Exp Brain Res (2004) 159: 310–318
DOI 10.1007/s00221-004-1952-8

RESEARCH ARTICLES

Michael D. McClean · Stephen M. Tasko

Correlation of orofacial speeds with voice acoustic measures in the fluent speech of persons who stutter

Received: 23 May 2003 / Accepted: 14 April 2004 / Published online: 10 July 2004

Abstract Stuttering is often viewed as a problem in coordinating the movements of different muscle systems involved in speech production. From this perspective, it is logical that efforts be made to quantify and compare the strength of neural coupling between muscle systems in persons who stutter (PS) and those who do not stutter (NS). This problem was addressed by correlating the speeds of different orofacial structures with vowel fundamental frequency (F0) and intensity as subjects produced fluent repetitions of a simple nonsense phrase at habitual, high, and low intensity levels. It is assumed that resulting correlations indirectly reflect the strength of neural coupling between particular orofacial structures and the respiratory-laryngeal system. An electromagnetic system was employed to record movements of the upper lip, lower lip, tongue, and jaw in 43 NS and 39 PS. The acoustic speech signal was recorded and used to obtain measures of vowel F0 and intensity. For each subject, correlation measures were obtained relating peak orofacial speeds to F0 and intensity. Correlations were significantly reduced in PS compared to NS for the lower lip and tongue, although the magnitude of these group differences covaried with the correlation levels relating F0 and intensity. It is suggested that the group difference in correlation pattern reflects a reduced strength of neural coupling of the lower lip and tongue systems to the respiratory-laryngeal system in PS. Consideration is given to how this may contribute to temporal discoordination and stuttering.

M. D. McClean (✉) · S. M. Tasko
Department of Speech Pathology and Audiology, Western Michigan University,
Kalamazoo, MI 49008-3825, USA
e-mail: michael.mcclean@wmich.edu
Tel.: +1-269-387-8049
Fax: +1-269-387-8044

Present address:
M. D. McClean · S. M. Tasko
Department of Speech Pathology and Audiology, Western Michigan University,
Kalamazoo, MI 49008-3825, USA

Keywords Stuttering · Speech motor coordination · Jaw · Lips · Tongue · Respiratory-laryngeal

Introduction

Normal human speech production involves precisely coordinated activation of distinct muscle systems to control the movements of different orofacial, laryngeal, and respiratory structures. It can reasonably be argued that this process involves behavior-specific linkages or coupling between the neural systems involved in controlling the movements of these structures (Grillner 1982; Lenneberg 1967; MacNeilage 1998). Empirical support for this general notion is provided by recent studies showing significant associations among orofacial, laryngeal, and respiratory system output during speech (Dromey and Ramig 1998; McClean and Tasko 2002; Watson et al. 2003). These effects were attributed in part to central neural rather than biomechanical or acoustic interactions between muscle systems.

McClean and Tasko (2002) attempted to quantify the strength of speech-related neural coupling between the orofacial and respiratory-laryngeal motor systems. The method involved correlating the speeds of the lips, tongue, and jaw with voice fundamental frequency (F0), intensity, and speech-breath inspiratory volume across repetitions of a simple speech utterance produced over a wide range of rates and intensities. The experimental approach assumes that orofacial speeds, F0, and intensity are strongly associated with underlying levels of muscle activity. This assumption is supported by studies showing strong associations of muscle activation levels with orofacial speed (McClean and Tasko 2003), voice F0 (Atkinson 1978; Sapir et al. 1984), and voice intensity (Finnegan et al. 2000). It is further assumed that variations in F0 are indicative of laryngeal muscle activity and that intensity reflects the combined activity of laryngeal and respiratory muscle systems. Therefore, significant correlations relating orofacial speeds to F0 and intensity were interpreted as reflecting the strength of coupling between the neural systems controlling motor output of orofacial and respiratory-laryngeal systems. The principal findings of the McClean and Tasko study were that the speeds of all four orofacial structures showed significant positive correlations with F0 and intensity, with the jaw showing the strongest correlations. The present study represents an extension of the methods and results of this initial investigation to the problem of stuttering.

Stuttering is often characterized as a problem in coordinating the movements of different structures (Caruso et al. 1988; Kent 1984; McClean and Runyan 2000; Zimmermann 1980). For this reason, the issue of neural coupling between speech motor subsystems is a logical area to be explored. When considering how the strength of neural coupling among motor subsystems might vary between persons who stutter (PS) and those who do not (NS), at least two areas for hypothesis formulation present themselves. The first pertains to the overall strength of coupling between subsystems. The second concerns the pattern of correlations across subsystems (e.g., are group differences noted for some structures and not others?). As to the overall strength of neural coupling, there is little experimental data on fluent speech to support either reduced or elevated coupling strengths in PS relative to NS. There is evidence for high amplitude coherent tremor across muscle systems in some PS during speech disfluency (Denny and Smith 1992; Smith et al. 1993), but the relevance of these findings for understanding possible group differences in neural coupling during fluent speech is uncertain in light of how little is known about the neural mechanisms of speech disfluency. On the basis of computer modeling work, Kalvarem (1993) has argued that PS are likely to have elevated strengths of neural coupling between neural systems controlling speech motor subsystems. He suggests that this may be a source of inflexibility or reduced adaptability of the speech motor system, which would be reflected in reduced inter-trial variability in parameters of motor output in PS. However, experimental results pertaining to inter-trial variability do not show strong group effects in either direction (e.g., McClean et al. 1994; Smith and Kleinow 2000), and studies in this area provide little basis to predict overall differences between NS and PS in the strength of neural coupling across motor subsystems.

Jaw speeds were previously observed to be more strongly correlated with respiratory-laryngeal system output compared to lip and tongue speeds (McClean and Tasko 2002). This result was interpreted in terms of the frame/content theory on the evolution of speech production (MacNeilage 1998; MacNeilage and Davis 2001). The theory may have relevance for understanding possible differences between NS and PS in the pattern of correlations relating orofacial speeds to respiratory-laryngeal system output. Frame/content theory posits that synaptic linkages between the inferior-frontal cortical system regulating cyclic jaw movement for mastication and more medial cortical systems involved in vocalization are phylogenetically older than cortical systems that couple of the lips and tongue with the vocalization system. It is suggested that as the requirements for precise consonant production increased during the human evolutionary process, pressure for more refined coordination of the lips and tongue with the respiratory-laryngeal system ensued. This type of sequencing in the evolution of neural pathways linking the orofacial and vocalization systems may be paralleled in the speech motor development of children where jaw motion appears to be the dominant and most stable contributor to syllable production during the earliest phases of speech acquisition (Davis and MacNeilage 1995; Green et al. 2002). Assuming that the neural systems controlling the lips and tongue develop relatively late in terms of phylogeny and ontogeny, and given the genetic and developmental origins of stuttering (Yairi et al. 1996), stuttering could originate partially from phylogenetic and developmental anomalies in the relative strengths of neural coupling of orofacial structures with the respiratory-laryngeal system. If such anomalies are sustained into adulthood, one might expect to see differences between NS and PS in the relative correlation levels relating voice acoustic measures to jaw speed versus lower lip and tongue speed.

The present study was undertaken to explore the hypothesis that neural coupling between the orofacial and respiratory-laryngeal systems differs in adult PS and NS. The methodology was adapted from our prior study of NS in which speeds of the lips, tongue, and jaw were correlated with voice acoustic measures assumed to reflect motor output of the respiratory-laryngeal system (McClean and Tasko 2002). Given its general relevance to frame/content theory, a secondary goal of the present investigation was to replicate and extend the earlier finding in NS showing higher correlations for the jaw compared to the lips and tongue when relating orofacial speeds to voice acoustic measures.

Materials and methods

Subjects

All subjects gave their informed consent prior to participation in the study. Data acquisition was carried out at the Walter Reed Army Medical Center, Washington, DC, approved by the Center's Human Use Committee, and performed in accordance with the ethical standards of the 1964 Declaration of Helsinki. Data were acquired on 82 adult speakers of American English who reported having normal hearing. Forty-three subjects (42 males and one female; mean age 23 years) had normal speech production (NS), and 39 subjects (37 males and two females, mean age 26 years) reported a history of stuttering since childhood (PS). Fourteen of the NS subjects had participated in the previous study involving analyses similar to those employed here (McClean and Tasko 2002).

The stuttering severity levels of the PS were determined from videotape recordings of monologue and reading using the Stuttering Severity Index or SSI (Riley 1994). Based on the SSI results, 23 PS were rated as mild, 9 as moderate, and 5 as severe. In the case of two PS, videotapes required for SSI ratings were not available.

Data acquisition procedures

Kinematic and audio signal recording was performed while subjects produced a variety of speech tasks in a series of 30-s recording blocks. The present analysis focuses exclusively on three of these tasks. These involved producing the utterance “a bad daba” (/ə bæd dæbə/) at normal, loud, and soft vocal intensities, with approximately equal stress on the two /æ/ vowels. This represents a modification of the method employed in the earlier study where trials involving variations in speech rate were included in correlation analyses (McClean and Tasko 2002). The procedural change was motivated by recent EMG data indicating that lip and jaw muscle activity levels show relatively weak correlations with lip and jaw speeds across variations in speech rate, while strong linear associations between orofacial speeds and muscle activity were observed across variations in vocal intensity (McClean and Tasko 2003). The three intensity conditions studied here were the first, third, and fifth of 16 speech tasks performed in each experimental session. The primary purpose in varying vocal intensity was to obtain a wide output range in kinematic and acoustic measures of interest in order to facilitate correlation analysis. The test utterance “a bad daba” was selected primarily because it involves large amplitude orofacial movements that are amenable to automated software measurement. Only fluent productions of the test utterance were included in the analysis. Identification of speech disfluencies was performed at the time of recording and then verified in subsequent listening to each utterance during signal processing and measurement.

Prior to each intensity condition, subjects repeated the test utterance a number of times at a normal loudness. For the loud speech condition, they were then briefly trained to produce the test utterance at intensity levels approximately 6 dB above their normal intensity, and prior to the soft speech condition approximately 6 dB below their normal level.

Two-dimensional motions of the upper lip, lower lip, tongue blade, and jaw within the midsagittal plane were transduced and recorded 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 vermilion, and lower-lip vermilion with biomedical tape and to the tongue blade (1 cm from the tip) and base of the lower incisors with surgical adhesive (Isodent). The audio signal was transduced with a Shure M93 miniature condenser microphone positioned 7.5 cm from the mouth, and the microphone-amplifier setup was calibrated to permit measurement of absolute sound pressure levels. The orofacial movement signals were digitized to computer at 250 Hz per channel and the audio signal at 16 kHz.

Signal processing and measurement

Stuttering is most likely to occur on the initial stressed syllables of speech utterances (Brown 1938; Taylor 1966). From this it was reasoned that fluent productions of the initial stressed syllable in the test utterance would be most indicative of anomalies in motor output associated with the fluent speech of PS. Therefore, kinematic and acoustic measures were restricted to events associated with fluent productions of the syllable /baed/.

The upper lip, lower lip, jaw, and tongue movement signals were low-pass filtered at 8 Hz and the nose signal at 3 Hz. Filtering was performed with a fifth-order Butterworth filter with a correction that eliminated any phase lags between signals. Head movements during recording were slight (<1.0 mm), but 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. Methods described by Westbury et al. (2002), which take into account jaw rotation, were used to decouple the lower lip and tongue signals from the jaw signal.

A three-point central method was used to differentiate the vertical and horizontal displacement signals and obtain speed histories for the upper lip, lower lip, jaw, and tongue blade as indicated in Fig. 1. These speed histories showed well-defined peaks associated with upper and lower lip opening (LO, UO), tongue closing (TC), jaw opening (JO), and jaw closing (JC) in /baed/. Each of these peak speed measures was employed in correlation analyses on individual subjects, relating speed to F0 and intensity across trials.

Vowel F0 and intensity for each test utterance was measured by positioning a cursor at the approximate midpoint of the waveform display of the first /æ/ vowel. 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 soft speech, this interval was set to a smaller value but it was never less than 50 ms. Vowel intensity was calculated as the RMS level over the same mid-vowel interval used to measure fundamental frequency and was expressed as absolute sound pressure level.

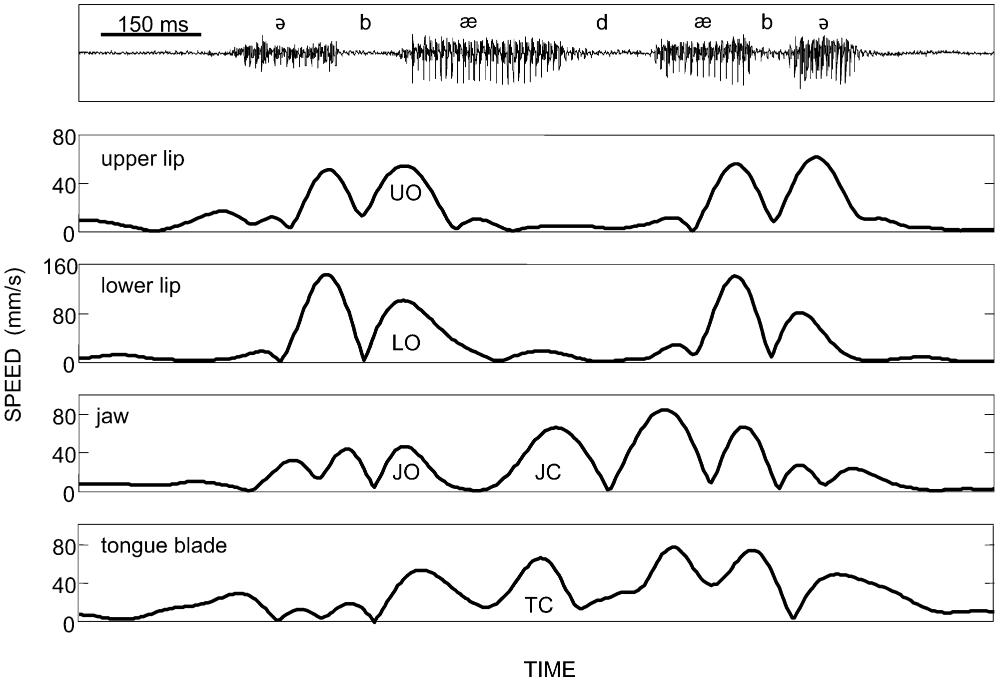

Fig. 1 Speed histories of the upper lip, lower lip, jaw, and tongue associated with a single production of the test utterance “a bad daba”. The acoustic waveform is shown at the top. The five speed peaks employed in the correlation analyses are labeled (see text). Measures of F0 and sound pressure level were obtained on the middle 80 ms interval of the first /æ/ vowel

Correlation analysis

Pearson product moment correlations were performed on the dataset of each subject to quantify the associations of the five orofacial speed measures (UO, LO, TC, JO, JC) with F0 and intensity. The mean and median number of trials used in these correlations was 56, with a range of 20–79 trials across the 82 subjects. Excluding the one subject with 20 trials, the minimum number of trials for a subject was 36. The mean number of trials for individual subjects was 58 for NS and 55 for PS.

In order to assess the pattern of outliers in the correlation datasets, linear regression analyses were performed for each of 820 cases (82 subjects × 10 correlations). For each subject the five speed measures were regressed independently on F0 and intensity. Cases where the standard residual of the predicted speed exceeded 2.0 were taken as outliers. The average number of outliers per correlation dataset were 2.4 and 2.5 respectively for NS and PS. All outliers as well as F0 and intensity values were within known ranges for the various measures and therefore were not excluded from the correlation analyses.

It is recognized that significant correlations between speed and intensity might be attributed in part to their common association with orofacial displacement as it affects the acoustic impedance of the vocal tract. In the previous study (McClean and Tasko 2002), partial correlation analysis was employed to address this issue. Partial correlation was not employed here, since it is unlikely that NS and PS differ in the physical acoustics of the vocal tract that contribute the association of displacement and intensity.

In addition to correlations relating orofacial speed to voice acoustic measures, correlations relating F0 and intensity were obtained. It is generally known that F0 and intensity tend to be positively correlated (Titze 1992). This is due in part to respiratory and laryngeal mechanisms that are common to the control of each (Finnegan et al. 2000; Titze et al. 1989). Preliminary analysis indicated that F0-intensity correlations covaried with correlations relating speed to F0. Hence, the F0-intensity correlation measure was employed as a covariate in group comparisons of speed-F0 and speed-intensity correlation levels. It should be emphasized that the F0-intensity correlations are not interpreted here as reflecting independent sources of neural drive (as with the speed correlations), since their control involves extensively shared anatomy and neurophysiology.

A Fisher *z* transform was performed in order to normalize the distribution of correlations for parametric statistical analysis. All statistical tests comparing the two groups were performed on these transformed data using the following equation taken from Edwards (1967).

$$z = 0.5[\log_e(1 + r) - \log_e(1 - r)]$$

Results

Group kinematics and voice acoustics associated with task performance

Before evaluating possible group differences in the association of orofacial speed and voice acoustic measures, it was important to determine the extent to which PS and NS performed the speech tasks in an equivalent manner as reflected in various output measures. Performance of the two groups was evaluated by computing the mean differences between the loud and normal intensity conditions, and the soft and normal intensity conditions for each subject. These two sets of difference measures were computed for the orofacial speeds, vowel intensity, and F0. Group comparisons also were made of the standard deviations of each subject's pooled data across the three intensity conditions. The means and standard errors of the various difference measures and the overall standard deviations are given in Table 1.

Both NS and PS showed consistent increases in all measures in the loud speech condition and decreases for the soft speech condition. *t*-Test comparisons showed a significant group difference ($p<0.05$) in only one instance where the extent of reduction in tongue speed was less in PS in the soft speech condition. The overall variability of the various measures, as reflected in their pooled standard deviations, was equivalent across the two groups.

It is known that movement durations during fluent speech tend to be longer in PS (e.g., Max et al. 2003), and movement speed and duration tend to show a nonlinear negative association across relatively constant movement distances (Tasko 1999). Therefore, we evaluated possible differences in movement durations between the NS and PS. Combining the three intensity conditions, average durations associated with the five movements ranged from 163 to 182 ms. Significant differences in movement duration between the groups were seen only for the lower lip, where PS showed a mean duration 179 ms compared to 168 ms for NS (t=2.09, p=0.04). For each orofacial structure, movement duration differences across the intensity conditions also were evaluated in the same manner as described above for speed, F0, and intensity. The absolute values of the mean differences in movement duration with increased and decreased intensity never exceeded 9 ms, and there were no significant differences between the groups on the duration difference measures.

Group differences in speed-F0 and speed-intensity correlations

Statistical comparisons, as described below, indicated significant reductions in PS relative to NS in the correlation levels relating lower lip and tongue speed to vowel F0 and intensity. This trend is illustrated on two

Table 1 Mean deviations from the normal intensity condition for the loud and soft speech conditions for the five speed measures, vowel intensity, and vowel F_0. Standard errors are given in parentheses. The pooled standard deviations are the means of the individual subject standard deviations across the three intensity conditions. Speed measures are in mm/s, intensity in dB, and F_0 in Hz. The asterisk indicates one case of a significant group difference ($p<0.05$)

	Loud speech		Soft speech		Pooled SD	
	NS	*PS*	*NS*	*PS*	*NS*	*PS*
Upper lip	5.0 (1.0)	4.6 (1.2)	−11.6 (2.6)	−7.1 (1.1)	10.7	8.1
Lower lip	12.4 (1.6)	8.4 (2.1)	−11.2 (2.0)	−11.3 (1.9)	15.1	15.1
Tongue	9.2 (1.9)	10.8 (2.8)	−10.0 (1.9)	−2.9 (2.2)*	16.5	18.1
Jaw open	13.9 (1.8)	15.6 (2.0)	−13.8 (1.9)	−10.6 (1.9)	15.4	14.9
Jaw close	19.7 (2.2)	21.3 (2.6)	−20.5 (2.6)	−18.1 (2.3)	19.8	20.3
Intensity	8.1 (0.4)	8.1 (0.3)	−7.0 (0.4)	−7.2 (0.4)	6.5	6.7
F_0	13.7 (8.1)	13.0 (1.3)	−6.6 (1.0)	−7.5 (1.2)	9.7	10.3

subjects in Fig. 2 which shows scatter plots relating lower lip speed to vowel intensity for an NS (upper plot) and PS (lower plot). Figure 2 also illustrates a characteristic feature of the data which is the clustering of points in relation to the three intensity conditions.

Two classes of response variable were considered when comparing speed correlation levels of NS and PS, the five sets of correlations relating orofacial speed to F0 and the five sets of correlations related orofacial speed to intensity. Because multiple response variables were involved in each case, two distinct MANOVA were performed, one for F0 and one for intensity. These MANOVA were employed to make overall statistical determinations as to whether mean correlation levels differed between the two groups of subjects irrespective of orofacial structure. As mentioned above, preliminary analysis suggested that the speed-F0 correlations might be associated with the correlation levels relating F0 and intensity. Therefore, the F0-intensity correlation was included as a covariate in the two MANOVA. Using Wilks' lambda, the speed-F0 MANOVA was significant with $p=0.01$ ($F=3.2$, $df=5{,}75$), and the speed-intensity MANOVA was significant with $p=0.03$ ($F=2.6$, $df=5{,}75$). The F0-intensity correlation was a significant covariate in the speed-F0 MANOVA ($p<0.0005$, $F=18.0$, $df=5{,}75$) but not the speed-intensity MANOVA.

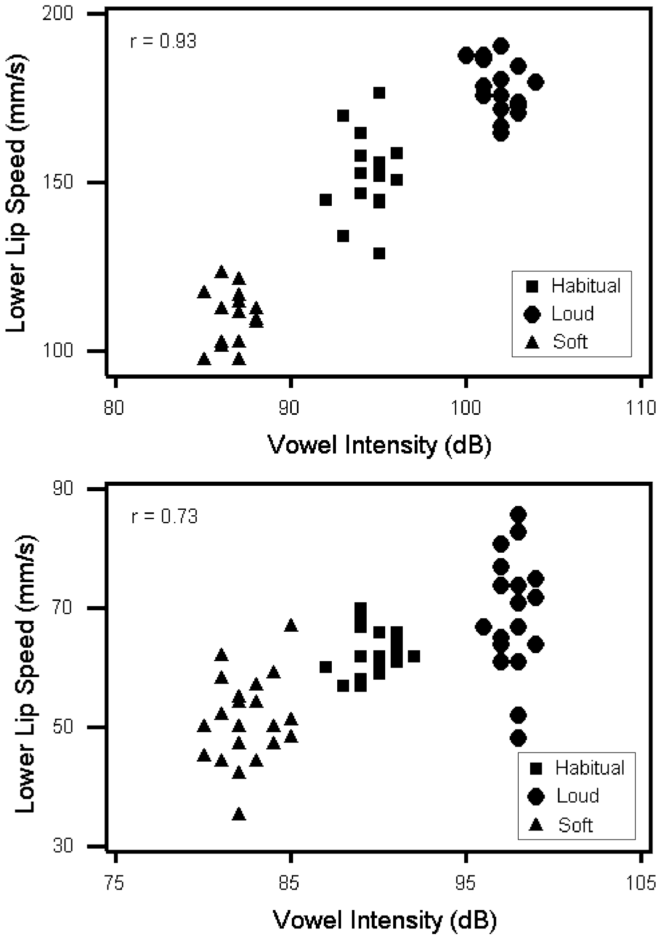

Fig. 2 Scatter plots showing the association of lower lip opening speed (*LO*) with vowel intensity for an NS (*upper plot*) and PS (*lower plot*) who generally showed strong correlations between speed and voice acoustic measures, and between intensity and F0

In light of the MANOVA results indicating overall differences between NS and PS in mean correlation levels, ten ANOVAs were performed in which the statistical significance of differences between NS and PS in mean correlation levels were evaluated for each orofacial speed measure in relation to F0 and intensity. The correlation of F0 with intensity was included as a covariate for the F0 ANOVA. ANOVA results indicated that correlation levels for the lower lip and tongue (LO and TC) were significantly less in PS compared to NS for both the F0 and intensity analyses, while significant group effects were not obtained for the upper lip or jaw. Overall results are summarized in Table 2 which shows the mean correlation levels relating orofacial speed to F0 and intensity for the NS and PS groups and the results of ANOVA with stuttering as the factor.

Dependence of speed correlations on F0-intensity correlations

For the speed-F0 ANOVA, the F0-intensity correlation was a significant covariate for UO, LO, JO, and JC ($p<0.0005$) but not TC ($p=0.08$). This association of the speed-F0 and F0-intensity correlations is illustrated in Fig. 3 for the lower lip (LO). Examination of the scatter for the two groups in Fig. 3 suggests that the strength of the group difference in the correlations relating lower-lip speed to F0 may depend on the level of the F0-intensity correlation. This has implications for interpretation of the group results and future application of the methods employed here.

To evaluate the group differences in speed-F0 and speed-intensity correlations in relation to the F0-intensity correlation, the distribution of F0-intensity correlations was broken into three subgroups representing low, mid, and high correlation levels. The ranges of these three

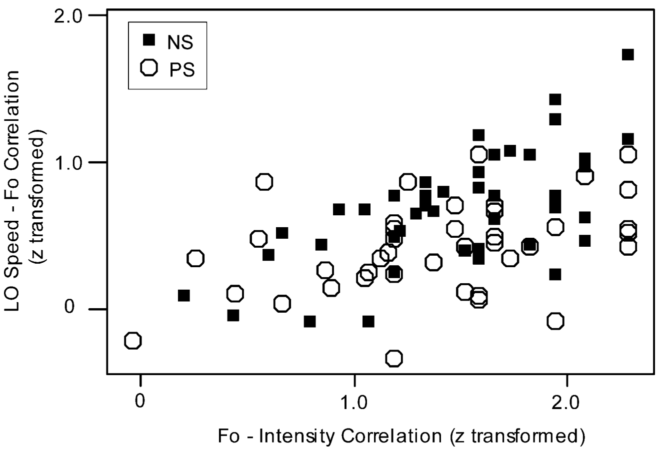

Fig. 3 Scatter plot relating the *z*-transformed correlations of lower lip speed to F0 (LO-F0) and F0 to intensity for the two subject groups

Table 2 Mean correlation values relating orofacial speeds to F_0 and intensity for the NS and PS groups are shown to the left. The standard errors of these mean correlations ranged from 0.03 to 0.05. The ANOVA results with group as the single factor are shown to the right. The degrees of freedom for F_0 analyses for group, covariate and error were 1, 1, and 79, and for intensity analyses the degrees of freedom for group and error were 1 and 80. Statistically significant cases (p<0.05) are indicated with asterisks

Correlation variables	NS—mean*r* (*N*=43)	PS—mean*r* (*N*=39)	*F* value	*p* Level
UO speed—F_0	0.49	0.48	0.0	0.952
LO speed—F_0	0.53	0.37	10.7	0.002*
TC speed—F_0	0.40	0.22	6.9	0.010*
JO speed—F_0	0.59	0.53	0.4	0.538
JC speed—F_0	0.68	0.61	1.2	0.278
UO speed—intensity	0.59	0.55	0.6	0.437
LO speed—intensity	0.61	0.48	6.8	0.011*
TC speed—intensity	0.46	0.31	7.3	0.009*
JO speed—intensity	0.69	0.66	0.1	0.743
JC speed—intensity	0.79	0.76	0.7	0.406

groups were based on examination of the distribution of F0-intensity correlations, which had a strong negative skew. The group with the lowest set of correlations included 21 cases (23% of NS and 28% of PS) comprising the negative tail of the distribution (range: 0–0.82). The remaining 61 cases were split, with the mid group having 30 cases (37% of NS and 36% of PS; range: 0.83–0.92) and the highest group 31 cases (40% of NS and 36% of PS; range 0.93–1.0). In Fig. 4, the mean LO speed-F0 correlations of the two subject groups have been plotted for each of the three F0-intensity subgroups. This figure shows clear tendencies for overall correlation levels and group differences to increase with the F0-intensity correlation level. The same trend for increased correlation levels across the F0-intensity subgroups also was seen for the other speed measures (UO, TC, JO, and JC), and group differences in TC-F0 correlation level were most pronounced in group 3.

The dependence of speed-F0 correlation group effects on F0-intensity correlation level suggests that there may have been qualitative differences among subjects in terms of respiratory-laryngeal control processes underlying changes in vocal intensity. From this one would expect different patterns of F0 and intensity variation for the F0-intensity subgroups. This possibility was evaluated by comparing the means and coefficients of variation (CV) of F0 and intensity across the three intensity conditions for the three subgroups. Four two-way repeated measures ANOVA were performed to test for significant differences in F0, intensity, and their CVs. CVs were calculated using the token repetitions within each intensity condition, and for statistical analysis, CVs were arcsine transformed to normalize their distributions. F0-intensity subgroup was the between-group factor, and intensity condition was the within-group factor.

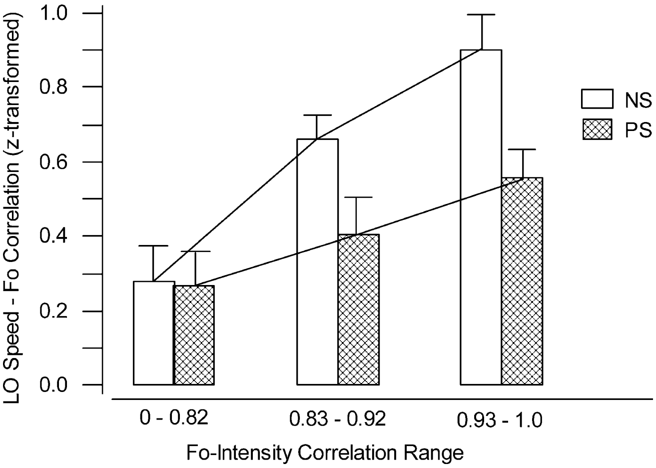

Fig. 4 Bar graphs showing the mean LO-F0 correlation levels (transformed) and standard errors for NS and PS across three F0-intensity correlation subgroups. From *left to right,* the numbers of subjects in the three subgroups were 10 NS and 11 PS, 16 NS and 14 PS, 17 NS and 14 PS

ANOVA results suggested that there were distinct differences in the voice control processes of the three F0-intensity subgroups. This can be seen in Table 3, which summarizes the ANOVA results for the subgroup main effects and interactions related to subgroup by condition. Statistics for condition are not shown, but these were highly significant and in the expected directions for both intensity and F0. Significant main effects were seen for F0 and intensity CVs, and significant interactions were noted for F0 and F0 CV. F0 and intensity CVs were elevated in the mid-*r* and low-*r* correlation groups, with ranges of 0.014–0.026 in the high-*r* group, 0.016–0.034 in the mid-*r* group, and 0.017–0.051 in the low-*r* group.

The elevated inter-trial variability in the mid-*r* and low-*r* subgroups may be related in part to differences in vocal intensity range and respiratory-laryngeal control strategy employed. This is suggested by data presented in Fig. 5, which shows the mean intensities and F0s for the three subgroups at the three loudness conditions. There it may be seen that the low-*r* group tended to operate closer to the low end of their intensity range, while showing marked elevations in F0 particularly in the habitual and soft conditions. Figure 5 also makes clear the basis for the strong interaction between subgroup and condition for F0 as indicated in Table 3.

Table 3 ANOVA results associated with analysis of F_0-intensity subgroups and subgroup interaction with condition in terms of mean F_0, intensity, and their coefficient of variation (CV). Statistics for condition are not shown, but these were highly significant for all response variables

	Subgroup *F* value	Subgroup *p* level	Subgroup × Condition *F* value	Subgroup × Condition *p* value
F_0	1.4	0.241	16.8	<0.0005*
Intensity	2.8	0.066	1.0	0.399
F_0CV	5.7	0.005*	3.6	0.007*
Intensity CV	4.5	0.014*	1.2	0.326

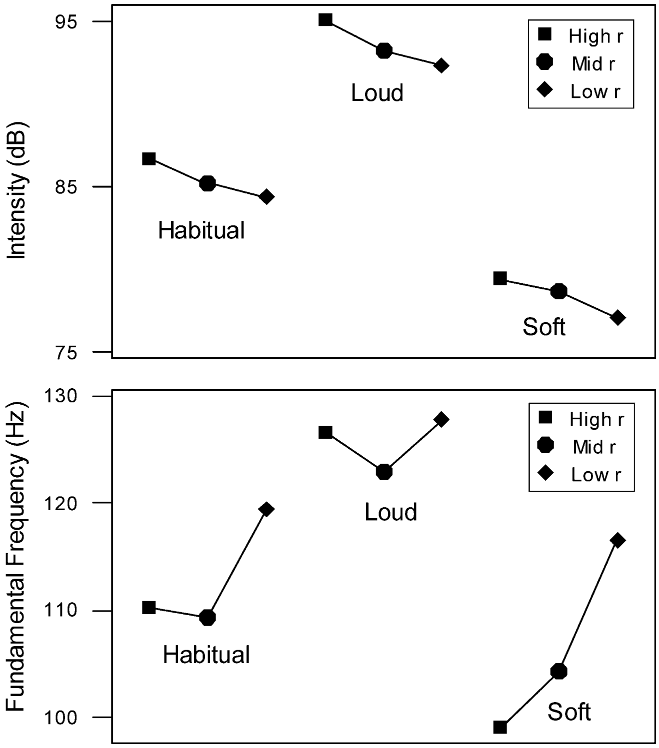

Fig. 5 Mean intensity (*top*) and mean F0 (*bottom*) for three F0-intensity correlation subgroups (high *r,* mid *r,* low *r*) for the three intensity conditions (Habitual, Loud, Soft)

Variations in speed correlations across orofacial structures

As observed in the previous study (McClean and Tasko 2002), correlation levels relating speed to F0 and intensity for NS were greater for the jaw relative to the lips and tongue. This can be seen in Table 2 which shows the means of the various correlation measures. For the NS group dataset, repeated measures ANOVA were performed on the speed-F0 and speed-intensity correlations with articulatory structure as the single factor. ANOVA results indicated that orofacial structure was significant for both the speed-F0 correlations (F=15.7, df=214, p<0.0005) and the speed-intensity correlations (F=23.2, df=214, p<0.0005). Post hoc Bonferroni comparisons of mean correlation levels further showed that for both F0 and intensity, JO and JC correlations were significantly greater than that for TC (p<0.0005), the JC correlation was greater than those for UO and LO (p<0.0005), and the LO correlation was greater than that for TC (p=0.01).

Discussion

The chief finding of the present study is that correlations relating lower lip and tongue speeds to F0 and intensity were reduced in PS compared to NS, while no group differences were seen for similar correlations obtained on the jaw and upper lip. Comparisons of the output measures of the two groups showed equivalent levels of variation in F0 and intensity across the three intensity conditions. While there was evidence for slightly less pronounced changes in tongue speed in the low intensity condition for PS, the variance of tongue and lower lip speed was equivalent across groups. Thus, given that the groups performed the speech tasks in approximately equivalent manners, we interpret the reduced correlation levels in PS as reflecting relatively weak connectivity in neural pathways linking the tongue and lower lip to those of the respiratory-laryngeal system. However, this interpretation is presented with certain precautions. First, it seems likely that the strength of neural coupling across muscle systems varies during the time course of speech utterances, and therefore, the correlation patterns and group differences obtained here are likely to reflect conditions within a restricted phase of a dynamic process. Also, it is uncertain whether the observed group differences in correlation level would generalize to more diverse and natural speech samples. The speech sample employed was selected in order to elicit a wide range of vocal tract excursions in association with vocal tract opening and closing. It may be that the type of speech task used here is required to generate a necessary distribution of output measures for correlations to be sufficiently strong and sensitive to group differences.

In agreement with the results of a prior study, correlations in NS relating jaw speed to F0 and intensity were significantly greater than similar correlations for the lips and tongue (McClean and Tasko 2002). This finding was replicated in the present study with a much larger group of subjects and more refined statistical methods. In addition, it was found that the lips tend to show greater speed correlation levels compared to the tongue, a trend that also was noted in the previous study when comparing the lower lip and tongue. The differences in correlation levels are taken to reflect variations in the strength of neural coupling between orofacial subsystems and the respiratory-laryngeal system. The principal assumption underlying this interpretation is that the various output measures reflect muscle activation levels within the muscle subsystems considered. New support for this assumption is provided by a recent study showing strong linear associations between orofacial speeds and EMG levels across variations in vocal intensity using the same test utterance employed here (McClean and Tasko 2003). It is notable that equivalent correlation levels were obtained for the lower lip and jaw when relating muscle activation level and movement speed. This suggests that in spite of obvious biomechanical and muscle-property differences between the lower lip and jaw, speed tends to covary with neural drive in a similar manner for the two muscle systems.

Group differences in correlations relating orofacial speeds to F0 were most evident when comparing NS and PS who showed strong positive correlations between F0 and intensity. Subjects showing reduced F0-intensity correlations displayed elevated inter-trial variability in F0 and intensity, lower overall intensity levels, and proportionately less reduction in F0 with decreased intensity across loudness conditions. This may indicate that the balance of laryngeal and respiratory muscle forces varied across subjects showing different F0-intensity correlation levels, and that F0 and intensity were less indicative of level of muscle activation as the F0-intensity correlation was reduced. The basis for the association of F0 and intensity generally is attributed to a complex array of interacting factors. Relevant acoustic variables that contribute to the association include the characteristics of mouth radiation, vowel spectral shape, laryngeal-source spectral slope and harmonic density (Titze 1992). Additionally, F0 and intensity both tend to covary with tracheal pressure and related variations in respiratory and laryngeal muscle activity (Atkinson 1978; Finnegan et al. 2000; Shipp 1982; Titze et al. 1989; Titze 1991).

The above interpretation assumes that the observed variation in correlations relating F0 and intensity is related to peripheral mechanical and acoustic factors. While there is considerable interaction between peripheral physiologic systems underlying F0 and intensity control, it is possible that some portion of their covariation is related to central neural control. Thus, the reduced F0-intensity correlations seen in some subjects may reflect reduced neural coupling between these independent control systems. There is evidence for substantial intersubject variability in the peripheral physiologic adjustments underlying covariation of F0 and intensity (Plant and Younger 2000). Further study is needed in order to develop more controlled methods for evoking variations in F0 and intensity. This could contribute to greater precision in efforts to quantify strength of neural coupling between orofacial muscle subsystems and the respiratory-laryngeal system as they are engaged in speech production.

Accepting the assumption that correlation levels relating orofacial speeds to F0 and intensity reflect the strength of neural coupling between the orofacial and respiratory-laryngeal systems, the elevated speed correlations observed for the jaw are consistent with the frame/content theory on the evolution and development of speech production (MacNeilage 1998). Because frame/content theory emphasizes the differences in the evolution and development of the neuromotor systems of the jaw relative to the lower lip and tongue, it may be relevant to the reduced correlation levels observed here for the lower lip and tongue in PS. Given the common neural innervation of the upper and lower lips, the lack of strong group differences in speed correlations obtained for the upper lip is not consistent with this perspective. This finding may be related to a reduced sensitivity of upper lip speed measures, which tend to have reduced magnitudes compared to the lower lip and tongue. Another possibility is that the evolution and ontogeny of the upper and lower lip neuromotor systems are relatively distinct. An ecological argument could be made for this given that the lower lip has a stronger mechanical linkage with the jaw, distinct reflex properties (McClean 1989; Wohlert 1996), and more extensive involvement in the phonetic repertoire of different languages (e.g., labiodental fricatives in English).

From the above interpretations, it is reasonable to suggest that for some PS the neural mechanisms of speech disfluency are related in part to reduced strengths of neural coupling of the lower lip and tongue with the respiratory-laryngeal system. A specific mechanism through which such reduced connectivity might operate to contribute to disfluency pertains to the temporal coordination of different orofacial structures with the respiratory-laryngeal system. General theory on the behavior of neural systems (e.g., Abeles 1991) suggest that reduced strength of connectivity between neural populations (i.e., lower synaptic gains or more intervening synapses) is likely to be associated with response delays and increased variability in the relative timing of system output. Hence, from the pattern of speed correlation results one would expect greater variability in the timing of lower lip and tongue movements relative to respiratory-laryngeal system output in PS compared to NS, with equivalent variability between the two groups in the timing of jaw movements relative to respiratory-laryngeal output. When viewed in terms of developmental sequencing of oromotor behaviors for speech acquisition in children, increased delays and excessive variability in the timing of lower lip and tongue output might naturally lead to hesitations in output and discoordination across muscle systems.

Acknowledgement The research was supported by NIH grant DC 03659.

References

Abeles M (1991) Corticonics: neural circuits of the cerebral cortex. Cambridge University Press, Cambridge

Atkinson JE (1978) Correlation analysis of the physiological factors controlling fundamental voice frequency. J Acoust Soc Am 63:211–222

Brown SF (1938) Stuttering with relation to word accent and word position. J Abnorm Soc Psychol 33:112–120

Caruso A, Abbs J, Gracco V (1988) Kinematic analysis of speech multiple movement coordination in stutterers. Brain 111:439–455

Davis BL, MacNeilage PF (1995) The articulatory basis of babbling. J Speech Lang Hear Res 38:1199–1211

Denny M, Smith A (1992) Gradations in a pattern of neuromuscular activity associated with stuttering. J Speech Hear Res 35:1216–1229

Dromey C, Ramig LO (1998) 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

Edwards AL (1967) Statistical methods. Holt, Rinehart & Winston, New York

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

Green JR, Moore CA, Reilly KJ (2002) The sequential development of jaw and lip control for speech. J Speech Lang Hear Res 45:66–79

Grillner S (1982) Possible analogies in the control of innate motor acts and the production of sound in speech. In: Grillner S, Lindblom B, Lubker J, Persson A (eds) Speech motor control. Pergamon Press, Oxford, pp 217–230

Kalvarem KT (1993) A neural-network model enabling sensorimotor learning: Application to the control of arm movements and some implications for speech motor control and stuttering. Psych Res 55:299–314

Kent R (1984) Stuttering as a temporal programming disorder. In: Curlee R, Perkins WH (eds) Nature and treatment of stuttering —new directions. Little, Brown and Co., Boston, pp 283–301

Lenneberg EH (1967) Biological foundations of language. John Wiley, New York

MacNeilage PF (1998) The frame/content theory of evolution of speech production. Behav Brain Sci 21:499–546

MacNeilage PF, Davis BL (2001) Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications. Curr Opin Neurobiol 11:696–700

Max L, Caruso AJ, Gracco VL (2003) Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. J Speech Lang Hear Res 46:215–232

McClean MD (1989) Gain and spatial characteristics of human lip-muscle reflexes. Brain Res 503:16–21

McClean MD, Runyan CR (2000) Variations in the relative speeds of orofacial structures with stuttering severity. J Speech Lang Hear Res 43:1524–1531

McClean MD, Tasko SM (2002) Association of orofacial with laryngeal and respiratory motor output during speech. Exp Brain Res 146:481–489

McClean MD, Tasko SM (2003) Association of orofacial muscle activity and movement during changes in speech rate and intensity. J Speech Lang Hear Res 46:1387–1400

McClean MD, Levandowski DR, Cord MT (1994) Intersyllabic movement timing in the fluent speech of stutterers with different disfluency levels. J Speech Hear Res 37:1060–1066

Plant RL, Younger RM (2000) The interrelationship of subglottic air pressure, fundamental frequency, and vocal intensity during speech. J Voice 14:170–177

Riley G (1994) Stuttering Severity Index. Pro-Ed, Austin, TX

Sapir S, McClean M, Luschei E (1984) Time relations between cricothyroid muscle activity and the voice fundamental frequency (F0) during sinusoidal modulation of F0. J Acoust Soc Am 75:1639–1641

Shipp T (1992) Aspects of voice production and motor control In: Grillner S, Lindblom B, Lubker J, Persson A (eds) Speech motor control. Pergamon Press, Oxford, pp 105–112

Smith A, Kleinow J (2000) Kinematic correlates of speaking rate changes in stuttering and normally fluent adults. J Speech Lang Hear Res 43:521–536

Smith A, Luschei E, Denny M, Wood J, Hirano M, Badylak S (1993) Spectral analyses of activity of laryngeal and orofacial muscles in stutterers. J Neurol Neurosurg Psychiatry 56:1303–1311

Tasko SM (1999) Kinematic factors underlying timing in connected speech. Unpublished Doctoral Dissertation, University of Wisconsin

Taylor IK (1966) The properties of stuttered words. J Verb Learn Verb Behav 6:112–118

Titze IR (1991) A model for neurologic sources of aperiodicity in vocal fold vibration. J Speech Hear Res 34:460–472

Titze IR (1992) Acoustic interpretation of the voice range profile (phonetogram). J Speech Hear Res 35:21–34

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

Watson PJ, Ciccia AH, Weismer G (2003) The relation of lung volume initiation to selected acoustic properties of speech. J Acoust Soc Am 113:2812–2819

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

Wohlert AB (1996) Reflex responses of lips muscles in young and older women. J Speech Lang Hear Res 39:578–589

Yairi E, Ambrose N, Cox N (1996) Genetics of stuttering: a critical review. J Speech Lang Hear Res 39:771–784

Zimmermann G (1980) Articulatory dynamics of fluent speech utterances of stutterers and nonstutterers. J Speech Hear Res 23:95–107