a subject-by-subject basis. Distinct sign tests were per-

formed for each orofacial structure evaluating the frequency of cases in which correlations with fundamental frequency, intensity, and inspiratory volume were greater or less than those of the other orofacial structures. Performing these sign tests involved making nine comparisons for each structure. For example, the correlations relating jaw speed to fundamental frequency, intensity, and inspiratory volume were compared with the same correlations for the upper lip, lower lip, and tongue. Making the same set of comparisons for each of the 14 subjects resulted in 126 comparisons for each orofacial structure. The results of the sign tests for each of the four orofacial structures are shown in Table 2. A highly significant effect was obtained for the jaw, which showed the highest correlation in 90 of 126 correlations. Moderate effects in the opposite direction also were obtained for the upper lip and tongue, which showed reduced correlations compared to other structures.

The tendency for the jaw to show the highest correlations raises a question as to whether, compared to the jaw, there was minimal involvement of the lips and tongue in controlling speech rate and intensity. This possibility was evaluated by performing a number of one-way ANOVAs and post hoc comparisons on speed for each subject and structure. In each ANOVA, speech

Table 2 Two-tailed sign tests results evaluating the hypothesis that the median difference between correlations for a particular structure compared to the other three structures does not equal zero. Columns 3–5 show the number of cases where the orofacial speed correlation for the particular structure was below, equal to, or above the corresponding correlations for the other three structures

Structure	<i>N</i>	Below	Equal	Above	<i>P</i>
Upper lip	126	76	0	50	0.0259
Lower lip	126	63	1	62	1.0000
Jaw	126	35	1	90	<0.0001
Tongue	126	77	0	49	0.0162

Table 3 Number of subjects out of 14 showing significant increases, no change, or significant decreases in orofacial speed for the particular condition (loud, soft, fast, slow) relative to the normal speech condition. Statistical significance was determined from a Tukey procedure in one-way ANOVAs using a 5% error rate

		Upper lip	Lower lip	Jaw	Tongue
Loud	Increase	5	11	10	5
	No change	8	3	4	8
	Decrease	1	0	0	1
Soft	Increase	0	2	0	0
	No change	3	5	2	7
	Decrease	11	7	12	7
Fast	Increase	2	5	1	5
	No change	8	6	7	6
	Decrease	4	3	6	3
Slow	Increase	0	0	3	1
	No change	4	6	3	4
	Decrease	10	8	8	9

condition (normal, loud, soft, fast, and slow) was treated as a single factor, and post hoc Tukey tests were performed using a 5% error rate. Particular attention was given to the frequency and pattern of significant post hoc tests, which compared the loud, soft, fast, and slow conditions to the normal speech condition. For the individual ANOVAs, significant overall *F* values ($P < 0.0005$) were obtained in all 56 cases (14 subjects \times four structures). The results of the relevant subset of post hoc comparisons are given in Table 3. This table shows the number of subjects for each structure and condition where there was a significant increase, significant decrease, or no change in speed relative to the normal condition. Examination of these data indicates that the pattern of significant effects was generally similar across the four orofacial structures, and it is clear that variations in speech rate and intensity were accompanied by significant increases and decreases in lip and tongue speeds.

Discussion

Significant correlations were seen in numerous cases when relating lip, tongue, and jaw movement speeds to fundamental frequency, vocal intensity, and inspiratory volume during subjects' repetition of a simple speech

utterance. Given that variations in these output measures are strongly determined by muscle activation levels, we interpret significant correlations as reflecting coupling between neural circuits that affect motoneuron discharge during speech. The strength of this interpretation is limited by the use of a single speech utterance and the fact that subjects' performance was not tightly controlled in terms of speech rate and intensity. Future studies might usefully employ a range of speech utterances and more rigorous control of rate and intensity.

One possible source of the neural coupling suggested by the observed correlations is efferent drive from a common region of the brain to the motoneurons innervating orofacial, laryngeal, and respiratory muscles. Obviously the present data cannot address directly what specific brain regions might underlie such common drive during speech production. However, a potential candidate may be the caudal midbrain. Neurophysiologic data indicate that this region is critical for specific animal calls (Jurgens 1998), and it has been suggested that it provides a neural substrate for coordination of orofacial, laryngeal, and respiratory muscle activation (Dusterhoft et al. 2000; Jurgens 2000; Larson 1985; Zhang et al. 1994). The potential importance of the caudal midbrain to coordination of multiple muscle systems during speech is further suggested by case reports of mutism following lesion of the periaqueductal gray (e.g., Esposito et al. 1999).

Another general source of neural coupling that may have contributed to the observed correlations involves sensory input from the mechanoreceptors of one muscle system projecting over multisynaptic pathways to motoneurons of another muscle system. For example, it is known that perioral mechanical stimulation in humans evokes excitatory reflex responses in the laryngeal system as reflected in increases in fundamental frequency during sustained phonation (Larson and Sapir 1995). Given that perioral reflexes are especially responsive to the velocity of stretch (Larson et al. 1978), the positive correlations between lip speed and fundamental frequency may have been due in part to sensory drive from velocity-sensitive trigeminal afferents projecting indirectly to laryngeal motoneurons. Orofacial speeds tend to be maximal near the onset and offset of vowel segments that are bounded by stop consonants (see Fig. 1 for example). This would allow for a sufficient afferent delay time following vowel onsets for perioral mechanoreceptors to influence vowel fundamental frequency.

There were relatively few significant correlations relating orofacial speed to inspiratory volume, suggesting relatively weak coupling between the orofacial and respiratory muscle systems. However, vocal intensity is regulated to a greater extent by respiratory compared to laryngeal muscle forces (Finnegan et al. 2000). Thus, unlike fundamental frequency, which is controlled predominantly by laryngeal muscle activity, the correlations with intensity are interpreted as partially reflecting neural coupling between the orofacial and respiratory motor systems. The few instances where negative correlations

were observed involved inspiratory volume and vocal intensity. This result implicates inhibitory or suppressive neural pathways between the orofacial and respiratory musculature during speech in some individuals.

A primary finding of this study was that among the orofacial structures, jaw speed showed the highest correlations with laryngeal and respiratory output measures. This result suggests that compared to those of the lips and tongue, the neural system underlying jaw motor control for speech is more strongly coupled to neural systems involved in vocalization (i.e., laryngeal and respiratory control). This interpretation is consistent with the frame/content theory of speech production (MacNeilage 1998), which posits that neural circuitry regulating cyclic jaw movements for syllable production has evolved in a way that establishes especially strong coupling with neural systems involved in vocalization. It is also notable that vocal tract opening and closing movements during the babbling phase of speech development (about 6–12 months) tend to be driven predominantly by the jaw (Davis and MacNeilage 1995; Green et al. 2000). Thus, long-term experiential effects could facilitate relatively strong coupling between the neural systems regulating jaw motion and vocalization during the first years of life.

Acknowledgements This research was supported by NIH grant DC 03659 and approved by the Walter Reed Army Medical Center's Human Use Committee, under Department of Clinical Investigation Work Unit 2585. All subjects enrolled in the study voluntarily agreed to participate and gave written and informed consent. The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

References

- Atkinson JE (1978) Correlation analysis of the physiological factors controlling fundamental voice frequency. *J Acoust Soc Am* 63:211–222
- Baker KK, Ramig LO, Sapir S, Luschei ES, Smith ME (2001) Control of vocal loudness in young and old adults. *J Speech Lang Hear Res* 44:297–305
- Barlow SM, Farley GR (1989) Neurophysiology of speech. In: Kuehn DP, Lemme ML, Baumgartner JM (eds) *Neural basis of speech, hearing, and language*. College-Hill Press, Boston
- Davis BL, MacNeilage PF (1995) The articulatory basis of babbling. *J Speech Lang Hear Res* 38:1199–1211
- Dromey C, Ramig LO (1998a) The effect of lung volume on selected phonatory and articulatory variables. *J Speech Lang Hear Res* 41:491–502
- Dromey C, Ramig LO (1998b) Intentional changes in sound pressure level and rate: their impact on measures of respiration, phonation, and articulation. *J Speech Lang Hear Res* 41:1003–1018
- Dusterhofs F, Hausler U, Jurgens U (2000) On the search for the vocal pattern generator. A single-unit recording study. *Neuroreport* 9:2031–2034
- Esposito A, Demeuisse G, Alberti B (1999) Complete mutism after midbrain periaqueductal gray lesion. *Neuroreport* 10:681–685
- Finnegan E, Luschei E, Hoffman HT (2000) Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech. *J Speech Lang Hear Res* 43:934–950
- Fregosi RF, Fuller DD (1997) Respiratory-related control of extrinsic tongue muscle activity. *Respir Physiol* 110:295–306
- Gandevia SC, Gorman RB, McKenzie DK, DeTroyer A (1999) Effects of increased ventilatory drive on motor unit firing rates in human inspiratory muscles. *Am J Respir Crit Care Med* 160:1598–1603
- Gay T, Hirose H (1973) Effect of speaking rate on labial consonant production. *Phonetica* 27:44–56
- Gracco VL, Löfqvist A (1994) Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *J Neurosci* 14:6585–6597
- Green JR, Moore CA, Higashikawa CA, Steeve RW (2000) The physiologic development of speech motor control: lip and jaw coordination. *J Speech Lang Hear Res* 43:239–255
- Grill SE, Hallett M (1995) Velocity sensitivity of human muscle spindle afferents and slow adapting type II cutaneous mechanoreceptors. *J Physiol* 489:593–602
- Hixon TJ, Mead J, Goldman MD (1973) Dynamics of the chest wall during speech production: function of the thorax, rib cage, diaphragm, and abdomen. *J Speech Hear Res* 19:297–356
- Jurgens U (1998) Neuronal control of mammalian vocalization, with special reference to the squirrel monkey. *Naturwissenschaften* 85:376–388
- Jurgens U (2000) Localization of a pontine vocalization-controlling area. *J Acoust Soc Am* 108:1393–1396
- Lansing W, Solomon NP, Kossev AR, Andersen AB (1991) Recording single motor unit activity of human nasal muscles with surface electrodes: applications for respiration and speech. *Electroencephalogr Clin Neurophysiol* 81:167–175
- Larson CR (1985) The midbrain periaqueductal gray: a brainstem structure involved in vocalization. *J Speech Hear Res* 28:241–249
- Larson KK, Sapir S (1995) Orolaryngeal reflex responses to changes in affective state. *J Speech Hear Res* 38:990–1000
- Larson C, Folkins J, McClean M, Muller E (1978) Sensitivity of the human perioral reflex to parameters of mechanical stretch. *Brain Res* 146:159–164
- Lieberman P (1984) *The biology and evolution of language*. Harvard University Press, Cambridge, MA
- MacNeilage PF (1998) The frame/content theory of evolution of speech production. *Behav Brain Sci* 21:499–546
- McClean MD (2000) Patterns of orofacial movement velocity across variations in speech rate. *J Speech Lang Hear Res* 43:205–216
- McClean MD, Clay JL (1995) Activation of lip motor units with variations in speech rate and phonetic structure. *J Speech Hear Res* 38:772–782
- Munhall KG, Löfqvist A, Kelso JAS (1994) Lip-larynx coordination in speech: effects of mechanical perturbation to the lower lip. *J Acoust Soc Am* 95:3605–3616
- Ostry DJ, Munhall KG (1985) Control of rate and duration of speech movements. *J Acoust Soc Am* 77:640–648
- Sapir S, McClean M, Luschei E (1984) Time relations between cricothyroid muscle activity and the voice fundamental frequency (F_0) during sinusoidal modulation of F_0 . *J Acoust Soc Am* 75:1639–1641
- Sauerland EK, Mitchell SP (1975) Electromyographic activity of intrinsic and extrinsic muscles of the human tongue. *Tex Rep Biol Med* 33:444–455
- Schulman R (1989) Articulatory dynamics of loud and normal speech. *J Acoust Soc Am* 85:295–312
- Stathopoulos E, Sapienza C (1993) Respiratory and laryngeal function of women and men during vocal intensity variation. *J Speech Hear Res* 36:64–75
- Titze IR, Luschei ES, Hirano M (1989) Role of the thyroarytenoid muscle in regulation of fundamental frequency. *J Voice* 3:213–224
- Tuller B, Shao S, Kelso JA (1990) An evaluation of an alternating magnetic field device for monitoring tongue movements. *J Acoust Soc Am* 88:674–679

- Turner RS, Grafton ST, Votaw JR, DeLong MR, Hoffman JM (1998) Motor subcircuits mediating the control of movement velocity: a PET study. *J Neurophysiol* 80:2162–2176
- Vatikiotis-Bateson E, Kelso JA (1993) Rhythm type and articulatory dynamics in English, French and Japanese. *J Phonet* 21:231–265
- Westbury JR, Lindstrom MJ, McClean MD (2002) Tongues and lips without jaws: a comparison of methods for decoupling speech movements. *J Speech Lang Hear Res* 45:651–662
- Zhang SP, Davis PJ, Bandler R, Carrive P (1994) Brain stem integration of vocalization: role of the midbrain periaqueductal gray. *J Neurophysiol* 72:1337–1356