JSLHR

Supplement

Interspeaker Variability in Hard Palate Morphology and Vowel Production

Adam Lammert,[a] Michael Proctor,[a] and Shrikanth Narayanan[a]

Purpose: Differences in vocal tract morphology have the potential to explain interspeaker variability in speech production. The potential acoustic impact of hard palate shape was examined in simulation, in addition to the interplay among morphology, articulation, and acoustics in real vowel production data.
Method: High-front vowel production from 5 speakers of American English was examined using midsagittal real-time magnetic resonance imaging data with synchronized audio. Relationships among hard palate morphology, tongue shaping, and formant frequencies were analyzed. Simulations were performed to determine the acoustical properties of vocal tracts whose area functions are altered according to prominent hard palate variations.
Results: Simulations revealed that altering the height and position of the palatal dome alters formant frequencies. Examinations of real speech data showed that palatal morphology is not significantly correlated with any formant frequency but is correlated with major aspects of lingual articulation.
Conclusion: Certain differences in hard palate morphology can substantially affect vowel acoustics, but those effects are not noticeable in real speech. Speakers adapt their lingual articulation to accommodate palate shape differences with the potential to substantially affect formant frequencies, while ignoring palate shape differences with relatively little acoustic impact, lending support for acoustic goals of vowel production.

Key Words: morphology, anatomy, physiology, articulation, speech production, speech motor control

Speech production research has long been concerned with explaining variability in the acoustic and articulatory domains, both within and across speakers. An essential consideration for explaining this variability is vocal tract morphology. The morphology of vocal tract structures is fundamentally linked to speech articulation and acoustics through a complex interplay. The overall shape of the vocal tract—so crucial for determining its acoustical properties (e.g., resonant characteristics)—is determined not only by actively controlled shaping mechanisms but also by the inherent shape of vocal tract components. Elements of vocal tract morphology that vary across speakers should then result in some combination of articulatory and acoustic differences across those same speakers, even when producing identical segments. Achieving the same acoustic output implies articulatory differences in compensation for morphological variation, unless those specific articulatory differences happen to not result in acoustic differences. Conversely, if speakers do not differ in their articulations, then their morphological differences will be manifested in the acoustics. Several key questions arise from this interplay. First, do certain active articulatory strategies reflect morphological characteristics? Also, which morphological differences have the potential to affect the acoustics? Finally, which aspects of morphology are evident in the acoustic signal?

Perhaps the most extensively studied aspect of the structure–function interplay concerns the role of vocal tract length in vowel production variability. Vocal tract length varies considerably through ontogenesis, whereby its average length doubles from around 8 cm to 16 cm (Fitch & Giedd, 1999; Vorperian et al., 2005, 2009). Adult vocal tracts also vary substantially across individuals, ranging from approximately 13 cm to as many as 20 cm. Lengthening the vocal tract has the potential to lower all formant frequencies (Fant, 1960; Stevens, 1998), and this effect has been observed in real vowel acoustics (Lee, Potamianos, & Narayanan, 1999; Peterson & Barney, 1952). Differences in articulation cannot easily compensate for this acoustic effect, should speakers try, except to a limited extent through lip protrusion and laryngeal raising. In addition, articulatory strategies may be influenced by the proportional length of the oral to pharyngeal cavities. The pharynx becomes proportionally longer through the course of development (Arens et al., 2002; Chiba & Kajiyama, 1941; King, 1952), and adult speakers display

[a]University of Southern California
Correspondence to Adam Lammert: lammert@usc.edu
Editor: Jody Kreiman
Associate Editor: Ewa Jacewicz
Received July 2, 2012
Revision received December 6, 2012
Accepted June 19, 2013
DOI: 10.1044/1092-4388(2013/12-0211)

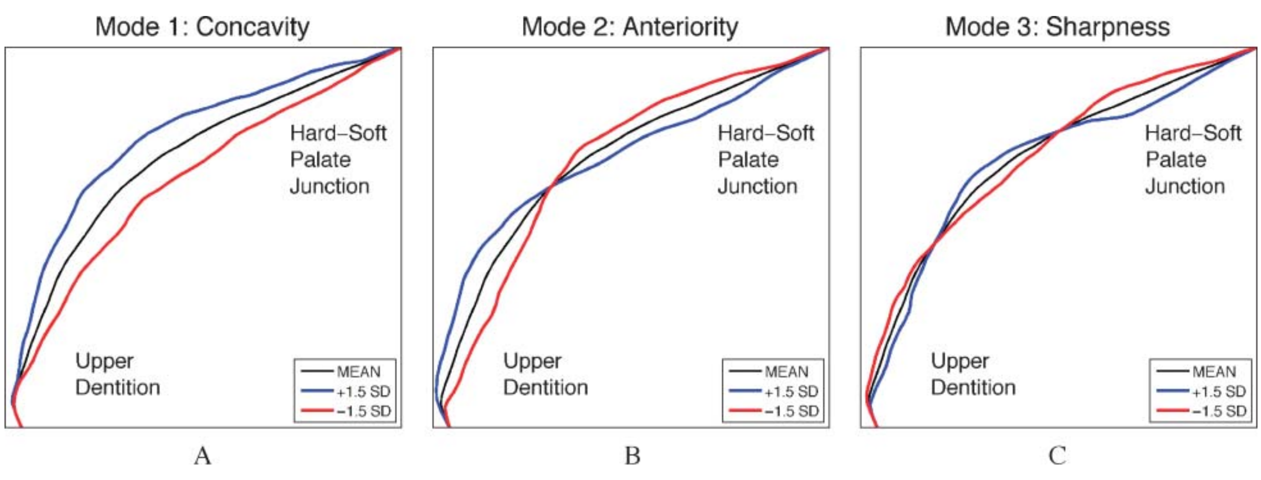
Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

sexual dimorphism, such that the pharynx of males is proportionally longer (Vorperian et al., 2011; Vorperian & Kent, 2007). Differences in proportions can influence articulatory strategies by forcing speakers into specific production patterns (Fuchs, Winkler, & Perrier, 2008; Ménard, Schwartz, Boë, & Aubin, 2007; Nissen & Fox, 2009; Winkler, Fuchs, & Perrier, 2006; Winkler, Fuchs, Perrier, & Tiede, 2011) and are also evident in the acoustics as vowel-specific scaling of acoustics with vocal tract length (Fant, 1966, 1975; Nordström, 1975).

Studies have investigated the role of hard palate morphology in speech production as well, which is also the focus of the present work. As shown in Figure 1, hard palate shape varies in three prominent ways: (a) the height of the palatal dome, (b) the position of the dome's apex in the oral cavity, and (c) the angularity of the dome around the apex (Lammert, Proctor, & Narayanan, 2013). Several studies have shown that the first of these variations—whether the palate is highly domed or relatively flat—has a substantial impact on articulatory strategies. Speakers with flat palates have been shown to exhibit less articulatory variability during vowel production than speakers with domed palates (Brunner, Fuchs, & Perrier, 2005, 2009; Mooshammer, Perrier, Geng, & Pape, 2004; Perkell, 1997). Apical versus laminal articulation of sibilant fricatives also depends on palate shape (Dart, 1991). Furthermore, artificially flattening palate shape forces corresponding changes in jaw height and the positioning of the tongue during coronal fricative production (Honda, Fujino, & Kaburagi, 2002; Thibeault, Ménard, Baum, Richard, & McFarland, 2011), which quickly minimizes acoustic differences (Baum & McFarland, 1997). Brunner, Hoole, and Perrier (2007) showed that vowel articulation also adapts over time to artificial changes in hard palate shape, suggesting that palatal morphology accounts for some vowel articulation variability across speakers. However, the specific aspects of vowel articulation and acoustics that are affected by morphology have not been rigorously investigated. It is clear that differences in palatal morphology have the potential to affect the resonant properties of the vocal tract and thereby alter the acoustic output (Lammert, Proctor, Katsamanis, & Narayanan, 2011), but whether articulation varies in compensation for palatal differences has not been well studied.

The present study focused on the interplay between hard palate morphology and vowel production behavior. Specifically, interspeaker variations in hard palate morphology were considered in regard to explaining variability in vowel articulation and acoustics. The investigation involved two parts that provide complementary insights: (a) analysis of real speech data to assess the extent of systematic relationships between morphological variation and vowel production variation and (b) acoustic vocal tract simulations to assess the potential of different morphological variations to affect the acoustics. The first part is essential for analyzing the interplay in question, but it does not provide a complete picture. With real speech data, it is not possible to factor out the effect of articulation, which creates ambiguity in the interpretation of some results. For instance, articulatory variability might be viewed as an attempt by speakers to minimize acoustic variability in the face of morphological differences. However, without knowing what acoustic variability results from those morphological differences irrespective of the accompanying articulatory variability, it is not possible to draw that conclusion with certainty. It may simply be that the particular morphological differences between those speakers have no

Figure 1. The three largest modes of variation in hard palate shape, previously determined. Modes reflect differences in concavity, anteriority of the apex, and sharpness of the palate around the apex. The overall mean hard palate shape is shown in black, and the blue and red lines show the nature of deviations from the mean according to each mode. The magnitude of the deviations shown reflects the magnitude of variations seen in the subject pool, at precisely ±1.5 *SD*s from the mean shape. Because these modes account for over 85% of the overall variance, it is possible to well represent arbitrary hard palate shapes using only these three modes. Panel A: 51% of variance. Panel B: 25% of variance. Panel C: 10% of variance. From "Morphological Variation in the Adult Hard Palate and Posterior Pharyngeal Wall," by A. Lammert, M. Proctor, and N. Narayanan, 2013, *Journal of Speech, Language, and Hearing Research, 56,* p. 524.



acoustic impact. Scenarios like this make it necessary to consider simulations alongside analysis of real data.

The present article reflects an expansion of initial work to quantify morphological variations in the hard palate across speakers using real-time magnetic resonance imaging (rtMRI) data (Lammert, Proctor, & Narayanan, 2011). Subsequent work has examined the nature of morphological variation in this crucial structure (Lammert et al., 2013) and the theoretical consequences of those variations on vowel acoustics (Lammert, Proctor, Katsamanis, & Narayanan, 2011). The current work comprises an investigation of the palate's impact on speech production using rtMRI paired with complementary acoustic simulations. By combining parametric analysis of palatal morphology with acoustic simulations of different vocal tract shapes, the effect of palatal variation on vocal tract resonances is studied. Palate shapes, lingual articulation, and formant frequencies of several American English speakers are also assessed in order to examine their interplay. Finally, the relevance of this work to current understanding of morphological variation, interspeaker variability, and goals of speech production is discussed.

Method

Speech Data

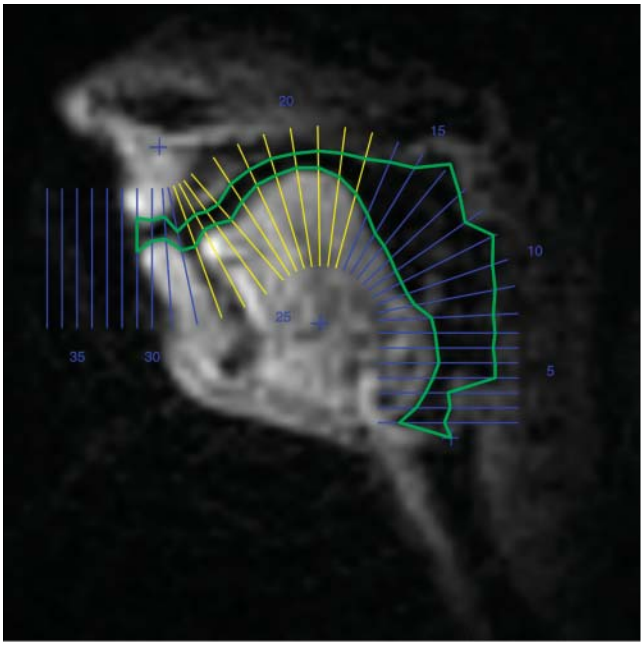
Articulatory and acoustic data were taken from five male speakers in the recently collected MRI-TIMIT corpus (Narayanan et al., 2011), a collection of rtMRI data (Narayanan, Nayak, Lee, Sethy, & Byrd, 2004) of continuous read speech from 10 native speakers of American English. Images were reconstructed using a sliding-window procedure with a step size of one repetition time (TR) (6.164 ms), resulting in an effective frame rate of 162.23 frames per second, with a spatial resolution of 68 × 68 pixels over 20 × 20 cm (approximately 2.9 cm pixel width). These images show full midsagittal views of the subjects' upper airways, including articulatory dynamics and morphological characteristics in the midsagittal plane. Speech acoustics were simultaneously recorded using an optical microphone and subsequently processed according to the method described by Bresch, Nielsen, Nayak, and Narayanan (2006) to remove scanner-generated audio noise.

For each of the five subjects, five tokens were selected of the high-front vowel /i/ as spoken during the word *people*.[1] Tokens produced in interlabial contexts were chosen in order to minimize lingual co-articulatory effects on the vowel of interest. Use of high-front vowels was motivated by previous modeling work, which suggested that high-front vowels emphasize any potential effect of palate morphology on formant frequencies differences (Lammert, Proctor, Katsamanis, & Narayanan, 2011), presumably because of the relative narrowing of the vocal tract in the palatal region. Each vowel token was considered to extend from the first and last peak of periodicity in the acoustic signal. Three points in each vowel were marked for further analysis, corresponding to 25%, 50%, and 75% of the total vowel interval.

Formant frequencies were estimated from the linear predictive coding spectrum, calculated over a 25-ms window centered at each specified time point. Positions of the three lowest peaks in the spectrum were identified as Formants 1 through 3. The mean formant values across all measurement points in a given vowel were used to represent the acoustics for that token. Linear prediction order was initialized to 14 and—in situations where this order failed to return formants in the broad frequency ranges expected for a high-front vowel (200–500 for Formant 1 [F1], 1800–2500 for Formant 2 [F2], and 2300–4000 for Formant 3 [F3], with no bandwidth criteria)—the order of analysis was refined using linear predictive coding models varying in order between 12 and 18 until a set of three formants could be robustly identified. The order number closest to 14 that returned formant values in the specified ranges was taken to be the optimal one.

For each token, a composite semipolar analysis grid was superimposed onto the midsagittal plane, extending from the glottis to the lips with gridlines spaced at approximately 5-mm intervals (Maeda, 1979; Öhman, 1967). An example grid for one subject can be seen in Figure 2. The grid was manually positioned relative to four anatomical landmarks: (a) the glottis, (b) the highest point on the palate, (c) the alveolar ridge, and (d) the lips. Proctor, Bone, Katsamanis,

Figure 2. Midsagittal image of one male subject that was used in the analysis. The image shows the subject during production of a high-front vowel. Automatically derived traces of the vocal tract outlines have been overlain, along with the gridlines used for analysis.



[1] Subject mm2 had only three usable tokens due to problems at acquisition time.

and Narayanan (2010) described this placement method, along with a technique for automatically tracing the vocal tract outlines in rtMRI data, by identifying air–tissue boundaries intersecting with the gridlines. This tracing method was used to produce midsagittal vocal tract outlines, which were subsequently inspected for accuracy and manually corrected when necessary.

Through use of the midsagittal vocal tract outlines, palate traces, tongue traces, and midsagittal distance functions were extracted. *Palate traces* were defined as the segment of the upper vocal tract outlines extending from the gridline passing closest to the upper dentition to the gridline passing closest to the hard–soft palate junction (i.e., the posterior nasal spine). This definition was used to be consistent with the traces identified by Lammert et al. (2013). These same gridlines bounding the palate were also used to delimit the relevant tongue traces for articulatory analysis. Tongue traces were extracted from three-frame intervals centered at the time points marked for acoustic analysis. This provided a total of nine tongue traces for each vowel token, the means of which were used to represent the tongue trace for that token. The overall mean tongue trace, across all five tokens, was used to represent the tongue trace for each subject. Midsagittal distance functions were extracted from the same three-frame intervals as the distance functions, and the means were taken in the same manner to provide a distance function for each subject. These midsagittal distance functions were calculated by finding the distance between the upper and lower outlines along each gridline. The resulting distance functions can be seen in Figure 3.

Vocal tract length and hard palate length were also measured for each token using the midsagittal vocal tract outlines. Gridlines that intersected with the glottis and the most anterior protrusion of the lips were identified and were used to delimit the extent of the vocal tract. The vocal tract midline was then defined as the series of line segments whose endpoints lie along neighboring gridlines, halfway between the outer and inner vocal tract outlines. Vocal tract length was calculated as the distance along the vocal tract midline between these delimiting gridlines. Hard palate length was taken to be the distance, along the vocal tract midline, between the gridlines delimiting the hard palate. The mean vocal tract length and palate length measurements for each subject in this study are shown in Table 1. These measurements facilitate normalization of formant frequencies for acoustic analysis, as well as subject-specific acoustic simulations, discussed in the Simulations subsection in the Method section.

Measurements of vocal tract length were used to normalize formant frequencies, which otherwise may vary with vocal tract length in addition to lingual articulation. Meaningful comparison of articulation-relevant formant variation is possible only after factoring out the effect of vocal tract length variation. Having vocal tract length measurements for each token affords very detailed normalization of formant frequencies, accounting for changes in vocal tract length due to, for instance, differences in lip rounding or vocal tract posture. Given measurements of vocal tract length, L_{obs}, and observed formant frequencies, F_{obs}, the normalized formant frequencies were obtained as follows: $F_{norm} = (L_{obs} F_{obs})/17.5$. This procedure scales all formant frequencies to those produced by a 17.5-cm vocal tract.

Descriptions of hard palate and tongue shape were based on the parameterization of hard palate morphology described by Lammert et al. (2013). A key result of that work was that most of the variation (approximately 85%) observed across subjects could be represented by a small number of modes in shape variation. The three largest modes are shown in Figure 1. The first mode, accounting for 51% of the variance in the data, represents the degree of concavity of the palate (i.e., whether it is flat or domed). The second mode, which accounts for another 25% of the variance, is related to the anteriority of the palate: whether the apex of the dome is positioned toward the anterior or posterior portion of the oral cavity. An additional 10% of the variance can be attributed to the sharpness/flatness of the palate at its apex. These modes are referred to as *concavity, anteriority,* and *sharpness,* respectively, for the remainder of this article.

The first step in this shape parameterization is to align the palate and overall mean tongue shapes of the subjects by their endpoints through rotation, translation, and uniform scaling (as shown in Figure 4). This allows each trace to be

Figure 3. Midsagittal distance functions, taken from the five subjects producing high-front vowels. These five distance functions were also used, in addition to one of uniform diameter, as template vectors in the acoustic simulation experiments, where their palate regions were de-formed according to the three major modes of palate shape variation. aa2, al1, and so forth, are subject identification numbers.

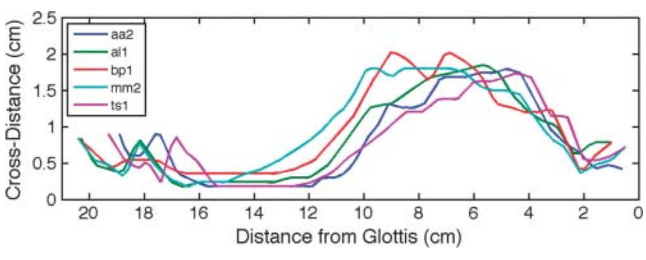


Table 1. Lengths of relevant vocal tract structures for each subject.

Subject ID	Mean vocal tract length (cm)	Palate limits (cm)	Palate length (cm)
aa2	18.9	11.1–15.7	4.6
al1	20.3	12.0–17.0	5.0
bp1	20.4	11.1–17.0	5.9
mm2	20.4	11.5–17.1	5.6
ts1	19.3	10.5–15.4	4.9

Note. Distances were calculated along the midsagittal vocal tract midline in all cases. Limits of the palate are given relative to the position of the glottis and refer to the position of the posterior nasal spine (posterior) and the upper dentition (anterior), respectively.

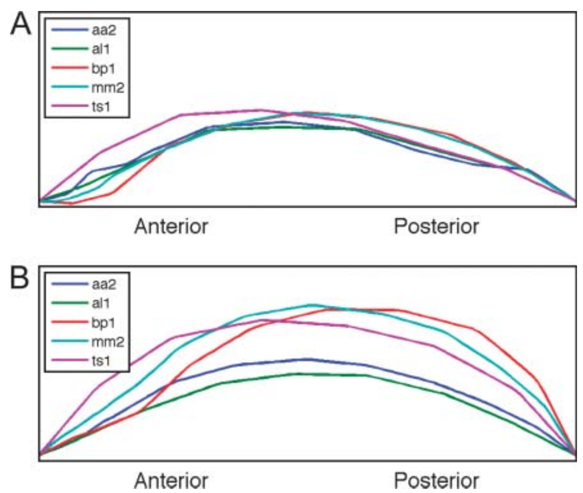


Figure 4. Midsagittal hard palate (Panel A) and tongue shapes (Panel B) of the five subjects, from the upper dentition to the hard–soft palate junction. Hard palates vary substantially in terms of the height of the palatal apex (concavity) and the anterior–posterior position of the apex as well as sharpness of the palate around the apex. Tongue shapes also vary, primarily in terms of concavity and anteriority.

regarded as a single vector of distance measurements, along the line defined by its delimiting points (i.e., the perpendicular distance). The resulting vectors for each speaker i are referred to as $\mathbf{P}_i$ and $\mathbf{T}_i$, respectively. A given speaker's palate trace, $\mathbf{P}_i$, can then be approximated as follows:

$$\hat{\mathbf{P}}_i = c_i\mathbf{C} + a_i\mathbf{A} + s_i\mathbf{S} + \mathbf{P}_\mu, \quad (1)$$

where $\mathbf{P}_\mu$ is the overall mean palate shape of all individuals. The vectors $\mathbf{C}$, $\mathbf{A}$, and $\mathbf{S}$ represent the modes of palate shape variation in terms of palatal concavity, anteriority, and sharpness, respectively; that is, they represent unit deviations in palate shape according to those three modes. The coefficients, c_i, a_i, and s_i reflect the contribution of each mode of shape variation to the specific palate under consideration. These modes are shown in Figure 1, where the coefficients are manipulated independently. It is crucial to note that the coefficients themselves can be used as a low-dimensional parameterization (i.e., three parameters for each individual) of a palate's shape. For a novel and arbitrary palate trace, shape parameterization can be done by projecting individual palate shapes into the vector space defined by the modes of variation. For instance, the coefficient representing concavity, c_i, can be calculated for palate i in the following way:

$$c_i = \mathbf{C}^\mathbf{T}(\mathbf{P}_i - \mathbf{P}_\mu). \quad (2)$$

This same form of projection can be done to obtain a_i or s_i by substituting either $\mathbf{A}$ or $\mathbf{S}$, respectively, for $\mathbf{C}$ in Equation 2. Moreover, this parameterization can be obtained for a speaker's tongue trace in a similar way, as well, by substituting $\mathbf{T}_i$ for $\mathbf{P}_i$. The result of this quantification is three coefficients for each subject that represent the concavity, anteriority, and sharpness of the palate and can be used to represent or reconstruct the palate, as shown in Equation 1.

To ensure that this parameterization did, in fact, accurately represent palate and tongue traces, the proportion of variance explained by application of the parameterization among the subjects was calculated. It was found that 85.75% of palate shape variance was explained by these three modes of variation, which is highly consistent with the findings presented by Lammert et al. (2013). Moreover, it was found that 95.13% of tongue shape variance was explained by the same representation, with 77.83% represented by concavity, 13.45% by anteriority, and 3.86% by sharpness. Note that this high degree of explanatory power for tongue shape is not necessarily expected because these modes were developed to represent palate shapes. The primary objective of parameterizing tongue contours using this method is simply to facilitate a meaningful comparison between palate shape and tongue shape, but this level of explanatory power indicates that the comparison will be relatively complete as well.

To observe the articulatory and acoustic effects of palate shape, it is necessary that the speakers under consideration exhibit substantial variation along the major modes of palatal variation. The palate shape parameter values were used to quantify the amount of variation in the present sample, and the range of variation in the present sample was compared to the range of values in the much larger and more diverse sample analyzed in Lammert et al.'s (2013) study. It was observed that concavity values in the current sample ranged across 18% of the larger sample's range (0.83 *SD*), whereas anteriority ranged across 46% of the range exhibited by that larger sample (2.16 *SD*s), and sharpness ranged across 13% (0.52 *SD*). These data indicate that palate morphology variation in the present sample is sufficiently large to allow examination of its influence on vowel production within the present speaker population.

Simulations

Vocal tract simulations were based on modeling the vocal tract as a series of lossless, cylindrical, concatenated tubes, a method that has been extensively studied with respect to vocal tract modeling and synthesis (Fant, 1960; Kelly & Lochbaum, 1962; Rabiner & Schafer, 1978; Stevens, 1998). Building such an acoustic model can begin by defining a vector, $\mathbf{D}$, representing the midsagittal distance at equally spaced intervals along the length of vocal tract. In this case, the equally spaced intervals—or, equivalently, the lengths of the tubes—were fixed to 2.5 mm in length. Longer vocal tracts were represented by adding more tubes to the model.

For the purposes of acoustic simulation experiments, the tubes can subsequently be de-formed according to the modes of palate shape variation used here for analysis. In particular, the midsagittal distance vector can be represented in the following way:

$$\mathbf{D}_i = \mathbf{M} + \mathbf{F}_i, \quad (3)$$

where **M** is a template vector of midsagittal distances that is fixed for a particular deformation experiment and $\mathbf{F}_i$ is a de-formation vector that can be made to represent arbitrary changes to the template shape. Assuming that the mean palate shape, $\mathbf{P}_\mu$, is already reflected in the template vector, palate shape can be represented as $\hat{\mathbf{P}}_i - \mathbf{P}_\mu$, and the de-formation vector can subsequently be represented as $\mathbf{F}_i = [0, \hat{\mathbf{P}}_i - \mathbf{P}_\mu, 0]$, with zero vectors used to pad the de-formation vector on either side of the palate. In the present experiments, each shape coefficient from Equation 1 was varied independently, while setting the other two equal to zero. This allowed the acoustic impact of each mode of shape variation to be examined individually.

A total of six template vectors were used in the simulation experiments. The first of these vectors simply assumes a uniform vocal tract diameter of 1 cm—that is, $\mathbf{M} = 1$. This uniform template vector further assumes a 17-cm vocal tract, making 68 total tubes from the glottis to the lips, with the hard palate extending 6 cm in length from 1 cm behind the open end of the tube (i.e., the lips). The other five template vectors were based on the midsagittal distance functions of subjects producing high-front vowels, already discussed in the Speech Data subsection in the Results section and shown in Figure 3. These latter five template vectors further utilize the vocal tract lengths and palate positions described in Table 1.

To calculate the resonance frequencies of this multi-tube model, it is necessary to convert the midsagittal distances into cross-sectional areas. This conversion was done according to the following formula:

$$\mathbf{A}_i = \pi(\mathbf{D}_i/2)^2. \tag{4}$$

The assumptions behind this conversion are coarse compared to methods used in recent efforts to accurately model vowel acoustics (see, e.g., Jackson & McGowan, 2012). However, there are at least two reasons to believe that such assumptions are more appropriate for this study. First, because morphology is the focus here, conversion techniques based on information from other speakers with different morphology may confound the simulations with shape information that is not appropriate to the subjects considered in this study. Second, assumptions behind any conversion technique must be simultaneously coherent for both the hard palate and the tongue. Even if certain conversion techniques are more accurate for the overall area function, assuming a similar conversion for the hard palate by itself may not be appropriate because parametric descriptions of the hard palate's three-dimensional morphology are not well known. Therefore, this study remains neutral about three-dimensional hard palate morphology by assuming a uniform projection from midsagittal profile to a normalized acoustic tube model of each speaker's vocal tract.

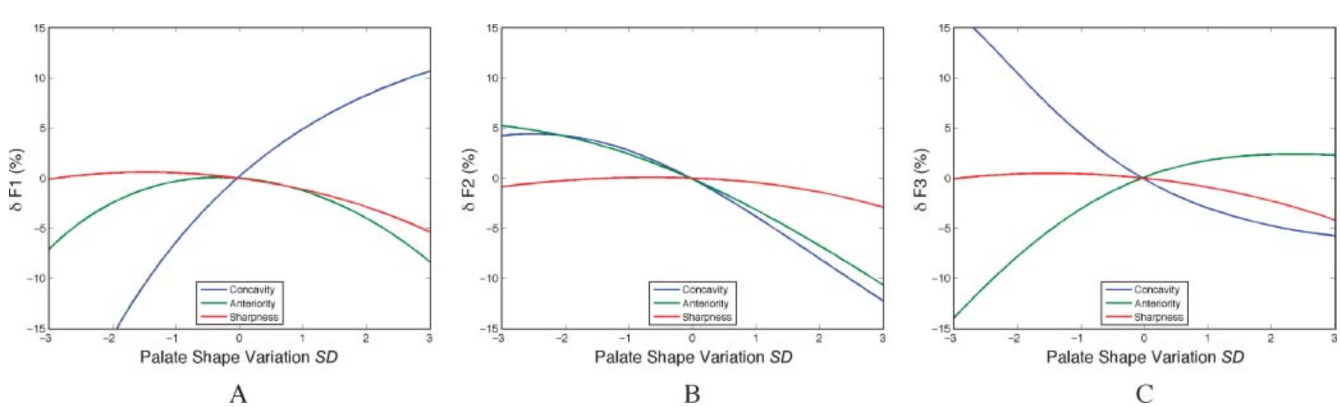
From the area vectors, the formants can easily be computed by first calculating the reflection coefficients between each adjacent tubes: $\Gamma_j = (\alpha_{j+1} - \alpha_j)/(\alpha_{j+1} + \alpha_j)$, where α_j is the value of element j in **A**. Reflection coefficients are then used to compute the coefficients of the prediction filter polynomial. This is done using Levinson's recursion, as described by Kay (1988) and implemented in the MATLAB Signal Processing Toolbox (Version 7.8.0). Finally, the formants can be found by taking the roots of the prediction filter polynomial.

Results

Simulation

The results from acoustic simulations using a uniform template vector are shown in Figure 5. Shown in the figure are the effects on the first three formant frequencies of varying each mode of palatal variation. Palatal variations along the three principal modes are plotted in terms of standard deviations (*SD*s) from the mean shape, where the *SD*s are

Figure 5. The acoustic impact of varying palate shape in a tube of uniform width and area. Palates were varied according to the three major modes of variation: concavity, anteriority, and sharpness. Change in frequencies of the first three formants are plotted as a function of variation along these three modes. Palate shape changes are represented in terms of the *SD*s.



defined over the distributions described in Lammert et al. (2013)—specifically, the following *SD*s: *SD* = 0.219 for concavity, *SD* = 0.154 for anteriority, and *SD* = 0.095 for sharpness. Formant frequencies are plotted in terms of percentage frequency change in Hz from the frequency at the mean shape.

Figure 6 shows the results from acoustic simulation using a template vector representing the high-front vowel posture from the five subjects in this study. Shown in the figure is the mean effect—across all five subjects—of varying each mode of palatal variation on the first three formant frequencies. Palatal variations along the three principal modes are plotted in terms of *SD*s from the mean shape, again using the values described by Lammert et al. (2013). Note that it was not possible to de-form the tube through as large a range when beginning with this template vector because de-formations near the already-narrow constriction in the palatal region quickly lead to negative midsagittal distances. Thus, only ±1 *SD* of de-formation are shown. Formant frequencies are plotted in terms of percentage frequency change in Hz from the frequency at the mean shape, corresponding to the observed high-front vowel shape.

Speech Data

Pearson's correlation coefficients between each mode of palate variation and each of the first three normalized formant frequencies are shown in Table 2. The statistical significance of these values was tested with a two-tailed hypothesis test ($n = 5$) based on Student's t distribution at a significance level of $p = .05$. None of the correlations are significant, and the correlation values are all rather small. The largest observed correlation value is between palatal anteriority and F3, which displays a correlation of –.80 ($p = .10$).

Pearson's correlation coefficients between each mode of palate variation and each mode of tongue shape variation are shown in Table 3. The statistical significance of these values was tested with a two-tailed hypothesis test ($n = 5$) based on Student's t distribution at a significance level of $p = .05$. The results indicated that three correlation coefficients that are largest in magnitude were statistically significant. First, there is a positive correlation between concavity of the palate and concavity of the tongue, and there is also a positive correlation between anteriority of the palate and anteriority of the tongue. In addition, a negative correlation exists between concavity of the tongue and sharpness of the tongue. All other correlations were nonsignificant.

Table 2. Correlation values between palate shapes (P) and formant frequencies (F).

Palate shape	F1	F2	F3
P1: Concavity	0.38	–0.41	–0.19
P2: Anteriority	–0.40	0.09	–0.80
P3: Sharpness	–0.05	0.47	0.53

Note. Palate shapes were parameterized according to the major modes of shape variation. High values reflect systematic relationships between hard palate shape variation and formant frequency variation across subjects. No correlations were statistically significant, as determined using a two-tailed hypothesis test at the $p = .05$ level ($n = 5$).

Discussion

Without any articulatory compensation, variations in palatal concavity and anteriority have the potential to substantially alter the resonant properties of the vocal tract. Simulations indicate that F1 increases with palatal concavity and that F2 and F3 decrease with concavity. By comparing Figure 5 and Figure 6, one can see that the magnitude of this effect is somewhat amplified in a high-front vowel posture when compared to a more uniform vocal tract shape. Any

Figure 6. The acoustic impact of varying palate shape in tubes of nonuniform width, averaged across subjects, corresponding to the articulation of a high-front vowel from the five speakers in this study. Palates were varied according to the three major modes of variation: concavity, anteriority, and sharpness. Change in frequencies of the first three formants are plotted as a function of variation along these three modes. Palate shape changes are represented in terms of the *SD*s.

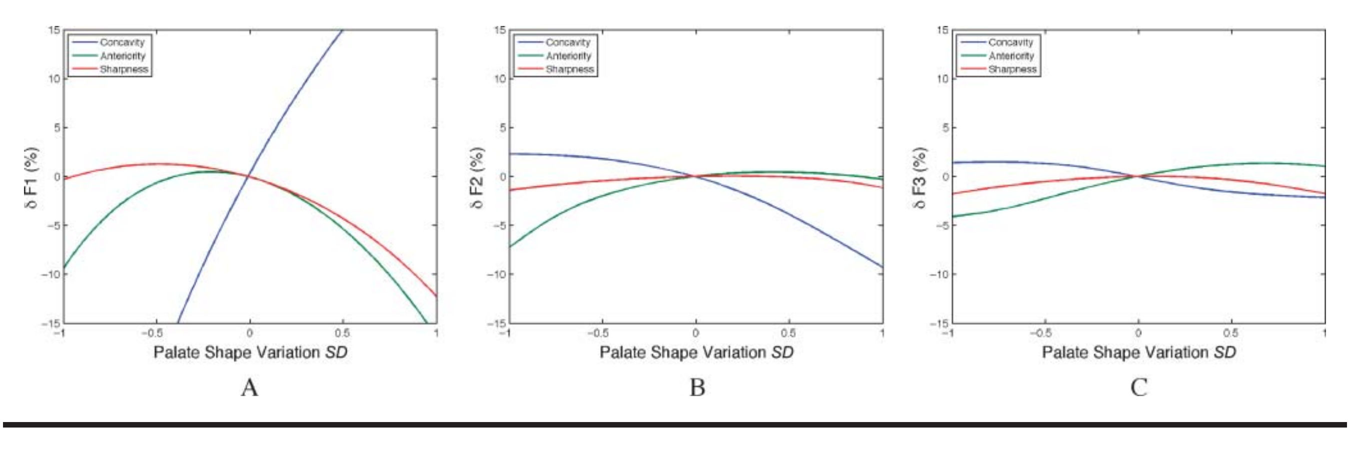


Table 3. Correlation values between palate and tongue (T) shapes, according to the major modes of shape variation (numbered 1 through 3 and labeled at each row).

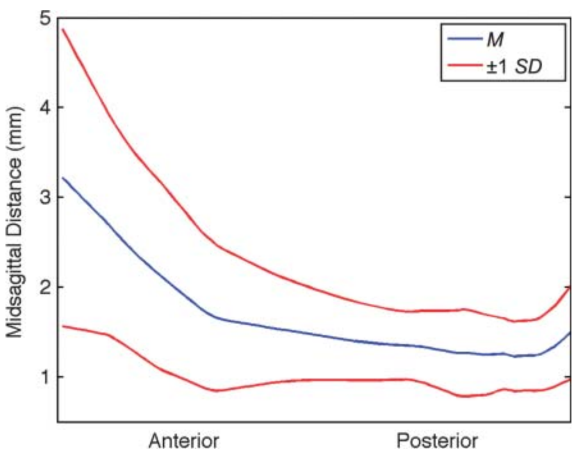
Palate shape	T1	T2	T3
P1: Concavity	**0.96**	–0.03	**–0.95**
P2: Anteriority	–0.25	**0.91**	0.18
P3: Sharpness	–0.55	–0.54	0.57

Note. High values reflect systematic relationships between hard palate shape variation and tongue shape variation across subjects. Numbers in boldface indicate statistical significance of a two-tailed hypothesis test at the p = .05 level (n = 5).

variation in anteriority from the mean palate shape causes F1 to decrease. Anteriority also has a substantial impact on F2 that is posture dependent: F2 is an increasing function of anteriority for high-front vowel postures and a decreasing function for a neutral posture. Furthermore, F3 increases with anteriority, which is nearly the mirror image of concavity's effect. Note that the effect of sharpness is generally marginal. An effect of sharpness can be observed on F1 only in high-front vowel posture and only at extremely sharp palate shapes.

Despite the predicted acoustic impact, differences in palatal morphology are not reflected in the human vowel acoustics in any statistically significant way. This apparent discrepancy would appear to be a direct result of the articulatory compensation observed during vowel production. Subjects appear to adapt their lingual contours to emulate the specific concavity and anteriority of their palates, resulting in midsagittal distance functions for all subjects that are relatively uniform throughout the palate region (see Figure 7). This similarity of vocal tract shapes across subjects leads to similar formant frequencies as well. Note that, even with these

Figure 7. Mean midsagittal distance function across the five subjects, from the upper dentition to the hard–soft palate junction. Standard deviations (±1 *SD*) along each line of the analysis grid are also shown.



findings, it is still not possible to say whether the production goal is to achieve a particular constriction shape or specific resonant characteristics because the two are likely isomorphic in this situation and subjects appear to be doing both in the situation considered.

Deeper insight into the question of goals comes from considering the case of palatal sharpness more closely. Subjects do not emulate palatal sharpness in their tongue shapes; neither are variations in palatal sharpness evident in acoustic variations. These findings are consistent with the assessment that palatal sharpness has marginal potential to affect the acoustics. Furthermore, they indicate that subjects will not account for certain morphological variations in their articulation if those variations do not alter the acoustics, which implies an acoustic production goal. This interpretation is reinforced by the significant negative correlation between palatal concavity and tongue sharpness, which implies that when subjects are faced with the choice between making a concave tongue shape and a sharp tongue shape, they choose the former, which has a more substantial impact on the acoustics.

Conclusions and Future Work

Examining the interplay among variations in articulation, acoustics, and morphology holds promise for explaining interspeaker variability. More comprehensive examination of these complex interrelationships will require detailed knowledge of four things: (a) the variety and extent of morphological variations, (b) the theoretical acoustic impact of these variations, (c) the observed variation in articulation, and (d) the observed variation in acoustics. This work represents a first attempt to leverage recent findings regarding the first point to gain insights into the latter three points with respect to vowel production. The results indicate that, although palatal morphology has the potential to substantially affect vowel acoustics, this effect of palatal morphology may not be observed because speakers appear to adjust their articulations to match the key components of their palatal morphology and minimize potential acoustic variation.

The analysis presented here will be extended further in an ongoing study as more subjects with more diverse vocal tract morphologies are included in the subject pool. Future work will investigate the interplay of other morphological characteristics with speech production behavior, including the posterior pharyngeal wall, which has been parameterized using a similar method to that presented here. Detailing morphological characteristics off the midsagittal plane is also of major importance for the analysis of speech behavior and for acoustic vocal tract simulations. Three-dimensional aspects of morphology are assumed to be particularly important both for analyzing real speech data and for accurately modeling vocal tract acoustics, but they also pose substantial practical challenges, given the current limitations of rtMRI acquisition. In addition, investigations are ongoing regarding the impact of structural variation on other classes of speech sounds as part of a broader research program into the influence of morphological variation on all aspects of speech production. The ultimate goal of this work is to predict individual speakers' production patterns, both articulatory and acoustic, from their morphological characteristics.

Another important avenue of inquiry is inverting the acoustic effect of vocal tract morphology—that is, assessing the extent to which information about a speaker's vocal tract morphology can be recovered from the speech signal. The results of the current study cast doubt on the feasibility of accurately predicting hard palate morphology from vowel acoustics, but they do not entirely eliminate the possibility that accurate statistical methods for prediction may be developed. For instance, the correlation value between palatal anteriority and F3 ($r = –.80$), though nonsignificant, still indicates that 64% of the variance in F3 is explained by palatal anteriority in the current data set (i.e., $r^2 = .64$). Several other correlation values are large enough in magnitude to indicate that they may be useful as well. Many morphological characteristics (e.g., vocal tract length) are inherently more difficult to compensate for with articulatory changes as well, which may facilitate their accurate prediction from the acoustic signal.

Acknowledgments

This publication was made possible by National Institute on Deafness and Other Communication Disorders (NIDCD) Grant R01 DC007124-01. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health/NIDCD.

References

Arens, R., McDonough, J., Corbin, A., Hernandez, M., Maislin, G., Schwab, R., & Pack, A. I. (2002). Linear dimensions of the upper airway structure during development assessment by magnetic resonance imaging. *American Journal of Respiratory and Critical Care Medicine, 165,* 117–122.

Baum, S., & McFarland, D. (1997). The development of speech adaptation to an artificial palate. *The Journal of the Acoustical Society of America, 102,* 2353–2359.

Bresch, E., Nielsen, J., Nayak, K., & Narayanan, S. (2006). Synchronized and noise-robust audio recordings during realtime MRI scans. *The Journal of the Acoustical Society of America, 120,* 1791–1794.

Brunner, J., Fuchs, S., & Perrier, P. (2005). The influence of the palate shape on articulatory token-to-token variability. *ZAS Papers in Linguistics, 42,* 43–67. Retrieved from www.zas.gwz-berlin.de/fileadmin/material/ZASPiL_Volltexte/zp42/zaspil42-fuchs-perrier-brunner.pdf

Brunner, J., Fuchs, S., & Perrier, P. (2009). On the relationship of palate shape and articulatory behavior. *The Journal of the Acoustical Society of America, 125,* 3936–3949.

Brunner, J., Hoole, P., & Perrier, P. (2007). Articulatory optimisation in perturbed vowel articulation. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 497–500). Retrieved from www.icphs2007.de/conference/Papers/1408/1408.pdf

Chiba, T., & Kajiyama, M. (1941). *The vowel—Its nature and structure*. Tokyo, Japan: Kaiseikan.

Dart, S. (1991). Articulatory and acoustic properties of apical and laminal articulations. *Working Papers in Phonetics* (Vol. 79). Los Angeles, CA: UCLA Department of Linguistics.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague, the Netherlands: Mouton.

Fant, G. (1966). A note on vocal tract size factors and non-uniform f-pattern scalings. *Speech Transmission Laboratory Quarterly Progress Status Report, 7*(4), 22–30.

Fant, G. (1975). Non-uniform vowel normalization. *Speech Transmission Laboratory Quarterly Progress Status Report, 16*(2–3), 1–19.

Fitch, W., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 106,* 1511–1522.

Fuchs, S., Winkler, R., & Perrier, P. (2008). Do speakers' vocal tract geometries shape their articulatory vowel space? In *Proceedings of the International Seminar on Speech Production* (pp. 333–336). Retrieved from http://issp2008.loria.fr/Proceedings/PDF/issp2008-77.pdf

Honda, M., Fujino, A., & Kaburagi, T. (2002). Compensatory responses of articulators to unexpected perturbation of the palate shape. *Journal of Phonetics, 30,* 281–302.

Jackson, M.-T., & McGowan, R. (2012). A study of high front vowels with articulatory data and acoustic simulations. *The Journal of the Acoustical Society of America, 131,* 3017–3035.

Kay, S. (1988). *Modern spectral estimation*. Englewood Cliffs, NJ: Prentice Hall.

Kelly, J. L., & Lochbaum, C. C. (1962, September). Speech synthesis. *Proceedings of the Fourth International Congress on Acoustics*, *Copenhagen*, pp. 1–4.

King, E. (1952). A roentgenographic study of pharyngeal growth. *Angle Orthodontist, 22,* 23–37.

Lammert, A., Proctor, M., Katsamanis, A., & Narayanan, S. (2011). Morphological variation in the adult vocal tract: A modeling study of its potential acoustic impact. In *Proceedings of Interspeech* (pp. 2813–2816). Retrieved from http://sail.usc.edu/aigaion2/index.php/publications/show/453

Lammert, A., Proctor, M., & Narayanan, S. (2011). Morphological variation in the adult vocal tract: A study using rtMRI. In *Proceedings of the International Seminar on Speech Production*. (pp. 309–410). Retrieved from http://sail.usc.edu/aigaion2/index.php/publications/show/441

Lammert, A., Proctor, M., & Narayanan, S. (2013). Morphological variation in the adult hard palate and posterior pharyngeal wall. *Journal of Speech, Language, and Hearing Research, 56,* 521–530.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105,* 1455–1468.

Maeda, S. (1979). Un modèle articulatoire de la langue avec des composantes lineaires [An articulatory model of the tongue with linear components]. In *10ème Journées d'Etude sur la Parole* (pp. 1–9). Grenoble, France: Groupement des Acousticiens de Langue Française.

Ménard, L., Schwartz, J.-L., Boë, L.-J., & Aubin, J. (2007). Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics, 35,* 1–19.

Mooshammer, C., Perrier, P., Geng, C., & Pape, D. (2004). An EMMA and EPG study on token-to-token variability. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel, 36,* 47–63.

Narayanan, S., Bresch, E., Ghosh, P., Goldstein, L., Katsamanis, A., Kim, Y., … Zhu, Y. (2011). A multimodal real-time MRI articulatory corpus for speech research. In *Proceedings of INTERSPEECH* (pp. 837–840). Baixas, France: ICSA.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *The Journal of the Acoustical Society of America, 115,* 1771–1776.

Nissen, S., & Fox, R. (2009). Acoustic and spectral patterns in young children's stop consonant productions. *The Journal of the Acoustical Society of America, 126,* 1369–1378.

Nordström, P.-E. (1975). Attempts to simulate female and infant vocal tracts from male area functions. *Speech Transmission Laboratory Quarterly Progress Status Report, 16*(2–3), 20–33.

Öhman, S. (1967). Numerical model of coarticulation. *The Journal of the Acoustical Society of America, 41,* 310–320.

Perkell, J. (1997). Articulatory processes. In W. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 333–370). Cambridge, MA: Blackwell.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *The Journal of the Acoustical Society of America, 24,* 175–184.

Proctor, M., Bone, D., Katsamanis, A., & Narayanan, S. (2010). Rapid semi-automatic segmentation of real-time magnetic resonance images for parametric vocal tract analysis. In *Proceedings of Interspeech 2010*. (pp. 1576–1579). Retrieved from http://sail.usc.edu/dbone/Proctor2010_Interspeech.pdf

Rabiner, L., & Schafer, R. (1978). *Digital processing of speech signals*. Englewood Cliffs, NJ: Prentice Hall.

Stevens, K. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Thibeault, M., Ménard, L., Baum, S., Richard, G., & McFarland, D. (2011). Articulatory and acoustic adaptation to palatal perturbation. *The Journal of the Acoustical Society of America, 129,* 2112–2120.

Vorperian, H., & Kent, R. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research, 50,* 1510–1545.

Vorperian, H., Kent, R., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America, 117,* 338–350.

Vorperian, H., Wang, S., Chung, M., Schimek, E., Durtschi, R., Kent, R., … Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America, 125,* 1666–1678.

Vorperian, H., Wang, S., Schimek, E., Reid, B., Kent, R., Gentry, L., & Chung, M. K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research, 54,* 995–1010.

Winkler, R., Fuchs, S., & Perrier, P. (2006). The relation between differences in vocal tract geometry and articulatory control strategies in the production of French vowels: Evidence from MRI and modelling. In H. C. Yehia, D. Demolin, & R. Laboissière (Eds.), *Proceedings of the International Seminar on Speech Production* (pp. 509–516). Ubatuba, Brazil: CEFALA. Retrieved from http://scholar.google.de/citations?view_op=view_citation&hl=de&user=DL_eBI0AAAAJ&citation_for_view=DL_eBI0AAAAJ:zYLM7Y9cAGgC

Winkler, R., Fuchs, S., Perrier, P., & Tiede, M. (2011). Speaker-specific biomechanical models: From acoustic variability via articulatory variability to the variability of motor commands in selected tongue muscles. In D. Ostry, S. R. Baum, L. Ménard, & V. L. Gracco (Eds.), *Proceedings of the International Seminar on Speech Production* (pp. 219–226). Montréal, Québec, Canada: UQAM. Retrieved from http://scholar.google.de/citations?view_op=view_citation&hl=de&user=DL_eBI0AAAAJ&citation_for_view=DL_eBI0AAAAJ:YsMSGLbcyi4C