EFFECT OF INDIVIDUAL VOCAL TRACT GEOMETRY ON THE TONGUE SHAPING FOR AMERICAN ENGLISH /ɹ/

Yijing Lu, Haley Hsu, Louis Goldstein, Asterios Toutios

University of Southern California
yijinglu@usc.edu, haleyhsu@usc.edu, louisgol@usc.edu, toutios@gmail.com

ABSTRACT

Individual differences in the geometry of vocal tract structures have been found to correlate with inter-speaker articulatory variations. However, there has been a lack of agreement in whether the tongue shaping for American English /ɹ/ is influenced by vocal tract geometry. Different ways to quantify /ɹ/ tongue shapes might have contributed to the divergence of previous findings. The current study compares the results of regressing three different tongue shaping measurements on the same set of vocal tract geometry measurements. Three tongue shaping variables show marginally to moderately significant linear relationships with two specific parameters of vocal tract geometry: the degree of mandibular inclination and the horizontal length of the oral cavity. But overall, there is no strong effect of individual vocal tract geometry on the tongue shaping for /ɹ/.

Keywords: articulation, anatomy, tongue shaping, inter-speaker variability, vocal tract MRI

1. INTRODUCTION

The peripheral morphology of the human vocal tract seems to play a key role in speech production. Rigid structures such as the palate, the posterior pharyngeal wall and the mandible delimits the space in which articulation takes place. Hence, the geometric properties of these vocal tract structures are expected to influence the patterns of deformation of the tongue during speech production. Individual speakers appear to adjust their articulatory strategies according to the morphology of vocal tract structures, possibly in order to generate acoustic consequences that are close enough to be perceived as the same phoneme [1].

A number of studies have examined the relationship between vocal tract geometry and inter-speaker articulatory variability in the production of various speech sounds. For example, the vertical displacement of tongue body from /a/ to /i/ was found to be negatively correlated with the pharyngeal distance (the distance from the anterior nasal spine to the pharyngeal wall) [2] and positively correlated with the pharynx length [3]. The apical and laminal variants of /s/ [4] were found to be partially conditioned by the speaker's palatal height [5, 6] and the slope of the anterior palate [7]. As for American English /ɹ/, which exhibits considerable variation in tongue shaping from "bunched" to "retroflex" [8], different findings have been reached as to whether the tongue shapes are influenced by vocal tract morphology. Westbury and colleagues [9] found no relationship between the /ɹ/ tongue shapes and various measurements of oral cavity size, whereas Dediu and Moisik [10] found that /ɹ/ tongue shapes were influenced by the width and height of the hard palate, the overall size of the mouth, and the prominence of the alveolar ridge. Another study [11] found that speakers using retroflex /ɹ/ and bunched /ɹ/ did not show significant difference in their palate doming degrees.

Given that these previous studies have used different quantitative or qualitative methods to characterize /ɹ/ tongue shapes, these methods may have captured different aspects of the tongue shaping, and therefore may have played a role in the divergent findings. In this paper, we examine the effect of vocal tract morphology on three different measurements of /ɹ/ tongue shapes. To do so, we compare the regression results using different tongue shaping measurements as the dependent variable and the same vocal tract geometry measurements as the independent variables.

2. METHOD

2.1. Data

The data analysed in the present study are a subset of the data in a publicly available multi-speaker MRI corpus [12]. Readers are referred to that work for full technical details on data collection and curation. Data from 32 native speakers of American English from various rhotic-dialect regions of the US (16 female, 16 male; mean age = 26; age range = 18–59) were used. The data consisted of midsagittal RT-MRI speech videos showing the vocal tract movements during two repetitions of /ɑɹɑ/ and high-resolution static anatomical T2-weighted MRI of the upper airway.

2.2. Articulatory analysis

A semi-automatic method [13] was used to segment tissues from airway and track the contours of articulators in the RT-MRI frames. The target frames of /ɹ/ that underwent further analysis were frames in which the distance between the tongue and the palate is the shortest, i.e., the palatal constriction degree presents a minimum.

On the selected /ɹ/ frames, 20 gridlines perpendicular to the line connecting the start and end points of the hard palate were superimposed, intersecting the contours of the palate and the tongue (Fig. 1). The distances between intersection points on the palate and the tongue (red lines in Fig. 1) were measured as the aperture function representing the deformation of the area between the tongue and the palate. Skewness and kurtosis of the aperture function were calculated and used as the first two tongue shape measurements in this study. Since skewness and kurtosis reflect the degrees of asymmetry and tail extremity in the shape of a distribution, these two measurements approximately indicate the location and the length of the constriction formed in the tongue-palate area. A negative skewness means that the shape leans towards the right, therefore representing a relatively posterior constriction location. Conversely, a positive skewness represents a relatively anterior constriction location. A larger kurtosis means a greater amount of deviation and therefore a shorter constriction length. On the other hand, a smaller kurtosis means a longer constriction.

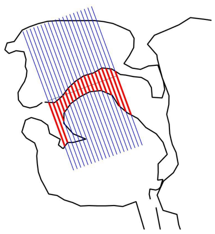

Figure 1: A segmented /ɹ/ frame with 20 superimposed gridlines. The red part of the gridlines constitutes the aperture function that represents the area between the tongue and the hard palate.

Additionally, each /ɹ/ frame was rated as either bunched or retroflex by three human raters. Linear support vector machine (SVM) was used to perform a classification experiment based on the skewness and kurtosis measurements and the human rating results (aggregated using the majority rule). 10 out of 64 /ɹ/ tokens were misclassified. The classification error rate was 15.62%. As shown in Fig. 2, retroflex /ɹ/ and bunched /ɹ/ are largely separated by the decision boundary. The signed distances from each token of /ɹ/ to the decision boundary were calculated as the classification score for each /ɹ/. The classification score is the third measurement of tongue shape in this study, which can be thought of as a measurement combining skewness and kurtosis, indicating a degree of “bunchedness” or “retroflexion”.

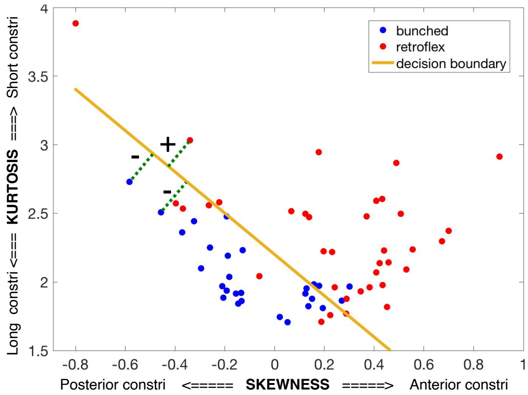

Figure 2: Skewness and kurtosis for each token of /ɹ/, which was colored according to the human rating results. The decision boundary (yellow line) was determined by a linear SVM. The green dotted lines are examples of the distances from /ɹ/ tokens to the decision boundary (classification scores). Tokens above the decision boundary were assigned positive classification scores, and tokens below were assigned negative classification scores.

2.3. Anatomical analysis

Measurements of vocal tract morphology were obtained using the method of [2] to define a geometrical space called Articulatory space (A-space) in the midsagittal vocal tract (Fig. 3). Using the softwere OsiriX Lite v12.0.0 (Pixmeo, Geneva, Switzerland), the anatomical landmarks (i.e., anterior nasal spine, posterior nasal spine, menton) were manually located, based on which the A-space was manually defined for each speaker. The measurements of the geometrical properties of the A-space were directly output from OsiriX. The top border of an A-space is a line that connects the anterior nasal spine (ANS) to the posterior nasal spine (PNS) and eventually stops at the posterior pharyngeal wall. This distance is called pharyngeal distance in [2]. It represents the length of the oral cavity roof. The bottom border is a line parallel to the top boarder, starting from the menton and extending rearward to the posterior pharyngeal wall. The anterior border is the line that connects the ANS to the menton. The posterior border is the outline of the posterior pharyngeal wall between the top and bottom borders. Apart from the four borders, the vertical distance between the top border and the bottom border and the angle between the anterior border and the bottom border were also measured. They respectively represent the lower facial height and the degree of inclination of the mandibular symphysis. Additionally, the ratio of the lower facial height to the top border was calculated, representing the aspect ratio of the oral cavity. In total, seven anatomical measurements were taken for each speaker: lengths of the top, bottom, front, back borders of the A space, the lower facial height (LFH), the aspect ratio (AR), and the angle of the mandibular inclination (MI).

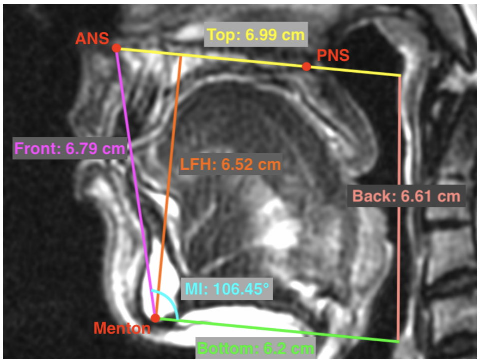

Figure 3: An example of A-space, relevant anatomical landmarks, as well as seven anatomical measurements shown on a T2 weighted midsagittal MRI slice.

3. RESULTS

First, single anatomical features were tested to see if any can form a predictive relationship with either one of the three tongue shape measurements. Simple linear regression models were fitted using each one of the anatomical measurements as the independent variable, and skewness, kurtosis or classification score as the dependent variable. When skewness was the dependent variable, mandibular inclination (MI) was the only predictor yielding a marginally significant relationship (see Fig. 4a, R^2 = 0.05974, $F(1, 62)$ = 3.939, p = 0.0516). When kurtosis was used as the dependent variable, the length of the top border (top), which approximated the horizontal length of the oral cavity, was the only significant predictor (see Fig. 4b, $R^2 = 0.06722$, $F(1, 62) = 4.468$, p = 0.03856). When classification score was used as the dependent variable, mandibular inclination (MI) was the only predictor yielding a marginally significant relationship (see Fig. 4c, R^2 = 0.05873, $F(1, 62) = 3.868$, $p = 0.05369$).

Multiple linear regression models were also fitted to predict each one of the three tongue shape measurements using all the anatomical measurements. None of the models yield significant overall regression. The cumulative R^2 for each model are shown in Table 1.

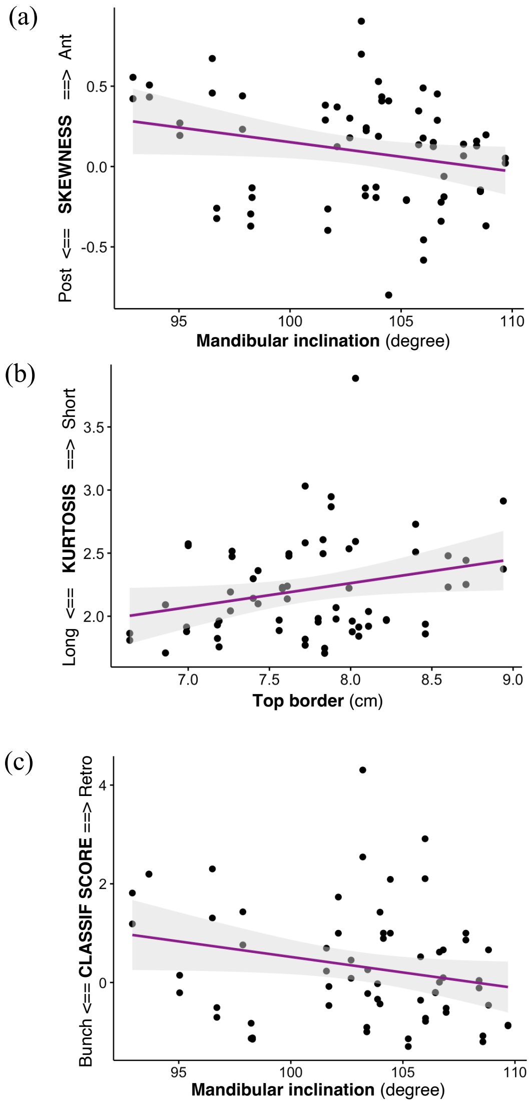

Figure 4: Scatterplots with each dot representing the values of (a) mandibular inclination degree and skewness, (b) top border length and kurtosis (c) mandibular inclination degree and classification score for each token of /ɹ/. The regression line (purple line) and the 95% confidence interval of the regression coefficients (grey area) are superimposed.

DV	R^2
Skewness	0.1033
Kurtosis	0.1532
Classification score	0.1387

Table 1: R^2 for the regression models using different tongue shaping measurements as the dependent variable (DV) and all the anatomical measurements as the independent variables.

4. DISCUSSION

Overall, we did not find a strong effect of vocal tract geometry on the tongue shapes that individual speakers use during the production of American English /ɹ/. For all three quantitative measurements that characterize different aspects of the tongue shape, multiple linear regression models with all the anatomical features as predictors all yield non-significant relationships and low R^2 values (Table 1).

But some marginally significant relationships emerge when certain anatomical features were used individually as the single predictor. Skewness, which characterizes the location of the palatal constriction achieved during the production of /ɹ/, has a weak negative association with the degree of mandibular inclination. A smaller degree of mandibular inclination results in a greater value of skewness, which means a more anterior constriction location. This echoes the findings about /t/ in [2]: English speakers with a smaller degree of mandibular inclination tended to show an apico-dental articulation for /t/, whereas speakers with a relatively larger degree of mandibular inclination tended to use lamino-alveolar /t/. The constriction location for apico-dental /t/ is more anterior compared to lamino-alveolar /t/. For speakers with smaller mandibular inclination, their tongue tends to sit further forward with respect to the palate. Therefore, it is natural for them to make a more anterior coronal constriction.

Kurtosis, which characterizes the length of the palatal constriction, has a positive association with the top border of A-space, which represents the length of the oral cavity roof. A longer oral cavity results in a greater value of kurtosis, which means a shorter constriction length. This is consistent with one of the findings in [10]: the tongue shape for /ɹ/ was significantly related to the length of the anterior mouth. When the mouth is longer, which means the oral cavity is longer, the tongue posture is more likely to have an apical stricture, which means a shorter constriction length. [14] found that for speakers whose palatal constriction degree contributed to F3 lowering more than pharyngeal or labial constriction, they preferred tongue postures with a shorter and more anterior palatal constriction. It is likely that this preference is linked to some morphological feature, i.e., a longer oral cavity. This needs to be further tested.

Classification score, which indicates a degree of "retroflexion" or "bunchedness", was negatively influenced by mandibular inclination. Speakers with a smaller degree of mandibular inclination (prognathic speakers) are more likely to produce "retroflex-flavored" /ɹ/. Prognathic speakers, compared to retrognathic speakers (speakers with a larger degree of mandibular inclination), have a relatively larger sublingual space [2], which may help the elevation of the tongue tip to form a retroflex posture.

In the current study, we characterize the /ɹ/ tongue shapes only using the properties of the palatal constriction. Future work should include measurements of the pharyngeal constriction, which is another factor that can contribute to the difference in tongue shaping. Also, the articulation of /ɹ/ in only one phonetic context was analyzed here. Future research will need to include more phonetic contexts and investigate the interaction between vocal tract morphology and phonetic context in determining the articulatory strategies of individual speakers. If the articulatory variation of /ɹ/ is truly marginally attributable to vocal tract morphology, as suggested by the current study, it is possible that this variation stems from individual differences in the cognitive representations of phonological units [15].

5. REFERENCE

[1] A. Lammert, M. Proctor, and S. Narayanan, "Interspeaker Variability in Hard Palate Morphology and Vowel Production," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 6, pp. 1924–1933, Dec. 2013.

[2] K. Honda, S. Maeda, M. Hashi, J. Dembowski, and J. R. Westbury, "Human palate and related structures: their articulatory consequences," in *Proceeding of 4th International Conference on Spoken Language Processing*, vol. 2. IEEE, 1996, pp. 784–787.

[3] S. Fuchs, R. Winkler, and P. Perrier, "Do speakers' vocal tract geometries shape their articulatory vowel space?" in *Proceeding of 8th International Seminar on Speech Production*, 2008, pp. 333–336.

[4] N. Dart, *Articulatory and Acoustic Properties of Apical and Laminal Articulations*. University of California, Los Angeles, 1991.

[5] M. Stone, S. Rizk, J. Woo, E. Z. Murano, H. Chen, and J. L. Prince, "Frequency of apical and laminal/s/in normal and postglossectomy patients," *Journal of Medical Speech-Language Pathology*, vol. 20, no. 4, 2012.

[6] D. L. Grimm et al., "The Effects of Palate Features and Glossectomy Surgery on /s/ Production," *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 12, pp. 3417–3425, Dec. 2017.

[7] M. Stone, A. D. Gomez, J. Zhuo, A. L. Tchouaga, and J. L. Prince, "Quantifying tongue tip shape in apical and laminal /s/: contributions of palate shape," *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 9, pp. 3149–3159, Sep. 2019.

[8] P. Delattre and D. C. Freeman, "A dialect study of American *r*'s by x-ray motion picture," *Linguistics*, vol. 6, no. 44, 1968.

[9] J. R. Westbury, M. Hashi, and M. J. Lindstrom, "Differences among speakers in lingual articulation for

American English /ɹ/," *Speech Communication*, vol. 26, no. 3, pp.203−226, 1998.

[10] D. Dediu and S. R. Moisik, "Pushes and pulls from below: Anatomical variation, articulation and sound change," *Glossa: a journal of general linguistics*, vol. 4, no. 1, Jan. 2019.

[11] S. Bakst, "Differences in the relationship between palate shape, articulation, and acoustics of American English /r/ and /s/," *UC Berkeley Phonology Lab Annual Report*s, vol. 12, 2016.

[12] Y. Lim et al., "A multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images," *Scientific Data*, vol. 8, no. 1, p. 187, Jul. 2021.

[13] E. Bresch and S. Narayanan, "Region Segmentation in the Frequency Domain Applied to Upper Airway Real-Time Magnetic Resonance Images," *IEEE Transactions on Medical Imaging*, vol. 28, no. 3, pp. 323–338, Mar. 2009.

[14] S. Harper, L. Goldstein, and S. Narayanan, "Variability in individual constriction contributions to third formant values in American English /ɹ/," *The Journal of the Acoustical Society of America*, vol. 147, no. 6, pp. 3905–3916, Jun. 2020.

[15] S. Harper, "Individual differences in phonetic variability and phonological representation," Ph.D. dissertation, University of Southern California, Los Angeles, 2021.