Modeling the effect of palate shape on the articulatory-acoustics mapping

Sarah Bakst and Keith Johnson[1]

[1]Department of Linguistics, University of California, Berkeley
`bakst@berkeley.edu`

Abstract

Previous research [1, 2] shows that articulatory variability is reduced for people with flatter palates. It has been hypothesized [1] that this is because the mapping between articulation and acoustics is more linear for flatter than for more domed palates. A combination of two synthesizers were used to model how vocal tract anatomy influences the mapping of articulation onto acoustics, using American English /r/ as a test case. A retroflexable tongue tip was added to the articulatory parameters. Two additional palate shapes and a sublingual cavity that appears during /r/ production were also added to the synthesizer. A Python script searched the articulatory-acoustic space for vocal tract configurations that resulted in a low F3 (the hallmark acoustic cue for /r/) for each palate. Palate shape influences not only the overall sensitivity of the articulatory-acoustic mapping, but also the effect of each individual articulatory parameter on F3.

1. Introduction

1.1. Background

This research tests a hypothesis that explains how vocal tract anatomy influences variability in speech production. One segment that is known to vary both within and across individuals is the American English rhotic /r/. This segment can be either retroflex, bunched, or in between [3], and individuals may have more than one type of /r/, typically varying by phonological context [4]. This type of variation is an example of the many-to-one mapping between articulation and acoustics: multiple qualitatively different articulatory configurations may exist that will result in similar if not identical acoustics [5].

Even within one variant (such as bunched or retroflex) there may be variability. This is because the mapping between articulation and acoustics is also quantal [6], meaning that for some regions of the vocal tract, a given difference in articulation will have a greater effect on acoustics than the same difference in a different region. Some regions of the vocal tract have relatively stable acoustic regions. Between regions of stability might be regions where small differences in articulation result in comparatively large changes in acoustics.

The degree of nonlinearity may not be the same for all individuals. In their electropalatographical study of front vowels, [1] found that people with flatter palates exhibit less articulatory variability than people with more domed palates. They hypothesize that this is because the mapping between articulation and acoustics is more linear for flatter palates, but for domed palates, there are greater regions of the articulatory space that are acoustically stable. Assuming speakers aim to maintain a degree of acoustic consistency, speakers with flatter palates must be more precise in their articulations than speakers with more domed palates.

1.2. Questions and Hypotheses

This study seeks to answer two main questions about the role of palate shape in variability in speech production. First, the modeling broadly examines how the mapping between articulation and acoustics varies by exploring the F3 acoustic space for different articulatory configurations for the different palate shapes. Second, the modeling assesses differences in the influence of various articulatory parameters on F3 for these different palate shapes. Specifically, this study tests the hypothesis in [1] that the increased articulatory precision observed in people with flatter palates is a result of a more linear mapping between articulation and acoustics for such palates. This hypothesis predicts a greater range of F3 space for flatter palates and a stronger relationship between each articulatory parameter and F3.

2. Methods

The hypothesis is that the shape of vocal tract (in particular the hard palate) plays an important role in the quality and variability of articulation in production. The work specifically considers whether, how, and to what extent articulatory configurations might differ in producing a low F3 for different palate shapes.

The original intent of the Maeda synthesizer [7] was to model French vowels and includes a single palate based off a real speaker. The articulatory parameters are principal components based off of X-ray data from this speaker. There is an apex position parameter that controls the proximity of the apex to the palate, but this principal component also affects the tongue root. Because French vowels do not typically include retroflex tongue configurations, we created a new tip-curling parameter, which controls the orientation of the tongue tip only. We also created two new palates, one flatter and one much more domed than the default. We tested each of these palates with a spectrum of tongue shapes. Figure 1 shows the implementation of all three palate shapes and the tongue tip parameter.

The active articulators are driven by the user of the synthesizer. There are a number of parameters, such as the shape and position of the tongue dorsum or the protrusion and aperture of the lips, that the user adjusts to make different phones. The user manipulates the shape of the vocal tract by indicating a setting for each articulator. This setting is a multiplier for the principal component that represents the articulatory parameter. The four active articulator parameters considered here are the dorsum position and shape, the protrusion of the lips, and the orientation of the tip (whether it is pointing up or down), which represents an important difference in retroflex versus bunched articulations.

The Maeda synthesizer models the vocal tract as a series of cross-sectional areas. To calculate the area from the width in the sagittal plane we have to assume something about the shape of the tract at that point (if the vocal tract is a cylinder then the cross-sectional area is $A(x) = \pi r^2$, if the vocal tract

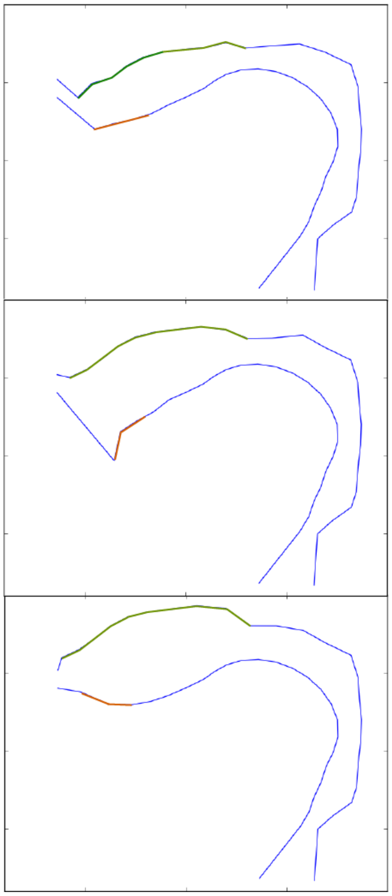

Figure 1: The three different palate shapes (from top: flat, default, domed) and examples of tongue tip configurations: neutral, extreme tip-down, and extreme tip-up.

is a square the cross-sectional area is $A(x) = d^2$, etc.). Within the oral cavity (as opposed to at the lips or in the pharynx) the Maeda synthesizer assumes that cross-sectional area is a function of palate doming, so that the cross-sectional area *A(x)* at a point with a given width in the sagittal plane *x* is calculated based on the formula in (1), where and were determined from real production data from a single speaker and hard-coded into the original model. These α values are a special ratio of the width and depth of the palate and actually correspond to the same metric of domedness as in [1].

$$A(x) = \alpha x^{\beta} \tag{1}$$

The α values for the sections corresponding to the hard palate were set to empirically-derived values for the domed and flat palates based on [2] in order to reflect realistic differences in cross-sectional area.

We used the Maeda model in conjunction with the Manzara tube model[1]. The Manzara model is a series of tubes of varying widths that are joined together. We added a short tube as a side-branch (after [8]) to model the sublingual cavity that emerges in /r/ production.

The model was run with each of the three palates. We wrote a program which cycled over the range of settings for each of the four articulatory parameters that would result in a possible articulatory configuration for /r/. The script produced cross-sectional area values, which were used as inputs for the Manzara tube synthesizer [9]. The script also called a program to perform acoustic analysis [10] over the synthesized output and rejected tokens that were silent or not speech-like based on their RMS amplitude (amplitude $<$ 1200). The program also recorded F2 and F3 measurements from the midpoint of the sound file.

	Domed	Default	Flat
α	1.3	(1.7)	2.7
Minimum F3	1789	1371	1578
Maximum F3	2789	2713	3428
Min F2	788	704	684
Max F2	2252	2227	2035

Table 1: F2 and F3 ranges for each palate shape, considering all articulations.

3. Results

A summary of the results is in Table 1. The flattest palate has the widest range of F3 values, suggesting that articulatory-acoustic mapping may indeed be more sensitive for a flatter palate than a more domed palate, given that the same range of articulation was used for all palates.

Figure 2 shows the spread of F3 values for each palate. The generated sound files were sorted by F3 value; the spread (not value) along the x-axis corresponds to the number of sound files generated at a given F3 value. The closer to zero the slope is, the greater the region of acoustic stability, and the less sensitive the mapping between articulation and acoustics. There is a greater range of values for the flat palate, less so for the default palate, and the smallest range for the domed palate. The overall acoustic flexibility is similar for domed and default palates; for much of the graph, the slope of the line is shallow. This indicates a large region of acoustic stability, where many articulatory configurations can result in similar if not identical acoustics. In contrast, the flattest palate has the steepest slope in this acoustic region, indicating the least acoustic stability for this palate.

While the F3 values reported in Figure 2 all come from articulations that might have hypothetically produced an /r/, some of these values are far too high to correspond with a phone that could be perceived as /r/. If we restrict our view to only those files produced with F3 values under 2300Hz, which is a reasonable cutoff for an /r/, there is less stability for the domed and default palates, but still more than for the flattest palate.

	Tip curl	Backing	Bunching	Lip
flat	0.15*	-0.78**	-0.56**	-0.06 (n.s.)
regular	-0.05 (n.s.)	-0.7**	-0.33**	0.15*
domed	-0.31**	-0.45**	0.36**	0.15*

Table 2: Correlations between articulators and F3 for each palate. *$p < 0.05$, **$p < 0.001$

Table 2 shows the correlation of each parameter with F3 for each palate. For all palates, the position of the tongue dorsum had a greater lowering effect on F3 than any other articulator, while lip rounding had minimal if any effect. The shape of the dorsum (bunching) had a surprising effect: for flat and

---

[1] We used two synthesizers: the Maeda synthesizer is more faithful to articulation, and the Manzara synthesizer produces files of better sound quality.

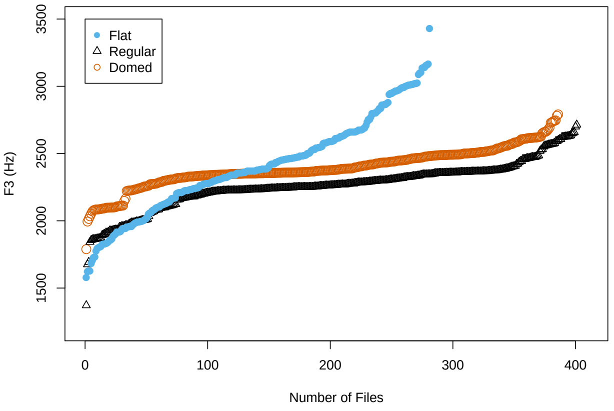

Figure 2: The F3 range for all three palates in increasing order of F3. The steady region for the flat palate is smallest and it is largest for the most domed palate.

regular palates, bunching of the tongue lowered F3, but for the domed palate, bunching actually raised F3. Conversely, raising the tongue tip slightly raised F3 for the flat palate, had almost no effect for the regular palate, and significantly lowered F3 for the domed palate. This result suggests that a flat palate would favor a bunched /r/, and a domed palate might favor a retroflex /r/. This relationship is weak, though: for all three palates, there was a wide range in F3 values for different settings of the tongue tip. Figure 3 shows that, given the other three articulatory factors tested here, a wide range of F3 values is possible for any setting of tip orientation.

For the flat palate, dorsum position and dorsum shape both had a great effect on F3. This supports the hypothesis that for a flatter palate, changes in articulation will generally have a greater effect on acoustics than for more domed palates. This holds true for all articulatory factors except retroflexion, where there is a much stronger relationship with acoustics for the most domed palate than for the other palates.

Figure 3 shows the articulatory settings for the five lowest F3 settings for each palate. For the first three parameters, settings range from 0 to 4, and for retroflexion, from -4 (tip down) to 4 (tip up). For all three palates, a low F3 was achieved primarily by retracting the tongue as much as possible, and secondarily by bunching the tongue. Lip protrusion appears to have no discernible effect in this table, even though lip protrusion is significantly correlated with F3 for regular and domed palates. Only the flat palate has a consistent pattern for the tongue tip; all five articulations for this palate have the tip either at a neutral or downward orientation.

4. Discussion

The modeling here suggests that the vocal tract sensitivity function is related to the flatness of the palate: the flatter the palate, the more acoustics are affected by a change in articulation. This was shown both in the range of F3 values produced with a flat palate in comparison with the default and more domed palates, and also in the large region of acoustic stability that was present for the more domed palates but not the flat palate.

The hypothesis proposed by [1] is that people with flat palates must reduce their articulatory variability to maintain acoustic consistency because their vocal tracts have smaller regions of acoustic stability. The hypothesis specifically applies to people with flat palates, and does not make predictions for the articulatory precision of people with domed palates. In the modeling here, the differences in the results from the three palates do not form a gradient. Rather, the domed and default palates have very similar results, with a large region of acoustic stability and similar slopes, but the flat palate has no regions of acoustic stability at all. This lack of progression between palate shapes is mirrored in behavioral data. [11] used ultrasound to compare production of /r/ and palate shape and found that articulatory precision sharply increases when palates reach a certain degree of flatness.

F3	lip protrusion	dorsum backing	dorsum bunching	retroflexion
Flat				
1578.7	1	4	3	-4
1622.2	2	4	2	-2
1625.8	3	4	2	-1
1627.4	4	4	2	0
1683.0	4	4	3	-2
Default				
1371.0	4	4	3	-4
1675.6	3	4	3	-4
1692.8	4	4	3	-3
1842.6	2	4	2	0
1856.0	4	4	2	3
Domed				
1789.1	3	3	4	-1
1994.5	1	4	3	4
2013.9	1	4	3	3
2033.7	0	4	3	2
2053.1	0	4	3	-3

Table 3: Parameter settings yielding the five lowest F3 values for each palate.

Palate shape not only influences the overall acoustic stability and flexibility of a vocal tract but also the effect of individual articulators on acoustics. Each of the articulators manipulated here had a different effect on F3. Most surprisingly, some factors (bunching of the tongue and orientation of the tongue tip) had opposite influences on F3 for the flat and domed palate shapes. This difference in effect of individual articulators provides a glimpse of an answer to the long-standing question of why some speakers have a retroflex /r/ and others have a bunched /r/. The shape of the palate is likely not the sole determining factor of a speaker's articulation, but it is certainly possible that the vocal tract is influential indirectly through this relationship between individual articulators and acoustics.

5. Conclusions

The models here test the hypothesis that the reason that people with flatter palates are articulatorily more precise is that the articulatory-acoustic mapping is most sensitive for such vocal tract shapes. The modeling shows a greater acoustic range overall for flatter palates. Changes in articulation are more closely correlated with acoustics for flatter palates than for more domed palates. This shown in how incremental changes in articulatory parameter settings have a greater effect on F3 for the flattest palate and the least on the most domed palate. Further, articulators seem to have different influences on acoustics in relation

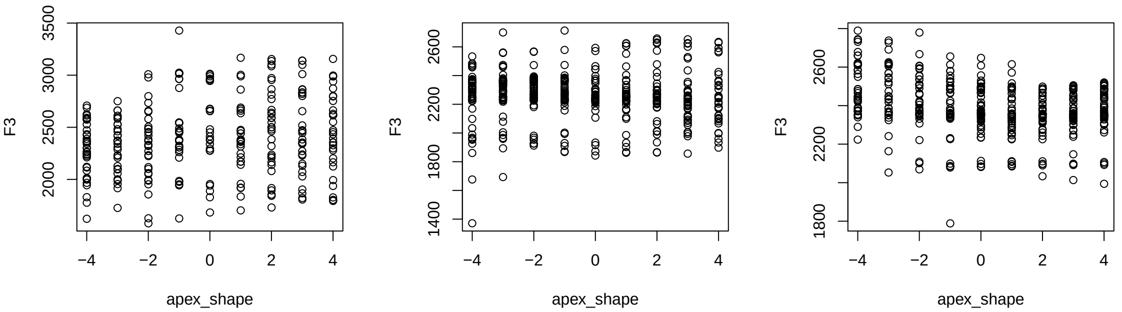

Figure 3: Spread in F3 for the flat, regular, and domed palates, respectively. Scale is the same although the range differs.

to each other for different palate shapes. This is seen in how closely linked each articulator is with acoustics; for example, while the position of the tongue dorsum has the strongest lowering effect on F3 for all three palates, the shape of the dorsum and orientation of the tongue tip have opposite effects for flatter and more domed palates.

The work supports the hypothesis in [1] that the articulatory-acoustics relationship is less quantal for flatter palates, and that this may be the reason that people with flatter palates are more articulatorily precise. The results also begin to answer how different palate shapes could influence articulatory variants for phones like /r/, which can have drastically different articulations.

Finally, the results here have implications for the organization of sound systems and may provide an explanation for the instigation of sound change. It provides evidence for the hypothesis in [12] that the phonemes of a language are attracted to regions of acoustic stability. In a hypothetical community with a high ratio of speakers with flatter palates (and therefore less acoustic stability), we might find higher rates of sound change.

6. Acknowledgements

This research greatly benefitted from feedback and technical help from Ronald Sprouse, Susan Lin, John Houde, Rich Ivry, the UC Berkeley Phon Lab, and attendees at ASA 2016 in Honolulu.

7. References

[1] J. Brunner, S. Fuchs, and P. Perrier, “On the relationship between palate shape and articulatory behavior,” *Journal of the Acoustical Society of America*, vol. 125, no. 6, pp. 3936–3949, 2009.

[2] S. Bakst and S. Lin, “An ultrasound investigation into articulatory variation in American /r/ and /s/,” in *Proceedings of the 18th International Congress of Phonetic Sciences*, The Scottish Consortium for ICPhS 2015, Ed. Glasgow, UK: The University of Glasgow, 2015.

[3] P. Delattre and D. C. Freeman, “A dialect study of American r’s by X-ray motion picture,” *Linguistics*, vol. 44, pp. 29–68, 1968.

[4] J. Mielke, A. Baker, and D. Archangeli, “Variability and homogeneity in American English /r/ allophony and /s/ retraction,” *Laboratory Phonology*, vol. 10, pp. 699–729, 2010.

[5] B. S. Atal, J. J. Chang, M. V. Matthews, and J. W. Tukey, “Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique,” *Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1535–1555, 1978.

[6] K. N. Stevens and S. E. Blumstein, “Quantal aspects of consonant production and perception: a study of retroflex stop consonants,” *Journal of Phonetics*, vol. 3, pp. 215–233, 1975.

[7] S. Maeda, “Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model,” in *Speech production and speech modeling*, W. J. Hardcastle and A. Marchal, Eds. Kluwer Academic Publishers, 1990.

[8] C. Y. Espy-Wilson, S. Boyce, M. Jackson, S. Narayanan, and A. Alwan, “Acoustic modeling of american english /r/,” *Journal of the Acoustical Society of America*, vol. 108, no. 1, pp. 343–356, 2000.

[9] L. Manzara, “The tube resonance model speech synthesizer,” Master’s thesis, University of Calgary, 1993.

[10] A. Watanabe, “Formant estimation method using inverse filter control,” *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, pp. 317–26, 2001.

[11] S. Bakst, “Differences in the relationship between palate shape, articulation, and acoustics of American English /r/ and /s/,” in prep.

[12] K. N. Stevens, *The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data*. New York: McGraw-Hill, 1972.