Morphological Variation in the Adult Vocal Tract: A Modeling Study of its Potential Acoustic Impact

Adam Lammert[1], *Michael Proctor*[2,3], *Athanasios Katsamanis*[3], *Shrikanth Narayanan*[1,2,3]

Departments of [1]Computer Science, [2] Linguistics and [3] Electrical Engineering,
University of Southern California, Los Angeles, CA, USA

{lammert , mproctor} @usc.edu , {nkatsam , shri} @sipi.usc.edu

Abstract

In order to fully understand inter-speaker variability in the acoustical and articulatory domains, morphological variability must be considered, as well. Human vocal tracts display substantial morphological differences, all of which have the potential to impact a speaker's acoustic output. The palate and rear pharyngeal wall, in particular, vary widely and have the potential to strongly impact the resonant properties of the vocal tract. To gain a better understanding of this impact, we combine an examination of morphological variation with acoustic modeling experiments. The goal is to show the theoretical acoustic effect of common inter-speaker differences for a set of English vowels. Modeling results indicate that the effect is indeed strong, but also surprisingly complex and context-specific, even when morphology varies in relatively straightforward ways.

Index Terms: speech production, vocal tract morphology, inter-speaker variability, acoustic modeling, speaker modeling

1. Introduction

Human vocal tracts display considerable variation in size, proportion and shape. Dramatic variations are present during development [1, 2, 3], but substantial differences also exist between adult individuals [4]. These morphological differences all have the potential to impact an individual's acoustic output by affecting the resonant and aerodynamic properties of the vocal tract. That impact may not necessarily be observed, however, if the specifics of articulation change in compensation. Thus, two possibilities are invited: (a) morphological variation is accompanied by alteration of the observed acoustic output, or (b) speakers compensate for morphological differences through alterations in production.

These possibilities are not necessarily mutually exclusive, and both may apply in natural situations. It has long been known that the acoustics of a particular phoneme will vary across speakers, and that morphological differences are sometimes the cause of this variation. For instance, overall vocal tract length is known to affect the formant frequencies of an individual's vowels [5]. At the same time, there is growing evidence to suggest that production behavior is sometimes altered systematically with differences in morphology. For example, palate shape explains certain differences in the production of sibilant fricatives [6, 7, 8].

Examining the relationships between variation in articulation, acoustics and morphology holds promise for explaining speaker-specific production patterns. It also gets at the heart of the longstanding debate over the nature of speech tasks (i.e., whether they are acoustical or articulatory). However, for a complete understanding we must know (1) the space of possible morphological variations, (2) the theoretical acoustic impact of these variations, (3) the observed variation in acoustics, and (4) the observed variation in production. We have been working to gain insights into the first point, which we presently combine with acoustic modeling to facilitate an understanding of the second.

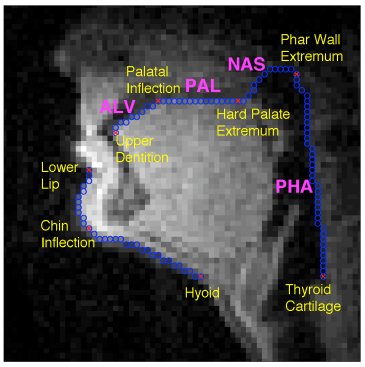


Figure 1: *Traces and anatomical landmarks for one subject.*

Our strategy is to take vocal tract configurations corresponding to English vowels, to deform those shapes according to observed morphological variations, and to calculate the acoustical properties of the resulting configuration. We focus on observed variations of the palate and rear pharyngeal wall, which are known to vary widely and which we expect to have a strong impact on critical vowel acoustics.

We view our methodology as asking what would happen to a speaker's vowels if he was suddenly given a new palate or pharyngeal wall and did not adjust his production whatsoever. We do not claim that this necessarily corresponds to any natural situation. Obviously, speakers do not experience abrupt alterations to their morphology, except through trauma or medical intervention (e.g., an orthodontic retainer). More to the point, we expect that speakers do alter certain aspects of their production based on their individual morphology. Nonetheless, we must understand the space of variations in both the acoustical and articulatory domains in order to understand intricacies of the natural situation.

In Section 2, we discuss our methodology and in Section 3 we provide an elaboration on the results of our modeling experiments. Our concluding remarks and future plans are presented in Section 4.

2. Methods

Our methodology involves: (1) collecting data regarding morphological variations, (2) categorizing those variations using

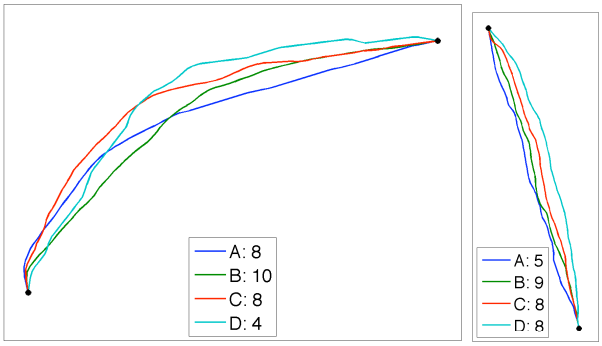


Figure 2: *Cluster centroids of the palatal (left) and pharyngeal (right) shapes. The legend shows the label and number of constituents of each cluster.*

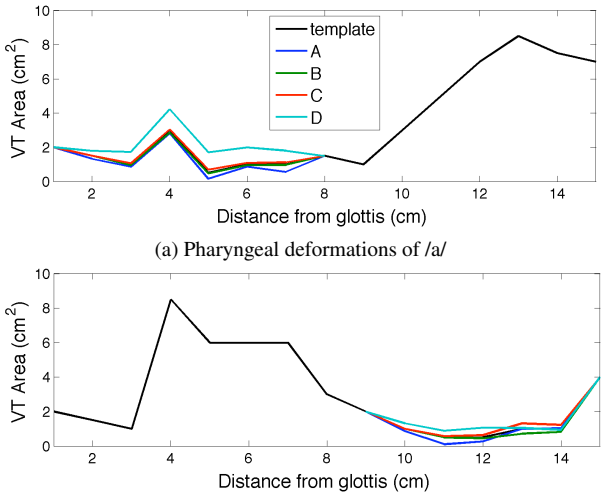
cluster analysis, (3) defining a set of template vocal tract shapes, (4) deforming the template shapes according to the observed morphological variations, and finally (5) using a tube model to determine the acoustical characteristics of the deformed vocal tract shapes. In this section, we provide detail about each of these steps. The goal is to determine the effect of morphology on formant frequencies. We examine these effects using the standard vowel chart and sensitivity functions, in the manner of Fant [9].

2.1. Data Collection & Processing

We collected real-time magnetic resonance imaging data (rtMRI) of 30 adult subjects (11 female, 19 male) from diverse linguistic (16 English, 8 German, 5 Mandarin, 1 Hindi) and racial (24 caucasian, 6 asian) backgrounds. Subjects spoke while lying supine in the scanner. The use of rtMRI allows subjects to speak naturally – i.e., there is no need for articulator posing. Midsagittal images of the subjects' vocal tracts were reconstructed at a resolution of 68×68 pixels, with a frame rate of 22.41Hz. The subjects' speech was simultaneously recorded, and will be analyzed in future extensions of this work. Further details of our acquisition protocol can be found in [10, 11], and sample videos can be found at http://sail.usc.edu/span.

For each subject, we identified 5 images representing absolute rest position during breathing, with clenched teeth. These images were averaged in order to reduce noise and to ensure a representative rest position. We used Canny edge detection [12] with manual linking and correction to trace the hard structures of the vocal tract, including (1) along the chin, and (2) along the passive articulators inside the vocal tract. The latter trace followed the upper dentition, the maxilla and palatine bones, along the posterior surface of the vomer, the pharyngeal tonsil and down the pharyngeal wall to the thyroid cartilage. Several anatomical landmarks were identified for the purposes of analysis (see Fig. 1).

Separating out the palate and pharyngeal wall contours, we aligned their end-points through rotation, translation and uniform scaling. This allowed us to regard each contour as a single vector of distance measurements, along the line defined by its end-points (i.e., the perpendicular distance). These vectors were resampled to 100 elements, and used to comprise the sets $\mathbf{x^{phar}} = \{x_{i=1}^{phar}, \ldots, x_{30}^{phar}\}$ and $\mathbf{x^{pal}} = \{x_{i=1}^{pal}, \ldots, x_{30}^{pal}\}$, for each subject i, and the pharyngeal and palatal shapes, respectively.



(a) Pharyngeal deformations of /a/

(b) Palatal deformations of /i/

Figure 3: *Example deformations of two template area functions.*

2.2. Shape Quantization & Characterization

To generalize over the variety of observed shapes, we partitioned $\mathbf{x^{phar}}$ and $\mathbf{x^{pal}}$ into 4 clusters each, using the k-means algorithm. The resulting cluster centroids, μ_c^{phar} and μ_c^{phar}, where $c = A, \ldots, D$, are displayed in Figure 2. The pharyngeal wall clusters reflect straightforward differences in curvature, from slightly convex to quite concave. Note that these differences are not related to spinal flexion or head orientation, but are inherent to a subject's morphology. The rear pharyngeal wall is a solid mass of connective tissue which does not readily distend [13, 4]. Palate shapes display a range of concavity as well, but contain substantial variation of other types as well.

In order to better understand the variations, we also performed a principal component analysis on $\mathbf{x^{phar}}$ and $\mathbf{x^{pal}}$. We observed that the largest component of pharyngeal wall variation was related to the degree of concavity, accounting for 78.5% of the variance. For the palate shapes, two major components of variation were observed, related to concavity and the position of the palatal inflection point – i.e., its position toward the front or back of the oral cavity. These accounted for 46.0% and 30.0% of the variance, respectively.

2.3. Area Functions for Template Vowel Shapes

We define one template area vector, denoted a_v^t for each vowel, v, from the set of English vowels /i, u, a/. For reference, we also define a completely neutral, uniform area function, denoted a_0^t. These vectors specify the area in cm^2 at regular, 1cm intervals along the vocal tract. The area specifications are derived from the radiographic data presented by Wood [14] in his study of English vowels.

2.4. Deformations to the Template Area Functions

We represent the morphological *differences* as deviation vectors, $d_c = \mu_c - \frac{1}{30}\sum_{i=1}^{30} x_i$. We also define a set of idealized deviation vectors, based on the observation that concavity is the major component of variation in both structures of interest (see Section 2.2). These vectors reflect a wide range of strictly parabolic deformations, which facilitates calculation of the sensitivity functions. We quantify the degree of concavity as the signed midsagittal area, in cm^2, represented by the deviation

vector.

It was necessary to convert the deviation vector values to areas in order to combine them with the template area vectors. By assuming that the vocal tract is cylindrical, this conversion can be done easily be done using the formula for the area of a circle: $a_c^d = \pi(d_c/2)^2 sgn(d_c)$. The final area functions, for acoustic modeling, are simply the sum: $a_{vc}^f = a_v^t + a_c^d$. Example area functions can be seen in Figure 3.

2.5. Acoustic Modeling

We implemented the most parsimonious acoustic model to fulfill our present purposes. Thus, we model the vocal tract in classical fashion, as a series of lossless, cylindrical, concatenated tubes (e.g., [9, 15, 16, 17]). Using this model, the formants can easily be calculated from the area vectors as follows:

1. Calculate the reflection coefficients between each tube i and the adjacent tube: $\Gamma_i = (\alpha_{i+1} - \alpha_i)(\alpha_{i+1} + \alpha_i)$, where α_i is the value of a_{vs}^f at location i.
2. Calculate the coefficients of the prediction filter polynomial from the reflection coefficients. This is done by Levinson's recursion, as described in [18] and implemented in the Matlab® Signal Processing Toolbox™.
3. Find the formants by taking the roots of the prediction filter polynomial, and converting to Hertz.

3. Results

Figure 4 displays a standard formant space with the vowel chart overlain. The position of the three English template vowels can be seen, as well as the position resulting from deformation of those templates with the observed morphological characteristics (see Figure 2). The crucial aspects to notice are the magnitude and direction shifts in position, and the shape formed by the progression of deformations. Note that these aspects are different between vowels, meaning that the effects of deformation are vowel-specific.

Shifts resulting from pharyngeal variation appear closely related to changes in concavity. The vowels /i/ and /u/ appear to shift linearly in one direction as concavity of the shape increases, though the magnitude is somewhat less for /u/. The movement for /i/ is toward the upper left-hand (i.e., more i-like) corner, and for /u/ it is almost directly upward. For /a/, the shift is straight in piecewise fashion on either side of the template. As the shapes become more concave than the template, the movement is toward the upper left-hand corner of the vowel chart. As the shapes become more convex than the template, the shift is toward the upper right-hand (i.e., more u-like) corner.

Palatal variation seems to only have a substantial impact on /i/, which has a constriction closest to the palate. Both /u/ and /a/ are barely affected by comparison, showing shifts approximately 10x less in magnitude. As the palate becomes less concave – essentially tightening the relevant constriction – /i/ shifts toward the upper left-hand corner of the vowel chart. There is also a smaller shift along the orthogonal direction in the vowel chart, corresponding to the position of the palatal inflection point.

Figures 5 displays the sensitivity of the first three formants to changes in concavity of the pharyngeal wall in a completely neutral vocal tract. Figure 6 shows the same, but for the palate. The smooth curve represents the sensitivity resulting from idealized, parabolic concavity deformations. The sensitivity related to the observed morphological shapes can also be seen.

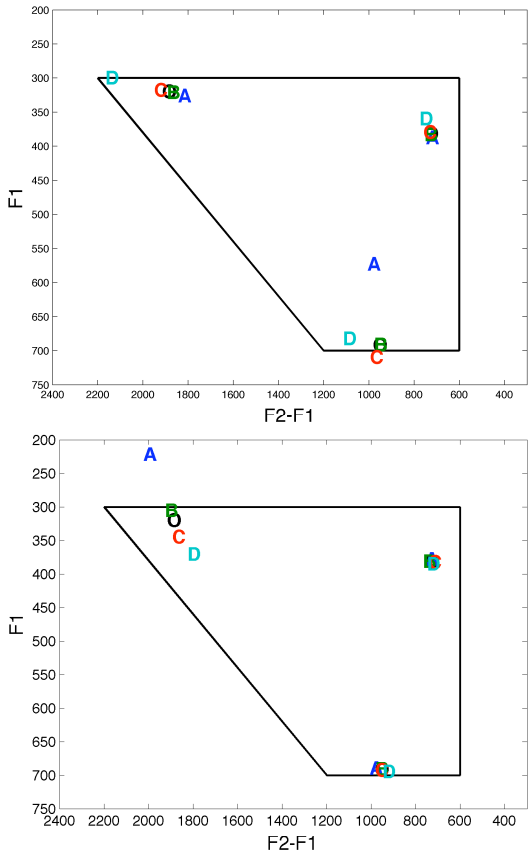


Figure 4: *Acoustic consequences of pharyngeal (top) and palatal (bottom) deformations to the template area functions.*

The sensitivity function is clearly nonlinear for all formants. There is very little sensitivity near the template shape, but it quickly accelerates as one moves away from that point. Note that the sensitivity of observed shapes can display large deviations from the idealized sensitivity function. This shows that concavity is not the only morphological variation which impacts the acoustics. This is is especially true of the palate, for which the position of the palatal inflection point is a major component of variation.

4. Conclusions

Common morphological variations in the palate and rear pharyngeal wall have the potential to strongly impact resonant properties of the vocal tract. This impact can be complex in terms of its shape in formant space. It can also be context-specific, in the sense that it will not impact all vocal tract configurations in the same way. This is the case even when morphology varies in relatively straightforward, interpretable and low-dimensional ways. In order to fully understand acoustic and articulatory variability, morphological variability must be considered as well. This will be possible with the use of Imaging technologies, like rtMRI, which allow for as complete data as possible.

Still, it is unclear whether these morphological variations are reflected in the acoustic signal during natural speaking situations. We expect that an individual's production will compensate for these differences, to some extent. The data collected for the present study will allow us to address this issue empirically, and that is our intention. Future work also will involve an extension of these modeling experiments to fricatives and other phonemes.

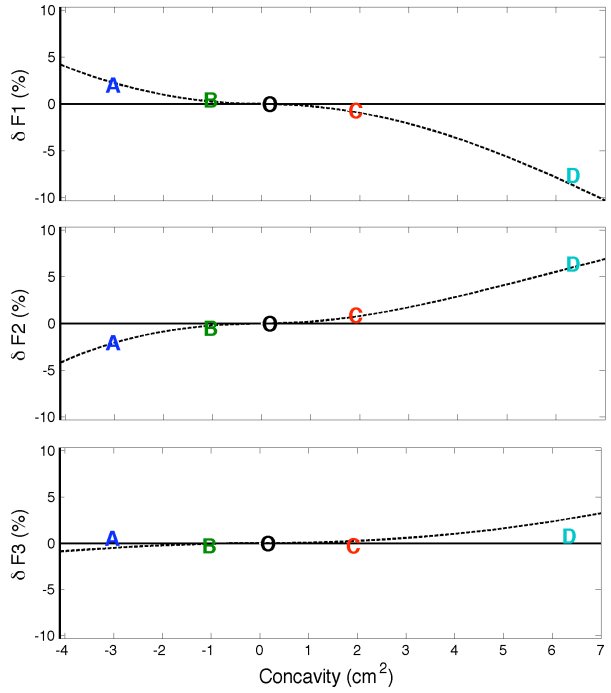


Figure 5: *Sensitivity of the first three formants with respect to idealized and observed concavity of the pharyngeal wall.*

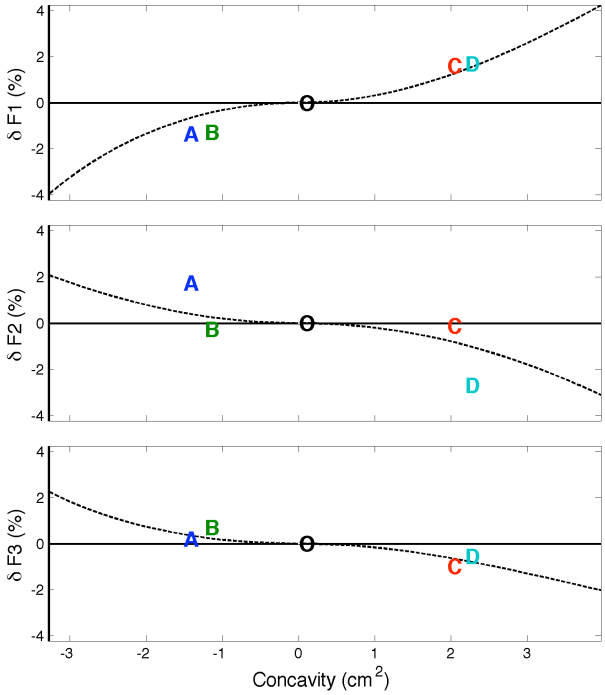


Figure 6: *Sensitivity of the first three formants with respect to idealized and observed concavity of the palate.*

5. Acknowledgements

This work was supported by NIH Grant DC007124.

6. References

[1] W. Fitch and J. Giedd, "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," *Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1511–1522, 1999.

[2] L. Boe, G. Captier, J. Granat, M. Deshayes, J. Heim, P. Birkholz, P. Badin, N. Kielwasser, and T. Sawallis, "Skull and vocal tract growth from fetus to 2 years," in *Eighth International Seminar on Speech Production*, 2008.

[3] H. Vorperian, S. Wang, M. Chung, E. Schimek, R. Durtschi, R. Kent, A. Ziegert, and L. Gentry, "Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study," *Journal of the Acoustical Society of America*, vol. 125, no. 3, pp. 1666–1678, 2009.

[4] A. Lammert, M. Proctor, and S. Narayanan, "Morphological variation in the adult vocal tract: A study using rtmri," in *9th International Seminar on Speech Production*, 2011.

[5] P. Ladefoged, *A Course in Phonetics: Fifth Edition*. Thomson Wadsworth, 2006.

[6] M. McCutcheon, A. Hasegawa, and S. Fletcher, "Effects of palatal morphology on /s, z/ articulation," *Journal of the Acoustical Society of America*, vol. 67, no. 1, pp. 94–94, 1980.

[7] K. Honda and C.-M. Wu, "Differences in speaker's articulatory space: Their contribution to vowel gesture and acoustic pattern," *The Journal of the Acoustical Society of America*, vol. 100, no. 4, pp. 2598–2598, 1996.

[8] S. Fuchs, P. Perrier, C. Geng, and C. Mooshammer, "What role does the palate play in speech motor control? insights from tongue kinematics for german alveolar obstruents," in *Towards a better understanding of production processes*, J. Harrington and M. Tabain, Eds. Psychology Press, 2006.

[9] G. Fant, *Acoustic Theory of Speech Production*. Mouton, 1960.

[10] S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production," *Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1771–1776, 2004.

[11] E. Bresch, J. Nielsen, K. Nayak, and S. Narayanan, "Synchronized and noise-robust audio recordings during realtime mri scans," *Journal of the Acoustical Society of America*, vol. 120, pp. 1791–1794, 2006.

[12] J. Canny, "Computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[13] L. Penning, "Radioanatomy of upper airways in flexion and retroflexion of the neck," *Neuroradiology*, vol. 30, pp. 17–21, 1988.

[14] S. Wood, "A radiographic analysis of constriction locations for vowels," *Journal of Phonetics*, vol. 7, pp. 25–43, 1979.

[15] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*. Prentice-Hall, 1978.

[16] K. Stevens, "On the quantal nature of speech," *Journal of Phonetics*, vol. 17, pp. 3–45, 1989.

[17] ——, *Acoustic Phonetics*. MIT Press, 1998.

[18] S. Kay, *Modern Spectral Estimation*. Prentice-Hall, 1988.