PREDICTING 3D TONGUE SHAPES FROM MIDSAGITTAL CONTOURS

Maureen Stone, Melissa A. Epstein & Martin W. Sutton

Dept. of Biomedical Sciences, University of Maryland Dental School

This study is interested in whether there exists a predictable relationship between midsagittal contour slope and 3D tongue surface shape. For any single language, a limited set of 3D tongue shapes are used. In American English, for example, sustained phonemes were adequately described by four 3D tongue shape categories (Stone and Lundberg, 1996). Moreover, all the 3D tongue surfaces were reconstructed very accurately from 6 coronal slices (Lundberg and Stone, 1999). The coronal slices were measured at the location of 6 midsagittal points that had optimally generated the midsagittal curve. Thus, there was a strong relationship between the shape of the midsagittal contour and the 3D surface with the coronal shapes serving as the link. Midsagittal height alone is not enough to predict 3D surface shapes; for example, both the arched (/k/) and the slightly grooved (/o/), have a high tongue body. However, we believe that that midsagittal displacement will correlate with coronal shape. Greater displacement will accompany an arch than a groove. We also believe that concatenated coronal slices will result in a predictable and small set of 3D surfaces. This is because in isolation a coronal slice has a large range of motion, but in a 3D surface it is constrained by its attachment to neighboring tissue, the tongue's muscular architecture, volume preservation, and language specific constraints.

INTRODUCTION AND BACKGROUND

The goal of this work is to map midsagittal tongue contours to 3D tongue surfaces, specifically 5 cross-sectional (coronal) contours. Because any single language has a finite number of phonemes and thus a finite number of 3D tongue surface shapes, a mapping of those shapes to midsagittal features could be executed for any language. In order to do this several questions must be answered. (1) Are there correlations between midsagittal displacement and coronal shape at five slices of the tongue? If so, one could predict coronal shapes from midsagittal displacements. (2) Can the coronal shapes, *i.e.*, 'slope signatures', be mapped to 3D shapes of known tongue shape categories, and differentiate them. (3) Can midsagittal contours for specific phonemes be uniquely mapped to specific 3D surface shapes? This paper will consider the first two questions. The oral presentation will explore the third as well.

Three-dimensional tongue surfaces are composed of a concatenation of coronal slices, a representative sample of which well represents the 3D surface (Stone & Lundberg, 1996). Coronal tongue shapes have been extensively studied and documented for American English (AE) (Stone, *et al*., 1988; Stone, 1990; Stone, *et al.*, 1992; Stone, 1995; Stone & Vatikiotis-Bateson, 1995; Stone, *et al*., 1997; Slud, *et al*., 2002). Coronal shapes primarily follow a continuum from midsagittal groove to midsagittal arch, with some left-to-right asymmetry. The present study is based on a single data set of sustained sounds spoken by a single speaker (Stone & Lundberg, 1996; Lundberg & Stone, 1999), whose tongue shapes are within the mainstream of those seen elsewhere.

Three-dimensional tongue surface shapes for static AE sounds have been classified into four shape-based categories (Stone & Lundberg, 1996). That study considered eighteen sounds of AE, whose shapes represented all the lingual sounds of the language. That is, homorganic sounds like /t,d,n/ were represented by /n/. The four categories are front-raising [ɪ,i,e,ɝ,n,ʃ], back-raising [u,ʊ,o,ɔ,ɑ,ŋ], continuous-groove [æ,ɛ,s,θ], two-point displacement [l] (see Figure 1). That study found that low front vowels, which at midline appear to be weak versions of front-raising (Harshman *et al*., 1977) actually have a midsagittal groove for the entire tongue. This data set, was based on 60 coronal ultrasound slices. However, since it is impractical to collect so much data, a 'sparse' data set of 6 slices was extracted. The sparse set was optimized to maximize surface coverage and minimize average error and maximum error. The six slice reconstructions had 80% coverage, an average error of .21 mm, and a maximum error of 1.40 mm. Some phoneme surfaces were shorter than others and were not captured in the first or last slices, usually the first. The sparse data set, excluding the first slice, is the basis of the present study.

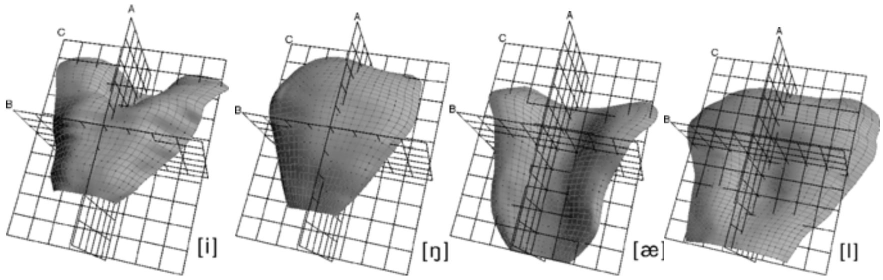

Figure 1. Representative three-dimensional tongue surface shapes for the four shape-based categories: front-raising (/i/), back-raising (/ŋ/), continuous groove (/æ/) and two point displacement (/l/). Tongue tip on lower left.

METHODS

Three measurements were made on each coronal contour: midline displacement (hereafter: height), midline-to-right slope (hereafter: right-slope), and midline-to-left slope (hereafter: left-slope). The slopes represented the coronal shape of the tongue; *i.e.*, on the left side a negative slope indicated a grooved tongue and a positive slope an arch. The slopes and midline displacement were calculated from three points on the coronal images. The same midline x-value was used in each contour even if a contour was asymmetrical, because the goal was to simulate a single midsagittal slice. A left and right x-value was chosen by visual examination of the coronal contours for all the phonemes overlaid first for each slice and then for all slices combined. The same x-values on the left and right were used for all five slices based on the peaks of the grooves and the slopes of the arches (see Figure 2).

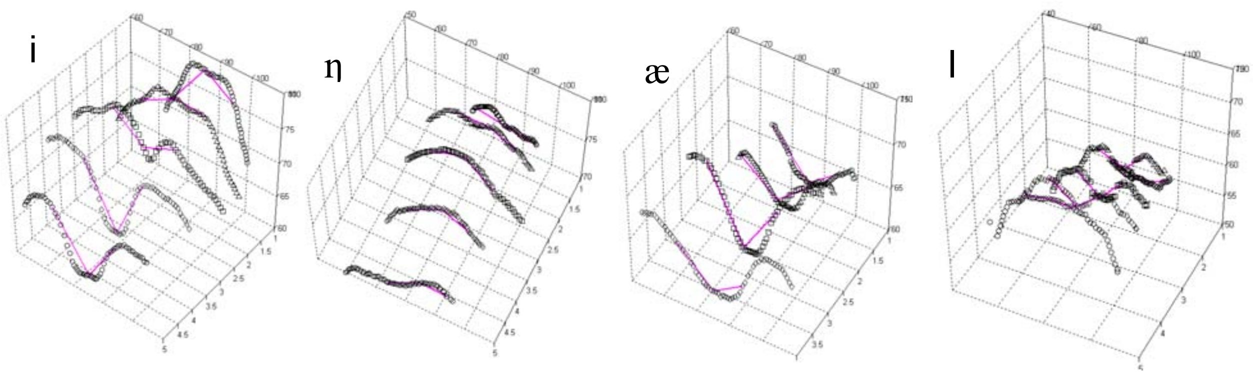

Figure 2. Coronal slices and extracted slopes for representative phonemes from each of the four shape-based categories: front-raising (/i/), back-raising (/ŋ/), continuous groove (/æ/) and two point displacement (/l/). Tongue tip is on upper right.

The right and left values were each about 1.0 cm from midline in the x-projection. Displacement is radial with the virtual pivot point inside the transducer. Since the tongue often was steeply sloped, the percent coverage was 40-100% of each contour. Some phonemes, with wider or narrower grooves, would have been better represented by different points. However, this study wished to simulate a condition where only the midsagittal data and not the coronal shapes were available. Uniform left and right x-values allowed the endpoints of the calculated lines to be identical. Pearson product moment correlation coefficients were then calculated, comparing left and right slopes, and midline height and slopes, for all 5 slices.

RESULTS AND DISCUSSION

The first question asked by this study is whether correlations exist between midsagittal displacement and cross-sectional shape (coronal slope) at each location of the tongue. The slopes, which represented partial tongue widths, did not capture all features of the coronal slices. However, as mentioned above, coronal tongue shapes for AE are well documented and consist of arches and grooves (see Figures 1-2). Arches are typically quadratic (concave down) in shape, sometimes with a level center. Anteriorly, arches occur when the tongue contacts the hard palate and the sides deform against it. Posteriorly arches occur when the tongue is pulled/pushed upward and backward by muscles for vowels or when it contacts the velum (/ŋ/). Grooves occur anteriorly due to genioglossus anterior contraction and contact with the mouth floor in low vowels or due to palatal bracing in grooved fricatives (/s/, /θ/). Anteriorly, grooves are characterized by a bimodal shape in which maximum height is reached at about the 1.0 cm point chosen in this study, after which the tongue typically slopes downward. Posteriorly, grooves are due to muscular activity and maximum height may occur lateral to the measured point. Thus, the selected points captured the key features of the arch or groove. In addition, extending the line downward from the selected left and right points provides a generally accurate completion of the contour.

Correlations between the midsagittal displacements and coronal slopes were calculated at each slice for all 18 phones. In each of the slices, greater midline displacement accompanied an arched coronal shape, and lower values a grooved shape. The transitional midline height displacement values between arch and slope were 65, 72, 80, 82, and 75 mm respectively for slices 1 to 5. Slices 4 and 5 had weak correlations between midsagittal displacement and coronal slope, but correlations for slices 1, 2 and 3 were strong. These three anterior slices contact the hard or soft palate and the tongue motion may be constrained by that contact, thus increasing the correlation (see Table 1).

Table 1. Pearson product moment correlation coefficients between midsagittal displacement and left / right slopes, as well as between left and right slopes (n=18).

	Correlations		
Slice	Mid-Left	Mid-	Left-
1	.67	-.70	-.91
2	.79	-.88	-.67
3	.71	-.78	-.85
4	.51	-.40	-.66
5	.64	-.55	-.70

Left-to-right asymmetries were determined by correlating left and right slopes. Previous research has revealed some left-to-right asymmetry in speech movements for specific speakers or movements (*cf.* Stone, 1990). When generalizing from midsagittal contour to coronal shape it is not possible to predict asymmetries. Therefore, a symmetrical prediction becomes the goal. In the present data set left-to-right correlations were fairly strong, indicating symmetry (see Table 1). We believe that steady-state sounds in any language will typically be symmetrical. Thus, mapping symmetrical representations of arched, level, or grooved shape onto midsagittal points would capture the essence of tongue shape quite well.

The second question asked by this study was how the slope signatures could be mapped to the four tongue shape categories, and differentiate amongst them. Figure 3 shows the slope patterns for each category. In front-raising coronal tongue shape progresses from a steep arch in front to a deep groove in back. For the three consonants /ʃ,ɝ,n/, the deepest groove is at slice 3, *i.e.*, a velar location. The vowels are more deeply grooved posteriorly. In back-raising, the tongue is arched for the consonant /ŋ/, and almost entirely grooved for the vowels. However, the arches and grooves in this category are very shallow compared to those of front-raising and continuous-grooving. In continuous-grooving the tongue is grooved throughout for consonants and vowels. However, for consonants the groove is much shallower in the back. In two-point displacement there is only one exemplar. The tongue progresses from a shallow groove to a deeper groove to a shallow arch. Naturally, differences within a category occur, since the sounds are distinctive, but the shape patterns cluster by category.

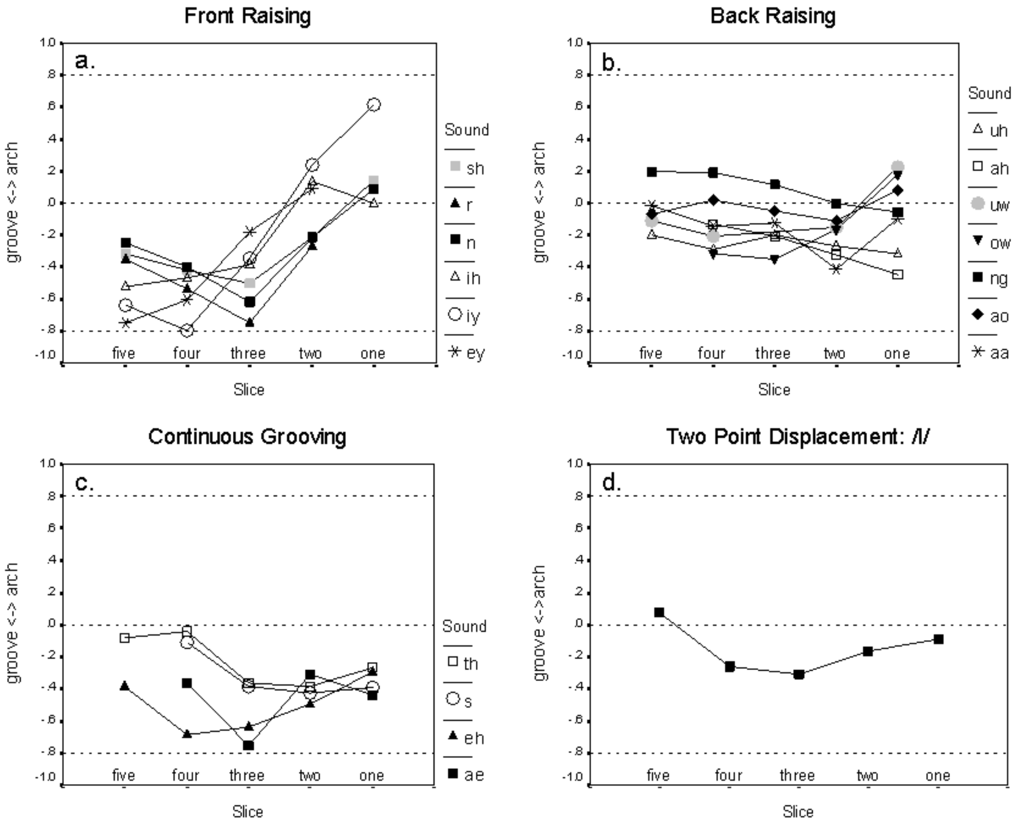

Figure 3. Left slopes (y-axis) for the coronal slices for each of the phonemes in the data set. Slice one is the most anterior; i.e. tongue tip is on right. Phonemes are grouped by tongue surface shape category.

This is a limited dataset, but it appears that midline displacement and category allegiance combine to provide more detailed information about 3D shape. Displacement value indicates arching or grooving. Knowledge of shape category and slice position may indicate the steepness of the groove or arch. For example, midline displacement at slice 4 was 70.8 mm for both /e/ and /u/. Both are grooved (*i.e.* below 82 mm) but, as predicted by category affiliation, there is a steeper slope for the front-raised /e/ (.60) than the back raised /u/ (.21). This is a first step toward differentiating specific phonemes.

CONCLUSIONS

This study asked whether midsagittal displacement data could predict coronal slope signatures representative of 3D surface shape, and whether the slope signatures could be mapped to the four tongue shape categories. It was found that in all segments of the tongue there was a transitional y-value above which the tongue shape was arched and below which it was grooved midsagitally. Thus general prediction of shape could be made. Better predictions are possible for the three anterior slices, where there was a strong correlation between midline displacement and groove depth - to - arch height. Coronal shapes clustered by category and within category (consonant vs. vowel). Knowledge of both category and midsagittal displacement allowed fairly good prediction of coronal shapes. These possibilities need to be explored further.

ACKNOWLEDGEMENTS

This study was supported in part by Grant No. R01-DC01758 from the National Institute of Deafness and Other Communication Disorders, NIH.

REFERENCES

Harshman, R.A., Ladefoged, P. & Goldstein, L. (1977) "Factor analysis of tongue shapes" *Journal of the Acoustical Society of America* **62** 693-707.

Lundberg, A. & Stone, M. (1999) "Three-dimensional tongue surface reconstruction: Practical considerations for ultrasound data" *Journal of the Acoustical Society of America* **106** 2858-2867.

Maeda, S. (1990) "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in W.J. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling*, The Netherlands: Kluwer, pp. 131-149.

Slud, E., Smith, P., Stone, M. & Goldstein, M. (2002) "Principal components representation of the two-dimensional coronal tongue surface" *Phonetica* **59** 108-133.

Stone, M. (1990) "A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data" *Journal of the Acoustical Society of America* **87** 2207-2217.

Stone, M. (1995) "How the tongue takes advantage of the palate during speech," in F. Bell-Berti & L.J. Raphael (eds.), *Producing Speech: Contemporary Issues: A Festschrift for Katherine Safford Harris*, New York: American Institute of Physics, pp. 143-153.

Stone, M., Faber, A., Raphael, L.J. & Shawker, T.H. (1992) "Cross-sectional tongue shapes and linguopalatal contact patterns in [s], [ʃ], and [l]" *Journal of Phonetics* **20** 253-270.

Stone, M., Goldstein, M. & Zhang, Y. (1997) "Principal component analysis of cross-sectional tongue shapes in vowels," *Speech Communication* **22** 173-184.

Stone, M. & Lundberg, A. (1996) "Three-dimensional tongue surface shapes of English consonants and vowels" *Journal of the Acoustical Society of America* **99** 3728-3737.

Stone, M., Shawker, T., Talbot, T. & Rich, A. (1988) "Cross-sectional tongue shape during the production of vowels" *Journal of the Acoustical Society of America* **83** 1586-1596.

Stone, M. & Vatikiotis-Bateson, E. (1995) "Trade-offs in tongue, jaw and palate contributions to speech production" *Journal of Phonetics* **23** 81-100.