AN INDEX FOR QUANTIFYING TONGUE CURVATURE

Slade Stolar and Bryan Gick
University of British Columbia, Dept. of Linguistics, Totem Field Studios
2613 West Mall, Vancouver, British Columbia, Canada, V6T 1Z4
gick@mail.ubc.ca

ABSTRACT

This study develops a method of quantifying tongue curvature by modeling the shape of the tongue surface in any anatomical plane using a polynomial approximation. In a validation experiment, the curvature indices of English vowel and consonant sounds were calculated across ten native speakers' productions based on midsagittal ultrasound images of the tongue. Indices confirm substantially higher curvature values for liquids /r/ and /l/ than for all other sounds in the inventory. This method is both more generalized and less dependent upon fixed locations than previous methods, and provides a simple, powerful metric for evaluating shape complexity with applications in areas such as motor development and aeroacoustics.

1. INTRODUCTION

In executing successful speech production, the tongue must often be able to take on quite complex shapes. In disordered or developing speech, speakers may not have achieved fully differentiated control of functionally independent parts of the tongue (see, e.g., Stone et al. 2004 for a discussion of functional tongue regions), limiting their ability to achieve adequate tongue shape complexity, and interfering with the achievement of speech goals (see Gibbon 1999, Cheng et al. 2007). Such is the case, for example, with lingually complex sounds such as the English liquids /r/ and /l/ (see Studdert-Kennedy and Goldstein 2003). Gick et al. (2008) argue that a reduced capacity for lingual shape complexity may account for some of the substitutions for English /r/ and /l/ commonly seen in speech disorders, children's speech, and L2 acquisition. The present study uses the complex tongue shape of these liquids as part of validating a novel method of quantifying tongue shape complexity.

While the complexity of liquid consonants may be easily viewed in the midsagittal plane, other postures appear more complex in the coronal plane (such as the deeply "grooved" shape for /s/; see Stone et al. 1988). Complex tongue curvature is likewise vital in aeroacoustics, where the degree of tongue grooving for sibilants, visible in coronal sections, corresponds with turbulence noise (see Stone et al. 1989).

Tongue shape complexity is not just important in achieving steady-state postures for specific sounds. In coarticulation between any anterior and posterior lingual sounds (as in the temporal overlap of consonants and vowels), different regions of the tongue are independently engaged, resulting in more complex tongue shapes. Noiray et al. (2013) provide evidence linking tongue shape to development, and Zharkova et al. (2012) report less consistency in tongue curvature during coarticulation of /sV/ sequences in children's than in adults' productions. The same presumably applies in longer-distance cases, as where Gafos (1999) treats tongue shape as an independent phonological parameter in understanding consonant harmony.

Despite the many instances where it arises, there have been few attempts to model tongue shape complexity, and fewer to develop practical indices for quantifying it. Stone and Lele (1992) use simple (second-order) shape functions to characterize 3-D tongue shapes and find "shape signatures" that simplify the tongue. Bressmann et al. (2005) use 3-D ultrasound to show that anteriority, concavity, and asymmetry could be used to quantitatively compare normal speakers to one partial glossectomee. While these approaches constitute important contributions to our understanding of tongue shape complexity, they are too data- and labor-intensive to provide practical indices.

A number of methods for quantifying aspects of tongue shape have been proposed in the last decade, although none of these quantify complexity per se, and all have limitations that prevent their being used for this purpose. Iskarous et al. (2003) fit conic arcs to the midsagittal tongue surface, enabling approximation of a variety of rounded shapes. Ménard et al. (2012) use superimposed triangles to extract important midsagittal tongue measures, such as high points or location and acuteness of curvature. Zharkova (2013) uses a tongue dorsum excursion index, a straight line between the ends of the midsagittal tongue curve rising to the highest point, to track variations in tongue shape during production. Whalen et al. (2011) use a similar approach (following a method used by Stone et al. 1988) to measure tongue groove depth in the coronal plane. Most of these previous methods rely upon having either rounded or grooved tongue shapes, but they are not able to quantify differences spanning both simple and complex shapes using a single measure. Further, most methods cannot compare data across different tongue morphologies or sizes, and several methods assume accurate knowledge of landmarks such as the tongue tip and root, which are often the least reliable parts of the tongue to image using ultrasound. These requirements limit the kinds of tongue postures and populations that existing methods can model. One additional approach, the SSANOVA statistical method described by Davidson (2006), also approximates tongue shapes based on edge tracking of ultrasound images. This method is designed to quantify the statistical differences between different tongue shape curves. However, it does not attempt to model the resulting curves, or to quantify degrees of curvature. Furthermore, a fixed probe location relative to the tongue is required, ruling this out as a preferred method for comparing multiple speakers or the same speaker over multiple sessions.

The present study describes an alternative method for quantifying tongue surface shape complexity. This method derives a one-dimensional curvature index from a 7^{th}-order polynomial curve fit to the tongue's surface as imaged in any plane (midsagittal, coronal, or transverse). This method has the advantage of being able to compare tongue shapes that have large variations in curvature along their length, orientation and image quality, and may be recorded from speakers with any range of tongue sizes or orofacial morphologies – an essential feature, as variation in gross tongue shape is associated with many disorders affecting speech.

Figure 1 shows a continuum of idealized tongue surface shapes, with curviness ranging from low (a) to high (d). These are canonical, idealized shapes that are based on observations of midsagittal and coronal ultrasound images from dozens of speakers. Note that the shapes in Figure 1 can represent sections of the tongue's surface in different anatomical planes equally well. For example, if the images represented a midsagittal slice, then shape (a) might correspond to a production of /ɔ/, whereas shape (d) corresponds to English /r/; if the images were coronal, shape (a) could correspond to /i/ or /ʃ/, whereas (d) corresponds to /s/. After presenting a derivation of the curvature index below, the following sections test the model, first against these idealized shapes, then through a validation experiment exploiting the known lingual shape complexity of the English liquids.

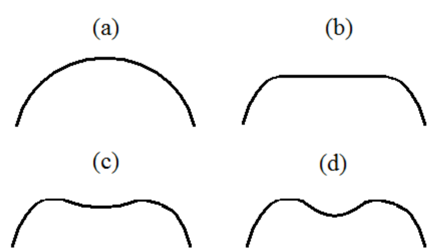

Figure 1. Idealized tongue shapes

2. LINGUAL CURVATURE INDEX

One mathematical formulation of lingual shape is the curvature of a line along the surface of the tongue. That is, one method to quantify lingual shape is to examine the ways in which (and degree to which) the tongue's surface curves upwards or downwards or remains flat in some standard cross-section (midsagittal, coronal, or transverse) of the tongue.

Once a tongue image has been collected and digitized, its surface can be traced either by automatic edge tracking or by a trained analyst placing representative points along its edge by hand. Using MATLAB, the representative points are approximated by a spline curve using a least-squares-fit. After applying the spline tracing to the tongue's surface, the program fits the curve using a 7^{th}-order polynomial fit (function f in equation (1)). The first and second derivatives of this function are used to calculate the radius of curvature r in equation (2) (at each point along its length), and the total curvature i, in equation (3).

The relevant equations are as follows:

$$f = ax^7 + bx^6 + cx^5 + dx^4 + ex^3 + fx^2 + gx + h \qquad (1)$$

where f is the polynomial fit to the spline of the tracing and constants a through h are the coefficients of the fit, and

$$r = (1 + f'^2)^{3/2} / (f''), \qquad (2)$$

$$i = \int_a^b r\,dx, \qquad (3)$$

where r is the radius of curvature, i is the total curvature, and a and b are the start point and end point of the spline, respectively.

The total curvature value i is referred to as the curvature index (CI) of a given tongue shape. This expression is unitless (that is, it is a ratio of the variation of curvature to the length) and is the same for a given shape, despite being rotated or scaled, such that the CI is essentially probe-location and rotation invariant. We found that the variability of CI values was ±0.04 (approximately 1-2%, due to slight variations in the placement of the points by hand, propagated through the spline and polynomial fits) for the same shape traced by the same judge. Once an image is loaded, it takes less than 30 seconds to manually trace and automatically compute the CI value. The above-discussed properties make this method a fast, reliable, easily-interpretable method of quantifying tongue shape.

Figure 2 shows the range of derived index values for the set of idealized shapes given in Figure 1, as well as values for two actual sample midsagittal tongue shapes.

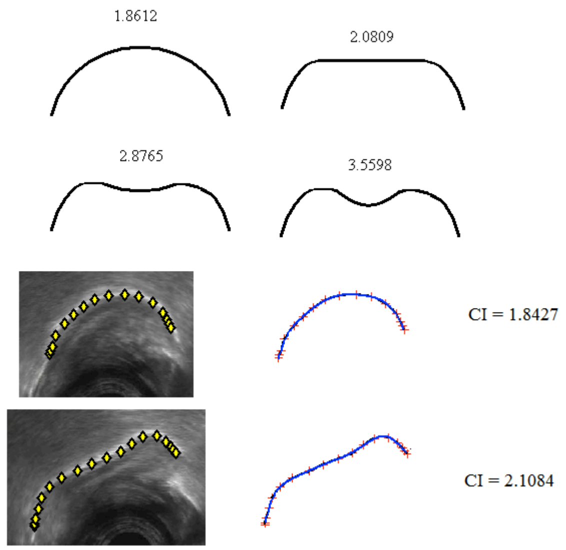

Figure 2. Curvature indices of idealized and recorded tongue shapes.

3. VALIDATION EXPERIMENT: SAGITTAL TONGUE DIFFERENTIATION IN THE SOUND INVENTORY OF ENGLISH

In order to verify the applicability of this index, we conducted a validation experiment using the index to quantify degrees of midsagittal lingual curvature in the sounds of English. This method is able to identify the various sounds as having high or low curvature values and as having a broad or narrow distribution of values.

3.1 Introduction

Past research has shown that certain sounds involve more complex tongue shapes than others, rendering them more difficult for speakers to produce, particularly for L1 or L2 learners and disordered speakers. For example, the late acquisition of English liquids has been ascribed to their having two simultaneous tongue constrictions (see Studdert-Kennedy and Goldstein 2003, McGowan et al. 2004 for /r/; Oh 2005 for /l/), giving them their characteristic shape in the midsagittal plane. This difficulty with producing English liquids relates to motor differentiation, whereby a speaker must learn to segregate and independently control anatomically coupled articulators during the course of motor development (as with the jaw and lower lip; see Green et al. 2000). Being a single muscular hydrostat, the tongue is subject to this same process of differentiation before its independently controllable parts can be separately manipulated for speech (Gibbon 1999). The present validation study uses the CI as a metric for shape complexity (and by extension, lingual differentiation), which is expected to be relatively high for the liquids /l/ and /r/ and relatively low for most other sounds of English.

3.2 Experiment

A study was conducted in which ten adult speakers with no known hearing or speech impairments produced a complete inventory of the sounds (the vowels in a /bVt/ context and consonants in a /aCa/ context) of English.

Participants were seated in a modified ophthalmic examination chair, with four-point head stabilization and a mechanical arm holding an ultrasound probe under the jaw. Midsagittal ultrasound images were recorded using an Aloka Pro-Sound SSD-5000 ultrasound machine. Audio recordings of speakers' productions were also taken.

Stimuli were constructed as follows. Consonants were presented in /aCa/ context; vowels were presented in /bVt/ context. Participants read words from a computer screen that briefly presented each token. Each subject read the prompts for three complete, randomized sets of tokens. In this way, an inventory of relevant sounds was taken, including consonants in context (ara, ala, awa, aga, anga, ana, aya, ata, asha, atha, asa, ada, aza) and vowels in context (boot, bought, bat, bit, beat, but, bet).

Single frames were extracted by pausing the ultrasound video in the middle of the relevant speech segment. Each extracted image provided, as much as possible, a clear midsagittal image of each sound's postural extremum. 15-20 points were manually selected along the surface of each tongue image and were converted into (x, y) coordinates. While it is possible to use fewer points, a 7^{th} order polynomial requires a theoretical minimum of 7 points. The CI value was then calculated for tongue shape using the algorithm described above.

Two-way ANOVAs were used to examine the differences among the curvatures. The sounds were further grouped by manner (liquids, glides, nasals, stops, vowels, and fricatives) to examine the similarities and differences among different categories of sounds.

3.3 Results

The following tables present the CIs for the tongue shapes averaged across subjects, plotted in descending order of median curvature.

Grouping the sounds according to their manner (liquids, glides, nasals, stops, vowels, and fricatives), as shown in Figure 3, two-way ANOVA results indicate a significant effect of manner on the curvature [$F(5, 208) = 7.168$, $p < .0001$]. Specifically, the liquids had a significantly different curvature from all other sounds taken together [$F(1, 212) = 30.25$, $p < .0001$].

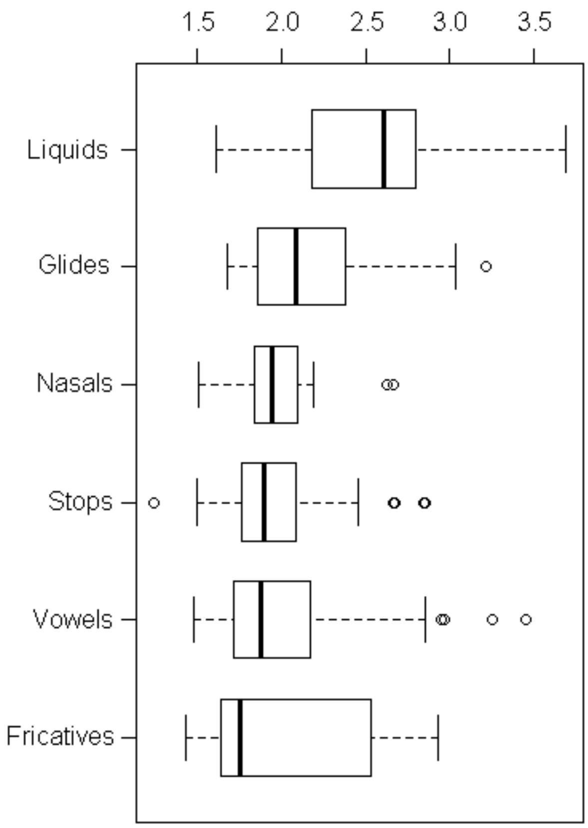

Figure 3. Curvature indices grouped by manner

Taken individually, the CI of /r/ was significantly higher than that of all other sounds taken together, both when excluding /l/ [$F(1, 202) = 33.87$, $p < .0001$] and when including /l/ [$F(1, 212) = 32.12$, $p < .0001$]. The CI of /l/ was distinct from that of all other sounds taken together (excluding /r/) [$F(1, 202) = 5.25$, $p = 0.0246$]. The CI of /l/ was also distinct from that of /r/ [$F(1, 18) = 5.95$, $p = 0.0253$]. /l/ was not, however, distinct from all other sounds taken individually. In particular, the CI for /w/ was similar to that of /l/ – and, indeed, all of the velar sounds were clustered at the high end of the scale – presumably due to the curvature introduced by advancing the tongue root for hydrostatic raising of the tongue body. Results by individual sound are given in Figure 4.

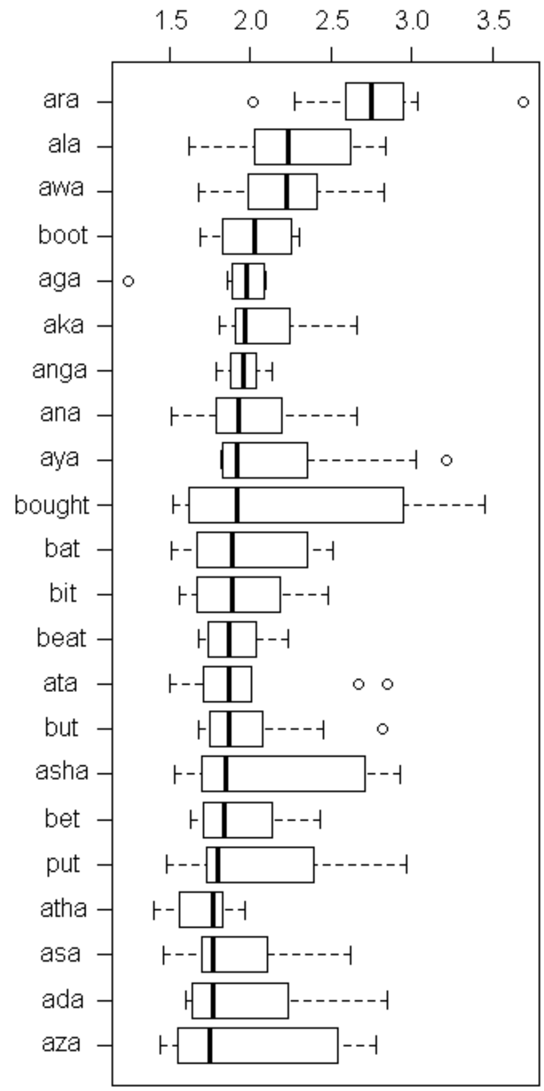

Figure 4. Ordered list of sounds by curvature index, averages within subject.

3.4 Discussion

Our results for the sound inventory of English indicate that the CI accurately differentiates /r/ (by a large margin) and /l/ (to a lesser extent) from the other sounds of English, capturing this known difference in degree of midsagittal tongue curvature.

It is worth noting that both retroflex and bunched tongue shapes appeared in the /r/ productions in our data. These two production variants were approximately equally curvy overall: whereas the retroflexed posture has a locally-high curvature at the tip and a relatively flat shape over the body of the tongue, the bunched /r/ was more curved over its entire length. Our data also indicate some variation in CI values for glides, vowels, and fricatives.

4. CONCLUSIONS

The curvature index thus offers one method of distinguishing more or less complex tongue surface shapes in any measurement plane, distinguishing liquid consonants from other tongue postures. This single metric can potentially be used to examine variation in a variety of speech tasks, such as with sibilant grooving or co-articulation, or to quantify differences in tongue motor differentiation across populations, such as in normal and disordered or developing speakers (see Gick et al. 2008).

In future research, a generalized curvature index of this kind can be implemented to track progress during a course of second language learning, speech intervention or remediation, or to track development of linguo-motor differentiation in normal or disordered populations.

REFERENCES

Bressmann, T., Thind, P., Uy1, C., Bollig, C., Gilbert, R.W., and Irish, J.C. (2005). Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: Data from 12 normal speakers and a partial glossectomee. Clinical Linguistics and Phonetics 19, 573-588.

Cheng, H., Murdoch, B., Goozée, J., and Scott, D. (2007). Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults. Journal of Speech, Language, and Hearing Research 50, 375-392.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. J. Acoust. Soc. Am. 120, 407-415.

Gafos, D. (1999). *The Articulatory Basis of Locality in Phonology*. Garland Publishing, New York.

Gibbon, F. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. Journal of Speech, Language, and Hearing Research 42, 382-397.

Gick, B., Bacsfalvi, P., Bernhardt, B.M., Oh, S., Stolar, S., and Wilson, I. (2008). A Motor Differentiation Model for Liquid Substitutions: English /r/ Variants in Normal and Disordered Acquisition. Proceedings of Meetings on Acoustics, Vol. 1, 060003 (9 pp.).

Green, J.R., Moore, C.A., and Higashikawa, M. (2000). The physiologic development of speech motor control: Lip and jaw coordination. Journal of Speech, Language, and Hearing Research 43, 239-255.

Iskarous, K., Goldstein, L.M., Whalen, D.H., Tiede, M., and Rubin, P. (2003). CASY: The Haskins Configurable Articulatory Synthesizer. M.J. Solé, D. Recasens, and J. Romero (eds.) Proceedings of the 15th International Congress of Phonetic Sciences. Universitat Autónoma de Barcelona, Barcelona. 185-188.

Ménard, L., Aubin, J., Thibeault, M., and Richard, G. (2012). Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model. Folia Phoniatrica et Logopaedica 64, 64-72.

Noiray, A., Ménard, L., and Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. J. Acoust. Soc. Am.133, 444-452.

Oh, S. (2005). *Articulatory characteristics of English /l/ in speech development.* Ph.D. Dissertation, University of British Columbia.

Stone, M., Raphael, L.J., and Faber, A. (1989). Cross-sectional tongue shapes and palatal contours during sibilant and lateral consonants. J. Acoust. Soc. Am. 86(S1), S113.

Stone, M., Epstein, M.A., and Iskarous, K. (2004). Functional segments in tongue movement. Clinical Linguistics & Phonetics 18, 507-521.

Studdert-Kennedy, M., and Goldstein, L.M. (2003). Launching language: The gestural origin of discrete infinity. M.H. Christiansen and S. Kirby (eds.) *Language Evolution: The States of the Art.* Oxford University Press, Oxford.

Stone, M. and Lele, S. (1992). Representing the tongue surface with curve fits. Proceedings of the International Conference on Spoken Language Processing 2, 875-878.

Whalen, D.H., Shaw, P., Noiray, A., and Antony, R. (2011). Analogs of Tahltan consonant harmony in English CVC syllables. Proceedings of the International Congress of Phonetic Sciences XVII, Hong Kong. 17-21.

Zharkova, N. (2013). Using Ultrasound to Quantify Tongue Shape and Movement Characteristics. The Cleft Palate-Craniofacial Journal 50, 76-81.

Zharkova, N., Hewlett, N., and Hardcastle, W.J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. Journal of the International Phonetic Association 42, 193-208.

ACKNOWLEDGEMENTS

We wish to thank Aislin Stott and Donald Derrick for their contributions to running the experiment and data acquisition. Research funded by NSERC Discovery grant #228600-09 to the second author.