Title

Differences in the relationship between palate shape, articulation, and acoustics of American English /r/ and /s/

Journal

UC Berkeley PhonLab Annual Report, 12(1)

Author

Bakst, Sarah

Publication Date

2016

DOI

10.5070/P7121040732

Differences in the relationship between palate shape, articulation, and acoustics of American English /r/ and /s/

Sarah Bakst

This ultrasound and acoustics study of American English /r/ and /s/ investigates whether variability in production is related to individual differences in vocal tract morphology, namely the shape of the palate.[1] There was reduced articulatory variability for flatter palates for both /r/ and /s/. There was a relationship between acoustics and palate shape for /s/ only, where flatter palates had increased acoustic variability, but for /r/, there was no relationship between palate shape and acoustic variability.

1 Background

Individuals vary in palate size and shape (Vorperian et al., 2005). Brunner et al. (2009) have proposed that there is a relationship between palate shape and articulatory variability. In their electropalatographical study of front vowels, they found that individuals with flatter, less domed palates exhibit less articulatory variability. There is a more quantal relationship between articulation and acoustics for domed palates (Stevens, 1972), which predicts that a given amount of tongue displacement should result in greater acoustic change for flatter palates than for domed palates. Given this relationship, we would expect that a flat palate shape requires greater articulatory precision to remain under a certain threshold of acoustic variability. The experiments presented here test whether the Brunner et al. (2009) finding for front vowels extends to the two consonants /r/ and /s/.

Stevens (1972) wrote that the quantal mapping between articulation and acoustics gives rise to the *invariability* of speech. This hypothesis relies on the idea that regions of the vocal tract vary in their acoustic stability. Stevens hypothesized that language "seeks out" such regions with greater acoustic stability, and from these regions "assembles an inventory of phonetic elements that are used to form the code for communication by language." This hypothesis links the shape of the vocal tract, and the resulting sensitivity of the mapping between articulation and acoustics, with the phonology of a language.

If we assume that the vocal tract does not differ across individuals, then these regions of acoustic stability also should not differ across individuals. This would predict that all languages draw from the same set of phones. The experiment presented here continues a line of investigation into the role of the shape of the vocal tract in speech production. If the size or location of these regions of acoustic stability *do* differ across individuals, it opens the possibility that communities of speakers with similarly-shaped vocal tracts may have a phonetic inventory that is skewed towards regions of stability for those palate shapes. This could potentially explain how physical differences between speakers could influence the phonological organization of different languages and language families.

[1] American English /r/ is known to have a continuum of variants, from retroflex, where the tongue tip points up, to bunched, where the tongue tip points down. No relationship was found between palate shape and whether a participant's /r/ is bunched or retroflex, nor was one necessarily expected.

2 Experimental Methods

2.1 Participants

Twenty-eight native speakers of Californian English, mainly students at the University of California, Berkeley, took part in this study (7 male, 21 female). Participants were compensated $15 for one hour of their time. Four (two female) were excluded from the final analysis due to experimental errors or ultrasound imaging that failed to capture a large portion of the tongue blade.

2.2 Stimuli

Stimuli for the experiment were English CV(C) words that contained /r/ or /s/ in onset or coda. The vowel environment was {a, i, o}. All words were presented in the carrier phrase "I'm a _____." Other words of the same shape containing /ʃ/ or /l/ as the segment of interest were collected but are not analyzed. There was also another set of words that served as fillers for this experiment that contained nasals and final coronal stops. All target stimuli and fillers are shown in Table 1, but only the /s/ and /r/ rows contain words analyzed in this paper.

Table 1: Stimuli

In this paper						
s	sob	sew	sea	boss	dose	piece
	/sab/	/soʊ/	/si/	/bas/	/doʊs/	/pis/
r	rah	Rome	ream	bar	bore	beer
	/ra/	/roʊm/	/rim/	/bar/	/bor/	/bir/
Not analyzed here						
ʃ	shaw	show	she	posh	gauche	quiche
l	lob	lobe	lee	ball	bowl	meal
Fillers						
wV(N)T	wad	Watt	wand	want	wan	
cV(N)T	cad	cat	canned	can't	can	
bV(N)T	bed	bet	bend	bent	Ben	

2.3 Procedure

Ultrasound images were recorded using an Ultrasonix SonixTablet using a C9-5/10 microconvex transducer, operating at 107 fps. Audio was recorded with an AKG 535 EB microphone at a sampling rate of 48kHz and digitized with a Steinberg UR22 USB audio interface. A second audio channel of synchronization pulses (one per ultrasound frame) was sent from the ultrasound to the UR22 and used to time-align the ultrasound frames with the speech audio signal.

A stabilization helmet from Articulate Instruments Ltd was used to hold the ultrasound probe in place under the chin in the midsagittal plane. A teleprompter displayed the test items, and the set of stimuli was randomized and repeated eight times. Participants also swallowed a small amount of water to generate a palate trace.

Each participant also provided a dental cast after the ultrasound session. Participants held a dental tray filled with alginate against the palate and upper dentition. Dental stone was poured into the resulting impression to form a permanent model of the participant's palate.

2.4 Analysis

A script using the Penn Phonetics Lab Forced Aligner used the synchronized audio and the corresponding stimulus file (Yuan and Liberman, 2008) to create Praat (Boersma and Weenink, 2009) TextGrids for each acquisition. The script combed the resulting TextGrids for /r/ and /s/ and extracted the corresponding ultrasound frames using the synchronization file.

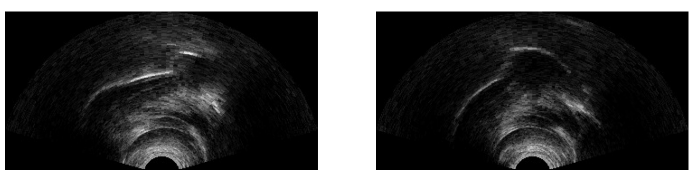
For each utterance, the frame corresponding to the acoustic midpoint of the segment was determined, and subsequent analyses were performed on those frames. The acoustic midpoint was chosen for consistency across tokens.

2.4.1 Measuring articulatory variability

EdgeTrak (Li et al., 2005) was used to trace the tongue contours in the selected ultrasound frames for four subjects (102, 103, 106, and 109). These traces were then visualized using the SSANOVA package (Gu, 2002; Davidson, 2006) in R and were used to make initial observations visually in order to corroborate and validate the subsequent principal components analysis (PCA)[2], which was then used to analyze the rest of the data. PCA was performed by-subject and by-phoneme.

Data from the first principal component (PC1), which by definition accounts for the majority of the variation in the data, was generally used. The exception was for /r/: all subjects who had a retroflex /r/ in some contexts also had a bunched /r/ in others. Therefore all subjects had a bunched /r/ in some context. A comparison between retroflex and bunched /r/ variants would likely result in an artificially large value of articulatory variability for those subjects who had both variants. Therefore, in order to maintain comparability across subjects, only variability in bunched /r/ is investigated in this study. For the few subjects who did have retroflex /r/ in some contexts, PC1 always separated out bunched from retroflex /r/, and PC2 showed differences within bunched /r/. For these subjects, further analysis was carried out using PC2.

Representative midpoint frames showing the difference in retroflex and bunched /r/ for subject 103 are shown in Figure 1a and 1b as examples. A retroflex /r/ is usually marked by an abrupt discontinuity in the tongue contour where the tongue tip is flipped up, whereas a bunched /r/ generally shows a smooth hill-like contour.

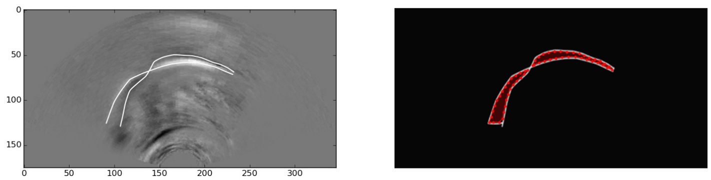


(a) 103 retroflex /r/ (b) 103 bunched /r/

Figure 1: Representative frames from 103's retroflex /r/ and bunched /r/. The bright white line in the middle of the fan is the air above the tongue, which for all intents and purposes here can be interpreted as the top surface of the tongue. Posterior is to the left and anterior is to the right.

Because PCA was performed over ultrasound images, the principal components themselves can also be represented as images, where the brightness of each pixel indicates how much that pixel loads on a PC, and lightness or darkness indicates positive or negative value for that PC. For example, a very white pixel indicates a strong positive loading for that PC, black a strong negative loading, and gray a loading closer to zero. In each image are two tongue contours that correspond

[2]PCA is a dimensionality reduction analysis technique for data such as numbers or images that finds components along which variance in the data can be accounted for.



(a) Tongue contours traced in Pixelmator

(b) Area computed in GraphClick

Figure 2: Contours corresponding to the highest positive and negative loadings of PC1 were traced, and then the area bound by the two curves was computed.

to strong positive and strong negative loading on that PC. The image representing the loadings for PC1 were loaded into the graphics program Pixelmator, and the two tongue contours were traced using an external tablet. The perimeter of the traces was automatically detected using the program GraphClick, and the area between the curves was calculated in pixels. The area was divided by the length of the tongue that was visible in order to account for individual differences in imageability. The resulting value was used as a metric to quantify individual variability in tongue shape. An example of a trace and the resulting area calculation is shown in Figures 2a and 2b.

2.4.2 Measuring acoustic variability

The acoustic measures for /r/ and /s/ were F3 and spectral peak, respectively, taken at the midpoint of the segment to match the point at which the ultrasound frame was chosen. A Praat script extracted F3, and a Python script calculated the spectral peak. The standard deviations of these measures were calculated over all of the utterances for each subject. These standard deviations were used as a measure of acoustic variability.

2.4.3 Measuring palate doming

Palates were measured following the method described in Brunner et al. (2009). They describe a value α that is correlated with the curvature of the palate and can be used to compare palate shapes. This value is calculated as in (1) by approximating a slice of the palate in the coronal plane as a parabola (shown in Figure 3).

(1) $$\alpha = \frac{4}{3\sqrt{|a|}}$$

In (1), a is the coefficient of the x^2 term of the equation describing the parabola, where half the width of the palate is the x-intercept of the parabola, and the depth of the palate is the distance between the x-axis and the vertex of the parabola. The α term is thus a kind of ratio of width to depth. The value of α is inversely correlated with domedness, so values closer to 3 are flatter, while values closer to 1 are more domed.

2.5 Results

2.5.1 Qualitative results

Domedness of the palate was not a significant predictor of retroflexion or bunching ($p = 0.09$). Of the 26 subjects included in the qualitative analysis (two excluded for experimental errors that precluded further quantitative analysis), seven had a retroflex /r/ in some context. The mean

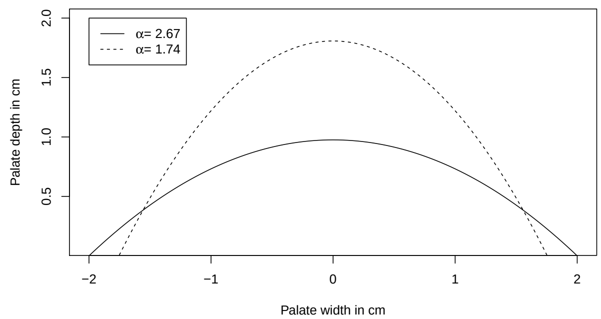


Figure 3: Approximation of the palate as a parabola for flat and domed palate shapes taken from participants in this study. For the derivation of α, see Brunner et al. (2009, p. 3948)

alpha values for speakers with retroflex and bunched /r/ were similar (2.09 and 2.07, respectively). Domedness also not a predictor of /s/ articulation ($p = 0.23$). The split between apical and laminal variants was more even; 11 speakers had apical articulations. Mean alpha values for apical and laminal /s/ were also similar (2.09 and 2.06) respectively. Therefore, the data in this study do not support the idea that palate shape is a significant predictor of articulatory variant.

2.5.2 Quantitative results

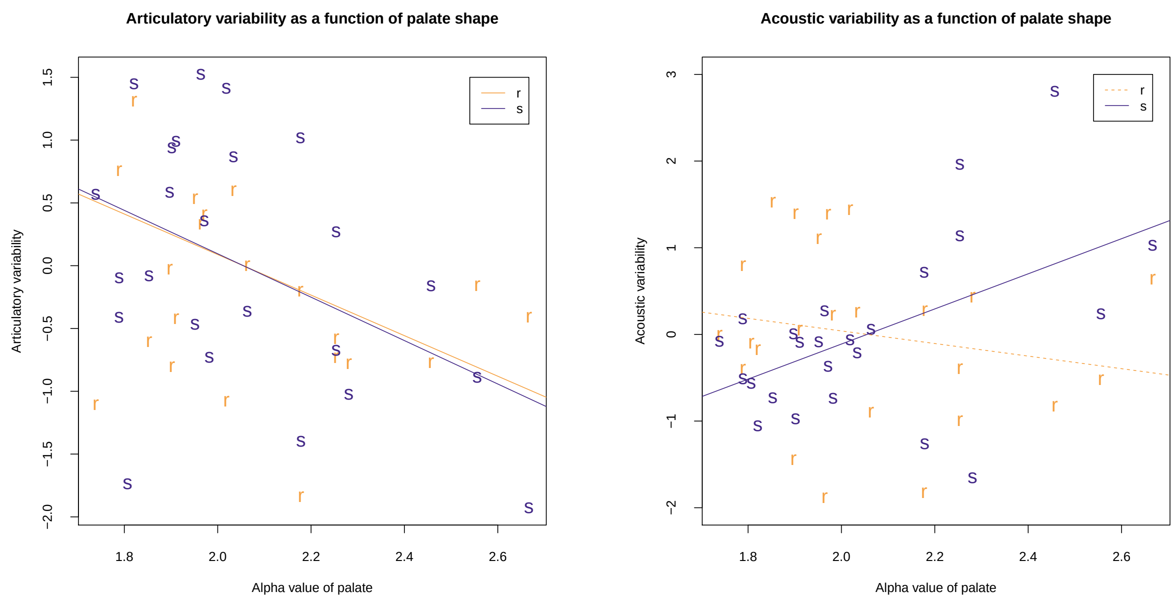
Figure 4 shows articulatory variability plotted against palate shape in (a) and acoustic variability plotted against palate shape in (b). The relationship between articulatory variability and palate shape was significant for both /r/ ($r = -0.43, p = 0.03$) and /s/ ($r = -0.41, p < 0.05$). The relationship between acoustics and palate shape, though, is significant only for /s/ ($r = 0.50, p = 0.01$). No relationship was found between acoustics and palate shape for /r/ ($r = -0.18, p = 0.39$). There was also no relationship between articulatory and acoustic variability for either /r/ or /s/. A summary of the results of the relationships between palate shape and variability in articulation and acoustics for /r/ and /s/ are summarized in Table 2.

The observations above reflect group dynamics, but there is substantial individual variation. A hierarchical clustering analysis was performed on the data to identify categories of speakers. To help understand the interaction of the different variables in forming these clusters, multidimensional scaling (MDS) analysis was performed. The four factors (articulatory and acoustic standard deviation data for both /r/ and /s/) were reduced to two. The palate data was intentionally not included in the MDS analysis in order to test how well it correlated with the aggregate production data.

Dimension 1 of the MDS analysis was strongly negatively correlated with variability in spectral peak for /s/ ($r = -0.87, p < 0.001$) but very positively with variability in F3 for /r/ ($r = 0.73, p < 0.001$). Dimension 2 was strongly correlated with /s/ articulatory variability ($r = 0.87, p < 0.001$), but there was virtually no relationship with articulatory variability in /r/ ($r = 0.23, p = 0.27$),

	Articulation-palate	**Acoustics-palate**	**Articulation-acoustics**
/r/	$\mathbf{r = -0.43, p = 0.03}$*	$r = -0.18, p = 0.39$	$r = -0.12, p = 0.57$
/s/	$\mathbf{r = -0.41, p < 0.05}$*	$\mathbf{r = 0.50, p = 0.01}$*	$r = -0.20, p = 0.33$

Table 2: Summary of the significant relationships found in the data. Significance indicated by "*".



(a) /r/: $r = -0.43, p = 0.03$
/s/: $r = -0.41, p < 0.05$

(b) /r/: $r = -0.18, p = 0.39$
/s/: $r = 0.50, p = 0.01$

Figure 4: Figure 4a shows articulatory variability graphed against palate shape for /r/ and /s/, and 4b shows the same for acoustic variability. All measures have been z-scored to allow for comparison across segment and variability type. Solid lines indicate significant correlations.

possibly due to the proportion of participants with domed palates and low articulatory variability. Although palate shape was not included in the MDS analysis, both dimensions significantly correlated with α value (Dimension 1: $r = -0.43, p = 0.04$, Dimension 2: $r = -0.49, p = 0.01$).

Several groups emerge (shown in Figure 5). The two larger groups (Group A and Group B in the figure) reflect the main ways in which people differ. Group A had low acoustic variability for /r/, though the source of this consistency was different for different members of the group: some had low articulatory variability, and others had very domed palates. The members of Group B were marked by their high articulatory variability in /r/ and low /s/ acoustic variability, and all had average to domed palates. The lower left group containing 126 and 104 had fairly flat palates and correspondingly high acoustic variability for /s/ despite low articulatory variability. The group containing 118, 121, and 125 had fairly domed palates and high articulatory variability for /s/, but their acoustic variability was average. In contrast, the group in the lower right (113 and 127) had very low articulatory variability despite not having particularly flat palates. In a sense, this group's articulation was more consistent than such a palate shape would predict to reach a particular level of acoustic consistency.

3 Discussion

Brunner et al. (2009) hypothesized that the motivation for individual differences in articulatory variability is that all speakers aim to minimize acoustic variability. Individuals may differ in the *value* of the F3 of their /r/ or the spectral peak of their /s/, but the standard deviations in these values are probably going to be relatively consistent from one individual to another. If there is

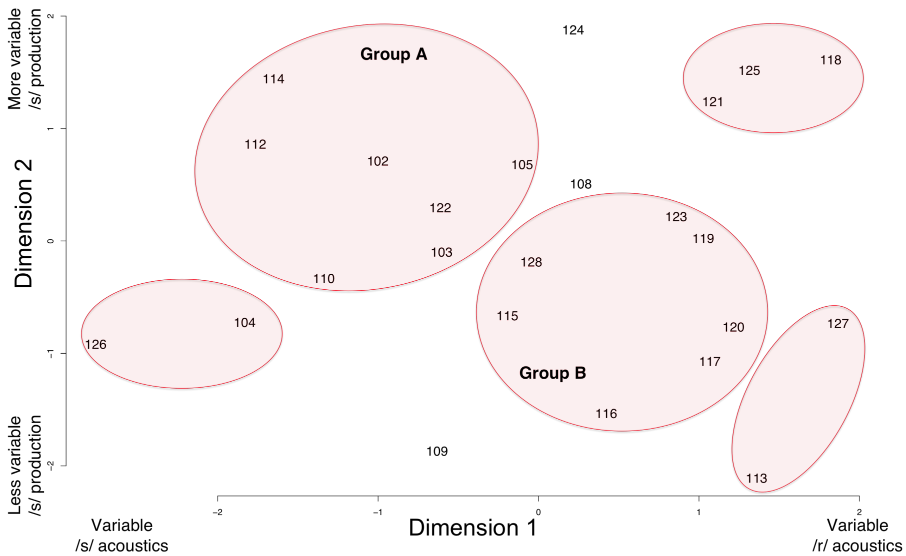


Figure 5: Participants grouped by cluster analysis performed on the data, plotted here on the resulting dimensions from MDS analysis. Dimension 1 is highly negatively correlated with variability in spectral peak for /s/ ($r = -0.87, p < 0.001$) but very positively with variability in F3 for /r/ ($r = 0.73, p < 0.001$). Dimension 2 is strongly correlated with /s/ articulatory variability ($r = 0.87, p < 0.001$).

a more linear mapping between articulation and acoustics for vocal tracts with flat palates than domed palates, then vocal tracts with flatter palates will have smaller regions of acoustic stability. Speakers with such vocal tracts must reduce their articulatory variability if they wish to stay within (or not deviate far from) such a region of acoustic stability. Therefore, we expect that while variability in articulation may change in accordance with palate shape, there might not be a relationship between articulatory and acoustic variability. This explains the main trends that we see in the data: there is a relationship between articulatory variability and palate shape, but no relationship between articulatory and acoustic variability.

Although there was significantly less articulatory variability for flat palates for both /r/ and /s/, the pattern differs between the two segments. For /r/, flatter palates are articulatorily highly consistent, but within the domed group (where α is under 2.0) there is less of a relationship between articulatory and acoustic variability, and there are more participants like 105, who have reduced articulatory variability despite a more domed palate. In other words, while a flat palate predicts articulatory consistency, a domed palate does *not* necessarily predict increased variability in the articulation of /r/. For /s/, the relationship is more linear, so that even within the domed group, there is a strong relationship between palate domedness and acoustic variability. The trend is therefore the same for both segments, but the extent to which the palate shape has an effect on acoustics for all speakers depends on the segment. While palate shape might be one interfering force that results in individual differences in articulatory variability, there are other factors that have different influences over the two segments. For some participants with domed palates (e.g. 121) who have high articulatory variability for /r/, it is possible that there is not a way to reduce articulation to the same degree for /s/ while also maintaining the constriction required for turbulent airflow. This possibility is corroborated by the fact that there is no correlation between an individual's articulatory variability in /r/ and /s/. While palate shape certainly influences articulatory and acoustic variability, on an individual level there are many other possible factors yet to be investigated that might also play important roles.

Despite the decrease in articulatory variability, there is greater acoustic variability in /s/ for participants with flatter palates. The difference in the significance of the relationships between palate shape, articulatory variability, and acoustic variability for /r/ and /s/ suggests that consistency in the acoustic goals are not equally attainable across individuals and phones. From the MDS analysis, it seems that there are different degrees of attainability for this consistency. While people with flatter palates may increase their articulatory precision in order to maintain acoustic consistency, their palate shape is a limiting factor in how much acoustic consistency they can achieve.

Individuals are also likely to differ in how much variation they are willing to tolerate in their own acoustics and articulation. Although the main motivating factor seems to be maintaining acoustic consistency, it is possible that some speakers weight their somatosensory feedback more strongly than others (Lametti et al. 2012). This might explain both why some some speakers are articulatorily very consistent despite having domed palates (e.g. participant 105), and it also might explain the small differences in acoustic consistency in /r/. Speakers may also differ in their sensitivity to certain acoustic cues, or in the knowledge of their own articulatory-acoustic mapping.

4 Conclusion

Palate shape does play a significant role in individual differences in articulatory and acoustic variability, but the weight of this role depends on the segment. For /r/, there was relative consistency within-individual in acoustics, but this did not extend to /s/. Acoustic variability did not correlate with articulatory variability for either segment, which is to be expected if speakers' articulatory variability is related to consistency in acoustics. For /s/, there was a strong relationship between both types of variability and palate shape, but for /r/ there was a strong relationship between palate shape and articulatory but not acoustic variability. There may be a difference in attainability of acoustic goals. For /r/, speakers seem to be adjusting their articulations in order to maintain a relatively consistent F3. The spectral peak of /s/ is much less consistent than the F3 for /r/. It is possible that the acoustics for /s/ are more strongly influenced by the shape of the palate than /r/, so acoustic consistency is more easily attained for /r/ than /s/.

The results presented here provide part of the answer to how morphological differences between individuals may influence speech production. The domedness of the hard palate is one such anatomical factor, but many other factors, such as the size of the oral cavity or the size and range of motion of the tongue, are likely to exert their own influences. Further research needs to take more of these factors into consideration in order to understand how differences in vocal tract morphology affect production.

References

Bladon, R.A.W., and F. Nolan. 1977. A video-flurographic investigation of tip and blade alveolars in English. *Journal of Phonetics* 5:185–193.

Boersma, Paul, and David Weenink. 2009. Praat: doing phonetics by computer (Version 5.1.05)[computer program].

Brunner, Jana, Susanne Fuchs, and Pascal Perrier. 2009. On the relationship between palate shape and articulatory behavior. *Journal of the Acoustical Society of America* 125:3936–3949.

Dailide, Saulius, and Aidas Dailide. 2014. Pixelmator. Version 3.2 Sandstone.

Dart, Sarah N. 1991. Articulatory and acoustic properties of apical and laminal articulations. Doctoral Dissertation, UCLA.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120:407–415.

Delattre, Pierre, and Donald C. Freeman. 1968. A dialect study of American r's by X-ray motion picture. *Linguistics* 44:29–68.

Gu, Chong. 2002. *Smoothing spline ANOVA models*. Springer Series in Statistics. New York: Springer-Verlag.

Inc., Arizona Software. 2010. Graphclick. http://www.arizona-software.ch/graphclick/.

Johnson, Keith. 2011. Retroflex versus bunched [r] in compensation for coarticulation. *UC Berkeley Phonology Lab Annual Report* .

Lametti, Daniel R., Sazzad M. Nasir, and David J. Ostry. 2012. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience* 32:9351–9358.

Lammert, Adam, Michael Proctor, and Shrikanth Narayanan. 2013. Morphological variation in the adult hard palate and posterior pharyngeal wall. *Journal of Speech, Language, and Hearing Research* 56:521–530.

Lawson, Eleanor, James M. Scobbie, and Jane Stuart-Smith. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics* 15:256–268.

Li, Min, Chandra Kambhamettu, and Maureen Stone. 2005. Automatic contour tracking in ultrasound images. *International Journal of Clinical Linguistics and Phonetics* 19:545–554.

Mielke, Jeff, Adam Baker, and Diana Archangeli. 2010. Variability and homogeneity in American English /r/ allophony and /s/ retraction. *Laboratory Phonology* 10:699–729.

Stevens, Kenneth N. 1972. The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data. In *Human Communication: A Unified View.*, ed. Edward E. David and Peter B. Denes, 51–66. New York: McGraw-Hill.

Stevens, Kenneth N., and Sheila E. Blumstein. 1975. Quantal aspects of consonant production and perception: a study of retroflex stop consonants. *Journal of Phonetics* 3:215–233.

Twist, Alina, Adam Baker, Jeff Mielke, and Diana Archangeli. 2007. Are 'covert' /ɹ/ allophones really indistinguishable? *University of Pennsylvania Working Papers in Linguistics* 13.

Vorperian, Houri K., Ray D. Kent, Mary J. Lindstrom, Cliff M. Kalina, Lindell R. Gentry, and Brian S. Yandell. 2005. Development of vocal tract length during early childhood: A magnetic resonance imaging study. *Journal of the Acoustical Society of America* 117:338–350.

Yuan, J., and M. Liberman. 2008. Speaker identification on the scotus corpus. *Journal of the Acoustical Society of America* 123.

Zhou, Xinhul, Carol Y. Espy-Wilson, Suzanne Boyce, Mark Tiede, Christy Holland, and Ann Choe. 2008. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *Journal of the Acoustical Society of America* 123:4466–4481.