Running head: VOWEL CATEGORY DEPENDENCE...

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

Mark Hasegawa-Johnson
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Shamala Pizza
Department of Dentistry
University of California at Los Angeles

Abeer Alwan and Jul Setsu Cha
Department of Electrical Engineering
University of California at Los Angeles

Katherine Haker
Department of Radiology
Cedars-Sinai Medical Center

Abstract

This paper evaluates inter-talker variance of oral area, oral log 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 five talkers, each producing eleven different vowels. Tongue height (left, right, and midsagittal), palate height, and oral area are measured in three coronal sections anterior to the oropharyngeal bend, and are subjected to multivariate analysis of variance (MANOVA), variance ratio analysis, and regression analysis. The primary finding of this paper is that oral area (between palate and tongue) shows less inter-talker 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 is positively correlated with palate height during production of palatal vowels, but not during production of non-palatal 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 in order 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.

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

The 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. This paper seeks to demonstrate that variability among multiple talkers also obeys the constraints suggested by the Stevens and House model.

This paper will 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, or MRI (Baer, Gore, Gracco, & Nye, 1991; Narayanan, Alwan, & Song, 1997; Story, Titze, & Hoffman, 1996). There is no inherent limit on the number of subjects in an MRI study, or the number of vowels produced per subject, but most published studies include no more than 2-4 subjects, 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 inter-token standard deviation of midsagittal tongue height during repeated productions of the same vowel by any given talker is $\sigma \approx 1.5 - 2.5$mm (Perkell, 1996), so inter-vowel and inter-talker differences in tongue height acquired from MRI should not be considered significant unless greater than at least $2\sigma \approx 3-5$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 seconds, or in 2-3 acquisition intervals of 8-13 seconds each. Most subjects are capable of either sustaining a phoneme for 8-13 seconds at a time, or of silently maintaining phoneme position ("miming" the phoneme) for 25 seconds at a time. Mimed production has not been previously analyzed, but Engwall (2000) studied the relationship between sustained and dynamic phonation. He concludes that sustained MRI production of /a,ɪ,ʊ,f,s,ʃ/ is not identical to dynamic production extrema observed in EPG and EMMA, but that MRI production may be usefully studied as an example of hyper-articulated production.

Wood (1992) analyzes X-ray tracings of midsagittal tongue shape during English and Arabic vowels, and argues 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,ɛ/). Wood does not discuss the organization of vowels within each place of articulation category, but the within-class organization is clarified by the work of Harshman, Ladefoged, and Goldstein (1977). Harshman et al. demonstrate 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, while 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 two, producing the list /i,e,ɪ,ɛ/.

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) demonstrate 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 and 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 piece-wise 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 and 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 demonstrate that any given talker produces the vowels /i/ and /ɑ/ with precisely repeatable tongue positions near the constriction, while allowing variability away from the constriction (Perkell & Cohen, 1989; Perkell & Nelson, 1985). Narayanan, Alwan, and Song (1997) observe that different talkers produce /i/ and /ɑ/ 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.

The goal of this paper is to identify phoneme-dependent patterns of inter-talker articulatory variability. Specifically, the paper seeks to demonstrate three points: (1) the place of articulation distinction is produced similarly by all five talkers, in the sense that inter-talker variability is not larger than inter-place variability, (2) oral cross-sectional area shows less inter-talker variability during production of palatal and velar vowels than it does during production of uvular and pharyngeal vowels, and (3) 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. The discussion section proposes a model of inter-talker differences that seems to explain the observed pattern of inter-talker variability. The proposed model extends the Stevens and 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 inter-talker variability near vocal tract constrictions.

Method

Image Acquisition and Acoustic Recording

Data were collected from five native speakers of American English, three males (m1-m3) and two females (f1 and f2). All talkers were undergraduate or graduate students at UCLA at the time of the study. All talkers were native speakers of the southern California dialect of American English except m2, who is 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.

Subjects 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 24cm x 24cm field of view. Anatomic regions of interest extend 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 3mm inter-slice increments, except that m2 was imaged in 2mm inter-slice increments. The tongue was separately segmented only in coronal images, therefore only coronal images were analyzed for this study. Subjects produced the English vowels /i,u,e,o,ɑ,æ,ɪ,ɛ,ʊ,ʌ,ɝ/. Subjects 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 /ɝ/ was excluded from all analyses of vowel categories, since it does not fit into any of the four place of articulation categories categories proposed in Wood (1992).

The first three subjects (f1, m1, 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 paper) 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, three subjects (f2, m2, m3) were imaged while holding vowel position ("miming" the vowels), without actually phonating. Subject m2 was imaged both while phonating and while miming the vowels, and statistical comparison of the two datasets is used in this paper to test for differences between phonated and mimed production. Phonated vowels were sustained for up to 12 seconds, depending on the comfort level of the subject. In order to acquire a complete image stack of a phonated vowel, it was necessary for the subject to repeat the vowel 2-3 times. Mimed vowels were produced for

20-25 seconds, again depending on the ability of the subject to comfortably hold his or her breath. The entire vocal tract was imaged from the lips to the trachea. The procedure required about 10 minutes per phoneme, or about 2.5 hours per subject.

In MRI images, the teeth of a subject 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 subject were made. Dental casts were submerged in water, and coronal and axial magnetic resonance image stacks were created with a 3mm 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 subject images.

Acoustic recordings were acquired in an acoustically isolated recording room one week after imaging (subjects f1, m1, and m2) or 2-6 months before imaging (subjects f2 and m3). Eight productions of each vowel were recorded: three while supine on a couch, in order to simulate subject posture during magnetic resonance imaging, and five while normally seated. During supine productions, subjects were asked to sustain the vowel for 10-15 seconds. 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 1/3, 1/2, and 2/3 of the way through the vowel, and these nine measurements (3 samples×3 tokens) were averaged for each vowel. Two talkers are 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 inter-dental 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 inter-dental space. During production of non-low vowels, in the mid-palatal and velar regions of the oral cavity, it is possible to locate the inter-dental regions as a pair of rectangles bounded by the maxillary and mandibular bulges of the cheeks and tongue (Fig. 1a). During production of low vowels, in the mid-palatal and velar regions, the tongue surface is located between the maxillary and mandibular occlusal planes, so only three of the four corners of each inter-dental rectangle are visible in the image (Fig. 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 inter-dental rectangle are visible in the image (Fig. 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 inter-molar distance.

Insert Figure 1 about here

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 (/ɝ/), the middle slice approximates the palatal place of articulation (/i,e,ɪ,ɛ/), and the posterior slice approximates the velar place of articulation (/u,ʊ/) (Wood, 1992). The two remaining vowel places of articulation, uvular (/o/) and pharyngeal (/ɑ,æ,ʌ/), were not accurately represented using coronal image data. Artifacts in the anterior images of subject m3, apparently caused by a dental restoration in the mandibular canines, made it impossible to locate the tongue surface in anterior images of this subject. Data from the anterior MRI slice of this subject were therefore excluded from statistical analysis. File formatting difficulties also precluded analysis of subject 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, Titze, and Hoffman (1996) failed near the oropharyngeal bend because the midline of the pharynx could not be reliably determined on the basis of coronal image sections.

In order to minimize the effect of inter-talker differences in head position relative to the coronal imaging plane, a tilt angle θ 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 subject 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 subjects 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 0.96 for talkers m1 and m2, 0.99 for talkers m3 and f1, and 0.86 for talker f2.

Tongue and palate contours were re-plotted using matlab (The Mathworks, Natick, MA), 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 (Fig. 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 1) midsagittally, 2) in a line displaced to the left of midsagittal by one-third of the inter-molar distance, and 3) 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.

Insert Figure 2 about here

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 ($R_{QQ} \geq 0.89$).

Inter-talker averages and standard deviations of all formant frequencies, midsagittal tongue height measurements and oral area measurements, after $\cos\theta$ correction, are given in table 1.

Insert Table 1 about here

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 five 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 five different talkers) to a maximum of $N_g = 20$ in the palatal group (productions of /i,ɪ,e,ɛ/ by five different talkers).

Four types of statistical analysis are used to analyze the means and covariances of measurement vectors. MANOVA (multivariate analysis of variance) 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. Inter-group 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 β 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 $1 \leq g \leq G$:

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

Results

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 0.3mm and 0.72mm (equal to the two-sided critical points of a univariate t statistic at levels $0.43 \leq \alpha \leq 0.77$). Average differences in oral area were between 0.012cm^2 and 0.31cm^2 ($0.76 \leq \alpha \leq 0.99$). Average magnitude difference between midsagittal tongue contours was less than 3mm, and the largest single tongue height difference was 9mm. These 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 one of the five subjects was able to perform the vowel-miming task with minimal articulatory error (less than 3mm 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 five talkers, and once using data only from the three 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 inter-talker differences are not larger than inter-vowel differences for any measurement dimension. A measurement vector of length $P = 21$ was constructed to include three formant frequencies (F1, F2, and F3), 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 (one talker per group) yields $\chi^2_{63} = 150$, significant at $\alpha = 1.9 \times 10^{-9}$.

Bonferroni simultaneous confidence intervals were calculated for all ten possible pairings of the five talkers f1, f2, m1, m2, and m3. Bonferroni analysis identified eight measurements that show a significant difference between at least one pair of talkers (simultaneous significance level $\alpha = 0.05$): palate height (anterior, middle, and posterior), posterior tongue height (left, midsagittal, and right), anterior jaw opening, and F3. Seven of the ten possible talker pairs differed in at least one articulatory measurement. These seven 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 inter-talker differences. Non-significant differences are marked “ns.” Seven of the significant inter-talker differences were differences in palate height.

Insert Table 2 about here

Additional qualitative details of palatal vault morphology can be observed in Figure 3, which shows all talkers producing the vowel /ɛ/. The palatal vaults of talkers m2 and f1 are noticeably wider than the palates of f2 and m1. The palatal vault of talker m1 appears to be narrower than that of any other talker.

Insert Figure 3 about here

F3 differences are not shown in Table 2, because there were too many of them: F3 of every male talker is less than F3 of every female talker, with average differences ranging from 318Hz (f2-m3) to 547Hz (f1-m1). Inter-talker differences in the other formants (F1 and F2) were not larger than inter-vowel differences. This finding supports the previously published observation that, independent of vowel identity, most non-rounded, non-retroflex vowels are produced with an F3 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 F3 for all vowels produced by male talkers m1, m2, and m3 was 2550Hz, implying a typical male vocal tract length of $L = 16.7$cm. Average F3 of talkers f1 and f2 was 2960Hz, 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 subject m2 producing the vowels

(/ɑ,æ,ʌ,e,ɪ,i,o,u,ʊ/). Qualitative results described in the next paragraphs were compiled by inspecting similar images from all five talkers, and by inspection of data in Table 1.

Insert Figure 4 about here

As shown in Fig. 4, the pharyngeal vowels /ɑ,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 /ɑ/ or /æ/ for every talker, with /ɑ/ lowest on average. In the middle and anterior slices, the lowest vowel for any given talker was always one of the set /ɑ,æ,o,ʊ,ɛ/, with /ɑ/ lowest on average and /o/ second-lowest on average.

The palatal vowels /i,ɪ,e,ɛ/ are 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 is always one of the set /e,ɪ,i,u/, with the vowel /i/ highest and /e/ second-highest on average. The vowel /ɛ/ is much lower than the other palatal vowels, with a tongue position that qualitatively resembles that of /æ/. Maximum average tongue height is in the middle slice for /i,ɪ,e/, and in the posterior slice for /ɛ/. Posterior tongue grooving is observed in most subjects' productions of /ɛ/, /e/, /ɪ/, and very slightly in /i/. The tongue extends laterally into the interdental space for all subjects in /ɛ/, and somewhat less so for /e/ and /ɪ/.

Most subjects produced /u/ with a tongue body position similar to that of /e/ in the middle and posterior slices, as shown in Fig. 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 Fig. 4) to nearly the height of /u/. In production of the uvular vowel /o/, the tongue is usually higher than /u/ and /ʊ/ in the posterior slice, but lower than any other vowel but /ɑ/ in the middle and anterior slices (m2 is an exception: his /o/ is relatively high in the middle slice, see Fig. 4). Qualitative inspection of the complete three-dimensional images suggests that the maximum constriction for /u,ʊ/ is achieved near the posterior slice, while that of /o/ is achieved approximately 2cm posterior to the posterior slice.

A measurement vector of length $P = 21$ was constructed to include three formant frequencies (F1, F2, and F3), 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 yields $\chi^2_{63} = 120$, significant at $\alpha = 0.00002$.

Bonferroni simultaneous confidence intervals identified ten measurements that show 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 three talkers who phonated during imaging (f1, m1, m2). Analysis of the 3-talker sample yields exactly the same list of significant pair-wise differences as analysis of the 5-talker sample, with three deletions: in the 3-talker sample, there is no significant F1 difference between uvular and pharyngeal vowels, and there is no significant difference among oral areas. With the available data, there is no way to determine whether the change in results is 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.

Inter-Talker 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 $(\alpha/2PG) = (0.05/24))$. The three significant differences marked in Fig. 5 are also found to be significant when analysis is performed using only the three talkers who phonated during image acquisition.

Insert Figure 5 about here

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:

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

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

The variance ratio test was used to evaluate inter-talker variance of both formant frequencies and log area as a function of vowel place of articulation. Inter-talker 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 are 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 3 MRI slices for 4 places of articulation), two are significant (significance level $(\alpha/6G) = (0.05/24)$). First, tongue height of palatal vowels is positively correlated with palate height in the posterior MRI slice. This result is repeated when posterior and middle-slice data are pooled as in Fig. 6. Second, oral area of pharyngeal vowels is positively correlated with palate height in the anterior slice, as shown in Fig. 6.

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

Insert Figure 6 about here

Discussion

The data presented in this paper suggest that talkers configure the tongue so as to minimize inter-talker area differences near a constriction. This result supports previous results (Perkell & Cohen, 1989; Perkell & Nelson, 1985) showing that inter-repetition variance in the utterances of any one 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 inter-talker variance of middle-slice area of palatal and velar vowels is smaller than inter-talker variance of uvular and pharyngeal vowels. Inter-talker variance of log area, 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 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

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

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

The results presented in this paper 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 are inter-talker alignment differences that are 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 inter-vowel 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 paper show that inter-talker 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 and House model (1955) 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 paper by showing reduced inter-talker variability of the mid-palatal 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 and House model, the vocal tract is represented by a series of concatenated 0.5cm 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, $R_t = 1.2$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:

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

In order to explore the effect of vocal tract size differences, four vocal tracts were simulated using the Stevens-House 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, $L = 16.7$cm for male models, corresponding to $5c/4F_3$ for average male and female F_3 measured in this paper. Small male and small female models used a vocal tract width given by $R_p = R_0 = 1.3$cm, and $R_t = 1.0$cm, while 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 five talkers analyzed for this paper, maximum oral area was just under 11cm^2, corresponding to a maximum oral radius (assuming circular cross-section) of just under 1.9cm. Second, the largest inter-talker palate height difference observed in this paper was 4.7mm, so differences between large-talker and small-talker radii were set to be in the range of 4-6mm. Third, large-talker parameters are 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 113Hz (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 five talkers in this paper: (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 /ɑ/ for all four simulated talkers. The area function of /ɑ/ shows considerably greater inter-talker variability far from the constriction than near the constriction. This finding is quantified in Table 3, which shows the simulated inter-talker 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 /ɑ,o,u,i/ respectively. The simulated inter-talker area variance of /ɑ/ 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/. While 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.

Insert Figure 7 about here

Insert Table 3 about here

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 “Small Male” vocal tract model. Beginning with the area function $A_v(x)$ optimized for vowel $v \in$/ɑ,i,o,u/, the area or log area of each tube section in turn was perturbed by ±0.05. Formant frequencies of each perturbed model were calculated, and used to estimate the derivative functions

$dF_i/dA_v(y)$ and $dF_i/d\log A_v(y)$:

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

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

where $F_i(A(x))$ is the ith formant frequency resulting from area function $A(x)$, and $\delta(x-y)$ is one when $x=y$ and zero 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 Fig. 9. 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,ɑ/). 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 /ɑ,o,u/, the derivative of F_1 is greatest for area changes near the glottis.

Insert Figure 8 about here

Insert Figure 9 about here

These simulations suggest an explanation for the empirical findings of this paper. The empirical results demonstrate that the area of the vocal tract near the palatal and velar places of articulation shows minimal inter-talker variance during production of palatal and 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,ɑ/ in this region of the vocal tract are somewhat less.

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 inter-talker oral area variance in the vicinity of the molars than uvular and pharyngeal vowels. Inter-talker 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 inter-talker 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 non-palatal vowels. These data suggest that talkers configure the tongue so that inter-talker oral area variance of any vowel is roughly proportional to the inter-talker average area of the same vowel: inter-talker variance of constriction area is smaller than inter-talker variance of an open oral cavity.

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(2), 799-828.

Cha, J. S. (2000). Articulatory speech synthesis of female and male talkers. Unpublished master's thesis, UCLA.

Chiba, T., & Kajiyama, M. (1941). The vowel—its nature and structure. Tokyo-Kaiseikan Publishing Co.

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. Proceedings of the International Conference on Spoken Language Processing (Vol. I, p. 17-20). Beijing, China.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. Journal of the Acoustical Society of America, 102(1), 655-658.

Harshman, R., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. Journal of the Acoustical Society of America, 62(3), 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. 21st annual International Conference of the IEEE/EMBS Society (p. 1170). Atlanta, GA.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America, 97(5), 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 (Third 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. Proceedings of EUROSPEECH (p. 1007-1009). Patras, Greece.

Perkell, J. S. (1996). Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modeling. J. 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/. J. 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(3), 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(1), 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.

Wood, S. (1992). A radiographic analysis of constriction locations for vowels. Journal of Phonetics, 20, 25-43.

Author Note

Mark Hasegawa-Johnson, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign; Shamala Pizza, Department of Dentistry, University of California at Los Angeles; Abeer Alwan, Department of Electrical Engineering, University of California at Los Angeles; Jul Setsu Cha, Department of Electrical Engineering, University of California at Los Angeles; Katherine Haker, Department of Radiology, Cedars-Sinai Medical Center. This research was supported by an NIH individual National Research Service Award (F23 DC 00323-01) granted to the first author, by a grant to the second author from the American Association of Orthodontics, and by an NSF CAREER award granted to the third author. Correspondence concerning this article should be addressed to Mark Hasegawa-Johnson, Beckman Institute, University of Illinois, Urbana, Illinois, 61801.

Table 1

Average measurements of formant frequencies (kHz), oral area (cm^2, posterior, middle, and anterior sections), and midsagittal tongue height (mm, posterior, middle, and anterior sections). Sign of tongue height measures is taken from measurement coordinate space shown in Fig. 2.

Measurement	/ɑ/	/æ/	/ʌ/	/ɛ/	/e/
Formant 1 (kHz)	0.72	0.72	0.60	0.62	0.43
Formant 2 (kHz)	1.28	1.66	1.29	1.88	2.24
Formant 3 (kHz)	2.67	2.66	2.64	2.79	2.78
Post Area (cm^2)	4.8	3.5	2.2	2.1	1.4
Mid Area (cm^2)	5.8	3.5	2.9	2.0	1.1
Ant Area (cm^2)	5.8	3.6	1.9	1.9	2.2
Post Ht (mm)	-0.3	0.0	2.7	4.6	6.7
Mid Ht (mm)	-4.8	-0.3	1.2	3.7	8.5
Ant Ht (mm)	-13.8	-9.6	-5.7	-5.2	-4.3

Measurement	/ɪ/	/i/	/o/	/ʊ/	/u/
Formant 1 (kHz)	0.48	0.35	0.44	0.47	0.37
Formant 2 (kHz)	2.00	2.43	0.93	1.08	0.99
Formant 3 (kHz)	2.77	2.93	2.61	2.64	2.57
Post Area (cm^2)	1.4	0.5	2.4	2.2	1.4
Mid Area (cm^2)	1.2	0.4	4.7	2.7	2.6
Ant Area (cm^2)	1.1	0.9	4.8	2.8	3.0
Post Ht (mm)	7.1	11.4	5.9	4.5	7.0
Mid Ht (mm)	8.5	11.5	-2.1	2.6	3.6
Ant Ht (mm)	-1.9	-2.3	-11.4	-8.0	-7.3

Table 2

Significant pair-wise differences, as demonstrated using Bonferroni post-hoc testing. Measurements include palate height, tongue height, jaw opening, oral area, and formant frequencies. Places of articulation include Ph=Pharyngeal, U=Uvular, V=Velar, Pa=Palatal. Sign of tongue height measures is taken from measurement coordinate space shown in Fig. 2.

Articulatory Differences Between Talkers

Measurement	f2-f1	m2-f1	m2-f2	m2-m1	m3-f1	m3-f2	m3-m1
Posterior Midsagittal Palate Ht (mm)	ns.	ns.	ns.	-4.7	ns.	ns.	ns.
Middle Midsagittal Palate Ht (mm)	-3.3	-2.8	ns.	ns.	-3.6	ns.	-2.4
Anterior Midsagittal Palate Ht (mm)	ns.	ns.	4.0	4.5	ns.	ns.	ns.
Posterior Midsagittal Tongue Ht (mm)	ns.	ns.	ns.	ns.	ns.	12.1	ns.
Posterior Left Tongue Ht (mm)	ns.	ns.	ns.	ns.	ns.	9.5	ns.
Posterior Right Tongue Ht (mm)	ns.	ns.	ns.	ns.	ns.	10.4	ns.
Anterior Jaw Opening (mm)	ns.	-6.6	-6.6	ns.	ns.	ns.	ns.

Differences Between Places of Articulation

Measurement	U-Ph	V-Ph	Pa-Ph	Pa-U	Pa-V
Middle Midsagittal Tongue Ht (mm)	ns.	ns.	9.6	10.2	ns.
Anterior Midsagittal Tongue Ht (mm)	ns.	ns.	6.7	ns.	ns.
Middle Left Tongue Ht (mm)	ns.	ns.	8.7	ns.	ns.
Anterior Left Tongue Ht (mm)	ns.	ns.	6.4	ns.	ns.
Middle Right Tongue Ht (mm)	ns.	ns.	9.1	9.9	ns.
Anterior Right Tongue Ht (mm)	ns.	ns.	6.8	ns.	ns.
Posterior Area (cm^2)	ns.	ns.	-2.3	ns.	ns.
Middle Area (cm^2)	ns.	ns.	-2.9	ns.	ns.
F1 (kHz)	-0.24	-0.27	-0.22	ns.	ns.
F2 (kHz)	ns.	ns.	0.72	1.20	1.10

Table 3

Simulated inter-talker oral area standard deviation (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 model.

Vowel	x=0.25L	x=0.43L	x=0.53L	x=0.68L
/ɑ/	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

Figure Captions

Figure 1. Posterior, middle, and anterior slices of talker f1 saying /o/; white rectangles mark the interdental spaces.

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

Figure 3. Middle-slice MRI of all subjects, including both phonated and mimed productions by subject m2. Prompt word "bet," field of view 135x90mm.

Figure 4. MRI of subject m2 producing 9 vowels (90mm field of view).

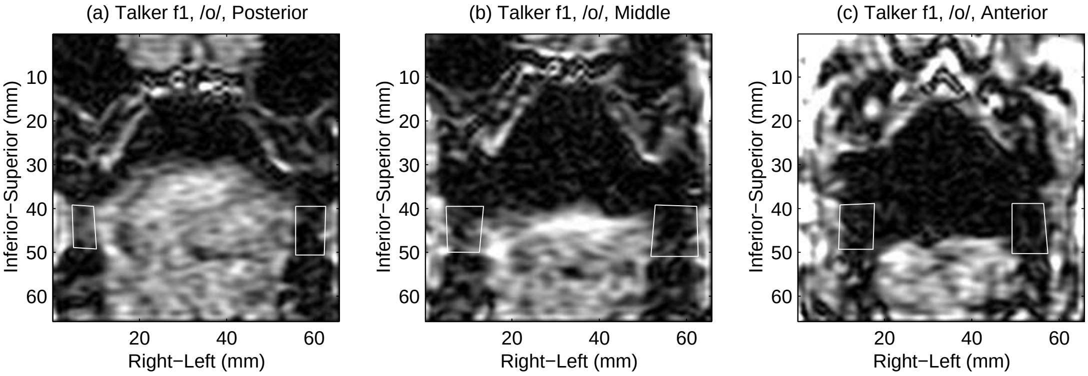
Figure 5. Within-class standard deviation of oral area, as a function of place of articulation and of MRI slice.

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.

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.

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

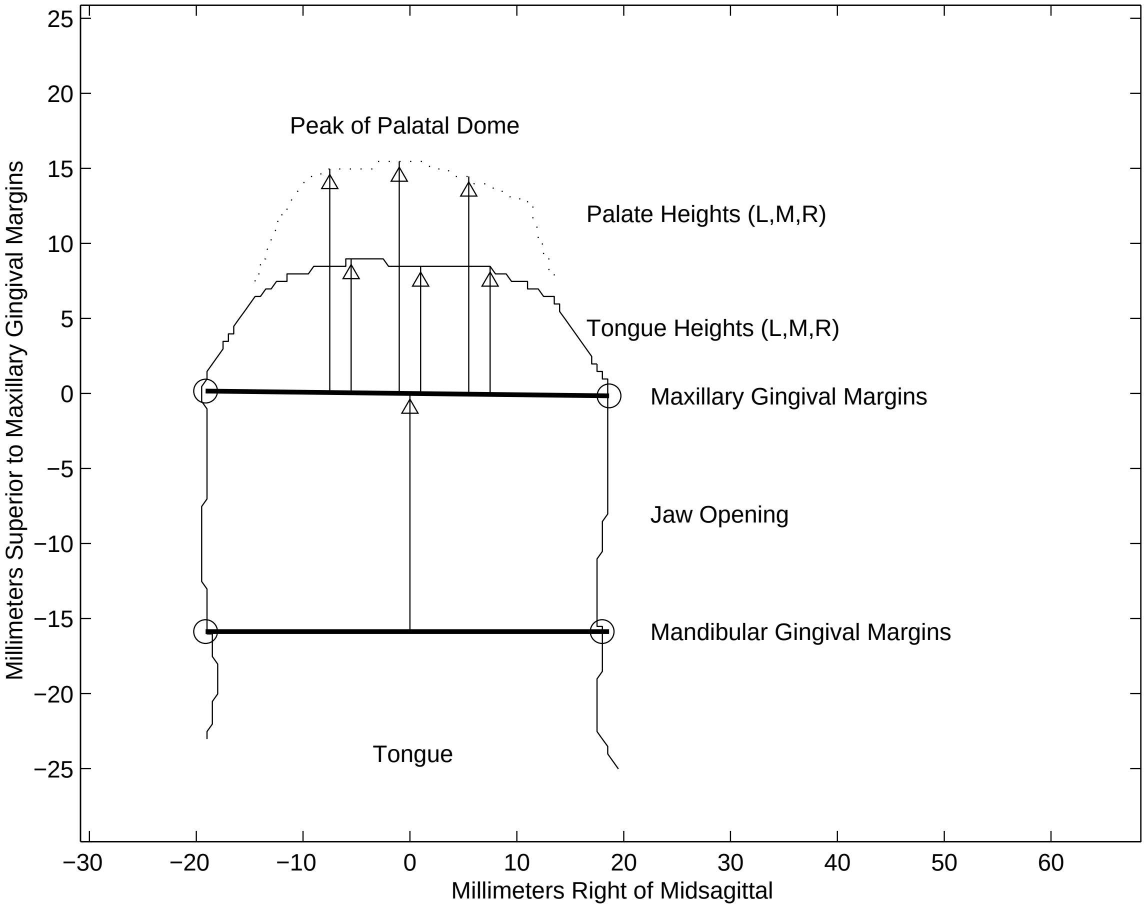
Figure 9. Derivative of formant frequencies with respect to log area changes, plotted as a function of the location of the change.

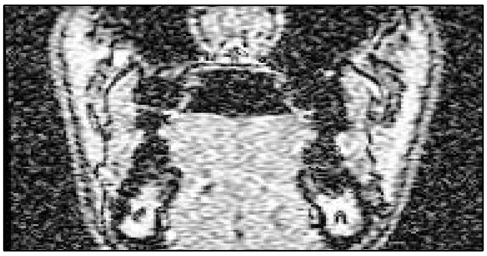


(a) Talker f1, /o/, Posterior

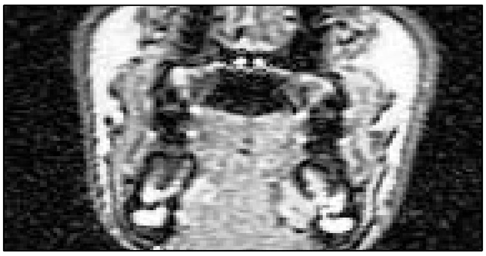
(b) Talker f1, /o/, Middle

(c) Talker f1, /o/, Anterior

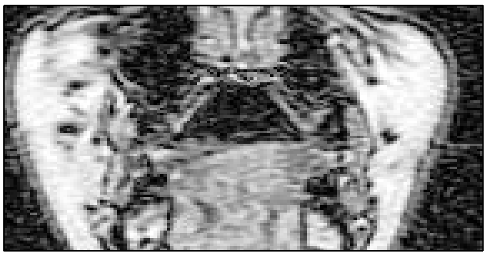

Peak of Palatal Dome
Palate Heights (L,M,R)
Tongue Heights (L,M,R)
Maxillary Gingival Margins
Jaw Opening
Mandibular Gingival Margins
Tongue
Millimeters Superior to Maxillary Gingival Margins
Millimeters Right of Midsagittal
25
20
15
10
5
0
−5
−10
−15
−20
−25
−30
−20
−10
0
10
20
30
40
50
60



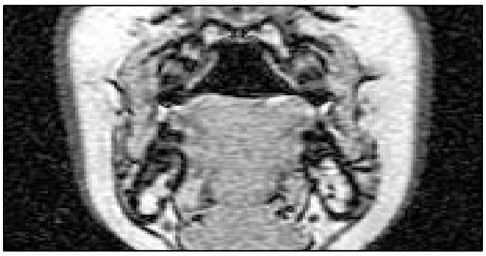
Subject m2 Phonating



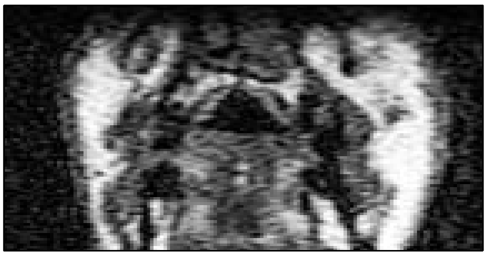
Subject m2 Miming



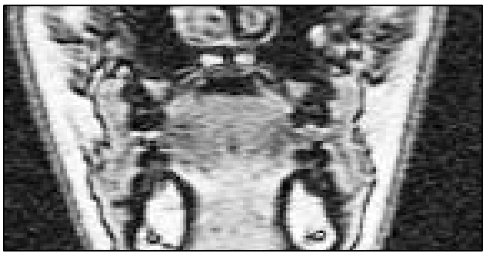
Subject f1



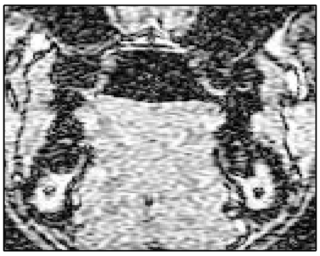
Subject f2

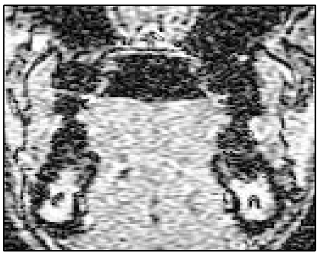


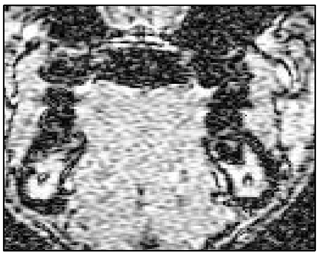
Subject m1

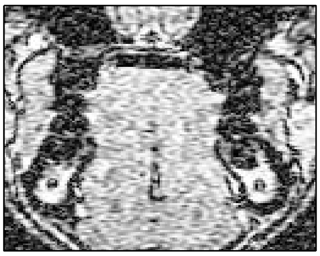


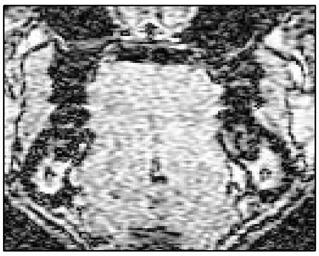
Subject m3


Prompt: father

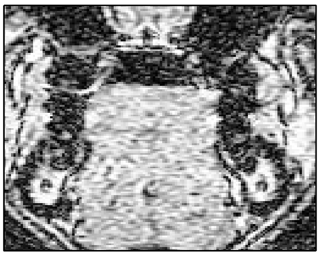

Prompt: bat

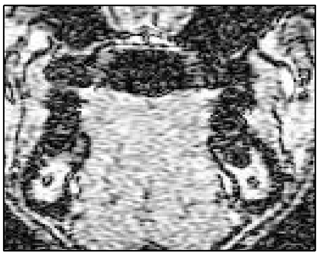

Prompt: but

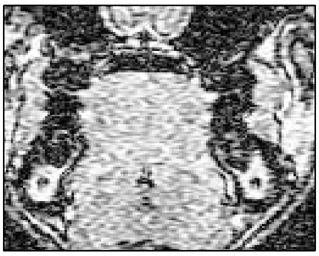

Prompt: bait

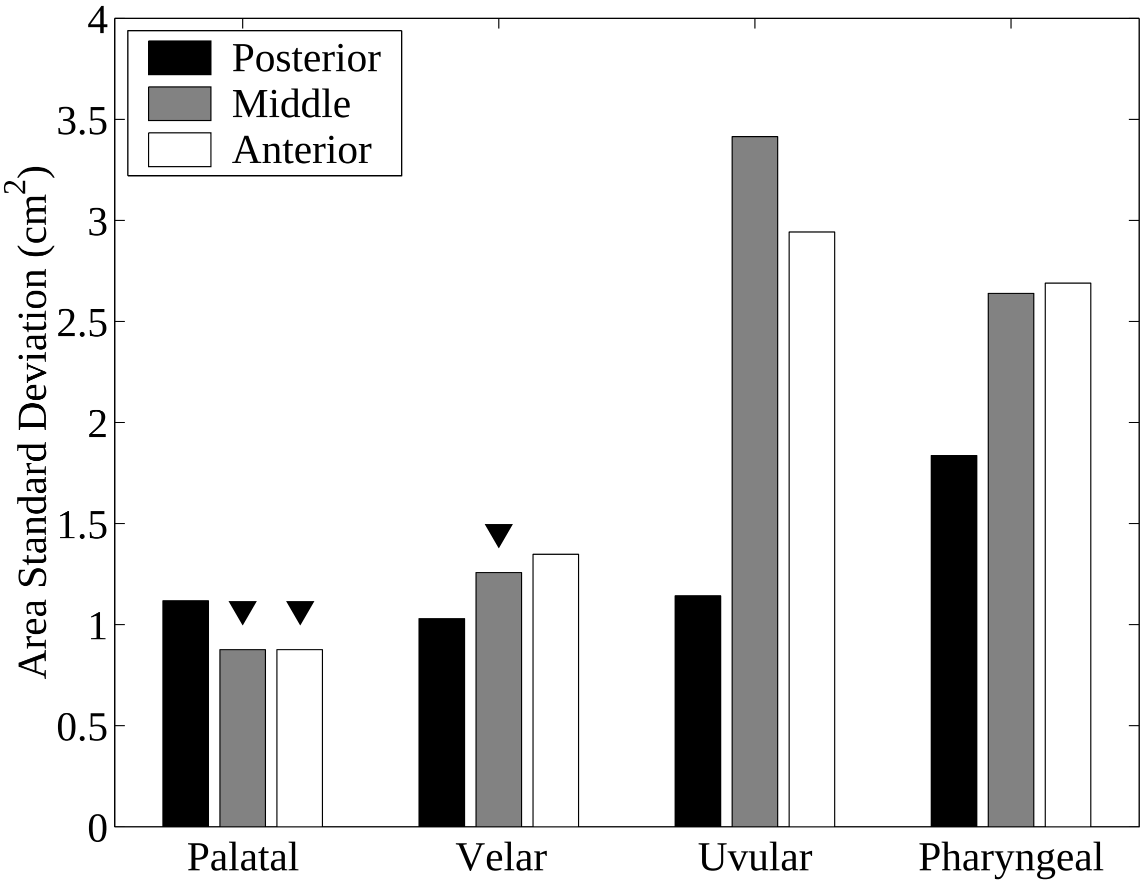

Prompt: bit

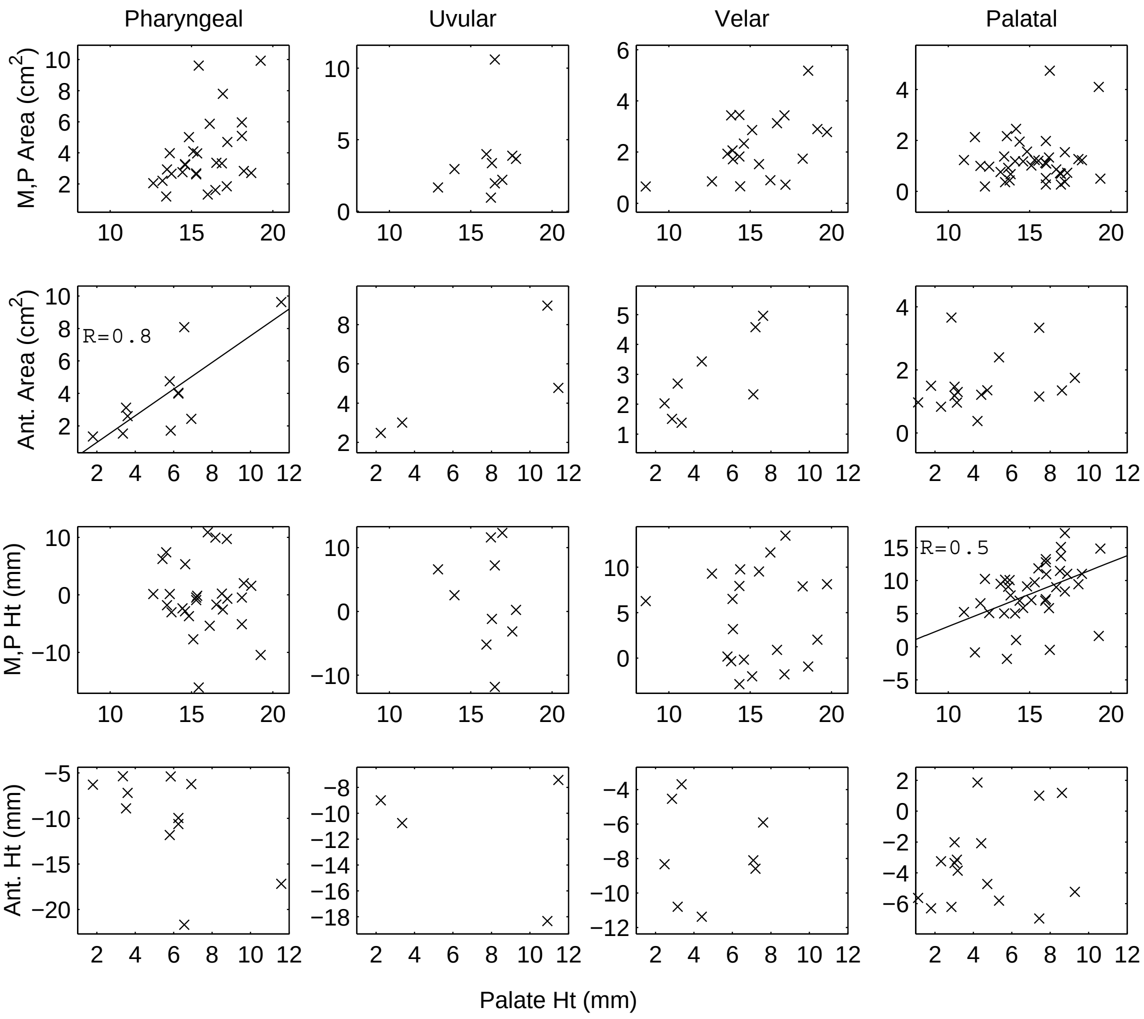

Prompt: beat

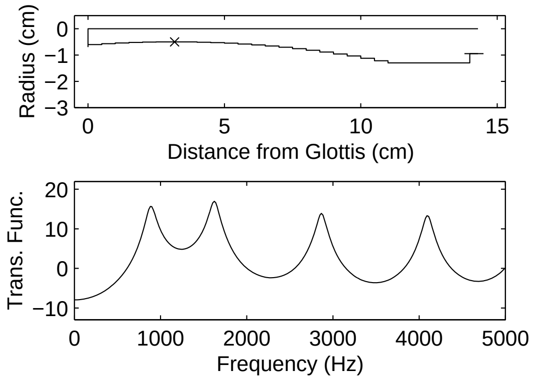

Prompt: boat


Prompt: put

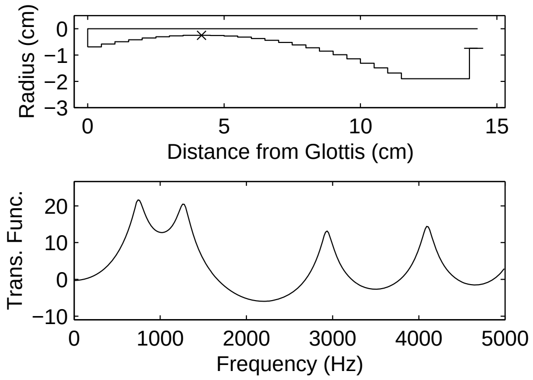

Prompt: boot


Area Standard Deviation (cm^2)
0
0.5
1
1.5
2
2.5
3
3.5
4
Posterior
Middle
Anterior
Palatal
Velar
Uvular
Pharyngeal

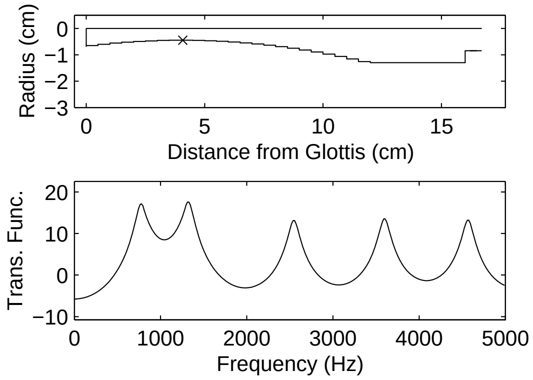

Pharyngeal
Uvular
Velar
Palatal
M,P Area (cm^2)
Ant. Area (cm^2)
M,P Ht (mm)
Ant. Ht (mm)
R=0.8
R=0.5
Palate Ht (mm)



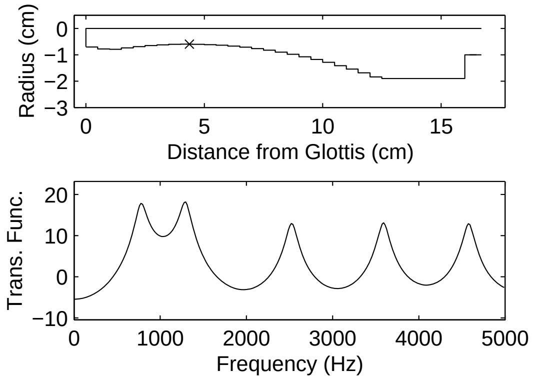
(a) Small Female



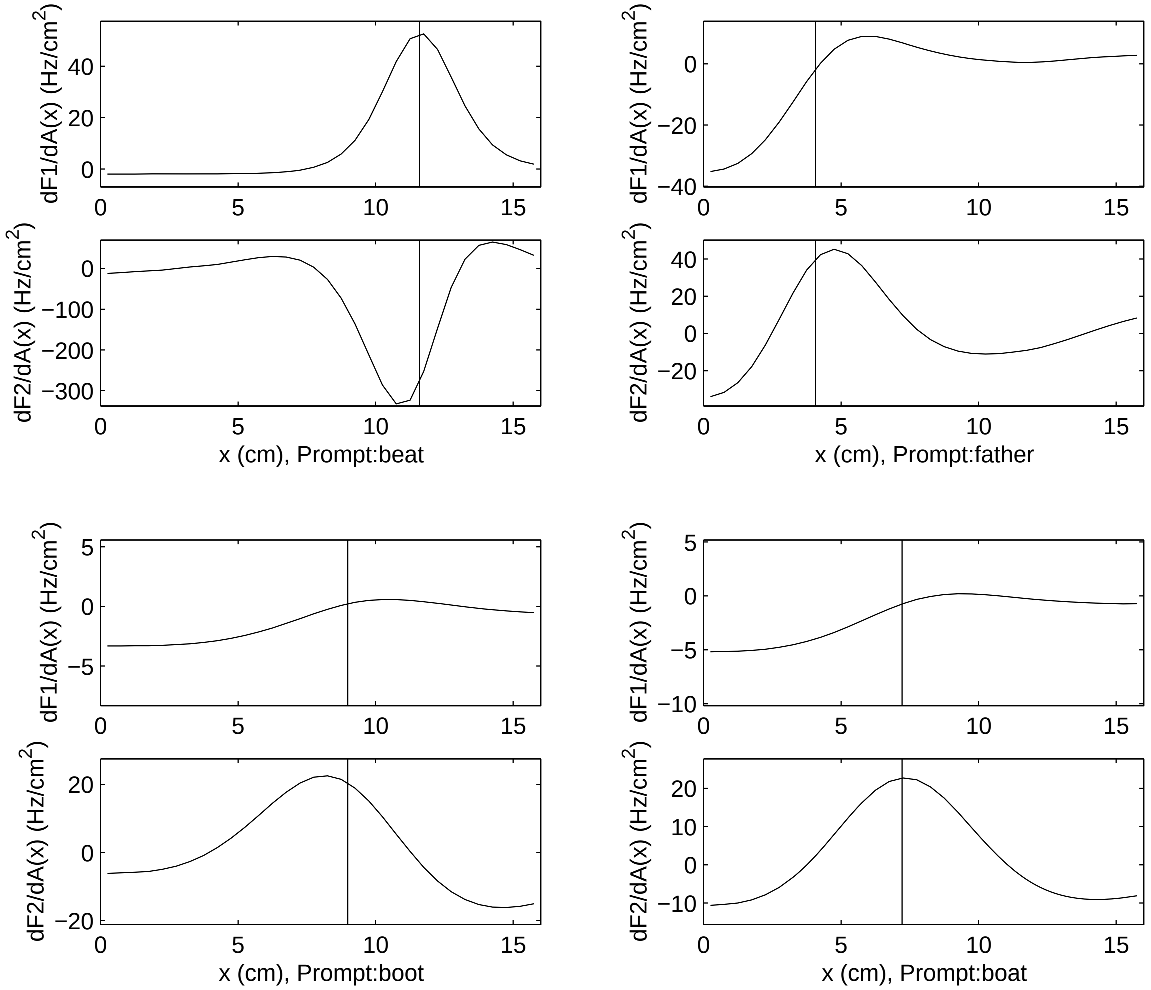
(b) Large Female

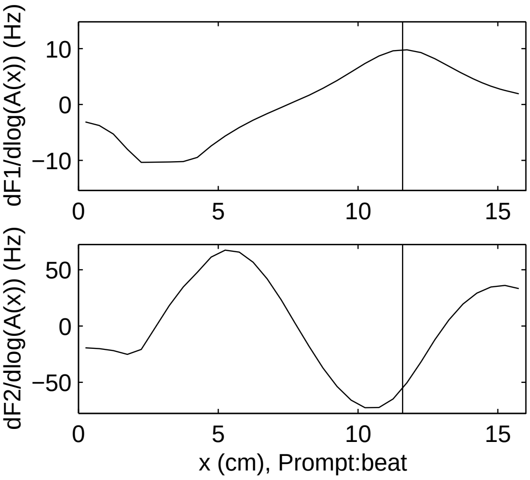


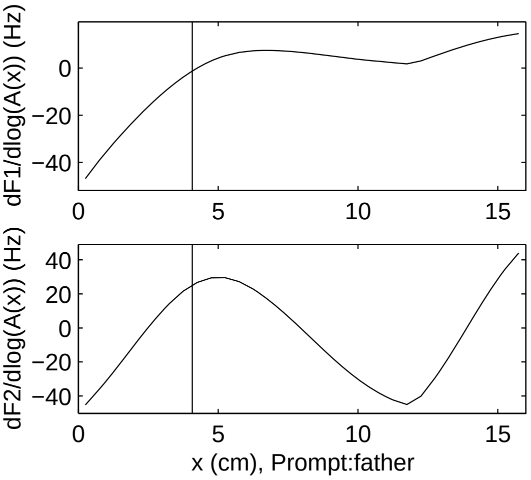
(c) Small Male

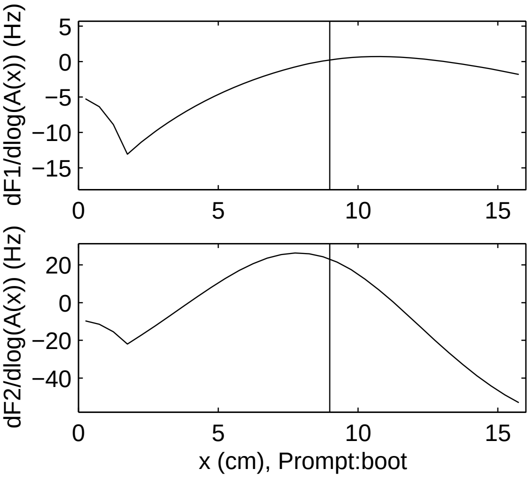


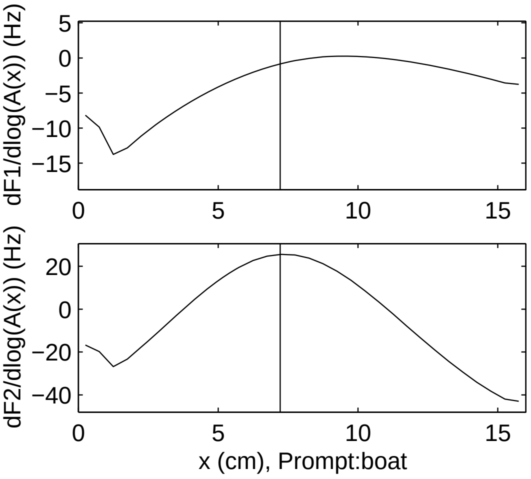
(d) Large Male


dF1/dA(x) (Hz/cm^2)
dF2/dA(x) (Hz/cm^2)
x (cm), Prompt:beat
40
20
0
−100
−200
−300
x (cm), Prompt:father
−20
−40
x (cm), Prompt:boot
5
−5
x (cm), Prompt:boat
−10
10
0
5
10
15


dF1/dlog(A(x)) (Hz)
10
0
−10
0
5
10
15
dF2/dlog(A(x)) (Hz)
50
0
−50
0
5
10
15
x (cm), Prompt:beat


dF1/dlog(A(x)) (Hz)
0
−20
−40
0
5
10
15
dF2/dlog(A(x)) (Hz)
40
20
0
−20
−40
0
5
10
15
x (cm), Prompt:father


dF1/dlog(A(x)) (Hz)
5
0
−5
−10
−15
0
5
10
15
dF2/dlog(A(x)) (Hz)
20
0
−20
−40
0
5
10
15
x (cm), Prompt:boot


dF1/dlog(A(x)) (Hz)
5
0
−5
−10
−15
0
5
10
15
dF2/dlog(A(x)) (Hz)
20
0
−20
−40
0
5
10
15
x (cm), Prompt:boat