# Articulatory Tradeoffs Reduce Acoustic Variability During American English /r/ Production

**Frank H. Guenther**[1,2], **Carol Y. Espy-Wilson**[3,2], **Suzanne E. Boyce**[4,2],
**Melanie L. Matthies**[5,2], **Majid Zandipour**[2,1], **and Joseph S. Perkell**[2,6]

Address correspondence to:
Prof. Frank H. Guenther
Boston University
Center for Adaptive Systems and
Department of Cognitive and Neural Systems
677 Beacon Street
Boston, MA, 02215
Fax Number: (617) 353-7755
Email: guenther@cns.bu.edu

## ABSTRACT

The American English phoneme /r/ has long been associated with large amounts of articulatory variability during production. This paper investigates the hypothesis that the articulatory variations used by a speaker to produce /r/ in different contexts exhibit systematic tradeoffs, or articulatory trading relations, that act to maintain a relatively stable acoustic signal despite the large variations in vocal tract shape. Acoustic and articulatory recordings were collected from seven speakers producing /r/ in five phonetic contexts. For every speaker, the different articulator configurations used to produce /r/ in the different phonetic contexts showed systematic tradeoffs, as evidenced by significant correlations between the positions of transducers mounted on the tongue. Analysis of acoustic and articulatory variabilities revealed that these tradeoffs act to reduce acoustic variability, thus allowing relatively large contextual variations in vocal tract shape for /r/ without seriously degrading the primary acoustic cue. Furthermore, some subjects appeared to use completely different articulatory gestures to produce /r/ in different phonetic contexts. When viewed in light of current models of speech movement control, these results appear to favor models that utilize an acoustic or auditory target for each phoneme over models that utilize a vocal tract shape target for each phoneme.

[1]Department of Cognitive and Neural Systems, Boston University
[2]Research Laboratory of Electronics, Massachusetts Institute of Technology
[3]Department of Electrical and Computer Engineering, Boston University
[4]Department of Communication Sciences and Disorders, University of Cincinnati
[5]Department of Communication Disorders, Boston University
[6]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

# 1. Introduction

The American English phoneme /r/ has long been associated with relatively large amounts of articulatory variability (Alwan, Narayanan, and Haker, 1997; Delattre and Freeman, 1968; Espy-Wilson and Boyce, 1994; Hagiwara, 1994, 1995; Ong and Stone, 1998; Westbury, Hashi, and Lindstrom, 1995, 1998). In fact, the endpoints of the articulatory continuum for /r/ can be analyzed as functionally different articulator configurations that use different primary articulators (tongue tip vs. tongue dorsum). These endpoints have been characterized in the literature as "bunched" (using the tongue dorsum) and "retroflexed" (using the tongue blade/tip). Often, the same speaker will use different types of /r/ in different productions, e.g., in different phonetic contexts. At the same time, the primary acoustic cue for /r/ is relatively simple and stable: a deep dip in the trajectory of the third spectral energy peak of the acoustic waveform, or third formant frequency (F3) (Boyce and Espy-Wilson, 1997; Delattre and Freeman, 1968; Westbury et al., 1995, 1998). Furthermore, no consistent acoustic difference between bunched and retroflexed /r/'s has been discovered.

How is it that a speaker can produce perceptually acceptable /r/'s despite using such variable vocal tract shapes? One possible answer to this question is that the variations in vocal tract shape for /r/ are not haphazard, but are instead systematically related in a way that maintains a relatively stable acoustic signal across productions despite large variations in vocal tract shape across productions. In other words, the different vocal tract shapes used to produce /r/ by a particular subject might involve articulatory tradeoffs, or *trading relations*. The concept of articulatory trading relations is illustrated by the following example. Assume that narrowing either of two constrictions at different locations along the vocal tract (call them location 1 and location 2) has the same effect on an important acoustic cue for a phoneme. Assume further that narrowing either constriction causes a reduction in F3. In such a case, one could use different combinations of the two constrictions to achieve the same acoustic effect. For example, to achieve a particular value of F3, one might form a very narrow constriction at location 1 and a less narrow constriction at location 2, or one might alternatively form a very narrow constriction at location 2 and a less narrow constriction at location 1. If a speaker used one of these options in one phonetic context and the other option in a second phonetic context, a negative covariance between the sizes of these two constrictions would be seen across phonetic contexts.

The primary purpose of the current study is to investigate the issue of whether the various vocal tract shapes used by an individual to produce /r/ in different phonetic contexts exhibit articulatory trading relations that act to maintain a relatively stable acoustic signal. As discussed at the end of this article, this issue has important implications for theories of speech motor control and speech production. Largely for this reason, several recent experiments have investigated the trading relations issue for phonemes other than /r/ (e.g., de Jong, 1997; Perkell et al., 1993; Perkell, Matthies, and Svirsky, 1994; Savariaux et al. 1995a), but the results have not been uniform across subjects: although most subjects exhibit expected articulatory trading relations, some others do not. A possible reason for this ambiguity is that these studies have primarily concentrated on one hypothesized trading relationship, and subjects who do not exhibit this trading relation may be exhibit other, unanalyzed trading relations that act to reduce acoustic variability. For example, Perkell et al. (1993) investigated an hypothesized trading relation between lip rounding and tongue body raising for the vowel /u/. Three of four subjects showed weak trading relations, but the fourth subject showed the opposite pattern. This fourth subject may have been using other trading relations that overrode the effect of the lip rounding/tongue body raising relationship. In the current study, we employ analysis procedures that allow us to assess the combined effects of multiple articulatory covariances on the variability of the acoustic signal. Furthermore, American English /r/ was chosen[1] because the large amount of articulatory variability associated with /r/ productions should make it easier to detect trading relations if they are present.

# 2. Methods

## 2.1. Data collection

An electromagnetic midsagittal articulometer (EMMA) system (Perkell et al., 1992) was used to track the movements of six small (5mm long x 2.5mm diameter) transducer coils. The coils were attached in the midsagittal plane to the tongue (3 coils), lips (2 coils), and lower incisor (1 coil) with bio-compatible adhesive. Transducers were also placed on the upper incisor and the bridge of the nose, for defining a coordinate system with a maxillary frame of reference. A directional microphone was suspended 14 inches from the subject's mouth and the acoustic signal was recorded simultenously with the EMMA signals. Standard EMMA calibration protocols were completed prior to each experiment (c.f. Perkell et al., 1992 for details). The current study focused on the positions of the three tongue transducers, which were located approximately 1, 2.5, and 5 cm back from the tongue tip (with the tongue in a neutral configuration).

The seven subjects were young adults, two females (Subjects 2 and 3) and five males. They had no history of speech, language, or hearing deficits or pronounced regional dialects. Each of the seven subjects produced 4-6 repetitions of the carrier phrase "Say ____ for me" for each of the five test utterances; /warav/, /wabrav/, /wadrav/, /wagrav/, and /wavrav/. The articulatory and acoustic data from these utterances were time-aligned to allow direct comparison between the two data types.

## 2.2. F3 extraction and alignment

The minimum measured F3 value during /r/ production, which can be thought of as the acoustic "center" of /r/, served as a landmark for time-alignment of the data across utterances for each speaker. Formant tracks were computed for all utterances using the ESPS/WAVES formant tracker and a 51.2 ms window and 3.2 ms frame rate. The F3 minimum was detected using an automatic procedure that first identified all sonorant regions, then located the point of minimal F3 from the relevant sonorant regions. F3 values and transducer positions within a 140 ms time window centered at the F3 minimum were extracted. Extracted F3 traces for some utterances were corrupted due to technical difficulties in automatically tracking low amplitude and low frequency values of F3 after stop consonants. Therefore, utterances whose F3 tracks changed by more than 200 Hz in a 3.2 ms time step were eliminated from the study, leaving 12 to 27 analyzed utterances per subject. After this elimination process, the tongue shapes at the F3 minimum of the remaining utterances were visually inspected, and two additional utterances (one each for Subjects 1 and 4) were identified as having articulations that were incorrectly labeled as /r/ by the automatic extraction process. These two utterances were also eliminated from the study.

## 2.3. Effects of vocal tract shape parameters on F3

The vocal tract shape for /r/ involves a palatal constriction formed by the tongue in the anterior third of the tract. Roughly speaking, the third formant frequency (F3) of /r/ is the resonance resulting from the cavities anterior to the palatal constriction (e.g., Alwan et al., 1997; Espy-Wilson, Narayanan, Boyce, and Alwan, 1997; Stevens, 1998). This part of the vocal tract consists of an acoustic compliance due to a large front cavity volume and two parallel acoustic masses due to natural tapering by the teeth/lips and the palatal constriction behind the front cavity. The resulting resonance is inversely proportional to the product of the total acoustic mass and the acoustic compliance. Because it is difficult to accurately infer lip aperture from EMMA data, we focus on the effects of the acoustic mass due to the size and location of the palatal constriction. From these considerations, we conclude that the frequency of F3 can be decreased by tongue movements that lengthen the front cavity (thereby increasing the acoustic compliance of the front cavity), lengthen the constriction (thereby increasing the acoustic mass of the constriction behind the front cavity), or decrease the area of the constriction (thereby increasing the acoustic mass of the constriction)[2].

The predicted effects of these movements on F3 were confirmed using vocal tract area functions derived from structural MRI scans of a speaker producing /r/[3]. Two area functions were derived: one representing a "bunched" /r/ configuration, and one representing a "retroflexed" /r/ configuration. Three manipulations were carried out on each area function to test the effects on F3 predicted from acoustic theory: (i) the palatal constriction was extended backward by narrowing the vocal tract area immediately behind the constriction, (ii) the front cavity was lengthened by displacing the palatal constriction backward, and (iii) the vocal tract area at the palatal constriction was decreased. For all three manipulations, an acoustic signal was synthesized (using S. Maeda's VTCALCS program; Maeda, 1990) and compared to the signal synthesized from the original area function. Each manipulation resulted in a lower F3 in both the bunched and retroflexed /r/ cases, as expected from the acoustic theory analysis.

Because all three manipulations act to lower F3, subjects could maintain a relatively stable F3 despite vocal tract shape variations across contexts if these variations involved tradeoffs between the different manipulations. When looking at the different vocal tract shapes for /r/ across contexts, these tradeoffs would be manifested by correlations between constriction length, front cavity length, and constriction area. Specifically, the following three correlations would be expected to aid in maintaining a relatively stable F3 across utterances while allowing variations in vocal tract shape:

(1) a negative correlation between constriction length and front cavity length, since increases in constriction length and front cavity length both act to reduce F3,

(2) a positive correlation between constriction length and constriction area, since increases in constriction length reduce F3 and decreases in constriction area reduce F3, and

(3) a positive correlation between front cavity length and constriction area, since increases in front cavity length reduce F3 and decreases in constriction area reduce F3.

## 2.4. Predicted articulatory covariances

To determine whether a subject uses any of the three trading relations hypothesized above, we must first describe the trading relations in terms of the x and y coordinates of the tongue transducers. For tongue configurations during /r/ production, a forward movement of the tongue front transducer generally corresponds to a shortening of the front cavity, an upward movement of the tongue front transducer generally corresponds to a decrease in the area of the palatal constriction for /r/, and, since the point of maximal constriction for /r/ is typically anterior to the tongue back transducer, an upward movement of the tongue back transducer generally corresponds to a lengthening of the palatal constriction and possibly a decrease in the area of the constriction. When determining the signs of the transducer coordinate correlations corresponding to the trading relations delineated above, we must take into account that increasing values of the tongue front horizontal position correspond to decreases in front cavity length, and increasing values of the tongue front vertical position correspond to decreases in constriction area. From these considerations, we can surmise that the three trading relation strategies described above should be evidenced by the following articulatory correlations:

(1) a positive correlation between tongue back height and tongue front horizontal position,

(2) a negative correlation between tongue back height and tongue front height, and

(3) a positive correlation between tongue front horizontal position and tongue front height.

Note that the use of all three trading relations by a single subject is unlikely given that they impose competing constraints; i.e., if tongue back height and tongue front horizontal position are positively correlated as in relation (1), and tongue front horizontal position and tongue front height are positively correlated as in relation (3), it is very likely that tongue back height and tongue front height will also be positively correlated, thus violating relation (2).

## 2.5. Analysis of articulatory and acoustic variances

To quantify the combined effects of articulatory covariances on F3 variability, an analysis was performed using both acoustic and articulatory data to estimate F3 variance as a function of articulatory variances. The relationship between transducer coordinates and F3 during /r/ can be written for each speaker as follows:

$$F3 = A_0 + \sum_{i=1}^{N} A_i\, c_i + E \tag{1}$$

where the $A_i$ are constants, the $c_i$ are the transducer coordinates, $N$ is the number of transducer coordinates considered in the analysis, and $E$ is a residual term that accounts for the effects on F3 due to all other sources, including articulators not included in the analysis, measurement errors, and nonlinearities in the relationship between F3 and the transducer coordinates. The equation relating F3 variance to articulatory variances at each point in time is then:

$$Var(F3) = \sum_i A_i^2 Var(c_i) + Var(E) + 2\sum_i \sum_{i<j} A_i A_j Cov(c_i, c_j) + 2\sum_i A_i Cov(c_i, E). \tag{2}$$

To determine the effects of articulatory covariances on F3 variability, we can compare the variance estimate of Equation 2 to the following variance estimate that excludes the covariances between the analyzed transducer coordinates:

$$Var(F3) = \sum_i A_i^2 Var(c_i) + Var(E) + 2\sum_i A_i Cov(c_i, E). \tag{3}$$

If the F3 variance estimate in the absence of articulatory covariances (Equation 3) is significantly larger than the variance estimate including the articulatory covariances (Equation 2), we conclude that the primary effect of the articulatory covariances is a reduction in the variance of F3.

Strictly speaking, a comparison of the F3 variance estimates in Equations 2 and 3 tells us only about the effects of the covariances of the *linear* component of each transducer's relation to F3. However, the relationship between F3 and transducer coordinates should be linear near a particular configuration of the vocal tract, since F3 is presumably a continuous nonlinear function of the vocal tract area function, and such functions are locally linear. One would further expect that the relationship is still approximately linear for the relatively limited range of vocal tract configurations utilized by a particular subject for /r/. The linear approximations reported below captured approximately 80% of the variance when using only three pellet coordinates, providing support for the assertion that the primary effect of articulatory covariances on F3 variance can be captured by considering only the linear component of each transducer's relationship to F3. Furthermore, the sign (positive or negative) of an articulatory covariance's contribution to F3 variance depends only on the sign of the corresponding $A_i$ terms, and we are primarily interested in the sign of the combined effects of articulatory covariances on F3 variance. The expected signs of the $A_i$ for tongue back height, tongue front horizontal position, and tongue front height can be deduced from acoustic theory considerations (Sections 2.3 and 2.4). $A_i$ values were estimated for each subject using muliple linear regression on the acoustic and articulatory data. As discussed in Section 3.4, all 21 estimated $A_i$ values (3 values for each of 7 subjects) were of the sign expected from these acoustic theory considerations.

# 3. Results

## 3.1. Temporal progression of tongue shapes

Figures 1 through 7 show sample lingual articulations used to produce /r/ in the five contexts by the seven subjects. For each context, two schematized tongue shapes and a palatal trace[4] are shown. The tongue shape schematics were formed by connecting the three tongue transducers with straight lines. The solid tongue shape corresponds to the point in time at which F3 reached its minimum value. The tongue shape 70 ms prior to the F3 minimum is indicated by dashed lines. The movement of the tongue can thus be roughly characterized as a transition from the dashed tongue shape to the solid tongue shape. This movement corresponds to the articulation toward the "acoustic center" of /r/; i.e., the portion of the movement up to the point in time of the F3 minimum.

Inspection of the lingual articulations for some subjects suggests that these subjects utilize different articulatory gestures, aimed at different vocal tract shapes, to produce /r/ in different phonetic contexts. For example, the backward movement of the tongue, with a slight downward movement of the tongue blade, used by Subject 1 to produce the /r/ in /wadrav/ does not appear to be aimed at the same vocal tract shape for /r/ as the upward movements of the tongue blade used by the same subject to produce /r/ in the /warav/, /wabrav/, and /wavrav/ contexts (Figure 1). Similarly, the downward movement of the tongue blade used by Subject 2 to produce the /r/ in /wadrav/ does not appear to be aimed at the same vocal tract shape as the upward movements of the tongue blade used by the same subject to produce /r/ in /warav/, /wabrav/, or /wavrav/ (see Figure 2). Additional examples of this phenomenon can be seen in Figures 1-7. The possible relevance of these observations to theories of speech motor control will be addressed in the Discussion section.

## 3.2. Tongue shapes at acoustic center of /r/

Figure 8 shows tongue configurations at the F3 minimum of /r/ for each of the seven speakers. For each utterance, the three tongue transducer positions are connected by a straight line. The tongue configurations for all repetitions in all phonetic contexts are superimposed for each speaker. Thus, the fact that different numbers of utterances were analyzed for different subjects and contexts is reflected in this figure. As previously reported elsewhere (e.g., Delattre and Freeman, 1968; Hagiwara, 1994, 1995; Ong and Stone, 1998; Westbury, Hashi, and Lindstrom, 1995), a wide range of tongue shapes is seen both within and across subjects. Also of note is the fact that, although most subjects seem to use an approximate continuum of tongue shapes (e.g., S2, S3, S6, and S7), others show a more bimodal distribution of tongue shapes (e.g., S4, S5). Finally, the tongue shapes across subjects appear to form an approximate continuum between a bunched configuration (e.g., S6) and a retroflexed configuration (e.g., S4). A more detailed indication of the effects of the different phonetic contexts on the tongue shapes for /r/ can be gained from Figure 9, which shows the average tongue shapes used by each subject in each phonetic context, coded by phonetic context. Figures 10 through 16 show the corresponding average F3 traces, starting from the point of the F3 minimum for /r/ and continuing for 70ms, for each speaker coded by phonetic context. With the exception of the /wadrav/ productions of Subject 2, which had considerably higher values of F3 than the other utterances for that subject, the subjects showed minimum F3 values well below 2000 Hz in all contexts, as expected from earlier studies of /r/ production.

Figure 17 shows the midsagittal palatal outline (thick solid line) and mean tongue shapes at the time of the F3 minimum for /r/ for each of the seven subjects. For each subject, mean configurations from two phonetic contexts (solid and dashed lines) are shown to illustrate the range of tongue shapes used by that subject. Tongue outlines were created by connecting the average positions of the three tongue transducers for a given utterance with a smooth curve to roughly approximate tongue shape[5]. A line was then extended
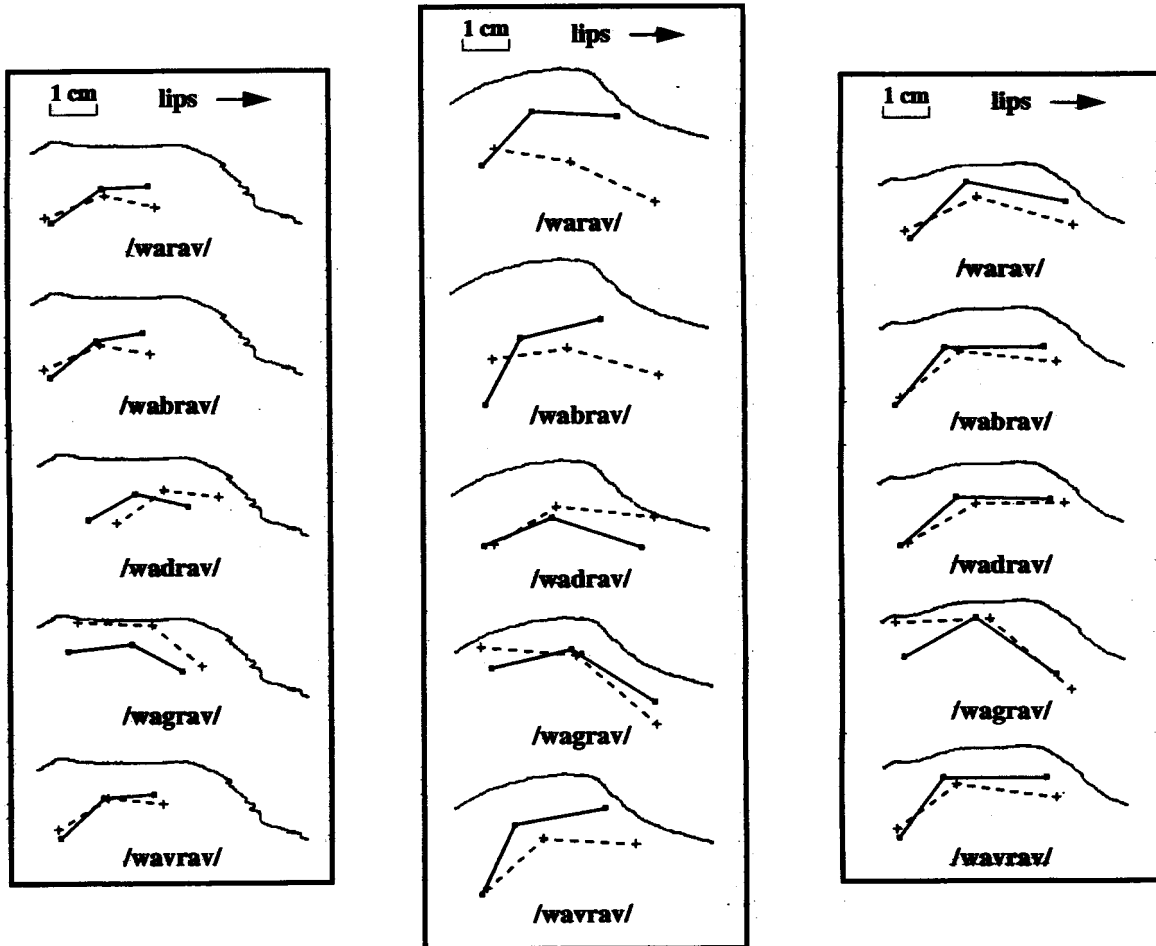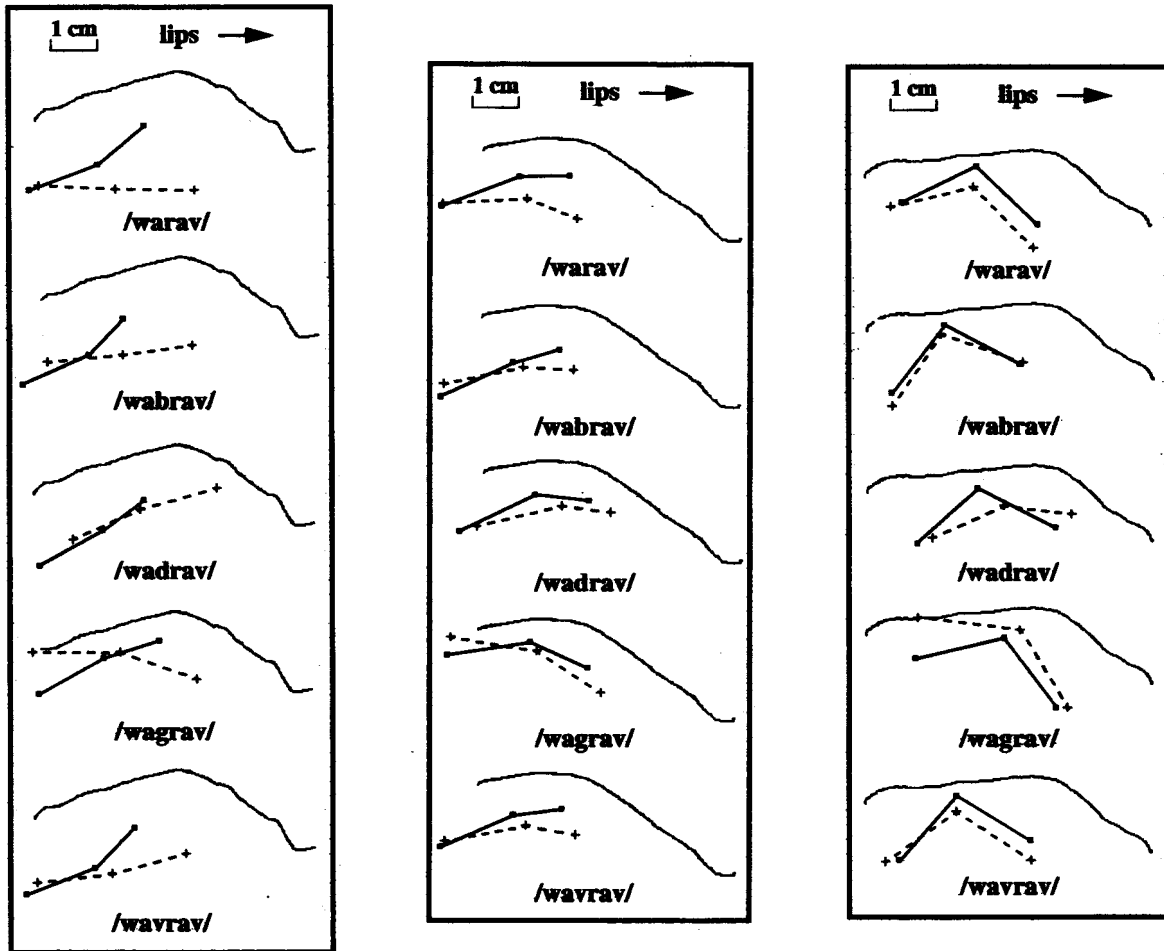
**Figures 1-3. Sample lingual articulations used by Subjects 1-3 to produce /r/ in the five phonetic contexts. For each context, two schematized tongues shapes and a palatal trace are shown. Each tongue shape schematic was formed by connecting the three tongue transducers with straight lines. The tongue shape at the F3 minimum for /r/ is drawn with solid lines. The tongue shape 70 ms prior to the F3 minimum is drawn with dashed lines.**

downward from the tongue front transducer position, then forward to the lower incisor transducer position, to provide a rough estimate of the relative size of the front cavity across contexts[6]. Also shown in the upper left corner of this figure are two superimposed, highly schematic vocal tract outlines that illustrate trading relations for maintaining a relatively stable F3. The effect on F3 of the longer front cavity of the dashed outline, which can be roughly characterized as a retroflexed /r/, is counteracted by the effects of the longer and slightly narrower constriction of the solid outline, which can be roughly characterized as a bunched /r/. Similarly, the vocal tract outlines for all subjects indicate that shorter front cavity lengths are accompanied by a compensating increase in constriction length and/or decrease in the constriction area. Furthermore, the tongue shapes during /wagrav/ (solid lines) are generally much closer in shape to tongue shapes for /g/ than are the /r/ shapes for /wabrav/ or /warav/ (dashed lines), suggesting that subjects utilize /r/ configurations that are reached relatively easily in the current phonetic context.

## 3.3. Articulatory trading relations

For each subject, Pearson correlation coefficients corresponding to the predicted covariances described in Section 2.4 were estimated across utterances at the point of F3 minimum and are listed in Table 1. All subjects showed a significant positive correlation between tongue back height (TBY in Table

**Figures 4-6. Sample lingual articulations used by Subjects 4-6 to produce /r/ in the five phonetic contexts. For each context, two schematized tongues shapes and a palatal trace are shown. Each tongue shape schematic was formed by connecting the three tongue transducers with straight lines. The tongue shape at the F3 minimum for /r/ is drawn with solid lines. The tongue shape 70 ms prior to the F3 minimum is drawn with dashed lines.**

1) and tongue front horizontal position (TFX), indicative of a trading relation between constriction length and front cavity length. Six of seven subjects also showed a second strong trading relation: five subjects showed a trading relation between constriction length and constriction area as evidenced by a negative correlation between TBY and tongue front height (TFY), and one subject showed a trading relation between front cavity length and constriction area as evidenced by a positive correlation between TFX and TFY. One subject (Subject 7) showed only very weak correlations other than the strong trading relation between tongue back height and tongue front horizontal position.

## 3.4. Analysis of acoustic and articulatory variabilities

The results in Section 3.3 indicate that most subjects exhibited two of three hypothesized articulatory trading relationships that act to reduce acoustic variability. Furthermore, as described in Section 2.4, it is unlikely or impossible for a subject to utilize all three trading relations because they counteract one another. However, it is still possible that the significant correlations that violate the trading relations could effectively "override" the beneficial articulatory tradeoffs, potentially nullifying or even reversing the

**Figure 7. Sample lingual articulations used by Subject 7 to produce /r/ in the five phonetic contexts. For each context, two schematized tongues shapes and a palatal trace are shown. Each tongue shape schematic was formed by connecting the three tongue transducers with straight lines. The tongue shape at the F3 minimum for /r/ is drawn with solid lines. The tongue shape 70 ms prior to the F3 minimum is drawn with dashed lines.**
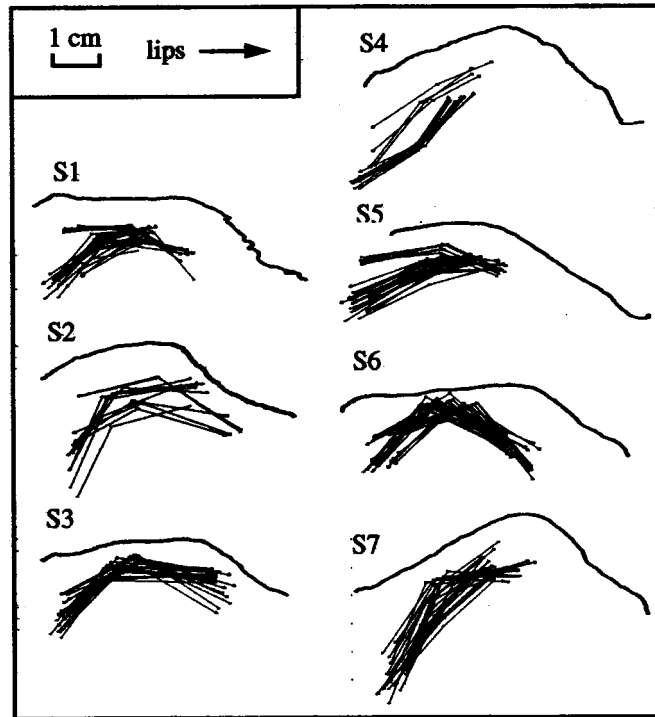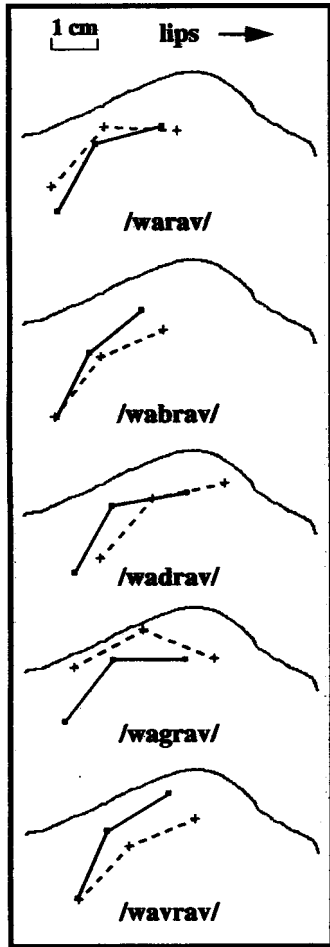




**Figure 8. Tongue configurations at the F3 minimum of /r/ for each of the seven speakers. For each utterance, the three tongue transducer positions are connected by straight lines. The tongue configurations for all repetitions in all phonetic contexts are superimposed for each speaker.**

TABLE I. Articulator correlation coefficients. Significant correlations that are consistent with hypothesized trading relations are shown in boldface. TBY=tongue back height; TFX=tongue front horizontal position; TTY =tongue front height.

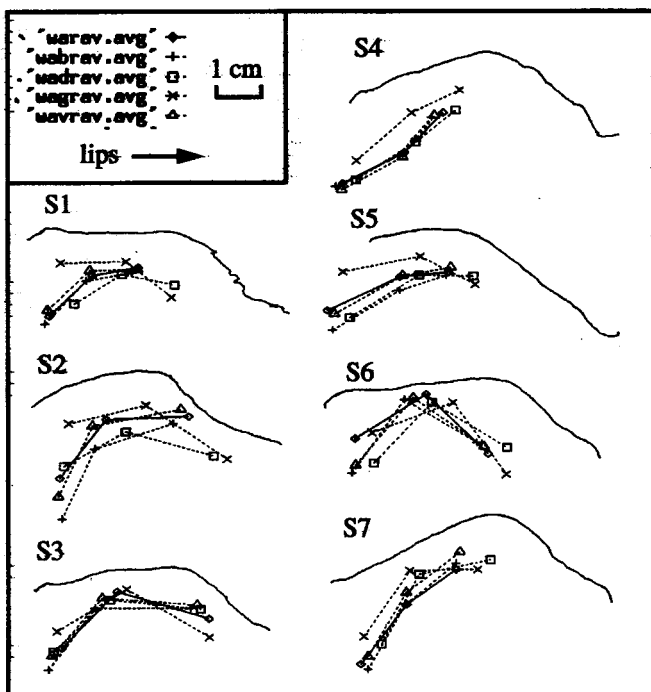| Subject | TBY-TFX | TBY-TFY | TFX-TFY |
|---------|---------|---------|---------|
| 1 | **0.74[a]** | **−0.75[a]** | −0.75[a] |
| 2 | **0.92[a]** | **−0.69[a]** | −0.88[a] |
| 3 | **0.77[a]** | **−0.74[a]** | −0.46 |
| 4 | **0.91[a]** | 0.89[a] | **0.82[a]** |
| 5 | **0.64[a]** | **−0.49[a]** | −0.57[a] |
| 6 | **0.55[a]** | **−0.81[a]** | −0.60[a] |
| 7 | **0.84[a]** | 0.05 | 0.06 |

[a]Statistically significant ($p<0.01$).

**Figure 9. Averaged tongue configurations at the F3 minimum of /r/ for each of the seven speakers. The averaged positions of the three tongue transducer positions for each of the five phonetic contexts are connected by straight lines.**

effect of the utilized trading relations on acoustic variability. It is therefore necessary to estimate the net effect of all three articulatory covariances, as outlined in Section 2.5.

F3 variance estimates with and without covariance terms (Equations 2 and 3, respectively) were calculated using the tongue back height, tongue front horizontal position, and tongue front height transducer coordinates. The corresponding F3 standard deviations were then averaged across subjects. The $A_i$ values for each speaker were estimated using multiple linear regression across utterances and time bins and are provided in Table 2; the value of $E$ for a particular time bin was simply the residual of the regression in that time bin. $R^2$ values for the F3 fit (without the residual term) ranged from 0.75 to 0.87 for the different subjects, with an average $R^2$ of 0.79. If covariances are high and the actual effect of an articulator's position on F3 is very low, the regression analysis can possibly result in estimates of transducer contributions that have the wrong sign, which could in turn cause some articulatory covariances to decrease estimated F3 variability when in reality they increase or have no significant effect on F3 variability. The fact that none of the transducer contribution estimates produced by the regression were of the opposite sign as expected from acoustic theory considerations and the MRI-based area function analysis indicates that this potential problem did not affect our results.

F3 standard deviation estimates with and without covariance terms are shown in Figure 18 as a function of time starting at the F3 minimum for /r/, averaged across subjects. (Standard deviations were plotted in place of variances to produce values whose units are Hz.) Also plotted is the standard deviation obtained from measured values of F3. When articulatory covariances are included, the F3 standard deviation estimate is equal to the measured F3 standard deviation; this is as expected because of the inclusion of the residual term in the variance estimate calculations. The solid line in the figure thus represents both the measured F3 standard deviation and the estimated F3 standard deviation including the covariance terms. When articulatory covariances are removed from the estimates, the estimated F3 standard deviation increases substantially. The dashed line in Figure 18 represents estimated F3 standard deviation without covariances using the three tongue transducer coordinates. According to this estimate, then, F3 standard deviation would be 105% higher at the acoustic center of /r/ if the articulatory tradeoffs had not been present.
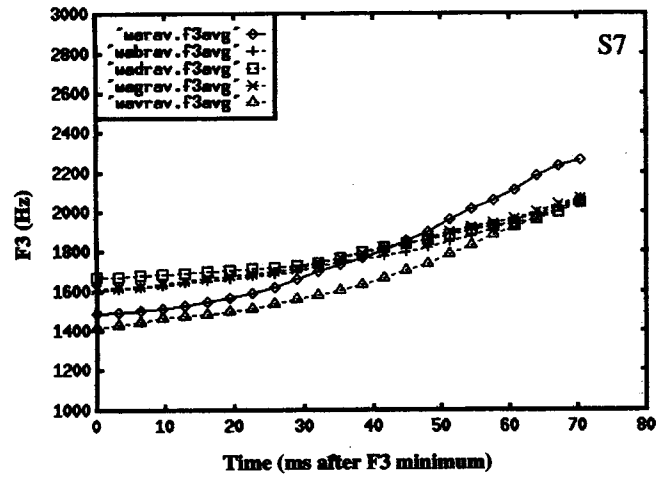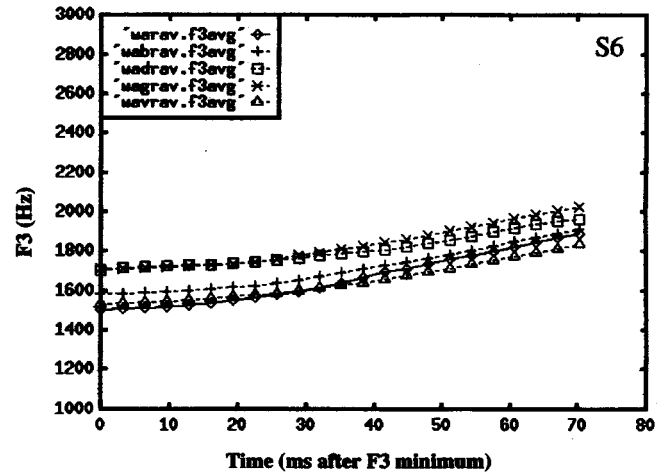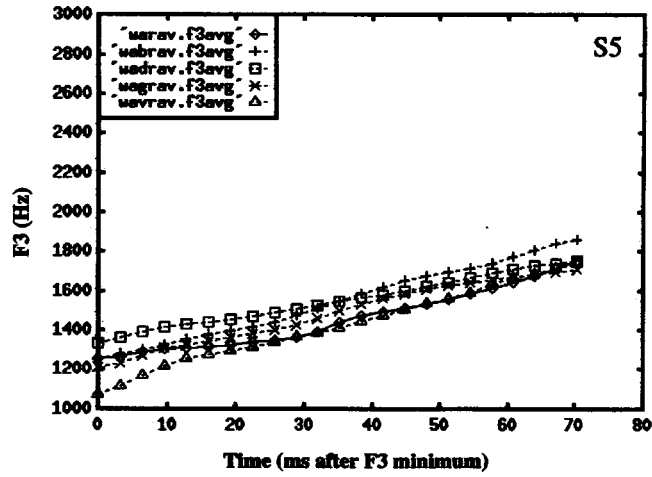
**Figures 14-16. Averaged F3 values for /r/ in the five phonetic contexts for Subjects 5-7. For each context, the averaged F3 is plotted as a function of time starting from the F3 minimum.**

TABLE II. Regression coefficients indicating the relationship between transducer coordinates and $F3$. Units are Hz/mm.

| Subject | $A_1$ (TBY) | $A_2$ (TTX) | $A_3$ (TTY) |
|---------|-------------|-------------|-------------|
| 1 | −28.25 | 15.20 | −35.61 |
| 2 | −81.13 | 92.77 | −35.25 |
| 3 | −12.28 | 25.93 | −51.50 |
| 4 | −36.83 | 77.34 | −36.55 |
| 5 | −24.04 | 21.44 | −30.49 |
| 6 | −21.68 | 10.99 | −30.95 |
| 7 | −46.29 | 33.71 | −31.87 |
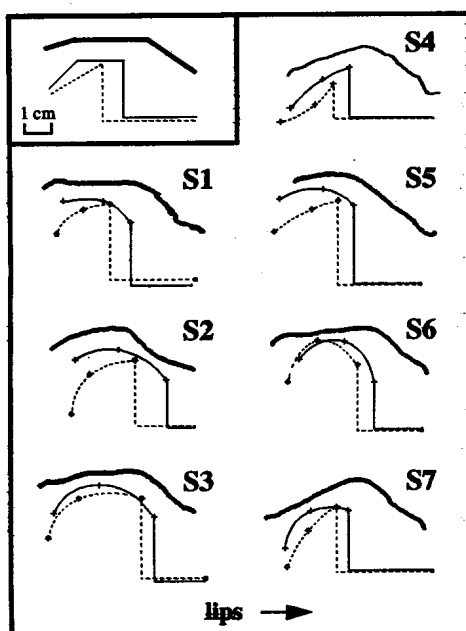


Figure 17. Trading relations during /r/ production. The upper left corner shows two superimposed, highly schematized vocal tract outlines (dashed and solid lines) illustrating trading relations between front cavity length and palatal constriction length and area. Also shown are vocal tract outlines that illustrate the range of tongue shapes used by each of the seven subjects to produce /r/ in different phonetic contexts. Thin solid lines correspond to the tongue shapes for /r/ in /wagrav/ (averaged across repetitions), and dashed lines correspond to the /r/ in /wabrav/ or /warav/, depending on the subject. Thick solid lines indicate palatal outlines. Each outline is formed by connecting the three tongue transducer positions with a smooth curve, then projecting downward and forward from the front-most tongue tongue transducer to the lower incisor transducer. All seven subjects show tradeoffs between the front cavity length and the constriction length and/or area when producing /r/ in the two different contexts.

The increase in the F3 variance estimate without covariances is seen at the F3 minimum for all subjects. This observation suggests that ambiguous results from previous studies may have been at least partly due to analyzing only one articulatory tradeoff at a time, since in our study no subject exhibited all three hypothesized trading relations, but all subjects showed a net decrease in acoustic variability due to the *combined* effects of the articulatory covariances. Assume, for example, that the data listed in each column of Table 1 were the result of an independent research study. Researchers investigating the trading relation in column 2 (in which 5 of 7 subjects used the trading relation) would sharply disagree with researchers investigating the trading relation in column 3 (in which only 1 of 7 subjects used the trading relation) as to whether or not trading relations are reliably exhibited, and both sets of researchers would report ambiguous results since in neither case do all subjects behave in the same way. A much clearer picture has emerged from the current study due to the analysis of the combined effects of the articulatory covariances.

Also evident in Figure 18 is a steady decrease with time of the effects of the covariance terms on F3 as the /r/ transitions into the following /a/. This decrease, evident in six of the seven subjects, is suggestive of

a decrease in the use of trading relations as vocal tract shape differences across utterances (due to the different phonemes preceding /r/ in different utterances) diminish.
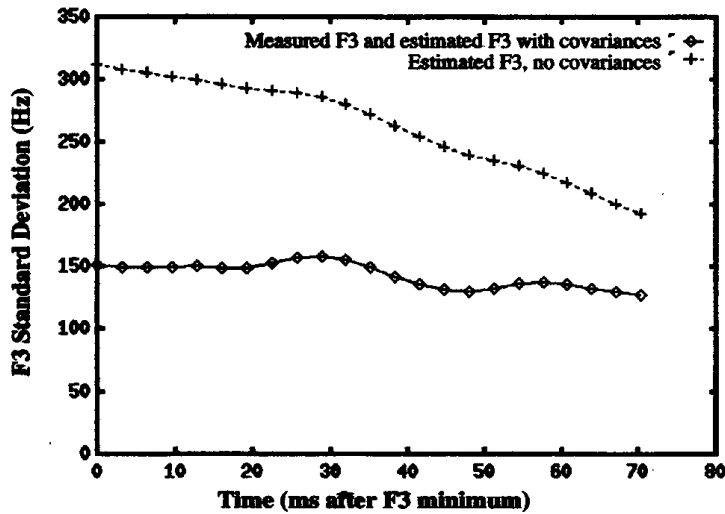


Figure 18. Comparison of the measured F3 standard deviation with estimates of F3 standard deviation derived from articulator transducer positions, averaged across subjects. Removal of the articulatory covariance terms results in a much higher estimate of F3 standard deviation, indicating that articulatory covariances greatly reduce acoustic variation.

## 4. Discussion

The results of this study indicate that the widely varying tongue shapes used by each subject to produce /r/ in different phonetic contexts exhibit systematic articulatory tradeoffs that act to reduce acoustic variability across contexts despite the large variations in vocal tract shape. The first piece of evidence arises from inspection of the tongue shape extremes for each subject (i.e., the averaged tongue shapes from the two phonetic contexts whose averaged tongue shapes differed the most), which show clear tradeoffs between the length of the front cavity and the length and/or area of the palatal constriction (Figure 17). An acoustic theory analysis of the effects of these aspects of vocal tract shape on F3 indicates that the systematic tradeoffs would be expected to reduce acoustic variability across contexts. Analysis of articulatory covariances indicates that most speakers exhibit two of three articulatory trading relationships that were hypothesized based on acoustic properties of the vocal tract (Section 3.3). Furthermore, the use of all three trading relationships is very unlikely if not impossible given that they counteract each other. Finally, analysis of the combined effects of these articulatory covariances indicates that they strongly influence F3 variability across contexts, effectively cutting F3 standard deviation in half compared to what it would have been without the articulatory covariances (Section 3.4; Figure 18).

Unlike earlier trading relations studies that reported mixed results across subjects (e.g., de Jong, 1997; Perkell et al., 1993; Perkell, Matthies, and Svirsky, 1994; Savariaux et al. 1995a), the reduction of F3 variability due to articulatory covariances was seen at the acoustic center of /r/ for all seven subjects in the current study. We believe that the following factors contributed to this difference. First, the current study investigated a phoneme known to exhibit a large amount of articulatory variability across contexts. Such a sound would be expected to exhibit stronger trading relations due to the larger overall articulatory variability. Second, the current study investigated the combined effects of multiple articulatory covariances. Although the combined effect of articulatory covariances was a reduction of F3 variability in all seven subjects according to the analysis of Section 3.4, different subjects used different combinations of the individual articulatory trading relations (Table 1). It is therefore not surprising that in earlier studies, which investigated articulatory covariances individually, some subjects did not use an hypothesized trading relationship. The results of the current report suggest that these subjects may well have used other, unanalyzed trading relations that reduced acoustic variability.
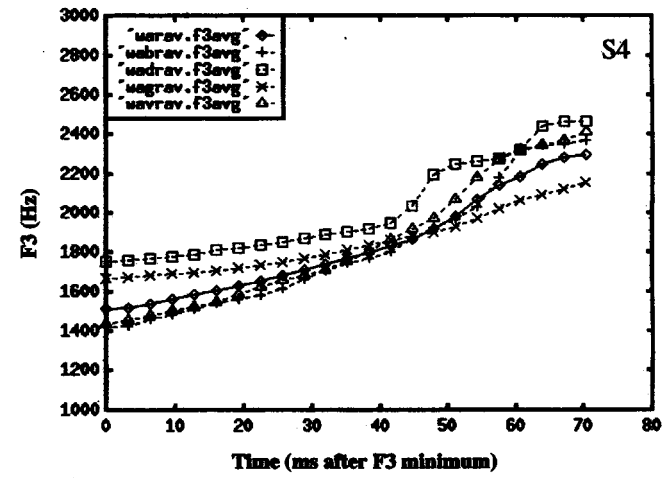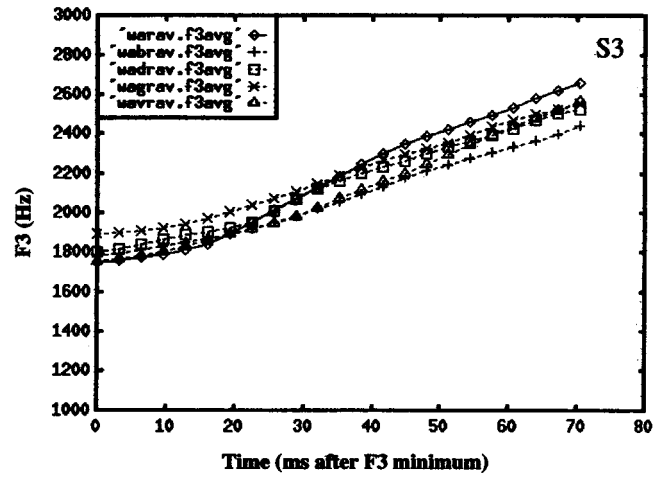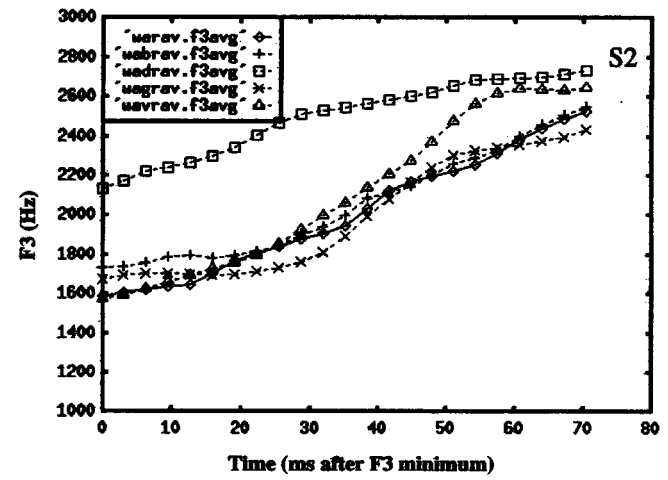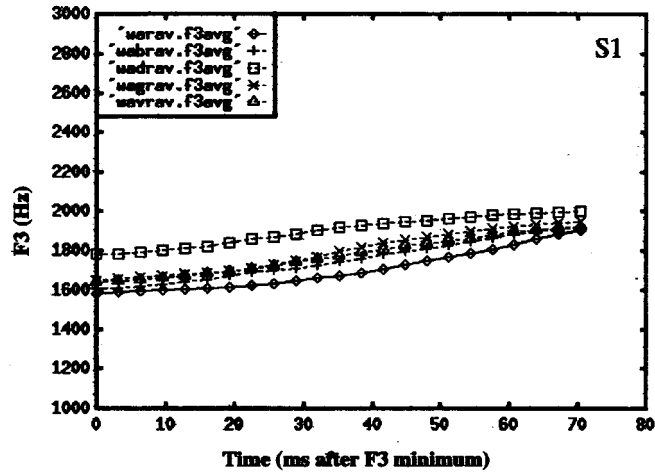
**Figures 10-13. Averaged F3 values for /r/ in the five phonetic contexts for Subjects 1-4. For each context, the averaged F3 is plotted as a function of time starting from the F3 minimum.**

The issue of articulatory trading relations in speech production is relevant to current theories concerning the control of speech movements. Roughly speaking, computational models of speech motor control can be classified according to the type of phonemic "targets" that they use to command movements of the speech articulators. One type of computational model, exemplified by the task-dynamic model of Saltzman and Munhall (1989), utilizes a target for each phoneme that specifies important aspects of the shape of the vocal tract for that phoneme. This "vocal tract shape target" view is closely related to theories of speech perception and production in which the articulatory gesture serves as the basic unit of speech. These include the motor theory of speech perception (Liberman and Mattingly, 1985; Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967), the direct realist theory of speech perception (Fowler, 1986, 1996), and the linguistic-gestural theory of phonology (Browman and Goldstein, 1990a,b). A second type of computational model, exemplified by the DIVA model of speech acquisition and production (Guenther, 1995; Guenther, Hampson, and Johnson, 1998), utilizes only an acoustic or auditory target for each phoneme, with no explicit vocal tract shape target. These models may use different shapes of the vocal tract to produce the same acoustic signal for a phoneme depending on things like phonetic context. Theories related to this "auditory target" view have been posited by various researchers (for some recent examples, see Johnson, Ladefoged, and Lindau, 1993; Perkell, Matthies, Svirsky, and Jordan, 1993, 1995; Perkell et al., 1997; Savariaux, Perrier, and Orliaguet, 1995a; Savariaux, Perrier, and Schwartz, 1995b), and many of the roots for this line of thinking can be traced to the influential work of Jakobson, Fant, and Halle (1951).

A major difference between the auditory target and vocal tract shape target computational model classes is that the former explicitly predict the existence of articulatory trading relations when producing the same phoneme in different contexts, whereas the latter do not. Because the current results show the existence of trading relations in all seven subjects, they appear to favor acoustic target models over vocal tract shape target models. A potential reason for the use of articulatory tradeoffs is that they can reduce the amount of effort required to move the articulators through a set of acoustic targets. For example, the tongue shapes for /r/ in /wagrav/ were generally closer to the tongue shapes for /g/ than the tongue shapes for /r/ in other contexts, suggesting that, to a first approximation, subjects moved to the closest vocal tract shape that could be used to produce the appropriate sound in the prevailing phonetic context. The acoustic target control scheme used by the DIVA model has this property.

It is possible, however, that the trading relations found in the current study are artifactual, rather than the result of a movement control process that is making use of them to minimize effort while maintaining an acceptable acoustic signal. For example, one might imagine that the tongue shape variations simply reflect a controller that undershoots the vocal tract shape target for /r/ (thus maintaining some aspects of the vocal tract shape from the preceding phoneme at the point of F3 minimum), rather than one that purposefully utilizes a different vocal tract shape for /r/ in different contexts because it "knows" how to use different shapes to achieve the same acoustic/auditory result in different contexts. However, the fact that some subjects appear to use articulatory gestures toward different vocal tract shapes to produce /r/ in different contexts argues against an "artifact" interpretation. For example, the backward movement of the tongue, with a slight downward movement of the tongue blade, used by Subject 1 to produce the /r/ in /wadrav/ does not appear to be a truncated movement aimed at the same vocal tract shape for /r/ as the upward movements of the tongue blade used by the same subject to produce /r/ in the /warav/, /wabrav/, and /wavrav/ contexts (see Figure 1). Instead, it appears that the /r/ gesture in /wadrav/ is toward a different vocal tract shape that nonetheless suffices for /r/. (See Section 3 and Figures 2-7 for further examples.) Further research into this issue will be carried out in a subsequent modeling study that will address other possible explanations for the trading relations found in the current study.

# 5. Acknowledgements

# 6. References

Alwan, A., Narayanan, S., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *Journal of the Acoustical Society of America*, **101**, pp. 1078-1089.

Boyce, S. & Espy-Wilson, C. Y. (1997). Coarticulatory stability in American English /r/. *Journal of the Acoustical Society of America,* **101**, 3741-3753 (1997).

Browman, C., & Goldstein, L. (1990a). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M. E. Beckman (Eds.), *Papers in laboratory phonology. I: Between the grammar and physics of speech* (pp. 341-376). Cambridge, UK: Cambridge University Press.

Browman, C., & Goldstein, L. (1990b). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, **18**, pp. 299-320.

de Jong, K. J. (1997). Labiovelar compensation in back vowels. *Journal of the Acoustical Society of America,* **101**, pp. 2221-2233.

Delattre, P., & Freeman, D. C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, **44**, pp. 29-68.

Espy-Wilson, C. Y., & Boyce, S. E. (1994). Acoustic differences between "bunched" and "retroflex" variants of American English /r/. *Journal of the Acoustical Society of America*, **95**, Pt. 2, p. 2823.

Espy-Wilson, C. Y., Narayanan, S., Boyce, S. E., and Alwan, A. (1997). Acoustic modeling of American English /r/. *Proceedings of Eurospeech, September 1997, Rhodes, Greece* (vo1 1., pp. 393-396). Grenoble, France: ESCA.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, **14**, pp. 3-28.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, **99**, pp. 1730-1740.

Guenther, F. H. (1995). A modeling framework for speech motor development and kinematic articulator control. *Proceedings of the XIIIth International Conference of Phonetic Sciences (*vol. 2, pp. 92-99). Stockholm, Sweden: KTH and Stockholm University.

Guenther, F. H., Hampson, M., and Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, in press.

Hagiwara, R. (1994). Three types of American /r/. *UCLA Working Papers in Phonetics*, **88**, pp. 63-90.

Hagiwara, R. (1995). Acoustic realizations of American /r/ as produced by women and men. *UCLA Working Papers in Phonetics*, **90**.

Jakobson, R., Fant, G., and Halle, M. (1951). Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, MA: MIT Press.

Johnson, K. Ladefoged, P., Lindau, M. (1993). Individual differencs in vowel production. *Journal of the Acoustical Society of America*, **94**, pp. 701-714.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, **74**, pp. 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech revisited. *Cognition*, **21**, pp. 1-36.

Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W.J. Hardcastle and A. Marchal (Eds.), *Speech production and speech modelling* (pp. 131-149). Boston: Kluwer Academic Publishers.

Ong, D., & Stone, M. (1998). Three-dimensional vocal tract shapes in /r/ and /l/: A study of MRI, ultrasound, electropalatography, and acoustics. *Phonoscope*, 1, pp. 1-13.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (1992). Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical. Society of America,* **92**, pp. 3078-3096.

Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F. H., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic segmental goals, saturation effects, auditory feedback and internal models. *Speech Communication*, **22**, 227-250.

Perkell, J. S., Matthies, M. L., & Svirsky, M. A. (1994). Articulatory evidence for acoustic goals for consonants. *Journal of the Acoustical Society of America*, **96**(5) Pt. 2, p. 3326.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel [u]: A pilot "motor equivalence" study. *Journal of the Acoustical Society of America*, **93**, pp. 2948-2961.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1995). Goal-based speech motor control: A theoretical framework and some preliminary data. *Journal of Phonetics,* **23**, pp. 23-35.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, **1**, pp. 333-382.

Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995a). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *Journal of the Acoustical Society of America*, **98**, pp. 2428-2442.

Savariaux, C., Perrier, P., & Schwartz, J. L. (1995b). Perceptual analysis of compensatory strategies in the production of the French rounded vowel [u] perturbed by a lip tube. *Proceedings of the XIIIth International Congress of Phonetic Sciences (*vol. 3, pp. 584-587).  Stockholm, Sweden: KTH and Stockholm University.

Stevens, K. (1998). *Acoustic phonetics*. In press.

Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1995). Differences among speakers in articulation of American English /r/: An x-ray microbeam study. *Proceedings of the XIIIth International Conference of Phonetic Sciences (*vol. 4, pp. 50-57).  Stockholm, Sweden: KTH and Stockholm University.

Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation of American English /ɹ/. *Speech Communication*, in press.

# FOOTNOTES

1. In the International Phonetic Alphabet, this phoneme is indicated by the symbol /ɹ/. We use the symbol /r/ here because it is more widely recognized and is unambiguous with regard to American English.

2. Although the direction of the effect on F3 for all of these vocal tract shape manipulations should be the same, the magnitude of the effect on F3 will be different for each manipulation. We account for this by including the magnitudes of the effects when analyzing the combined acoustic effect of these movements; this is done through the $A_i$ terms in Equations 1-3 below.

3. The vocal tract area functions were provided by Abeer Alwan and colleagues from the Electrical Engineering Department at the University of California, Los Angeles.

4. The palatal traces for Subjects 2 and 4 were slightly misaligned relative to the tongue transducer data. To correct for this, the palatal trace for these subjects have been raised approximately 3 mm relative to the tongue transducer positions in all figures.

5. Because any form of curve through the three tongue transducer points would represent only a rough approximation to the actual tongue shape, the curves in Figure 17 were simply hand-drawn using computer drawing software.

6. The lower incisor location was not available for Subject 2, so the vocal tract outlines for this subject in Figure 17 are based on a lower incisor position estimated from the lower lip position.