18

Predicting 3D Tongue Shapes from Midsagittal Contours

MAUREEN STONE, MELISSA A. EPSTEIN, MIN LI, AND CHANDRA KAMBHAMETTU

ABSTRACT

This study is interested in whether there exists a predictable relationship between the midsagittal tongue contour and its related 3D tongue surface shape during speech. The assumption is that for any single language, a limited set of phonemically based 3D tongue shapes are used. If these shapes can be delineated and mapped to specific midsagittal displacements and cross-sectional (coronal) shapes, then predictions from midsagittal displacements to coronal shapes and then to 3D shapes can be made for specific speech sounds. The present study examined two ultrasound data sets: (1) the 3D static tongue surface reconstructions from a single subject (Stone & Lundberg, 1996), and (2) five coronal slices for two sentences spoken by a second subject (Yang & Stone, 2002). The midsagittal-to-coronal relationship for the 3D surfaces was extracted and applied to the continuous speech data. The predicted midsagittal-to-coronal relationship was well captured. This result supports the idea that a knowledge of the 3D shapes of a language, even based on a single speaker, can then be used to transform 2D midsagittal data into a 3D surface for other data sets.

INTRODUCTION AND BACKGROUND

Any single language has a finite number of lingual phonemes and thus a limited set of 3D static tongue surface shapes. An inventory of these 3D tongue surface shapes can be made for all the lingual phonemes of the language. Each 3D surface can be deconstructed into a “chain” of 2D coronal contours, and a complete map of 2D-contour to 3D-surface shapes can be made. The number of coronal contours required to adequately represent the 3D surface can be fairly small; it has been found that six coronal contours well represent 3D surface shapes in English (Stone & Lundberg, 1996).

Although the relationship between 3D surfaces and 2D coronal contours is straightforward, only a few instruments are able to directly collect coronal contours and thus reconstruct 3D tongue surface motions (c.f. Yang & Stone, 2002). Many instruments, however, can collect midsagittal displacement data. Midsagittal contours can be deconstructed into a chain of midsagittal points in the same way that 3D surfaces can be deconstructed into a chain of coronal contours. Five to six appropriately placed points well represent the midsagittal contour (Lundberg & Stone, 1999). Moreover, coronal contours measured at these points can be reconstructed into low-error representations of the 3D surface. Therefore, the first goal of this paper is to determine whether there is a straightforward relationship between 2D coronal tongue shapes and their 1D midsagittal displacements (hereafter, midline heights).

The second goal of this work is to determine whether the 1D to 2D mapping of steady-state sounds can be used to estimate a continuous speech data set. Tongue shapes are more variable for continuous speech than steady-state sounds, due to coarticulation, prosody, and speaking rate, to name a few effects. However, if the extreme tongue shapes are known, it is possible that the transitional shapes will be contained within that range. Figure 18.1 (top) shows four

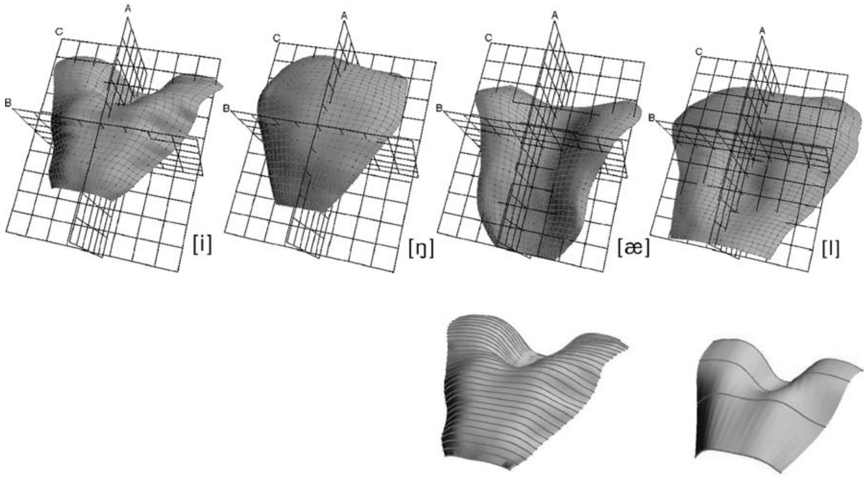

FIGURE 18.1. Top: Representative 3D tongue surface shapes based on 55 coronal contours for the four shape-based categories: front-raising (/i/), back-raising (/ŋ/), continuous groove (/æ/), and two-point displacement (/l/)); Bottom: reconstructions for /i/ based on 55 slices and 4 slices.

extreme tongue shapes for US English. The surfaces can each be described, or mapped, as a chain of assorted coronal arches and grooves. The key to using these maps in a predictive manner is the relationship between the midline heights and the coronal shapes at each tissue slice. The relationship needs to be physiologically constrained, which would create universal limits, or linguistically constrained so that it holds for all data in the language. Badin, Bailly, Reveret, Baciu, and Segebarth (2002) found that tongue surface shape was linked to midsagittal contour shape as well as to jaw height. They studied the tongue tip, body, dorsum, root (advancement), and jaw height using 3D tongue surfaces reconstructed from MRI data. Using Linear Component Analysis they found that jaw height had a predictable effect on tongue surface shape because a higher jaw increases tongue–palate contact. In addition, they found that 72% of the variance of the 3D geometric features was accounted for by midsagittal contours.

In the present study, Experiment 1 used 18 3D static tongue surfaces, represented by six coronal contours, to map the relationship between midline heights and related coronal shapes (data from Lundberg & Stone, 1999; Stone & Lundberg, 1996). Coronal shape was represented by two quantities: (1) left slopes and (2) quadratic fits. Correlations between midline height and these two quantities determined whether there was a sufficient relationship to justify using this map in Experiment 2. Experiment 2 explored whether the map could be used to estimate coronal shapes from midline heights for continuous speech from a second speaker (data from Yang & Stone, 2002). ~~RMS~~ errors in Experiment 2 were calculated between the coronal contour, the true quadratic fit, and the estimated quadratic fit, and were used to determine the quality of the estimations.

AQ1

EXPERIMENT 1: MAPPING MIDLINE HEIGHTS TO CORONAL SHAPES FOR STATIC 3D SURFACES

In order to determine the relationship between midline height and coronal shape, it is useful to understand the nature of coronal shapes. Coronal shapes primarily follow a continuum from midsagittal groove to midsagittal arch, with some lateral asymmetry (Slud, Smith, Stone, & Goldstein, 2002; Stone, 1990, 1995; Stone, Faber, Raphael, & Shawker 1992; Stone, Goldstein & Zhang, 1997; Stone, Shawker, Talbot, & Rich, 1988; Stone & Vatikiotis-Bateson, 1995). Arches are typically quadratic (concave down) in shape, sometimes with a level or grooved centre. Anterior arches occur when the tongue contacts the hard palate and the sides deform against it. Posterior arches are due to muscle activity that elevates the tongue upward and backward as in vowels, or when it contacts and takes the shape of the velum (/ŋ/). Grooves are usually characterized by a bimodal shape. Maximum tongue height of the two peaks is reached some distance lateral to midline after which the tongue typically angles downward

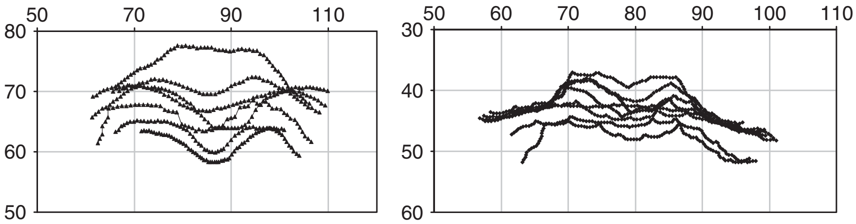

FIGURE 18.2. Representative contours for the third coronal slice for the static data set (left) and for the sentence "it ran a lot" (right).

(see Figure 18.2). Anterior grooves occur when genioglossus anterior contracts and lowers the anterior tongue, as in low vowels (/æ/,/ɛ/), or when the tongue is high and braced against the palate as in grooved fricatives (/s/,/θ/). Posterior grooves are muscularly determined as when genioglossus posterior pulls the tongue forward or when styloglossus pulls it up. Lateral asymmetries can occur because of palatal asymmetry, asymmetrical tongue shape, or motion.

The lateral tongue margins have a slightly different function in speech from the midsagittal tongue: They stabilize the tongue as well as shape the airstream. Therefore, lateral tongue motion may not correlate well with midsagittal motion. Alternatively, it may correlate very well because of the physiological linkages between the surface tissue points and the invariant boundary conditions of the vocal tract, especially the hard palate.

Methods–Experiment 1

Data set 1, spoken by a 24-year-old white, female, native of Baltimore, MD, consists of 3D tongue surfaces for 18 static sounds of US English. These 18 surfaces represent all the lingual sounds of English; for example, homorganic sounds like /t,d,n/ are represented by /n/. The original surfaces consisted of 55 coronal ultrasound slices. Figure 18.1 (bottom) shows the 55 slice and the 4 slice reconstruction for /i/, which is a short ~~contour~~ ^ with no data in Slices 1 and 6. Full methodology and additional detail can be found in Stone and Lundberg (1996). To determine a small, but efficient number of slices to be used in future reconstructions, that experiment examined sparse data sets consisting of three to nine slices, which were optimized to maximize surface coverage and minimize error and improvement in the cost function asymptoted at five to six slices. The six slice reconstructions had 80% coverage, a mean error of 0.21 mm, and a maximum error of 1.40 mm. Since additional slices added very little improvement to the cost function and added considerable time to collect, we used six-slice coronal sets for 3D surface reconstructions. Some phonemes had shorter tongue surfaces than others and were not captured in the first or last slices, usually the first.

Therefore, for Experiment 1, Slices 2–5 were used from the sparse data set, hereafter called Data set 1.

Four shape-based categories were defined from the data: front-raising (FR) /i, ɪ, e, ɝ, n, ʃ/, back-raising (BR) /u, ʊ, o, ɔ, ɑ, ŋ/, continuous-groove (CG) /æ, ɛ, s, θ/, and two-point displacement /l/ (see Figure 18.1, and Stone & Lundberg, 1996). The original study found that low front vowels, which at midline appear to be weak versions of FR (Harshman, Ladefoged, & Goldstein, 1977), actually have a midsagittal groove for the entire tongue.

Three measurements were made on each coronal contour of the sparse data set: midsagittal height, left slope, and quadratic fit. The midsagittal height was chosen at the same *x*-value in every contour in the data set to simulate a single midsagittal slice, even those whose midpoints appeared off centre (see Figure 18.2). For Data set 1, the midline *x*-value is the plane made by axis A in Figure 18.1. For the left slope, the tongue contours were cut to a uniform length that captured the arching and grooving features. The end point was determined by overlaying the coronal contours for all the phonemes and slices (see Figure 18.2). The best value was about 1 cm from midline in the *x*-projection. Some phonemes, with wider or narrower grooves, would have been better represented by different lateral points, particularly in the posterior tongue. However, the chosen lengths captured the key shape features. The left slope was calculated by applying the equation $y =$ ~~$a1x + a0$~~ to the line connecting the midline point and left point.

The quadratic functions were fit to the entire contour length. It is not possible to estimate asymmetries when generalizing from midline contour to coronal shape, so only the arching/grooving features are considered here. In this way, for a given quadratic fit, where $y =$ ~~2 + $a1x$ + a0, the $a2$~~ term represents contour shape, the quantity of interest in this study. The sign and magnitude of $a2$ indicate, respectively, the direction and degree of global arching or grooving of the contour. AQ2

Pearson product moment correlation coefficients were calculated to determine the relationships between (1) midline height and $a2$ and (2) midline height and left slope, for all five slices.

Results and Discussion–Experiment 1

The question asked by this study was whether correlations exist between midsagittal height and coronal shape at each segment of the tongue. In each of the slices, greater midline height accompanied an arched coronal shape, and lower heights a grooved shape. For each coronal segment, a specific midline height demarcated the difference between an arched and grooved shape. Correlations for Slices 1–3 were consistent and fairly strong. The three anterior slices may have been more easily mapped because palatal contact causes shape constraints that increase the correlation between shape and height (see Table 18.1).

TABLE 18.1. Pearson Product Moment Correlation Coefficients between Midline Height/Left Slopes and between Midline Height/*a*2 of the Quadratic Function ($n=18$)

	Correlations—Data set 1	
Slice	Mid- to left slope	Mid to *a*2
1	.67	.65
2	.79	.73
3	.71	.70
4	.51	.65
5	.64	.44

This experiment was also designed to determine whether there was a unique map between each 3D ~~shape~~ category and its coronal shape chain; that is, whether shape differences were smaller for phonemes within than between categories. The shapes in Figure 18.1 represent category shapes measured for a single speaker of US English and observed in many others. Although phonemes within these categories differ somewhat, they form a family of similar shapes (Stone & Lundberg, 1996). Figure 18.3 graphs each phoneme in each category on the basis of midline height and left slope. Fairly good correlations can be observed at each slice. In addition, the data can be further grouped by shape category. In Slice 2, FR is distinguished from CG, on the basis of height and shape. In Slice 3, CG is distinguished from BR and in Slice 4, BR is distinguished from FR and CG. Thus at any single slice, some categories clustered on the basis of left slope and/or midline displacement. In addition, some within-category patterns emerged. BR was distinguished from the other groups by very displaced midlines and little-to-no grooving at Slices 3–5. FR was high only anteriorly. Thus, the chains of contour shapes define 3D shape for each category, and further indicate 1D to 2D relationships that could be used for estimation. Similar patterns were seen for the quadratic functions; however, the left-slope data had slightly better correlations with midline height (Table 18.1); as a result the estimations in Data set 2 were done with cropped contours.

Data set 1 suggests that the information provided by midline height and category affiliation combine to provide substantial information about 3D shape. A higher anterior tongue (above 82 mm for this subject) indicates arching, a lower one grooving. Knowledge of shape category and slice position may further detail the steepness of the groove or arch. For example, midline displacement at Slice 4 was 70.8 mm for both /ɛ/ and /o/. Both are grooved (i.e., below 82 mm) but, as predicted by category affiliation, there is a steeper slope for the front-raised /ɛ/ (0.60) than the back-raised /o/ (0.21). In addition, back-raised sounds had a global coronal shape that was arched. This was not captured by the cropped contours; instead, the central groove was measured (see Figure 18.2).

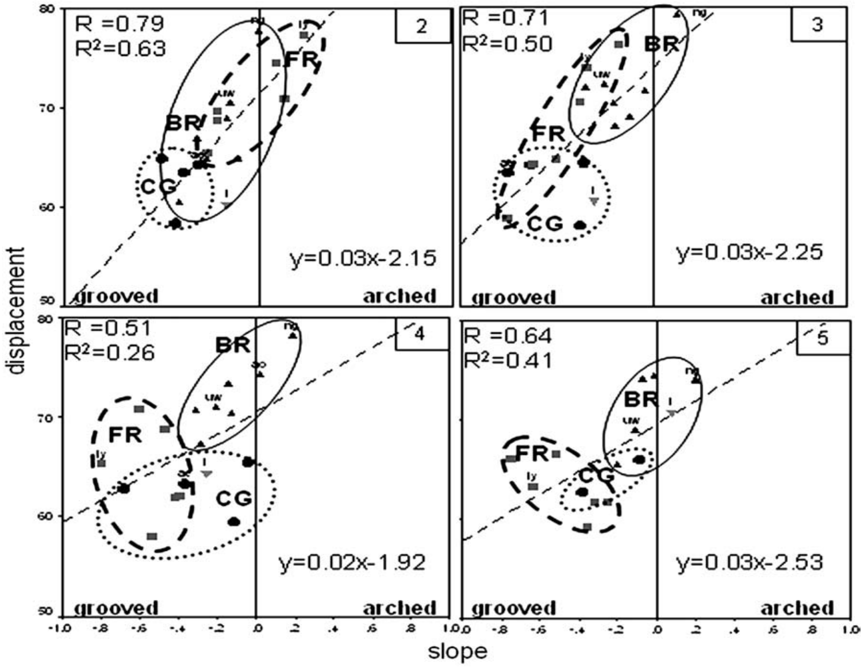

FIGURE 18.3. Correlations between left slope (x-axis) and midline height (y-axis) at the five tongue slices for all the static sounds. Categories are circled: front-raising (squares); back-raising (upright triangles); continuous-groove (circles); two-point displacement (inverted triangle).

For certain categories and slices, particularly BR and the velar slices, quadratic fits of the entire contour are a valuable shape supplement to left slopes.

EXPERIMENT 2: PREDICTING CORONAL SHAPES FROM MIDLINE HEIGHTS USING A PREDETERMINED MAPPING

There is more variability in tongue surface motions than in static shapes due to nonlinear tongue motion, and nonuniform front back timing (c.f. Yang & Stone, 2002). These factors make time-varying data sets more complex than static ones. Experiment 2, therefore, considered whether coronal shapes could be estimated, rather than mapped, from midline heights. In other words, whether the left slope or quadratic fits derived from 3D static shapes could be used in continuous speech to estimate coronal shapes from midsagittal heights. The strong correlations for the left-slope data suggested that the estimation should be for the medial 2 cm of the data. However, because the left-slope data were based on only two points (midline and 1 cm to the left), “cropped” data sets were made from Data set 1. The cropped data sets contained all the intervening points between the midline and left point, which was 2.0 cm long in the *x* projection. Quadratic functions were calculated for this cropped data set. The contours of Data set 1 were made symmetrical by measuring the points on the left side of the cropped contour and creating a mirror image on the right side. The symmetric contour was transformed so that its centre was 0, which means that its quadratic fit, *y* = ~~*a*2×2 + *a*1*x* + a0,~~ will have a midline height (*a*0) and a midline slope (~~*a*1)~~ of 0. This transformation made it possible to represent the quadratic fit entirely by its ~~*a*2~~ value.

Estimating coronal shape for Data set 2 from the cropped ~~*a*2~~ terms would be more successful if the height/shape correlations are strong and in the same direction for both data sets. In Data set 1, the relationships were strong anteriorly and moderate posteriorly between specific midline heights and left slopes. Given linguistic or physiological constraints as contributors to this relationship, it should be possible to take the new cropped ~~*a*2~~ values for Data set 1, insert the midline heights of Data set 2, and solve the equations to get estimated ~~*a*2~~ values for Data set 2. The transitional movements occurring in the sentences would hopefully follow the same height/shape relationships as the extreme positions.

Methods–Experiment 2

The speaker was a 31-year-old white, female, native of Toronto, Canada. The sentences “it rang a lot” and “it ran a lot” were spoken ~~five~~ times each while coronal ultrasound data sets were collected at six coronal slices of the tongue. The slice angles were not optimized, but were roughly equidistant. The coronal video sequences were temporally aligned using the acoustic wave so that comparable moments in the speech event would occur in the same frames (Yang & Stone, 2002). As in Data set 1, the anterior most slice had incomplete data due to occasional tongue retraction. In addition, the most posterior slice moved little and appeared to be the jaw muscles. Therefore, the medial four slices (2–5) were used in the experiment. Tongue segments 2–5 of Data set 2 corresponded roughly to Segments 2–5 of Data set 1 and estimations were made accordingly.

For Data set 2, the left tongue point was about 0.7 cm from the midline in the *x* projection. Differences from midline to left points between the two studies (0.7 cm vs. 1.0 cm) are probably due to different tongue sizes, speech tasks, subject/dialect, and possibly transducer locations.

The midline height was measured for each symmetrical contour at each slice, and left slopes and cropped quadratic fits were calculated. Correlations determined that the relationship between midline height and the two shape measures were quite similar since they covered the same end points (see Table 18.2). Subsequent analyses, therefore, were done on the quadratic fits only since they used the intervening points. The estimated quadratic fits were

TABLE 18.2. Pearson Product Moment Correlations between Midline Height and Coronal Shape Measures for both "Ran" and "Rang"

Ran	Correlations—Data set 2	
Slice	Mid- to left slope	Mid to *a*2
2	.66	.63
3	.05	.04
4	.94	.94
5	.88	.87
Rang		
2	.76	.72
3	.44	.44
4	.86	.84
5	.82	.80

calculated using y = ~~$a2\times 2 + a1x + a0$,~~ where the *a* values were from Data set 1. To determine the quality of the estimated shapes for Data set 2, the following RMS errors were calculated for each frame in the two sentences:

RMS1: the true quadratic fit and the true symmetric contour
RMS2: the estimated quadratic fit and the true quadratic fit
RMS3: the estimated quadratic fit and the true symmetric contour

Results and Discussion–Experiment 2

Table 18.2 shows a strong relationship between midline height and ~~*a*2,~~ except at Slice 3. The relationship, as with Data set 1, was that a high tongue was more arched and a low tongue more grooved. At Slice 3, however, the correlations were weak, especially for the 'ran' data set. Figure 18.4 shows the poor correlation between midline height and ~~*a*2~~ for Slice 3 of "ran" and the strong correlation for Slice 4 of "ran". The dashed line is the best-fit line and the solid line represents the linear formula obtained from Data set 1 and used to estimate the ~~*a*2~~ values from the midline heights~~.~~

Two features may contribute to poor correlations for Slice 3; both are due to its location in the palatal vault/velar region of the vocal tract where there is space for much tongue elevation. The first feature is that, posterior tongue elevation, which is critical to creating high back sounds, is usually overlaid by a flat or grooved region at midline. The tongue is checked from its upward motion, probably by genioglossus, which creates this flat or grooved region (see Figure 18.2). Contact with the palate will also flatten out the upper surface. The cropped ~~*a*2~~ values captured the central shape features, which did not correlate with midline height. A full-contour quadratic might have better correlated with the globally arched shapes.

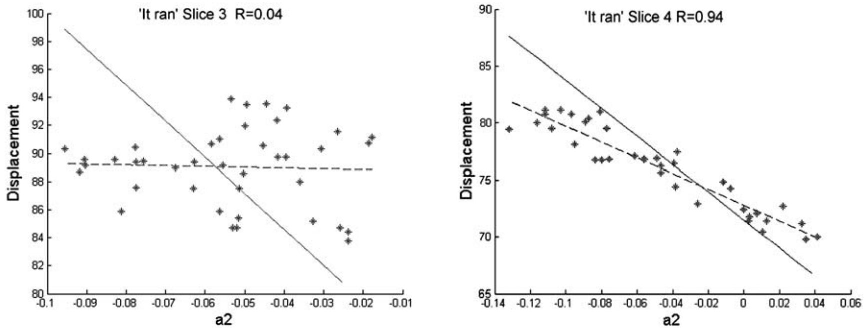

FIGURE 18.4. Correlation between midline height and *a*2 for Slices 3 and 4 of "ran".

The second reason for low height/shape correlation at Slice 3 may be due to the inclusion of motion in Data set 2. The region of the third slice often is the pivot point for front–back tongue rotation (Iskarous, 2001; Unser & Stone, 1992). As the pivot point, Slice 3's heights and shapes are influenced by Slices 2 and 4, whose oppositional motions may create a more variable height-to-shape relationship. Front–back rotation occurs in continuous speech, but not in static data, so Slice 3 had stronger correlations in Data set 1 (see Table 18.1).

Estimated shapes correlated strongly with the true shapes if the height/shape correlations in Data set 2 were high (Table 18.2). Figure 18.5 details the variability in the quality of the estimations. Slice 3 had a very low height/~~*a*2~~ correlation for "ran" (see Figure 18.4) and a somewhat better one for "rang" (*r*=.44). The ~~*a*2~~ values for the individual frames show that for "rang" the

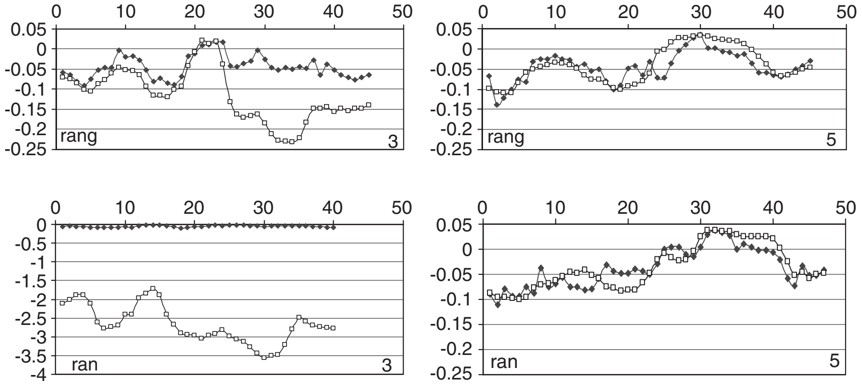

FIGURE 18.5. Comparison of predicted and original coronal shapes for slices 3 (left) and 5 (right) for the sentences "it ran a lot" (bottom) and "it rang a lot" (top). Original shape is represented by white squares; estimated shape is represented by black diamonds.

latter frames had little shape variation in the true data, but large variation in the estimated shapes. This means that the shapes were similar for these frames, but the midline heights varied considerably resulting in similar shape estimates (see Figure 18.2). Slice 5 had stronger correlations, and the similarity between true and estimated ~~*a*2~~ values is apparent.

RMS errors of Slices 2–5 for "ran" are shown in Figures 18.6–18.8, and Slices 2–5 for "rang" are shown in Figures 18.9–18.12. The low RMS1 errors indicate that the true quadratic fit is a good representation of the true symmetric contour. The similarity between true quadratic and true contour explains the similarity of the RMS2 and RMS3 errors, which respectively compare these two measurements to the estimated quadratic. Maximum RMS2 and RMS3 errors are about 2.2 mm for both data sets. The time frames with maximum RMS errors seem to be random. Mean RMS2 and RMS3 errors, shown in the figure legends, is less than 1 mm, which indicates fairly successful estimation of shape from height.

The goal of this study was to determine whether 3D surfaces could be predicted from midline contours, since midline data are so much more prevalent than coronal or 3D data. To do this for any specific language one would collect a static 3D corpus, determine shape measurements at reasonably placed (if not

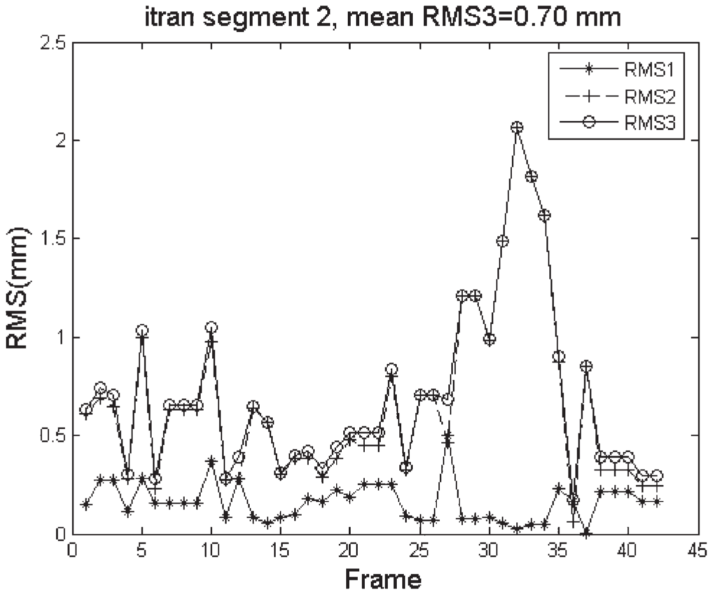

FIGURE 18.6. RMS errors of Segment 2 of "ran". RMS1 error compares the true symmetric contour to the true quadratic fit. RMS2 error compares the estimated quadratic fit to the true quadratic fit. RMS 3 error compares the estimated quadratic fit to the true symmetric contour.

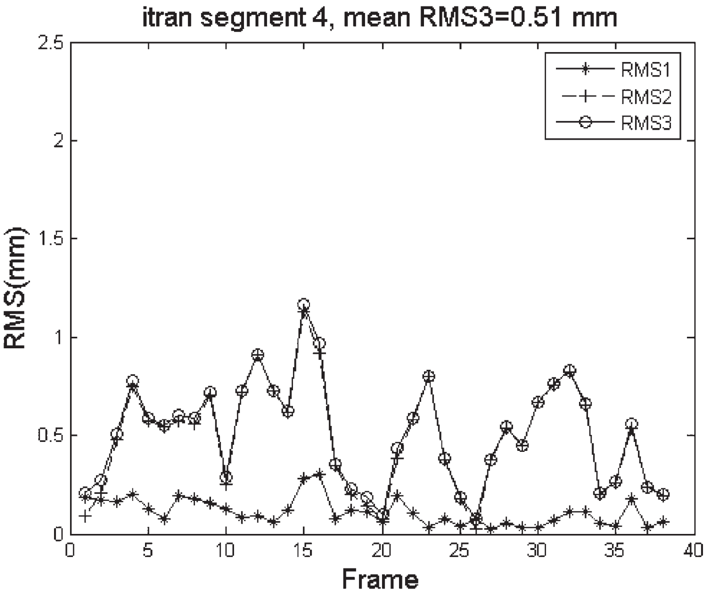

FIGURE 18.7. RMS errors of Segment 4 of "ran".

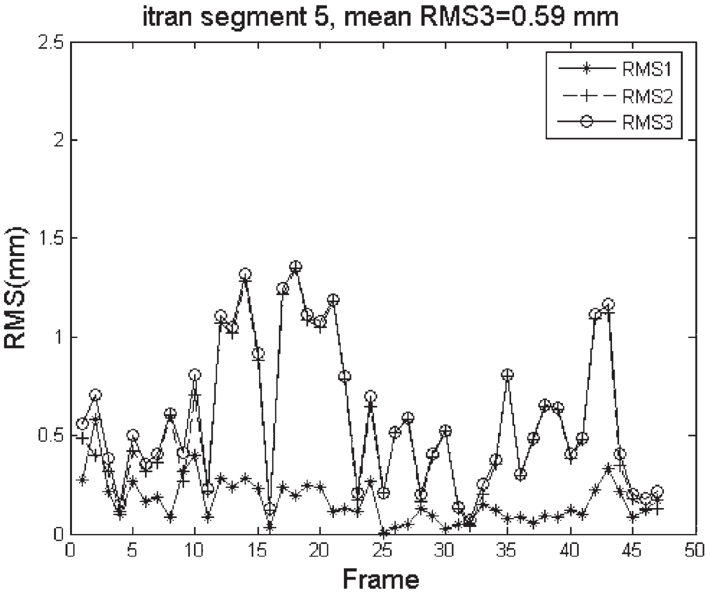

FIGURE 18.8. RMS errors of Segment 5 of "ran".

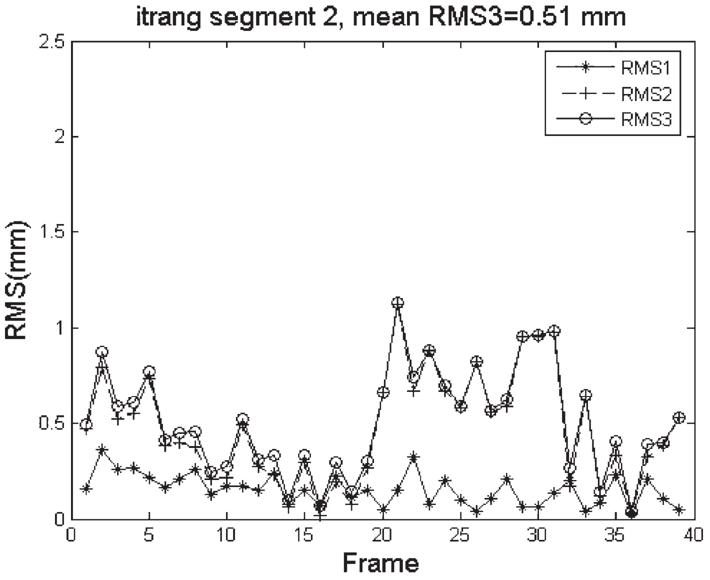

FIGURE 18.9. RMS errors of Segment 2 of “rang”.

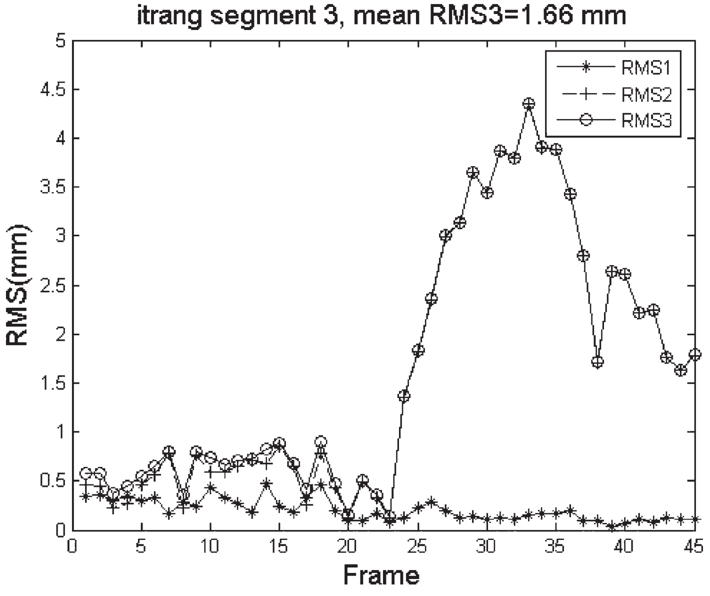

FIGURE 18.10. RMS errors of Segment 3 of “rang”.

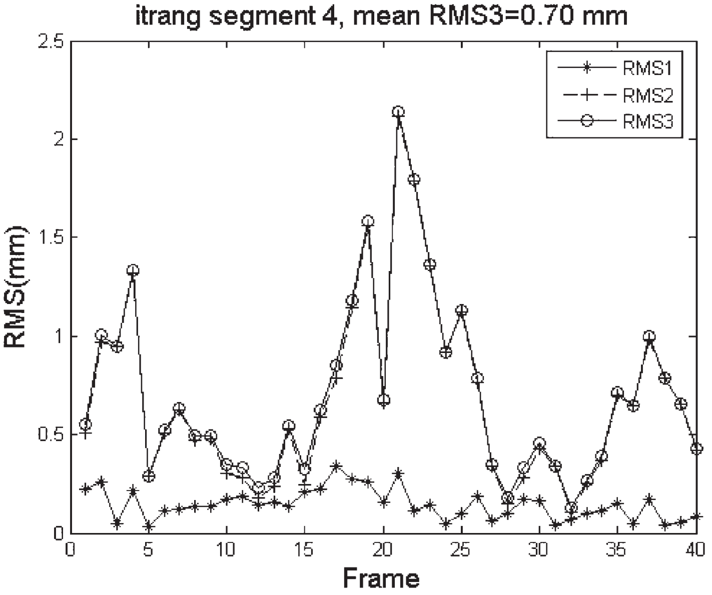

FIGURE 18.11. RMS errors of Segment 4 of "rang".

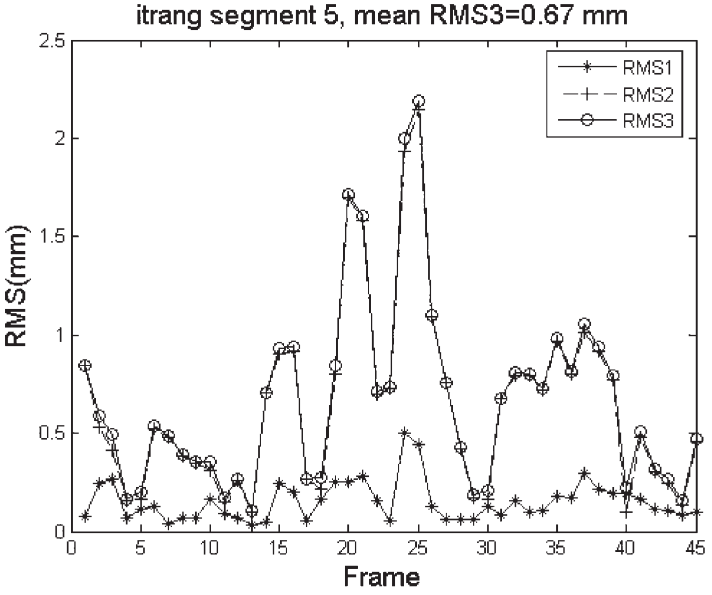

FIGURE 18.12. RMS errors of Segment 5 of "rang".

optimized) coronal slices, and calculate the linear relationships between midline height and left slope or quadratic fit. These linear and quadratic correlations could be used to predict coronal shapes at similar locations for other subjects for whom only midsagittal data are available. Chaining of these coronal shapes from front-to-back would yield a 3D tongue surface. Part of the convenience of this prediction is that the initial data set is static and can be collected via ultrasound or MRI using off-line and, if necessary, time-consuming 3D reconstructions. The subsequent data sets can be collected at midline using time-varying speech materials. The good correlations for most slices indicate that knowing the shape category of the initial sounds may not be critical.

CONCLUSIONS

This study asked whether coronal shapes could first be mapped to and then estimated from midline height data, and whether chains of coronal shapes characterized the four tongue shape categories. It was found that, in general, there was a transitional *y*-value at each coronal slice, above which the tongue shape was arched and below which it was grooved. Thus, general estimation of shape could be made. Better estimations are possible with better correlations in both data sets. At the hard palate, there was a strong correlation between midline height and groove shape due to the tongue taking the palate's shape on contact. In the velar region, however, tongue rotation and large elevations reduced the accuracy of shape estimation and quadratic fits of the entire contour are recommended as a supplementary shape parameter. Considering that these predictions were based on two different subjects, nonidentical tongue slices and different speech materials, the results are very promising for 3D tongue shape estimation from midline height.

ACKNOWLEDGMENTS

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

REFERENCES

Badin, P., Bailly, G., Reveret, L., Baciu, M., & Segebarth, C. (2002). Three dimensional articulatory modeling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*, *30*, 533–554.

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

Iskarous, K. (2001). *Dynamic acoustic–articulatory relations.* Ph.D. dissertation, University of Illinois at Urbana-Champaign.

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.

AQ3 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* (pp. 131–149). Dordrecht, The Netherlands: Kluwer.

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, & Raphael, L. J. (Eds.), *Producing speech: Contemporary issues: A festschrift for Katherine Safford Harris* (pp. 143–153). New York: American Institute of Physics.

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.

Unser, M., & Stone, M. (1992). Automated detection of the tongue surface in sequences of ultrasound images. *Journal of the Acoustical Society of America*, *91*, 3001–3007.

Yang, C. S., & Stone, M. (2002). Dynamic programming method for temporal registration of three-dimensional tongue surface motion from multiple utterances. *Speech Communication*, *38*, 199–207.

QUERY FORM

CRC Press

Speech Production/Harrington

JOURNAL TITLE:	**SP-Tabain**
ARTICLE NO:	**CH018**

Queries and / or remarks

Query No	Details required	Author's response
AQ1	Is the abbreviation RMS well known. If not, please provide expansion.	
AQ2	In the equation $y=a2\times2+a1x+a0$, please confirm the change from $a2x2$ to $a2\times2$.	
AQ3	Maeda (1990) is not cited in the text. Please cite the reference in the text.	