

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

Sarah Bakst, and Keith Johnson

Citation: [The Journal of the Acoustical Society of America](#) **144**, EL71 (2018); doi: 10.1121/1.5048043

View online: <https://doi.org/10.1121/1.5048043>

View Table of Contents: <https://asa.scitation.org/toc/jas/144/1>

Published by the [Acoustical Society of America](#)

ARTICLES YOU MAY BE INTERESTED IN

[Horizontal directivity patterns differ between vowels extracted from running speech](#)

[The Journal of the Acoustical Society of America](#) **144**, EL7 (2018); <https://doi.org/10.1121/1.5044508>

[Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences](#)

[The Journal of the Acoustical Society of America](#) **144**, EL33 (2018); <https://doi.org/10.1121/1.5044641>

[Perception and production in interaction during non-native speech category learning](#)

[The Journal of the Acoustical Society of America](#) **144**, 92 (2018); <https://doi.org/10.1121/1.5044415>

[Source characteristics of voiceless dorsal fricatives](#)

[The Journal of the Acoustical Society of America](#) **144**, 242 (2018); <https://doi.org/10.1121/1.5045345>

[Effects of consonantal context on the learnability of vowel categories from infant-directed speech](#)

[The Journal of the Acoustical Society of America](#) **144**, EL20 (2018); <https://doi.org/10.1121/1.5045192>

[Perception of relative pitch of sentence-length utterances](#)

[The Journal of the Acoustical Society of America](#) **144**, EL89 (2018); <https://doi.org/10.1121/1.5048636>

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

Sarah Bakst^{a)} and Keith Johnson

University of California, Berkeley, Berkeley, California 94720, USA
sbakst@uic.edu, keithjohnson@berkeley.edu

Abstract: Articulatory variability is reduced for people with flatter palates [Bakst and Lin (2015). *Proceedings of the 18th International Congress of Phonetic Sciences*; Brunner, Fuchs, and Perrier (2009). *J. Acoust. Soc. Am.* **125**(6), 3936–3949]. Brunner, Fuchs, and Perrier [(2009). *J. Acoust. Soc. Am.* **125**(6), 3936–3949] hypothesized that this is because the mapping between articulation and acoustics depends on palate depth. Articulatory synthesis was used with three different palate shapes to generate productions of /r/. The parameter spaces of the articulatory synthesizers were searched for vocal tract configurations that result in low F_3 (the hallmark acoustic cue for /r/). Palate shape influences not only the sensitivity of the articulatory-acoustic mapping, but also the effect of each individual articulatory parameter on F_3 .

© 2018 Acoustical Society of America

[AL]

Date Received: May 23, 2018 Date Accepted: July 10, 2018

1. Introduction

1.1 Background

The production of the American English rhotic /r/ is known to vary across individuals and is produced along a continuum from retroflex to bunched (Delattre and Freeman, 1968; Mielke *et al.*, 2010). 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 (Atal *et al.*, 1978). The experiment presented here models the influence of palate shape on the mapping between articulation and acoustics and assesses the extent to which palate shape determines an individual's /r/ variant.

Because the mapping between articulation and acoustics is also non-linear (Stevens, 1972), a given degree of vocal tract constriction will have an acoustic effect that depends on the specific location of the constriction. Some regions of the vocal tract have relatively stable acoustic output over a range of constriction locations, and between these there are 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, Brunner *et al.* (2009) 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 less quantal (more linear) for flatter palates, while domed palates show a more quantal map for place of articulation. 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 F_3 acoustic space for different articulatory configurations for the different palate shapes. A lowered F_3 is the hallmark acoustic cue for American English /r/, and different palate shapes may favor particular articulations for lowering F_3 . Second, the modeling assesses differences in the influence of various articulatory parameters on F_3 for these different palate shapes. Specifically, this study tests the hypothesis in Brunner *et al.* (2009) that the increased articulatory precision observed in people with flatter palates is a consequence of how an articulatory change affects the vocal tract area proportional to the height of the

^{a)}Author to whom correspondence should be addressed.

palate. An articulatory change of a given size will result in a smaller change in vocal tract cross-sectional area as a proportion of the total area for a more domed palate than that of a flat palate.

This hypothesis makes two main predictions for how a given set of articulation will result in different acoustics for flatter palates: (1) a wider range of produced $F3$ values, because regions of acoustic stability are smaller; (2) a stronger correlation between each articulatory parameter and acoustic measure, because a given change in articulations should result in greater acoustic change.

2. Methods

The hypothesis is that the shape of the hard palate plays an important role in the quality and variability of articulation in production. This 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 (Maeda, 1990) was to model French vowels and includes a single palate based on a real speaker. The articulatory parameters are principal components based on x ray data from this speaker. The vocal tract is divided into 32 discrete sections where the area function is calculated. The code is freely available in the Berkeley Phonetics Machine (Sprouse and Johnson, 2016). 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, by editing the width of the model at sections 18–26, which correspond to the palatal region. 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 user controls the synthesizer by indicating a setting for each articulatory parameter. The setting is a multiplier for the principal component representing the parameter. The four active articulator parameters considered here each represents an important difference in retroflex versus bunched articulations: the dorsum position and shape, the protrusion of the lips, and the orientation of the tip (whether it is pointing up or down).

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 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 the 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 Eq. (1), where α and β were determined from real production data from a single speaker and hard-coded into the original model. These α values are a ratio of the width and depth of the palate and actually correspond to the same metric of domedness as in Brunner *et al.* (2009).

$$A(x) = \alpha x^\beta. \quad (1)$$

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 Bakst and Lin (2015) in order to reflect realistic differences in cross-sectional area. Both the Maeda (1990) model and Brunner *et al.* (2009) model assigns $\beta = 1.5$, as we do here.

The model was run with each of the three palates. We wrote a program which cycled over the range of settings incrementally (step size = 1) for each of the four articulatory parameters in the Maeda model that would result in a possible articulatory configuration for /r/. There were nine settings for tip orientation (tip down to tip up,

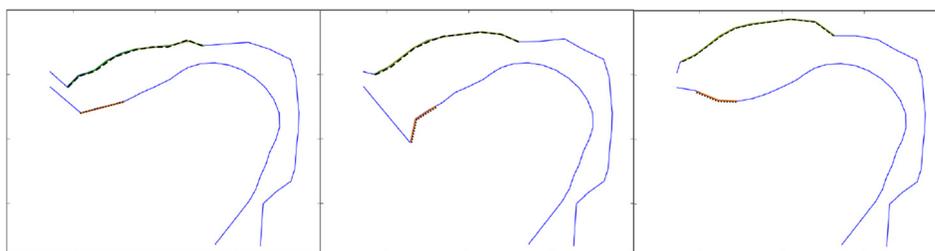


Fig. 1. (Color online) The three different palate shapes (dashed line; from left: flat, default, domed) and examples of tongue tip configurations (dotted line): neutral, extreme tip-down, and extreme tip-up.

range = -4 to 4), and five each for degree of lip protrusion, tongue bunching, and tongue backing (range = 0 to 4). From the area functions generated by the Maeda model we calculated acoustic waveforms with a flexible tube model synthesizer (Manzara, 1993) to which we added a short tube as a side-branch [after Espy-Wilson *et al.* (2000)] to model the sublingual cavity that emerges in /r/ production. The script then performed acoustic analysis (Watanabe, 2001) over the synthesized output. Tokens with amplitude less than 40% of the maximum value for tokens were excluded from further analysis; these were typically silent or not speech-like. $F2$ and $F3$ measurements were recorded from the midpoint of the sound file.

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 type of palate. The generated speech-like sound files were sorted by $F3$ 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 $F3$ values for the flat palate, less so for the original palate, and the smallest range for the domed palate. The overall acoustic flexibility is similar for domed and original 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. Figure 2 also shows that the lowest $F3$ values were possible with the flatter palate.

While the $F3$ values reported in Fig. 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 2300 Hz (dashed line), an empirically-derived cut-off [data from Bakst and Lin (2015)], there is less stability for the domed and original palates, but still more than for the flattest palate. Note also that the domed and original palates show similar sensitivity to each other but not the flat palate. This non-linear relationship between sensitivity and doming is consistent with behavioral data in Bakst and Lin (2015), where articulatory variability was reduced for flat palates, but articulatory variability did not distinguish different degrees of domedness among non-flat palates.

Table 2 shows the correlation of each parameter with $F3$ for each palate. The largest correlations between articulatory parameter and acoustics were in the flat palate. 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. 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 regular palates, bunching of the tongue lowered $F3$, but for the domed palate, bunching actually raised $F3$. In this model, bunching of the tongue is accompanied by increased tongue height, which decreases the cross-sectional area at that location. 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. For articulations producing very low $F3$ values with the flat palate, the tip was either at a neutral or downward orientation (i.e., non-retroflexed). This result suggests that a flat palate would favor a bunched /r/, and a domed palate

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

	Domed	Default	Flat
α	1.3	(1.7)	2.7
Minimum $F3$	1789	1371	1578
Maximum $F3$	2789	2713	3428
$F3$ range	1000	1342	1850
Min $F2$	788	704	684
Max $F2$	2252	2227	2035
$F2$ range	1464	1523	1351

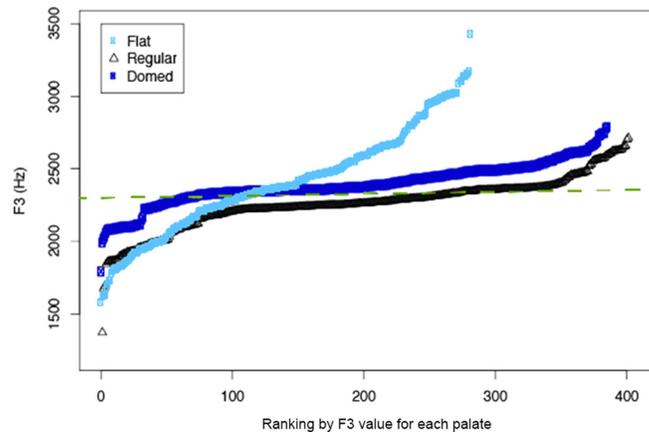


Fig. 2. (Color online) For each palate, the produced speech files were ranked by their F_3 values into ascending order. The x -axis gives the ranking of each file, illustrating the flexibility of the acoustic space for each palate. The steady region is smallest for the flat palate and largest for the most domed palate. Dashed line shows cutoff for $/r/$ consideration.

might favor a retroflex $/r/$. This relationship is weak, though: for all three palates, there was a wide range in F_3 values for different settings of the tongue tip.

For all three palates, a low F_3 was achieved primarily by retracting the tongue as much as possible, and secondarily by bunching the tongue. Lip protrusion was significantly correlated with F_3 for the original and domed palates, but in the articulations where F_3 was lowest, lip shape varied.

4. Discussion

The modeling here suggests that the vocal tract sensitivity function is related to the shape of the palate. For F_3 , flatter palates have a less quantal articulation-to-acoustics mapping than do domed palates. This was shown both in the range of F_3 values produced with a flat palate in comparison with the original 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 Brunner *et al.* (2009) 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 specific predictions for the articulatory precision of people with more domed palates. In the modeling here, the differences in the results from the three palates do not form a gradient. Rather, the domed and original palates have very similar results, with a large region of acoustic stability, but the flat palate has no regions of acoustic stability at all. This finding is corroborated by ultrasound data in $/r/$ production (Bakst and Lin, 2015; Bakst, 2016) showing that articulatory precision sharply increases when palates reach a certain degree of flatness, but that a domed palate does not predict articulatory precision.

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 F_3 . Most surprisingly, some factors (bunching of the tongue and orientation of the tongue tip) had opposite influences on F_3 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.

Table 2. Correlations between articulators and F_3 for each palate. * $p < 0.05$, ** $p < 0.001$. Note: n.s. = not significant.

	Tip curl	Backing	Bunching	Lip
flat (df = 280)	0.15*	-0.78**	-0.56**	-0.06 (n.s.)
regular (df = 399)	-0.05 (n.s.)	-0.7**	-0.33**	0.15*
domed (df = 384)	-0.31**	-0.45**	0.36**	0.15*

5. Conclusion

The models here test the hypothesis (Brunner *et al.*, 2009) that people who have flatter palates are articulatorily more precise because their articulatory-acoustic mapping is very sensitive to small perturbations. 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: 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 to each other for different palate shapes. The results thus suggest that different palate shapes could favor particular 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 some insight regarding language sound change. Given the quantal hypothesis (Stevens, 1972) that the phonemes of a language are attracted to regions of acoustic stability, our results suggest that 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.

Acknowledgments

This research greatly benefited 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.

References and links

- Atal, B. S., Chang, J. J., Matthews, M. V., and Tukey, J. W. (1978). "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique," *J. Acoust. Soc. Am.* **63**(5), 1535–1555.
- Bakst, S. (2016). "Differences in the relationship between palate shape, articulation, and acoustics of American English /r/ and /s/," UC Berkeley Phonology Lab Annual Report.
- Bakst, S., and Lin, S. (2015). "An ultrasound investigation into articulatory variation in American /r/ and /s/," in *Proceedings of the 18th International Congress of Phonetic Sciences*, edited by The Scottish Consortium for ICPhS 2015, The University of Glasgow, Glasgow, United Kingdom.
- Brunner, J., Fuchs, S., and Perrier, P. (2009). "On the relationship between palate shape and articulatory behavior," *J. Acoust. Soc. Am.* **125**(6), 3936–3949.
- Delattre, P., and Freeman, D. C. (1968). "A dialect study of American r's by X-ray motion picture," *Linguistics* **44**, 29–68.
- Espy-Wilson, C. Y., Boyce, S., Jackson, M., Narayanan, S., and Alwan, A. (2000). "Acoustic modeling of American English /r/," *J. Acoust. Soc. Am.* **108**(1), 343–356.
- Maeda, S. (1990). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic Publishers, The Netherlands).
- Manzara, L. (1993). "The tube resonance model speech synthesizer," Master's thesis, University of Calgary, <https://prism.ucalgary.ca> (Last viewed 10/18/2016).
- Mielke, J., Baker, A., and Archangeli, D. (2010). "Variability and homogeneity in American English /r/ allophony and /s/ retraction," *Lab. Phonology* **10**, 699–729.
- Sprouse, R., and Johnson, K. (2016). "The Berkeley Phonetics Machine," in *Proceedings of the Interspeech*, pp. 1623–1626.
- Stevens, K. N. (1972). "The quantal nature of speech: Evidence from articulatory-acoustic data," in *Human Communication: A Unified View*, edited by E. E. David and P. B. Denes (McGraw-Hill, New York), pp. 51–66.
- Watanabe, A. (2001). "Formant estimation method using inverse filter control," *IEEE Trans. Speech Audio Process.* **9**(4), 317–326.