

---

# Vowel Category Dependence of the Relationship Between Palate Height, Tongue Height, and Oral Area

**Mark Hasegawa-Johnson**  
University of Illinois  
at Urbana-Champaign

**Shamala Pizza**  
**Abeer Alwan**  
**Jul Setsu Cha**  
University of California,  
Los Angeles

**Katherine Haker**  
Cedars-Sinai Medical Center  
Los Angeles

---

This article evaluates intertalker variance of oral area, logarithm of the oral area, tongue height, and formant frequencies as a function of vowel category. The data consist of coronal magnetic resonance imaging (MRI) sequences and acoustic recordings of 5 talkers, each producing 11 different vowels. Tongue height (left, right, and midsagittal), palate height, and oral area were measured in 3 coronal sections anterior to the oropharyngeal bend and were subjected to multivariate analysis of variance, variance ratio analysis, and regression analysis. The primary finding of this article is that oral area (between palate and tongue) showed less intertalker variance during production of vowels with an oral place of articulation (palatal and velar vowels) than during production of vowels with a uvular or pharyngeal place of articulation. Although oral area variance is place dependent, percentage variance (log area variance) is not place dependent. Midsagittal tongue height in the molar region was positively correlated with palate height during production of palatal vowels, but not during production of nonpalatal vowels. Taken together, these results suggest that small oral areas are characterized by relatively talker-independent vowel targets and that meeting these talker-independent targets is important enough that each talker adjusts his or her own tongue height to compensate for talker-dependent differences in constriction anatomy. Computer simulation results are presented to demonstrate that these results may be explained by an acoustic control strategy: When talkers with very different anatomical characteristics try to match talker-independent formant targets, the resulting area variances are minimized near the primary vocal tract constriction.

**KEY WORDS:** articulation, oral cavity, physiologic acoustics

---

**T**he speech production model of Stevens and House (1955) proposed that speech can be synthesized with good quality using a three-parameter model that controls the position and cross-sectional area of the primary vocal tract constriction and the cross-sectional area of the lips; cross-sectional areas at all other positions in the vocal tract are passively interpolated based on position and area of the primary constriction. Perkell and his colleagues (Perkell & Cohen, 1989; Perkell & Nelson, 1985) have demonstrated that the Stevens–House model is also a plausible model of human speech motor control, in that variability among repeated productions of the same vowel is minimized near a vocal tract constriction. In this article, we seek to demonstrate that variability among multiple talkers also obeys the constraints suggested by the Stevens–House model.

In this article, we explicitly measure variability of oral area, tongue height, and palate height at several places in the vocal tract. The only medical imaging technology capable of providing large databases of linked tongue height, oral area, and formant frequency information is magnetic resonance imaging (MRI; Baer, Gore, Gracco, & Nye, 1991; Narayanan, Alwan, & Song, 1997; Story, Titze, & Hoffman, 1996). There is no inherent limit on the number of participants in an MRI study or on the number of vowels produced per participant, but most published studies include no more than 2–4 participants, apparently because of the time required to segment a large image database. Besides the time required for segmentation, MRI suffers two important limitations. First, acoustic noise inside the scanner precludes simultaneous imaging and acoustic recording, so images and acoustic recordings represent different productions of a phoneme by the same talker. The intertoken standard deviation of midsagittal tongue height during repeated productions of the same vowel by any given talker is  $\sigma \approx 1.5$  to 2.5 mm (Perkell, 1996), so intervowel and intertalker differences in tongue height acquired from MRI should not be considered significant unless greater than at least  $2\sigma \approx 3$  to 5 mm. The second limitation of MRI is the time required for image acquisition. MRI is able to acquire a complete three-dimensional image of the vocal tract in approximately 25 s, or in 2–3 acquisition intervals of 8–13 s each. Most participants are capable of either sustaining a phoneme for 8–13 s at a time or silently maintaining phoneme position (“miming” the phoneme) for 25 s at a time. Mimed production has not been previously analyzed, but Engwall (2000) studied the relationship between sustained and dynamic phonation. He concluded that sustained MRI production of /a, ɪ, u, f, s, ʃ/ is not identical to dynamic production extrema observed in electropalatography (EPG) and electromagnetic midsagittal articulometer (EMMA) studies, but that MRI production may be usefully studied as an example of hyperarticulated production.

Wood (1979) analyzed X-ray tracings of midsagittal tongue shape during production of English and Arabic vowels and argued that the observed tongue shapes are most naturally classified into four places of articulation: pharyngeal (/ɑ, a, ʌ, æ/), uvular (/o, ɔ/), velar (/u, ʊ/), and palatal (/i, e, ɪ, ε/). Wood did not discuss the organization of vowels within each place of articulation category, but the within-class organization was clarified by the work of Harshman, Ladefoged, and Goldstein (1977). Harshman et al. demonstrated that the midsagittal tongue shape of vowels can be represented using two talker-independent shape factors. The four vowels most heavily weighted positively or negatively on each tongue shape factor are the palatal vowel /i/ (positive Factor 1 weighting), the velar /u/ (positive Factor 2 weighting), the uvular /o/ (negative Factor 1

weighting), and the pharyngeal /ɑ/ (negative Factor 2 weighting). The vowel /æ/ is almost equally weighted between the palatal and pharyngeal categories, whereas the vowel /ʊ/ is almost equally weighted between velar and uvular. The four palatal vowels can be rank-ordered according to their weighting on Factor 2, producing the list /i, e, ɪ, ε/.

The change in formant frequencies caused by a small change in vocal tract area depends on the product of the percentage area change multiplied by a mode-dependent sensitivity function (Chiba & Kajiyama, 1941). In order to communicate effectively, therefore, precise control of small cross-sectional areas (i.e., constrictions) is more important than precise control of large cross-sectional areas. Stevens and House (1955) demonstrated that all of the vowels of English may be synthesized using a vocal tract analog controlled by just three parameters: the length/area ratio of the lips and the position and radius of the primary lingual constriction. The Stevens–House model assumes a vocal tract radius that varies parabolically away from the primary constriction, but similar acoustic results have been produced using spline interpolation (Iskarous, 2000) and piecewise constant area functions (Dunn, 1950): Apparently, the shape of the vocal tract away from a constriction does not have as much acoustic salience as the exact position and area of the constriction itself.

The Stevens–House (1955) model suggests an efficient strategy for speech motor control: Talkers should spend as much effort as possible to exactly reproduce a desired constriction location and area, possibly at the cost of increased variability away from the constriction. Perkell and his colleagues demonstrated that any given talker produces the vowels /i/ and /a/ with precisely repeatable tongue positions near the constriction, while allowing variability away from the constriction (Perkell & Cohen, 1989; Perkell & Nelson, 1985). Narayanan et al. (1997) observed that different talkers produce /i/ and /a/ with similar vocal tract areas near a constriction, but with quite different vocal tract areas away from the constriction. The height of the palatal dome, for example, varies a great deal from one talker to another, and this variation seems to have different effects depending on the place of articulation of the vowel. A palatal vowel is created by holding the lateral margins of the tongue tightly against the palate, with the result that vocal tract cross-sectional area in the palatal region is relatively talker-invariant during palatal vowels. Conversely, the area under the palatal dome is largest during production of a pharyngeal vowel, and small changes in area have little acoustic consequence, so talkers may choose to let the shape of the vocal tract in this region be determined by the natural morphological characteristics of the vocal tract.

In this article, our goal was to identify phoneme-dependent patterns of intertalker articulatory variability. Specifically, we sought to demonstrate three points: (a) the place of articulation distinction is produced similarly by all 5 talkers, in the sense that intertalker variability is not larger than interplace variability; (b) oral cross-sectional area shows less intertalker variability during production of palatal and velar vowels than it does during production of uvular and pharyngeal vowels; and (c) in order to maintain a relatively talker-independent constriction area during production of palatal vowels, talkers with a high palatal vault compensate by raising the tongue more than other talkers. In the Discussion section we propose a model of intertalker differences that seems to explain the observed pattern of intertalker variability. The proposed model extends the Stevens–House model to consider the effect of talker-independent acoustic constraints on talker-dependent anatomy. Simulation results demonstrate that if talkers with differently sized vocal tracts attempt to match talker-independent formant frequencies, the resulting area functions show minimal intertalker variability near vocal tract constrictions.

## Method and Results

### *Image Acquisition and Acoustic Recording*

Data were collected from 5 native speakers of American English, 3 men (M1–M3) and 2 women (F1 and F2). All talkers were undergraduate or graduate students at the University of California, Los Angeles, at the time of the study. All talkers were native speakers of the southern California dialect of American English, except M2, who was a native of Albany, New York. Hagiwara (1997) found that vowels in the southern California dialect are similar to the midwestern vowels reported by Hillenbrand, Getty, Clark, and Wheeler (1995), except that the rounded vowels /u, ʊ, o/ (as in *boot, put, boat*) are produced with higher formant frequencies, indicating possibly a fronted tongue body and possibly reduced lip rounding.

Participants were imaged on clinical equipment at Cedars–Sinai Hospital in Los Angeles. A Signa 1.5 Tesla scanner (General Electric) generated sagittal, axial, and coronal images with a 24 cm × 24 cm field of view. Anatomic regions of interest extended from above the hard palate to below the larynx and from the lips to the posterior pharynx. For each phoneme, 30–40 coronal image slices and 30–40 axial image slices were created at 3 mm interslice increments, except that M2 was imaged in 2 mm interslice increments. The tongue was separately segmented only in coronal images; therefore, only

coronal images were analyzed for this study. Participants produced the English vowels /i, u, e, o, a, æ, ɪ, ε, ʊ, ʌ, ɜː/. Participants were told to imagine the words *beat, boot, bait, boat, father, bat, bit, bet, put, tug*, and *Bert*, respectively, as a guide to pronunciation. The vowel /ɜː/ was excluded from all analyses of vowel categories, because it does not fit into any of the four place of articulation categories proposed in Wood (1979).

The first 3 participants (F1, M1, and M2) phonated during image collection, but vibration of the vocal folds and sympathetic vibration of the aryepiglottic folds made it very difficult to correctly locate the boundaries of the piriform sinus. One of the purposes of data collection (not reported in this article) was to estimate the effect of the piriform sinus on acoustic speech synthesis (Cha, 2000). In order to make study of the piriform sinus possible, 3 participants (F2, M2, and M3) were imaged while holding vowel position (“miming” the vowels), without actually phonating. Participant M2 was imaged both while phonating and while miming the vowels, and statistical comparison of the two datasets is used in this article to test for differences between phonated and mimed production. Phonated vowels were sustained for up to 12 s, depending on the comfort level of the participant. In order to acquire a complete image stack of a phonated vowel, it was necessary for the participant to repeat the vowel 2–3 times. Mimed vowels were produced for 20–25 s, again depending on the ability of the participant to comfortably hold his or her breath. The entire vocal tract was imaged from the lips to the trachea. The procedure required about 10 min per phoneme, or about 2.5 hr per participant.

In MRI images, the teeth of a participant are usually invisible because they contain little hydrogen. In order to obtain the best possible information about the location of the teeth, dental casts of each participant were made. Dental casts were submerged in water, and coronal and axial magnetic resonance image stacks were created with a 3 mm slice thickness; in these images, the dental cast is a dark region outlined by bright water. All of these images as well as the dental cast itself were later used to guide manual segmentation of participant images.

Acoustic recordings were acquired in an acoustically isolated recording room 1 week after imaging (Participants F1, M1, and M2) or 2–6 months before imaging (Participants F2 and M3). Eight productions of each vowel were recorded: three while supine on a couch, in order to simulate participant posture during MRI, and five while normally seated. During supine productions, participants were asked to sustain the vowel for 10–15 s. Formant frequencies of the supine vowels were computed using the Entropic formant tracker (Talkin, 1987). The formant frequency trajectory of each supine vowel was

measured approximately  $\frac{1}{3}$ ,  $\frac{1}{2}$ , and  $\frac{2}{3}$  of the way through the vowel, and these nine measurements (3 samples  $\times$  3 tokens) were averaged for each vowel. Two talkers were not characterized by supine vowels. M1 recorded only one supine and one seated production of each vowel, so both supine and seated productions were averaged. M3's supine vowels were too breathy for accurate formant measurement, so his seated vowels were analyzed instead.

Images and acoustic recordings collected for this work are available on the Web site of the first author (<http://www.ifp.uiuc.edu/speech/>).

## Image Analysis

Coronal MRI were segmented to isolate the tongue, oral airway, and gingival margins using custom software (Hasegawa-Johnson, Cha, & Haker, 1999). This analysis began immediately after the first MR imaging session.

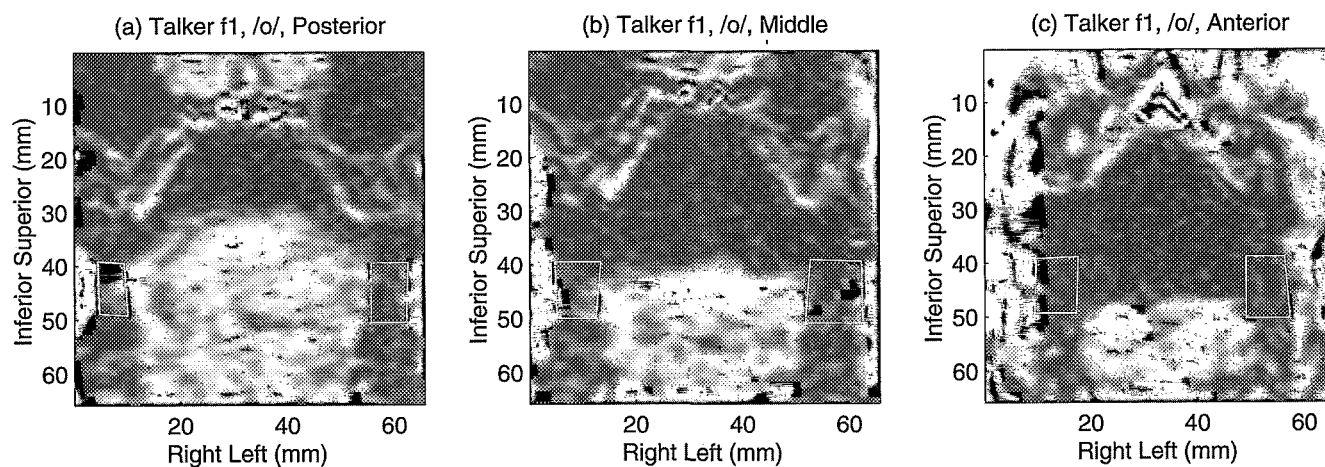
The interdental space was located in all images based on deformation of soft tissue near the dental crown. The tongue and cheeks always bulge slightly into the interdental space. During production of non-low vowels, in the midpalatal and velar regions of the oral cavity, it is possible to locate the interdental regions as a pair of rectangles bounded by the maxillary and mandibular bulges of the cheeks and tongue (see Figure 1a). During production of low vowels, in the midpalatal and velar regions, the tongue surface is located between the maxillary and mandibular occlusal planes, so only three of the four corners of each interdental rectangle are visible in the image (see Figure 1b). In the anterior region (near the canine teeth), the tongue surface is rarely but occasionally retracted below the mandibular occlusal plane, so only two corners of each interdental rectangle are visible in the image (see Figure 1c). Segmentation

of all three types of image plane is facilitated by frequently comparing different images of the same talker. The dental casts were available for reference as needed, and were particularly used to verify measurements of oral anatomy, including palatal vault morphology, dental arch form, tooth shapes, and intermolar distance.

Three image slices were selected for further analysis: an anterior slice (at the location of the canine), a middle slice (at the first molar), and a posterior slice (at the most distal surface of the distal molar). These three locations were chosen as reliable dental landmarks close to the reported constriction locations of the three oral places of vowel articulation: The anterior slice approximates the retroflex place of articulation ( $/s-/$ ), the middle slice approximates the palatal place of articulation ( $/i, e, \text{ɪ}, \text{ɛ}/$ ), and the posterior slice approximates the velar place of articulation ( $/u, \text{ʊ}/$ ) (Wood, 1979). The two remaining vowel places of articulation, uvular ( $/o/$ ) and pharyngeal ( $/ɑ, \text{æ}, \text{ʌ}/$ ), were not accurately represented using coronal image data. Artifacts in the anterior images of Participant M3, apparently caused by a dental restoration in the mandibular canines, made it impossible to locate the tongue surface in anterior images of this participant. Data from the anterior MRI slice of this participant were therefore excluded from statistical analysis. File formatting difficulties also precluded analysis of Participant F1's production of  $/ʌ/$ . Thus, there were a total of 49 tokens available for all posterior and middle-slice measurements (5 talkers  $\times$  10 vowels  $- 1$ ), and 39 tokens for all anterior-slice measurements (4 talkers  $\times$  10 vowels  $- 1$ ).

Image slices were oriented coronally, as originally acquired. Attempts to orient the image slices orthogonal to the vocal tract midline using the algorithm of Story et al. (1996) failed near the oropharyngeal bend because the midline of the pharynx could not be reliably determined on the basis of coronal image sections.

**Figure 1.** Posterior, middle, and anterior slices of Talker F1 saying  $/o/$ ; white rectangles mark the interdental spaces.

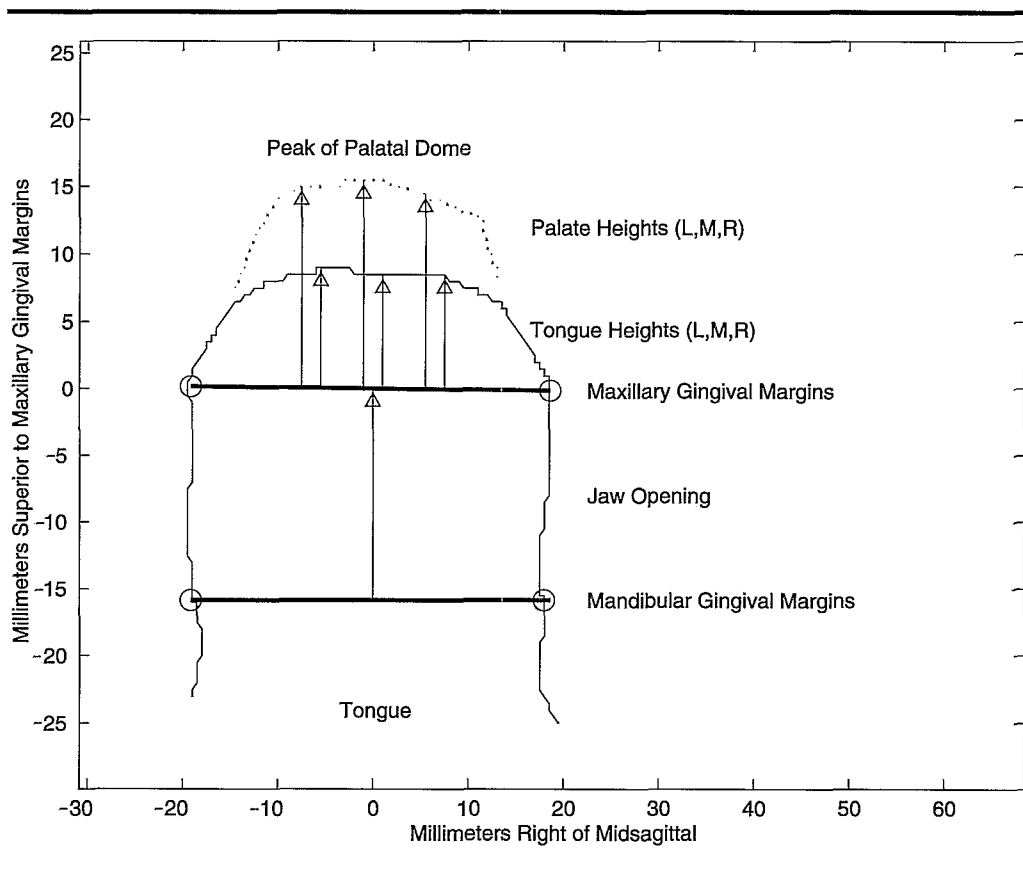


In order to minimize the effect of intertalker differences in head position relative to the coronal imaging plane, a tilt angle  $\theta$  was estimated for each talker, and all tongue height, palate height, and oral area measurements were multiplied by the factor  $\cos \theta$ . The tilt angle was estimated by drawing an alignment line connecting two different anatomical landmarks on a midsagittal image of the participant at rest breathing through the nose. Several different types of anatomical landmarks were considered for image alignment, including the nasion, the anterior maxillary gingival margin, the caudal edge of the upper lip, the anterior–superior line of the straightest part of the hard palate, the superior horn of the second cervical vertebra, the upper edge of the superior pharyngeal constrictor muscle, and the tip of the soft palate (the soft palates of most participants were bent sharply at the connection between the soft palate and the uvula, with the uvula extending forward and down toward the tongue dorsum; in such cases, the uvula was excluded, and the landmark was considered to be the bend in the soft palate). All soft tissue landmarks gave similar alignment results. Hard tissue landmarks gave remarkably variable results, apparently because many of the landmarks could not be reliably located; the cervical vertebrae, for example, are visible

in MRI only as a void between the soft tissues surrounding the spine. Alignment angles were estimated using the line between the caudal tip of the upper lip and the tip of the soft palate excluding uvula, but nearly identical angles were obtained using the maxillary anterior gingival margin and the superior pharyngeal constrictor. Resulting  $\cos \theta$  factors were .96 for Talkers M1 and M2, .99 for Talkers M3 and F1, and .86 for Talker F2.

Tongue and palate contours were replotted using MATLAB (The Mathworks, 1994), and the MATLAB plots were used to measure tongue height (left, right, and midsagittal), palatal vault height (left, right, and midsagittal), jaw opening, and oral area in each slice (see Figure 2). Tongue height and palatal vault height were defined to be the height, in millimeters, above a line connecting the maxillary gingival margins. Both tongue height and palatal vault height were measured (a) midsagittally, (b) in a line displaced to the left of midsagittal by one third of the intermolar distance, and (c) in a line displaced right of midsagittal by the same distance. Jaw opening was defined to be the average distance between the maxillary and mandibular gingival margins. Oral area was measured by automatically counting pixels within the segmented vocal tract outline.

**Figure 2.** Example of a segmented MRI image, demonstrating the anatomical measurements analyzed in this article.



The fit between each measurement and a univariate normal distribution was tested using quantile–quantile plots (plots of sorted measurements as a function of an equal number of quantiles from a unit normal distribution; Johnson & Wichern, 1992). All measurements are adequately modeled by a normal distribution ( $R_{QQ} \geq .89$ ).

Intertalker averages of all formant frequencies, midsagittal tongue height measurements, and oral area measurements, after  $\cos \theta$  correction, are given in Table 1.

## Statistical Analysis

Measurements analyzed in this article include three formant frequencies per vowel and the articulatory measurements schematized in Figure 2 (some of which are listed in Table 1). These measurements are grouped differently, in the sections that follow, for different types of analysis. When comparing the productions of the 5 talkers, there are  $G = 5$  groups, and  $N_g = 10$  different vowels in each group ( $N_g = 9$  for Talker F1,  $N_g = 0$  for Talker M3 in the anterior plane). On the other hand, when comparing different places of articulation, there are  $G = 4$  different groups (pharyngeal, uvular, velar, and palatal), and  $N_g$  ranges from a minimum of  $N_g = 5$  in the uvular group (productions of /o/ by 5 different talkers) to a maximum of  $N_g = 20$  in the palatal group (productions of /i, ɪ, e, ɛ/ by 5 different talkers).

Four types of statistical analysis are used to analyze the means and covariances of measurement vectors. Multivariate analysis of variance (MANOVA) was used to test the null hypothesis that groups are drawn from normal distributions with the same mean vector. If a MANOVA test rejects the null hypothesis, Bonferroni simultaneous confidence intervals may be used to identify significant pairwise differences in the mean. Intergroup differences in the variance of any particular measurement were analyzed using a variance ratio.

Finally, linear regression tests the null hypothesis that two sets of measurements are drawn from uncorrelated normal distributions. Details on all tests are given in standard textbooks (Johnson & Wichern, 1992), and all tests except the variance ratio are quite common in the literature.

A variance ratio tests the null hypothesis that measurements  $x_{pgi}$ ,  $1 \leq p \leq P$ ,  $1 \leq g \leq G$ ,  $1 \leq i \leq N_g$ , are drawn from  $G$  different  $P$ -dimensional normal distributions with the same variance. Let  $F_{m,n}(\beta)$  denote the critical point of the  $F$  distribution at significance level  $\beta$  with degrees of freedom  $m = N_g - 1$  and  $n = \sum_{h \neq g} (N_h - 1)$ . Under the null hypothesis, then with probability  $1 - \alpha$ , the following inequalities hold simultaneously for all  $1 \leq p \leq P$  and  $1 \leq g \leq G$ :

$$\frac{1}{F_{m,n}\left(\frac{\alpha}{2PG}\right)} \leq \frac{n \sum_{i=1}^{N_g} (x_{pgi} - \mu_{pg})^2}{m \sum_{h \neq g} \sum_{i=1}^{N_h} (x_{phi} - \mu_{ph})^2} \leq F_{m,n}\left(\frac{\alpha}{2PG}\right). \quad (1)$$

## Validation of the Vowel Miming Protocol

Talker M2 was imaged under two conditions: once while phonating and once while miming vowel production. Mimed productions by Talker M2 are used only in the tests reported in this section, and only for the purpose of validating the miming protocol.

Phonated and mimed articulations were compared using both average differences (averaged across all vowels) and average magnitude differences. Average tongue height differences were between 0.3 mm and 0.72 mm (equal to the two-sided critical points of a univariate  $t$  statistic at levels  $.43 \leq \alpha \leq .77$ ). Average differences in oral area were between  $0.012 \text{ cm}^2$  and  $0.31 \text{ cm}^2$  ( $.76 \leq \alpha \leq .99$ ). Average magnitude difference between midsagittal tongue contours was less than 3 mm, and the largest single tongue height difference was 9 mm. These

**Table 1.** Average measurements of formant frequencies (kHz), oral area ( $\text{cm}^2$ , posterior, middle, and anterior sections), and midsagittal tongue height (mm, posterior, middle, and anterior sections).

Measurement	/a/	/æ/	/ʌ/	/ɛ/	/e/	/ɪ/	/i/	/o/	/u/	/ʊ/
Formant 1 (kHz)	0.72	0.72	0.60	0.62	0.43	0.48	0.35	0.44	0.47	0.37
Formant 2 (kHz)	1.28	1.66	1.29	1.88	2.24	2.00	2.43	0.93	1.08	0.99
Formant 3 (kHz)	2.67	2.66	2.64	2.79	2.78	2.77	2.93	2.61	2.64	2.57
Posterior area ( $\text{cm}^2$ )	4.8	3.5	2.2	2.1	1.4	1.4	0.5	2.4	2.2	1.4
Middle area ( $\text{cm}^2$ )	5.8	3.5	2.9	2.0	1.1	1.2	0.4	4.7	2.7	2.6
Anterior area ( $\text{cm}^2$ )	5.8	3.6	1.9	1.9	2.2	1.1	0.9	4.8	2.8	3.0
Posterior height (mm)	-0.3	0.0	2.7	4.6	6.7	7.1	11.4	5.9	4.5	7.0
Middle height (mm)	-4.8	-0.3	1.2	3.7	8.5	8.5	11.5	-2.1	2.6	3.6
Anterior height (mm)	-13.8	-9.6	-5.7	-5.2	-4.3	-1.9	-2.3	-11.4	-8.0	-7.3

Note. Sign of tongue height measures is taken from measurement coordinate space shown in Figure 2.

measurements are consistent with the hypothesis that there is no difference between phonated and mimed vowel production.

The results of this section demonstrate that at least 1 of the 5 participants was able to perform the vowel-miming task with minimal articulatory error (less than 3 mm average magnitude tongue height difference), but the ability of M2 to perform the vowel miming task with minimal error does not necessarily imply that F2 and M3 were also able to perform this task. For this reason, all statistical analyses reported in the following sections were performed twice: once using data from all 5 talkers and once using data only from the 3 talkers who phonated during data acquisition (M1, M2, and F1).

### Differences Among Talkers

MANOVA and Bonferroni simultaneous confidence intervals were used to test the null hypothesis that intertalker differences are not larger than intervowel differences for any measurement dimension. A measurement vector of length  $P = 21$  was constructed to include three formant frequencies ( $F_1$ ,  $F_2$ , and  $F_3$ ), and anterior, middle, and posterior measurements of oral area; jaw opening; midsagittal palate height; and left, midsagittal, and right tongue height. Talker M3 was

included in Bonferroni analysis but not MANOVA, because image artifacts in the anterior plane precluded a complete 21-dimensional analysis of his vowels. MANOVA analysis with  $G = 4$  groups (1 talker per group) yielded  $\chi^2(63) = 150$ , significant at  $\alpha = 1.9 \times 10^{-9}$ .

Bonferroni simultaneous confidence intervals were calculated for all 10 possible pairings of the 5 talkers F1, F2, M1, M2, and M3. Bonferroni analysis identified eight measurements that showed a significant difference between at least 1 pair of talkers (simultaneous significance level  $\alpha = .05$ ): palate height (anterior, middle, and posterior), posterior tongue height (left, midsagittal, and right), anterior jaw opening, and  $F_3$ . Seven of the 10 possible talker pairs differed in at least one articulatory measurement. These 7 talker pairs are listed in the columns of Table 2; the rows of the table list the seven articulatory measurements with significant differences. In the table, values are given for all significant intertalker differences. Nonsignificant differences are marked *ns*. Seven of the significant intertalker differences were differences in palate height.

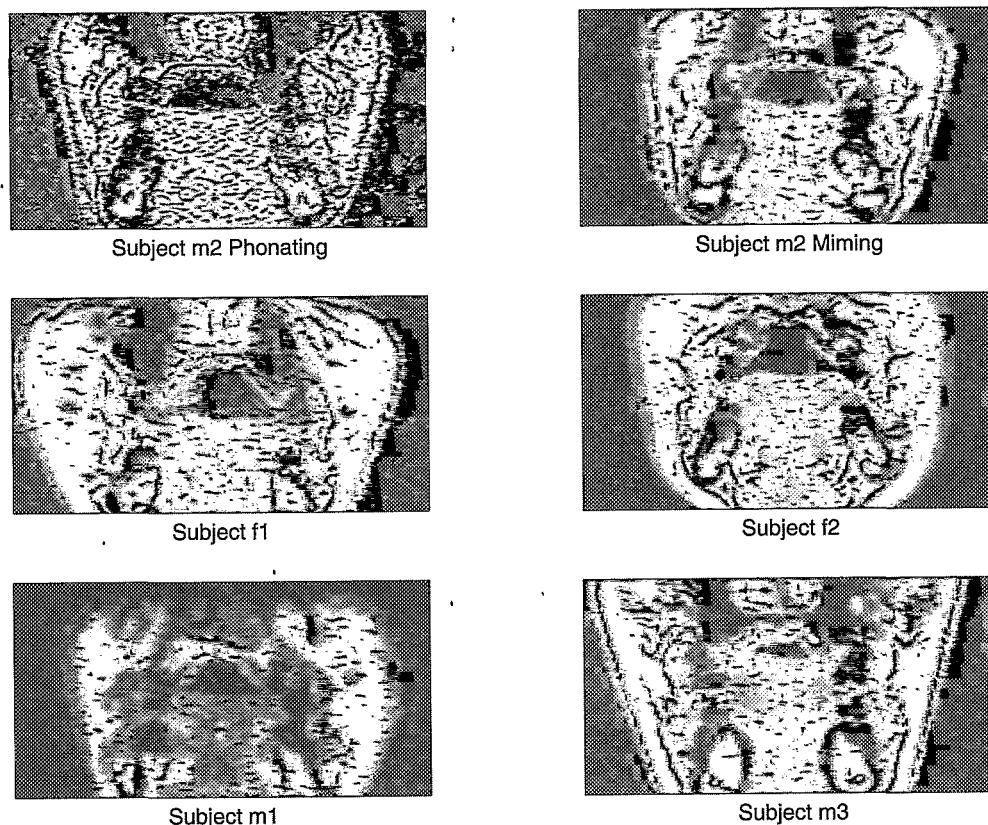
Additional qualitative details of palatal vault morphology can be observed in Figure 3, which shows all talkers producing the vowel /e/. The palatal vaults of Talkers M2 and F1 are noticeably wider than the palates

**Table 2.** Significant pairwise differences, as demonstrated using Bonferroni post hoc testing.

Measurement	Articulatory differences between talkers						
	F2-F1	M2-F1	M2-F2	M2-M1	M3-F1	M3-F2	M3-M1
Posterior midsagittal palate height (mm)	<i>ns</i>	<i>ns</i>	<i>ns</i>	-4.7	<i>ns</i>	<i>ns</i>	<i>ns</i>
Middle midsagittal palate height (mm)	-3.3	-2.8	<i>ns</i>	<i>ns</i>	-3.6	<i>ns</i>	-2.4
Anterior midsagittal palate height (mm)	<i>ns</i>	<i>ns</i>	4.0	4.5	<i>ns</i>	<i>ns</i>	<i>ns</i>
Posterior midsagittal tongue height (mm)	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	12.1	<i>ns</i>
Posterior left tongue height (mm)	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	9.5	<i>ns</i>
Posterior right tongue height (mm)	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	10.4	<i>ns</i>
Anterior jaw opening (mm)	<i>ns</i>	-6.6	-6.6	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>
Measurement	Differences between places of articulation						
	U-Ph	V-Ph	Pa-Ph	Pa-U	Pa-V		
Middle midsagittal tongue height (mm)	<i>ns</i>	<i>ns</i>	9.6	10.2	<i>ns</i>		
Anterior midsagittal tongue height (mm)	<i>ns</i>	<i>ns</i>	6.7	<i>ns</i>	<i>ns</i>		
Middle left tongue height (mm)	<i>ns</i>	<i>ns</i>	8.7	<i>ns</i>	<i>ns</i>		
Anterior left tongue height (mm)	<i>ns</i>	<i>ns</i>	6.4	<i>ns</i>	<i>ns</i>		
Middle right tongue height (mm)	<i>ns</i>	<i>ns</i>	9.1	9.9	<i>ns</i>		
Anterior right tongue height (mm)	<i>ns</i>	<i>ns</i>	6.8	<i>ns</i>	<i>ns</i>		
Posterior area (cm <sup>2</sup> )	<i>ns</i>	<i>ns</i>	-2.3	<i>ns</i>	<i>ns</i>		
Middle area (cm <sup>2</sup> )	<i>ns</i>	<i>ns</i>	-2.9	<i>ns</i>	<i>ns</i>		
$F_1$ (kHz)	-0.24	-0.27	-0.22	<i>ns</i>	<i>ns</i>		
$F_2$ (kHz)	<i>ns</i>	<i>ns</i>	0.72	1.20	1.10		

Note. Measurements include palate height, tongue height, jaw opening, oral area, and formant frequencies. Places of articulation include pharyngeal (Ph), uvular (U), velar (V), and palatal (Pa). Sign of tongue height measures is taken from measurement coordinate space shown in Figure 2.

**Figure 3.** Middle-slice MRI of all participants, including both phonated and mimed productions by Participant M2 (prompt word *bet*; field of view 135 × 90 mm).



of F2 and M1. The palatal vault of talker M1 appears to be narrower than that of any other talker.

$F_3$  differences are not shown in Table 2, because there were too many of them:  $F_3$  of every male talker was less than  $F_3$  of every female talker, with average differences ranging from 318 Hz (F2–M3) to 547 Hz (F1–M1). Intertalker differences in the other formants ( $F_1$  and  $F_2$ ) were not larger than intervowel differences. This finding supports the previously published observation that, independent of vowel identity, most nonrounded, nonretroflex vowels are produced with an  $F_3$  very close to the frequency  $5c/4L$ , where  $c$  is the speed of sound and  $L$  is the length of the talker's vocal tract (Stevens, 1999). Average  $F_3$  for all vowels produced by Talkers M1, M2, and M3 was 2550 Hz, implying a typical male vocal tract length of  $L = 16.7$  cm. Average  $F_3$  of Talkers F1 and F2 was 2960 Hz, implying a vocal tract length of  $L = 14.3$  cm.

### Statistical Analysis of Vowel Differences

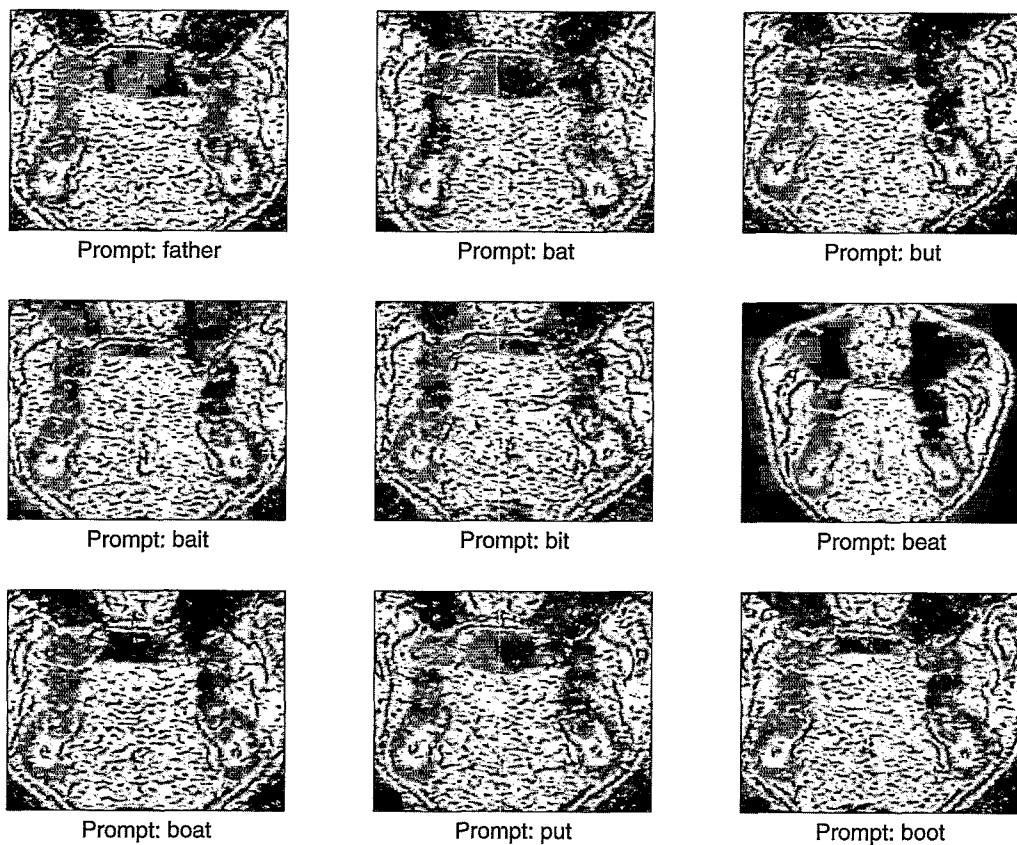
The null hypothesis that articulatory and acoustic measurements do not depend on vowel category was

tested qualitatively by inspection of images and quantitatively using MANOVA. Figure 4 shows middle image slices of Participant M2 producing the vowels (/a, æ, ʌ, e, ɪ, i, o, u, ʊ/). Qualitative results described in the next paragraphs were compiled by inspecting similar images from all 5 talkers and by inspection of data in Table 1.

As shown in Figure 4, the pharyngeal vowels /a, æ, ʌ/ are qualitatively characterized by a tongue low in the oral cavity, typically at about the level of the maxillary gingival margins. The tongue shape is flat or concave and extends laterally into the interdental space. In the posterior slice, the lowest vowel was /a/ or /æ/ for every talker, with /a/ lowest on average. In the middle and anterior slices, the lowest vowel for any given talker was always one of the set /a, æ, o, ʊ, e/, with /a/ lowest on average and /o/ second lowest on average.

The palatal vowels /i, ɪ, e, ε/ were characterized by an average tongue position higher than any other vowel in all three MRI slices. The highest vowel for any given talker in any given MRI slice was always one of the set /e, ɪ, i, u/, with the vowel /i/ highest and /e/ second highest on average. The vowel /ε/ was much lower than the other palatal vowels, with a tongue position that

Figure 4. MRI of Participant M2 producing nine vowels (90 mm field of view).



qualitatively resembled that of /æ/. Maximum average tongue height was in the middle slice for /i, ɪ, e/ and in the posterior slice for /ɛ/. Posterior tongue grooving was observed in most participants' productions of /ɛ/, /e/, /ɪ/, and very slightly in /i/. The tongue extended laterally into the interdental space for all participants in /ɛ/, and somewhat less so for /e/ and /ɪ/.

Most participants produced /u/ with a tongue body position similar to that of /e/ in the middle and posterior slices, as shown in Figure 4, but with a lower tongue body in the anterior slice. The vowel /ʊ/ was produced at heights ranging from the maxillary gingival margins (as in Figure 4) to nearly the height of /u/. In production of the uvular vowel /o/, the tongue was usually higher than /u/ and /ʊ/ in the posterior slice, but lower than any other vowel but /a/ in the middle and anterior slices (M2 was an exception: his /o/ was relatively high in the middle slice; see Figure 4). Qualitative inspection of the complete three-dimensional images suggests that the maximum constriction for /u, ʊ/ was achieved near the posterior slice, whereas that of /o/ is achieved approximately 2 cm posterior to the posterior slice.

A measurement vector of length  $P = 21$  was constructed to include three formant frequencies ( $F_1$ ,  $F_2$ ,

and  $F_3$ ), and anterior, middle, and posterior measurements of oral area, jaw opening, midsagittal palate height, and left, midsagittal, and right tongue height. Measurement vectors were categorized into  $G = 4$  different places of articulation. A MANOVA test yielded  $\chi^2(63) = 120$ , significant at  $\alpha = .00002$ .

Bonferroni simultaneous confidence intervals identified 10 measurements that showed a significant difference between at least one pair of place categories, as shown in Table 2.

Bonferroni analysis was repeated using only data from the 3 talkers who phonated during imaging (F1, M1, M2). Analysis of the 3-talker sample yielded exactly the same list of significant pairwise differences as analysis of the 5-talker sample, with three deletions: In the 3-talker sample, there was no significant F1 difference between uvular and pharyngeal vowels, and there was no significant difference among oral areas. With the available data, there was no way to determine whether the change in results was an artifact of the change in sample size, an indication of talker differences, or an indication of a difference between phonated and mimed vowel production.

## Intertalker Stability of Area Near a Constriction

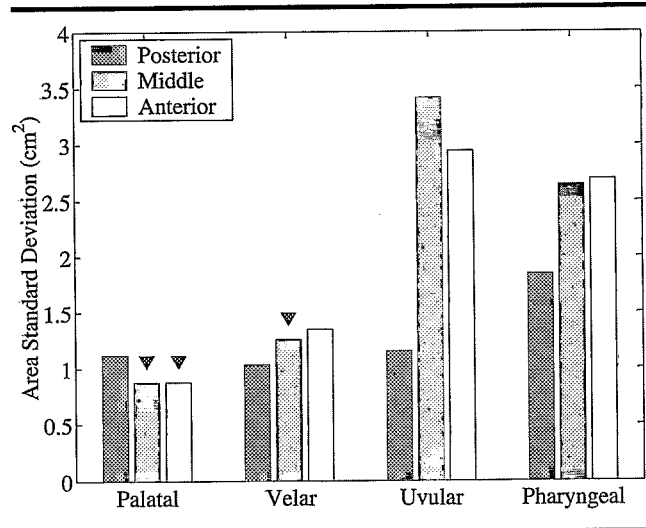
The relative variability of oral area during production of different vowels was tested using variance ratios. Figure 5 shows the within-class standard deviation of oral area as a function of place of articulation and as a function of MRI slice location. Each within-class sample variance was compared to the average of all three other within-class variances using Equation 1. Standard deviations significantly different from the others are marked with up arrows or down arrows, depending on the direction of the difference (2-sided  $F$  test at significance level  $\alpha/2PG = .05/24$ ). The three significant differences marked in Figure 5 were also found to be significant when analysis was performed using only the 3 talkers who phonated during image acquisition.

Chiba and Kajiyama (1941) demonstrated that formant frequency changes are proportional to the percentage change in cross-sectional area. Specifically, they demonstrated that for small perturbations in the area function  $A(x)$ , the corresponding formant frequency change is related to the percentage area change by a modal sensitivity function  $G_n(x)$ . They also demonstrated that percentage area function change is equal to the change of the log area function, thus:

$$\delta F_n = G_n(x) \left( \frac{\delta A(x)}{A(x)} \right) = G_n(x) \delta \log A(x) \quad (2)$$

According to Equation 2, intertalker formant frequency variance should be roughly proportional to intertalker variance of the log area function. If intertalker variance of formant frequencies is place independent, it should be the case that intertalker variance of log area function is also place independent.

**Figure 5.** Within-class standard deviation of oral area, as a function of place of articulation and of MRI slice.



The variance ratio test was used to evaluate intertalker variance of both formant frequencies and log area as a function of vowel place of articulation. Intertalker variance of both formant frequencies and log area were found to be independent of vowel place.

## Correlation Between Palate Height and Tongue Height

Figure 6 shows scatter plots of tongue height and oral area versus palate height for the four places of articulation. In order to improve legibility of the plot, data from the middle and posterior MRI slices were pooled. Data from the anterior slice were plotted separately, because anterior palate heights are quite different from middle and posterior palate heights. Correlation coefficients were computed separately for all three MRI slice locations. Of the 24 possible tongue height correlations (oral area and tongue height in three MRI slices for four places of articulation), 2 were significant (significance level  $\alpha/6G = .05/24$ ). First, tongue height of palatal vowels was positively correlated with palate height in the posterior MRI slice. This result was repeated when posterior and middle-slice data were pooled as in Figure 6. Second, oral area of pharyngeal vowels is positively correlated with palate height in the anterior slice, as shown in Figure 6.

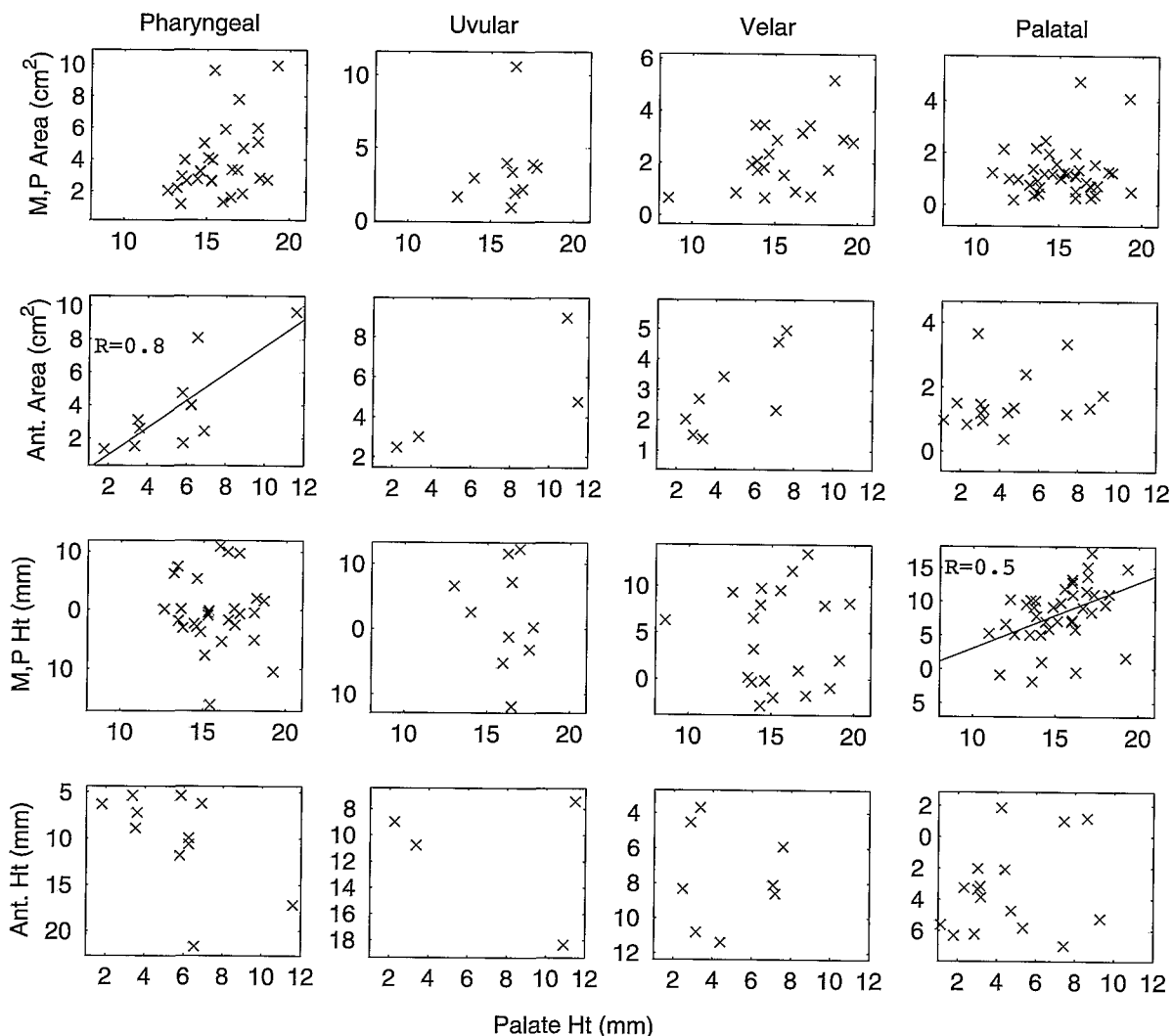
When correlation analysis was performed using data from only the 3 talkers who phonated during imaging, there was only one statistically significant correlation: Posterior tongue height was correlated with palate height during production of palatal vowels.

## Discussion

The data presented in this article suggest that talkers configure the tongue so as to minimize intertalker area differences near a constriction. This result supports previous results (Perkell & Cohen, 1989; Perkell & Nelson, 1985) showing that interrepetition variance in the utterances of any 1 talker is minimized near a constriction.

The primary constriction of a palatal vowel is near the middle MRI slice, and the primary constriction of a velar vowel is near the posterior MRI slice; other vowels are characterized by constrictions farther back in the vocal tract. Figure 5 shows that intertalker variance of middle-slice area of palatal and velar vowels is smaller than intertalker variance of uvular and pharyngeal vowels. Intertalker variance of the log area function, however, is not significantly place dependent, suggesting an elegant but unproven model for the positive result in Figure 5. Suppose that  $A(x, p, t)$  is the cross-sectional area at position  $x$  of talker  $t$  during production

**Figure 6.** Scatter plots of oral area (middle and posterior data pooled, anterior data separate) and tongue height (middle and posterior data pooled, anterior data separate) as functions of palate height.



of a vowel with place of articulation  $p$ . A first-order Taylor series approximation of  $\log(A(x, p, t)) - E_t[\log(A(x, p, t))]$  yields

$$\text{Var}_t(\log A(x, p, t)) \approx \frac{\text{Var}_t(A(x, p, t))}{E_t[A(x, p, t)]^2} \quad (3)$$

If it could be proven that  $\text{Var}_t(\log A(x, p, t))$  is independent of vowel place of articulation, then, as shown in Equation 3, the consequence would be that  $\text{Var}_t(A(x, p, t)) \propto E_t[A(x, p, t)]^2$ . Our empirical results demonstrate that  $\text{Var}_t(\log A(x, p, t))$  was not significantly dependent on  $p$ , but this was a negative statistical result and should not be construed as proof of the null hypothesis. The data analyzed in this article were not sufficient to prove that  $\text{Var}_t(A(x, p, t)) \propto E_t[A(x, p, t)]^2$ . The only statement that could be proven with certainty is that  $\text{Var}_t(A(x, p, t))$  depended on  $p$ , as shown in Figure 5.

Figure 6 demonstrates a physical mechanism that is probably related to the results in Figure 5. Positive correlation between palate height and tongue height suggests that talkers with a high palatal vault compensate by raising the tongue higher during palatal vowels, thus reducing intertalker variability of the oral area.

The results presented in this article are limited by the selection of coronal imaging planes for analysis. First, coronal imaging planes are not precisely orthogonal to the midline of the vocal tract. Although oral area and tongue height were adjusted using a talker-dependent  $\cos \theta$  factor, it is possible that there were intertalker alignment differences that were not captured by the  $\cos \theta$  factor. Such differences would appear as unexplained variability in this study. It is possible, therefore, that some of the intervowel differences reported as not significant in this study might have been

reported as significant if it were possible to construct reliable midline-orthogonal image planes. Second, and perhaps more important, it is not possible to reliably measure cross-sectional area and tongue height near the uvular and pharyngeal constriction locations on the basis of coronal images. Results in this article show that intertalker variance of oral area is reduced when a vowel with an oral constriction is produced; on the basis of these data, it is not possible to make any similar statement about pharyngeal area.

The following paragraphs propose a modification of the Stevens–House (1955) model in order to consider the effect of talker-independent acoustic targets on talker-dependent vocal tract anatomies. The model mimics the empirical results of this article by showing reduced intertalker variability of the midpalatal area during production of palatal and velar vowels. It remains for future studies to consider whether the predictions of the model are also empirically supported near the uvular and pharyngeal vocal tract constrictions.

In the Stevens–House (1955) model, the vocal tract is represented by a series of concatenated 0.5 cm cylindrical tubes. There are three vowel-dependent parameters:  $r_l$ , the radius of the lip tube;  $r_0$ , the radius of the vocal tract constriction; and  $d_0$ , the distance from the glottis to the constriction. In the experiments reported here, the size of the vocal tract is controlled by four talker-dependent parameters: ( $L, R_p, R_o, R_t$ ), representing the length of the vocal tract, maximum radius of the pharynx, maximum radius of the oral cavity, and resting radius of the tongue. Stevens and House used values of  $L = 17$  cm,  $R_p = 1.6$  cm, and  $R_t = 1.2$  cm, and did not constrain the maximum radius of the oral cavity. The radius of the vocal tract at all points is then given by  $r(x)$ , where  $x$  is the distance from the glottis, and all measurements are given in centimeters:

$$r(x) = \begin{cases} r_l & x = L \\ \min(R_o, r_0 + 0.025(R_t - r_0)(x - d_0)^2) & d_0 \leq x < L \\ \min(R_p, r_0 + 0.025(R_t - r_0)(x - d_0)^2, \\ \quad 0.7 + 0.144x^2) & x < d_0 \end{cases} \quad (4)$$

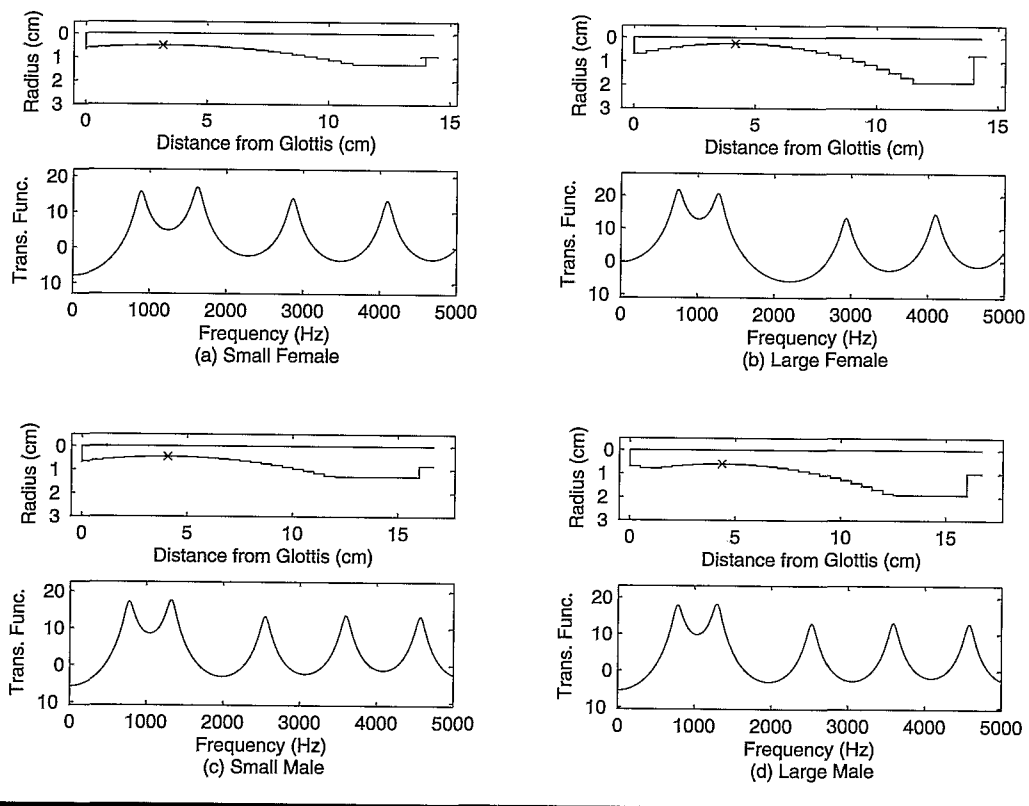
In order to explore the effect of vocal tract size differences, four vocal tracts were simulated using the Stevens–House (1955) model, representing a large male, a small male, a large female, and a small female. Vocal tract length was  $L = 14.3$  cm for female models and  $L = 16.7$  cm for male models, corresponding to  $5c/4F_3$  for average male and female  $F_3$  measured in this article. Small male and small female models used a vocal tract width given by  $R_p = R_o = 1.3$  cm, and  $R_t = 1.0$  cm, whereas for the large models, the parameters were  $R_p = R_o = 1.9$  cm and  $R_t = 1.4$  cm. Parameter choice is motivated by three considerations. First, among the 5

talkers analyzed for this article, maximum oral area was just under 11 cm<sup>2</sup>, corresponding to a maximum oral radius (assuming circular cross-section) of just under 1.9 cm. Second, the largest intertalker palate height difference observed in this article was 4.7 mm, so differences between large-talker and small-talker radii were set to be in the range of 4–6 mm. Third, large-talker parameters were set to be 18% larger, and small-talker parameters 18% smaller, than the values used by Stevens and House. Formant frequencies corresponding to each vocal tract configuration were computed by converting the radius function to an area function assuming circular cross section, converting the area function to reflection coefficients, and terminating the system with a zero-flow termination at the glottis and a zero-pressure termination at the lips. The resulting system is lossless, so transfer function plots in Figure 7 show the transfer function with formant bandwidths expanded by 113 Hz (bandwidths of digital resonances expanded by 0.01 radians/sample).

For each of the four models, the parameters  $r_0$ ,  $d_0$ , and  $r_l$  were adjusted using a gradient search procedure. The gradient search procedure was initialized using a manual approximation of the vocal tract shape for each vowel; for example, the gradient optimization of vowel /i/ began from the configuration  $r_l = 1.2$  cm,  $r_0 = 0.25$  cm,  $d_0 = 0.75L$ . Parameters were then adjusted separately for each vocal tract model in order to minimize mean-squared estimation error of the first two formants, where formant targets were the averages from the 5 talkers in this article: (354, 2429) for /i/, (725, 1284) for /a/, (367, 992) for /u/, and (442, 932) for /o/. Figure 7 shows the optimized vocal tract radius functions and transfer functions of the vowel /a/ for all four simulated talkers. The area function of /a/ shows considerably greater intertalker variability far from the constriction than near the constriction. This finding is quantified in Table 3, which shows the simulated intertalker oral area standard deviation at four places in the vocal tract:  $x = 0.25L$ ,  $0.43L$ ,  $0.53L$ ,  $0.68L$ , corresponding to the optimal place of articulation parameter  $d_0$ , averaged across the four vocal tract models, of /a, o, u, i/, respectively. The simulated intertalker area variance of /a/ is smallest of the four vowels at  $x = 0.25L$ , its place of maximum constriction. Likewise the smallest variance at  $x = 0.43L$  is that of /o/, and the smallest variance at  $x = 0.68L$  is that of /i/. Although the variance of /u/ at  $x = 0.53L$  is not smaller than that of all other vowels, it is smaller than the variance of /u/ at any other place in the vocal tract except the lips.

The derivative of formant frequencies with respect to changes in the oral area function, and with respect to changes in the log area function, was estimated by perturbing the four optimized vowel configurations of the

**Figure 7.** Vocal tract radius functions and transfer functions resulting from approximation of the formants  $(F_1, F_2) = (725, 1284)$  Hz using four differently sized vocal tract models.



small male vocal tract model. Beginning with the area function  $A_v(x)$  optimized for vowel  $v \in /a, i, o, u/$ , the area or log area of each tube section in turn was perturbed by  $\pm 0.05$ . Formant frequencies of each perturbed model were calculated and were used to estimate the derivative functions  $dF_i/dA_v(y)$  and  $dF_i/d \log A_v(y)$ :

$$\frac{dF_i}{dA_v(y)} \approx 10 (F_i(A_v(x) + 0.05\delta(x-y)) - F_i(A_v(x) - 0.05\delta(x-y))) \quad (5)$$

$$\frac{dF_i}{d \log A_v(y)} \approx 10 (F_i(A_v(x) e^{0.05\delta(x-y)}) - F_i(A_v(x) e^{-0.05\delta(x-y)})) \quad (6)$$

**Table 3.** Simulated intertalker oral area standard deviation ( $\text{cm}^2$ ) at four places in the vocal tract ( $x$  is distance from glottis,  $L$  is vocal tract length) computed using the Stevens and House (1955) model.

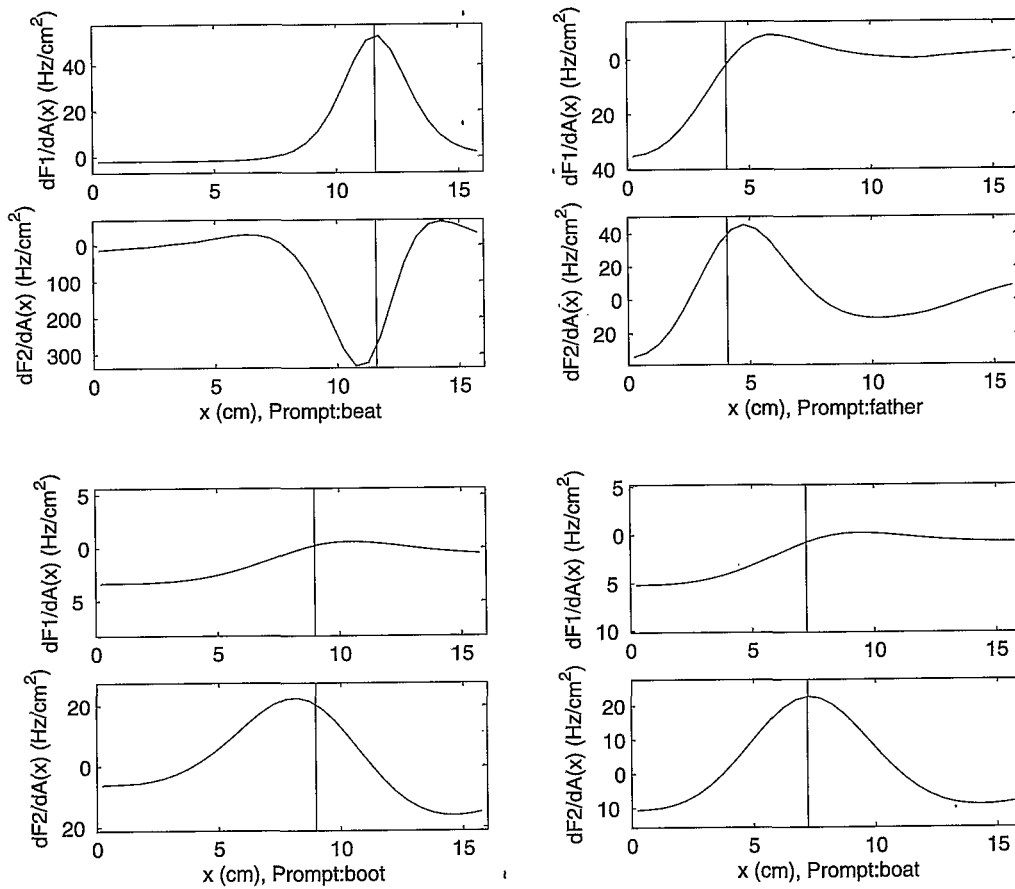
Vowel	$x = 0.25L$	$x = 0.43L$	$x = 0.53L$	$x = 0.68L$
/a/	0.39	0.57	0.77	1.83
/o/	0.53	0.25	0.26	0.58
/u/	0.99	0.37	0.29	0.36
/i/	2.57	0.49	0.21	0.13

where  $F_i(A(x))$  is the  $i$ th formant frequency resulting from area function  $A(x)$ , and  $\delta(x-y)$  is 1 when  $x = y$  and 0 elsewhere.

Derivatives of formants  $F_1$  and  $F_2$  with respect to area are shown in Figure 8, and derivatives with respect to log area are shown in Figure 9. The derivative with respect to lip area changes is not plotted. The derivatives with respect to log area ( $\partial F_i / \partial \log A(x)$ ) are similar to those plotted elsewhere (Chiba & Kajiyama, 1941):  $\partial F_2 / \partial \log A(x)$ , in particular, is approximately equal to half a sine wave when the lips are closed (/u, o/) and approximately equal to three quarters of a sine wave when the lips are open (/i, a/). The plots of  $\partial F_i / \partial A(x)$ , by contrast, are emphasized in the vicinity of the vocal tract constriction, so that, among all possible changes to the oral area behind the lips,  $F_2$  is maximally sensitive to area changes near the vocal tract constriction.  $F_1$  is also maximally sensitive to area changes near the constriction only for the vowel /i/. For /a, o, u/, the derivative of  $F_1$  is greatest for area changes near the glottis.

These simulations suggest an explanation for the empirical findings of this article. The empirical results demonstrate that the area of the vocal tract near the palatal and velar places of articulation shows minimal intertalker variance during production of palatal and

**Figure 8.** Derivative of formant frequencies with respect to area changes, plotted as a function of the location of the change.



velar vowels. Simulation results demonstrate that a similar result is obtained if differently sized vocal tract models are adjusted in order to optimally match talker-independent formant frequency templates. Empirical results also show that tongue height of palatal vowels is correlated with palate height, but tongue height of other vowels is not correlated with palate height. Simulation results suggest an explanation: both  $F_1$  and  $F_2$  of the palatal vowel /i/ are maximally sensitive to area changes in the palatal region, while formant derivatives of /o, u, a/ in this region of the vocal tract are somewhat less sensitive to changes.

## Conclusions

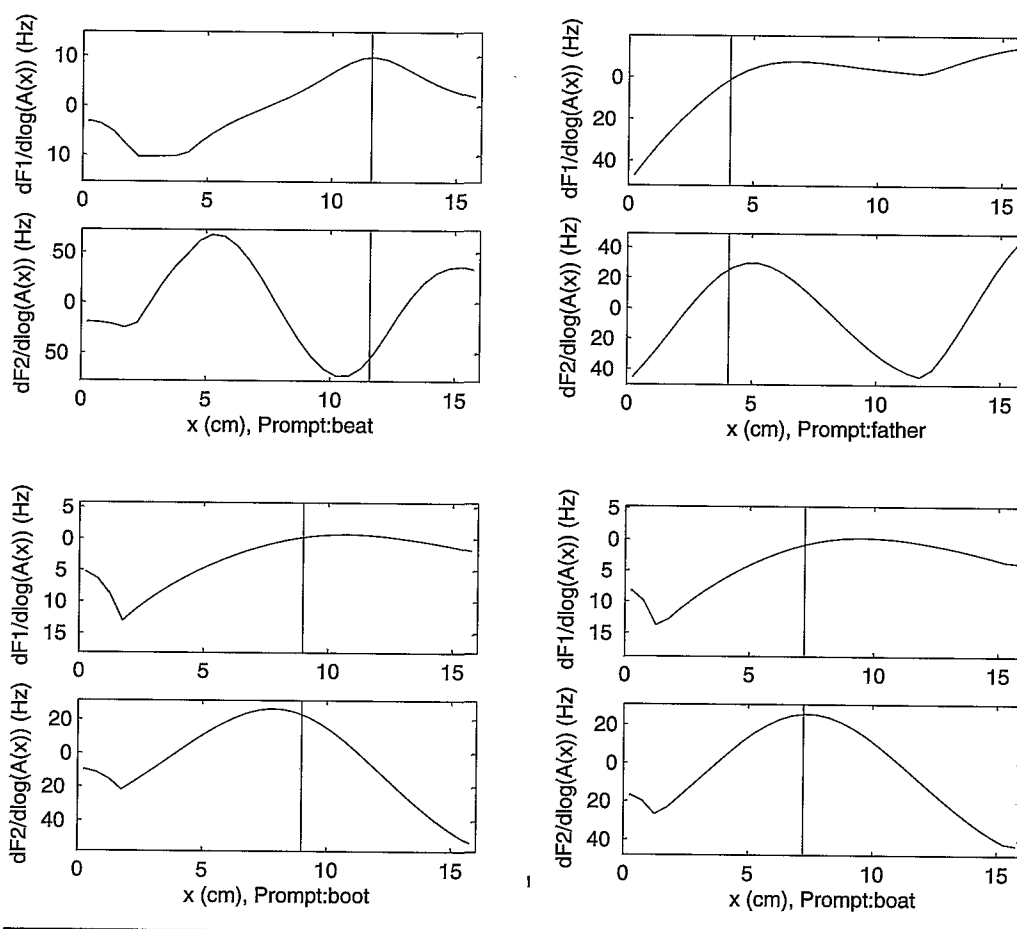
The mean, variance, and palate-height dependence of tongue height and oral area depend on vowel place of articulation. Palatal and velar vowels, articulated near the second molar and posterior molar, respectively, are produced with less intertalker oral area variance in the vicinity of the molars than uvular and pharyngeal vowels. Intertalker variance of the log area function is

not dependent on place of articulation, confirming the finding that vowels with a small oral area variance are also characterized by small intertalker area variance. Reduced area variance of palatal vowels is partly explained by a significant correlation between posterior tongue and palate height during production of palatal vowels. Tongue height is not correlated with palate height during production of velar, uvular, and pharyngeal vowels, suggesting that talkers do not actively compensate for palatal height differences during production of nonpalatal vowels. These data suggest that talkers configure the tongue so that intertalker oral area variance of any vowel is roughly proportional to the intertalker average area of the same vowel: Intertalker variance of constriction area is smaller than intertalker variance of an open oral cavity.

## Acknowledgments

This research was supported by National Institutes of Health Individual National Research Service Award F23 DC 00323-01 granted to the first author, by a grant to the second

**Figure 9.** Derivative of formant frequencies with respect to log area changes, plotted as a function of the location of the change.



author from the American Association of Orthodontics, and by a National Science Foundation CAREER award granted to the third author.

## References

- Baer, T., Gore, J., Gracco, L., & Nye, V. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *Journal of the Acoustical Society of America*, 90, 799–828.
- Cha, J. S. (2000). *Articulatory speech synthesis of female and male talkers*. Unpublished master's thesis, University of California, Los Angeles.
- Chiba, T., & Kajiyama, M. (1941). *The vowel—its nature and structure*. Tokyo: Tokyo-Kaiseikan Publishing.
- Dunn, H. (1950). The calculation of vowel resonances, and an electrical vocal tract. *Journal of the Acoustical Society of America*, 22, 740–753.
- Engwall, O. (2000). Are static MRI measurements representative of dynamic speech? Results from a comparative study using MRI, EPG, and EMA. In B. Yuan, T. Huang, & X. Tang (Eds.), *Proceedings of the International Conference on Spoken Language Processing* (Vol. 1, pp. 17–20). Beijing, China: China Military Friendship Publishing.
- Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*, 102, 655–658.
- Harshman, R., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, 62, 693–707.
- Hasegawa-Johnson, M., Cha, J. S., & Haker, K. (1999). CTMRedit: A MATLAB-based tool for segmenting and interpolating MRI and CT images in three orthogonal planes. *Proceedings of the IEEE Engineering in Medicine and Biology Society Annual Conference (EMBS)* (p. 1170). Atlanta, GA: IEEE Press.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustical characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099–3111.
- Iskarous, K. (2000). *Dynamic acoustic-articulatory relations*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

- Narayanan, S. S., Alwan, A. A., & Song, Y.** (1997). New results in vowel production: MRI, EPG, and acoustic data. In *The 5th European Conference on Speech Communication and Technology (EUROSPEECH)* (pp. 1007–1009). Patras, Greece.
- Perkell, J. S.** (1996). Properties of the tongue help to define vowel categories: Hypotheses based on physiologically-oriented modeling. *Journal of Phonetics*, *24*, 3–22.
- Perkell, J. S., & Cohen, M. H.** (1989). An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/, and /u/. *Journal of Phonetics*, *17*, 123–133.
- Perkell, J. S., & Nelson, W. L.** (1985). Variability in production of the vowels /i/ and /a/. *Journal of the Acoustical Society of America*, *77*, 1889–1895.
- Stevens, K. N.** (1999). *Acoustic phonetics*. Cambridge, MA: MIT Press.
- Stevens, K. N., & House, A. S.** (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, *27*, 401–493.
- Story, B. H., Titze, I. R., & Hoffman, E. A.** (1996). Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America*, *100*, 537–554.
- Talkin, D.** (1987). Speech formant trajectory estimation using dynamic programming with modulated transition costs. *Journal of the Acoustical Society of America*, *82*(S1), S55.
- The Mathworks, Inc.** (1994). *MATLAB: High-Performance numeric computation and visualization software* (Version 4.0) [Computer software]. Natick, MA: Author.
- Wood, S.** (1979). A radiographic analysis of constriction locations for vowels. *Journal of Phonetics*, *7*, 25–43.

---

Received April 29, 2002

Accepted December 9, 2002

DOI: 10.1044/1092-4388(2003/059)

Contact author: Mark Hasegawa-Johnson, PhD, Beckman Institute, University of Illinois, Urbana, IL 61801.  
E-mail: [jhasegaw@uiuc.edu](mailto:jhasegaw@uiuc.edu)

A vertical bar on the left side of the page, consisting of a series of horizontal segments in shades of yellow and orange, with a small red diamond at the top.

COPYRIGHT INFORMATION

TITLE: Vowel Category Dependence of the Relationship  
Between Palate Height, Tongue Height, and Oral Area  
SOURCE: J Speech Lang Hear Res 46 no3 Je 2003  
WN: 0315205579018

(C) The American-Speech-Language-Hearing Association is the publisher of this article and holder of the copyright. Further reproduction of this article in violation of copyright is prohibited without the consent of the publisher. To contact the publisher: <http://www.asha.org/>

Copyright 1982-2003 The H.W. Wilson Company. All rights reserved.