Modeling Tongue Surface Contours from Cine-MRI Images

Maureen Stone
Division of Otolaryngology, University of Maryland School of Medicine
16 S. Eutaw Street, Room 525, Baltimore, MD 21201
Tel. (410) 328-1888, FAX: (410) 328-5827,
E-mail: mstone@umaryland.edu

Edward P. Davis
Andrew S. Douglas
Dept. of Mechanical Engineering, Johns Hopkins University
3400 N. Charles St., Baltimore, MD, 21218

Moriel NessAiver
Rao Gullipalli
Department of Radiology, University of Maryland School of Medicine,
22 N. Greene St., Baltimore MD, 21201

William S. Levine
Dept. of Electrical Engineering, University of Maryland
A.V. Williams Bldg., College Park, MD, 20740

Andrew Jon Lundberg
Dept. of Computer Science, Johns Hopkins University
3400 N. Charles St., Baltimore, MD, 21218

Published in:
JOURNAL OF SPEECH, LANGUAGE, HEARING RESEARCH
OCTOBER, 2001

Abstract

This study demonstrated that a simple mechanical model of global tongue movement in parallel sagittal planes could be used to quantify tongue motion during speech. The goal was to represent simply the differences in 2D tongue surface shapes and positions during speech movements and in sub-phonemic speech events such as coarticulation and left-to-right asymmetries. The study used tagged Magnetic Resonance Images to capture motion of the tongue during speech. Measurements were made in three sagittal planes (left, midline, right) during movement from consonants (/k/, /s/) to vowels (/i/, /A/, /u/). MR image-sequences were collected during the C-to-V movement. The image-sequence had 7 time-phases (frames), each 56 ms in duration.

A global model was used to represent the surface motion. The motions were decomposed into translation, rotation, homogeneous stretch and in-plane shear. The largest C-to-V shape deformation was from /k/-to-/A/. It was composed primarily of vertical compression, horizontal expansion, and downward translation. Coarticulatory effects included a trade-off in which tongue shape accommodation was used to reduce the distance traveled between the C and V. Left-to-right motion asymmetries may have increased rate of motion by reducing the amount of mass to be moved.

Models of tongue motion and deformation differ widely depending on the goal of the model. Highly detailed physiological models (Wilhelms-Tricarico, 1995) are used to describe tongue physiology and function. Simpler, less detailed models are typically used in vocal tract models (cf., Kroger, 1992; Maeda, 1991; McGowan & Cushing 1999). Two-dimensional (2D) tongue motion has been modeled using simple interpolations between extreme phoneme positions (Mermelstein, 1973). Three-dimensional (3D) tongue motion is more complex (Wilhelms-Tricarico, 1995). To represent 3D tongue movement with only a few parameters, however, would be very valuable. Such models would add depth to vocal tract models and simplify the representation of tongue function in normal and disordered speech.

The goal of the present line of research is to develop simple models of 3D-tongue deformation to provide insight into speech motor control and oral function. The models should be as simple as possible, retain physiological accuracy, and infer muscle involvement of the tongue with little a priori expectation. Previously, dynamic models have been used to describe tongue surface deformation as a function of muscle contraction (Dang & Honda, 1997; Honda, Himi, & Dang, 1994; Ong & Stone, 1998; Perkell, 1974; Sanguinetti et al., 1997; Sanguinetti, Laboissiere, and Ostry, 1998; Wilhelms-Tricarico, 1995). The present work uses a novel approach. A global 2D mechanical model is applied to planar tongue surface (hereafter: tongue contour) motion observed in MRI image sequences. The aim is to determine a minimal meaningful set of parameters to describe the tongue contour. When deformable bodies, such as the tongue, move, this motion can be decomposed using continuum mechanics deformations. Zeroth order (or rigid body) deformation consists of uniform translation and rotation. First order (or homogenous) deformation is composed of uniform stretch and shear. Higher order (or inhomogeneous) deformations include bend and local concavities. This model used only the 2D zeroth and first

order deformations to describe tongue motion. The 6 parameters that describe these rigid body and homogeneous deformations essentially describe the global averaged motion. This allows a comparison of one deformation with another and one 2D region with another. In this approach each tongue slice is modeled as a single 2D object capable of only rotation, translation, and homogeneous stretch and shear. The deformations are homogeneous in that the stretch and shear values are assumed to be uniform throughout the tongue. The result is an organized global picture of tongue contour movement and deformation.

Tongue contour shapes are now available from MRI and ultrasound data. Recent MRI studies have reconstructed 3D static tongue and vocal tract shapes (Alwan, Narayanan & Haker, 1997; cf. Baer, Gore, Boyce & Nye, 1987; Baer, Gore, Gracco, & Nye, 1991; Moore, 1992; Narayanan, Alwan & Haker, 1997; Ong & Stone, 1998; Story, Titze, & Hoffman, 1996; Tiede & Yehia, 1996). MRI also has been used to capture tongue contour motion in a single plane (Crary, Kotzur, Gauger, Gorham, & Burton, 1996; Demoulin et al., 1997; Masaki et al., 1997). The present study builds on this work by examining time-motion tongue data from three sagittal slices, providing a more complete representation of the tongue volume than a single slice.

The process of spatially encoding the MRI signal requires multiple excitations and acquisitions (n=128 to 256). The time needed to create an MR image is typically slow, below the rates acceptable for speech motion. Motion can be examined, however, by using a technique consistent with ensemble averaging, that is "ensemble summation," instead of continuous summation. In ensemble summation, the subject repeats a motion multiple times, such as the enunciation of a syllable sequence. A short time-phase is measured out of the entire cycle yielding an image of that single moment in the cycle. This measurement is made in each repetition and the results accumulated. More than one time-phase can be measured in this manner allowing

reconstruction of the motion of the cycle. The cycles must be virtually identical to each other or the summed images are unclear. In MRI of the heart this is accomplished by timing the MRI time-phases to the R-wave peak of the ECG pulse. In speech the subject hears a metronome-like stimulus and strives to reproduce identical repetitions of the speech task.

Based on our previous work (Stone & Lundberg, 1996), the two consonants /k/ and /s/ were chosen for this study. The chosen consonants differ markedly in shape and position. The /s/ is a front consonant made with a lengthwise midsagittal groove. The /k/ is a back consonant with a lengthwise midsagittal arch. The point vowels, /i/ /A/ /u/, differ substantially in 3D shape and position. Tongue deformation was explored in left, middle, and right sagittal slices. It was expected that the varied shapes would produce a model capable of quantifying C-to-V motion as well as sub-phonemic events such as coarticulation and left-to-right (LR) asymmetries.

The question asked in this study is whether a global, homogeneous deformation model can resolve tongue shape changes that are due to C-to-V movement, and to the smaller shape effects of coarticulation and LR asymmetry.

Methods

Subject and Speech Material

The model was based on the speech of a 19-year-old male, native speaker of English. The subject had no dental fillings that might interfere with the MRI magnetic field. The speech materials were the six CV's /si,sA,su,ki,kA,ku/. Cine-MRI slice-sequences were collected at left, middle, and right sagittal slices. Each syllable was repeated once per second, 96 times in succession, that is, 32 times per slice. An ECG simulator emitted a beep to the subject at the R-wave peak of the simulated ECG signal. The simulator then triggered the tags 300 ms later. Images for the sounds in the midsagittal plane (summed from 32 repetitions) appear in Figures 1 and 2.

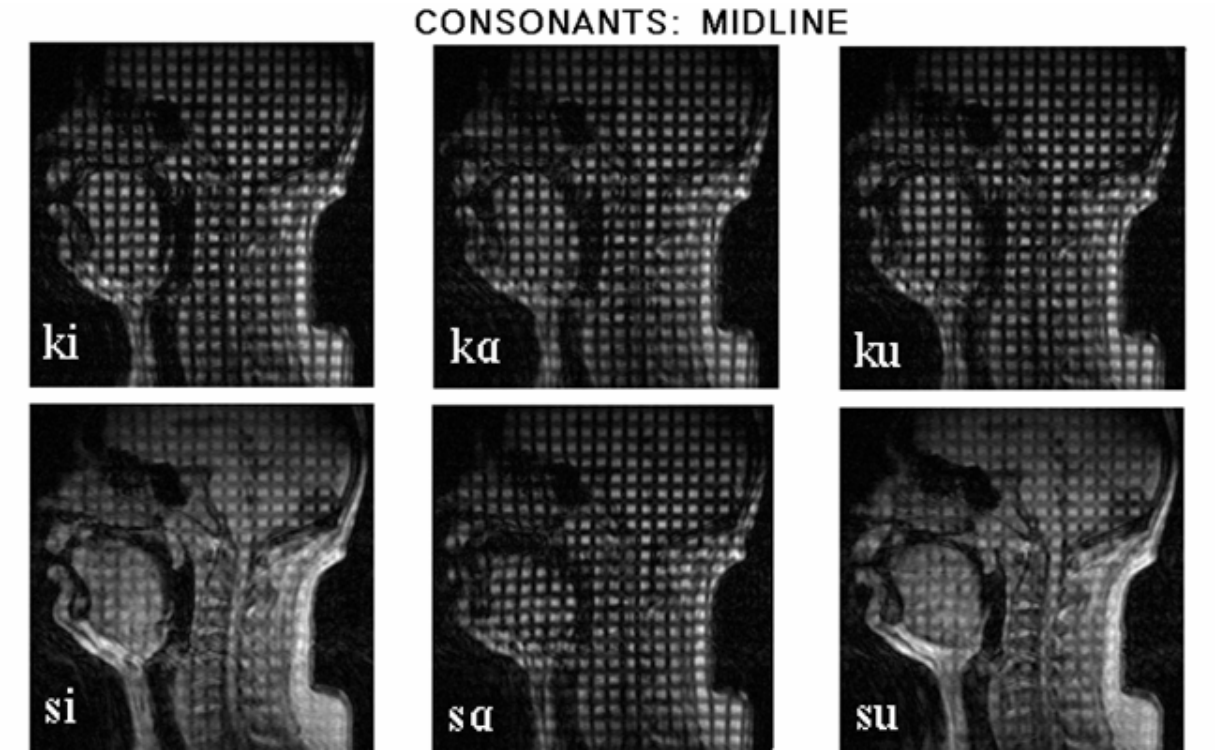

Figure 1. Raw MRI images of extreme consonants at midline. This image, and all images derived from the MRI data are oriented with the tongue tip facing left.

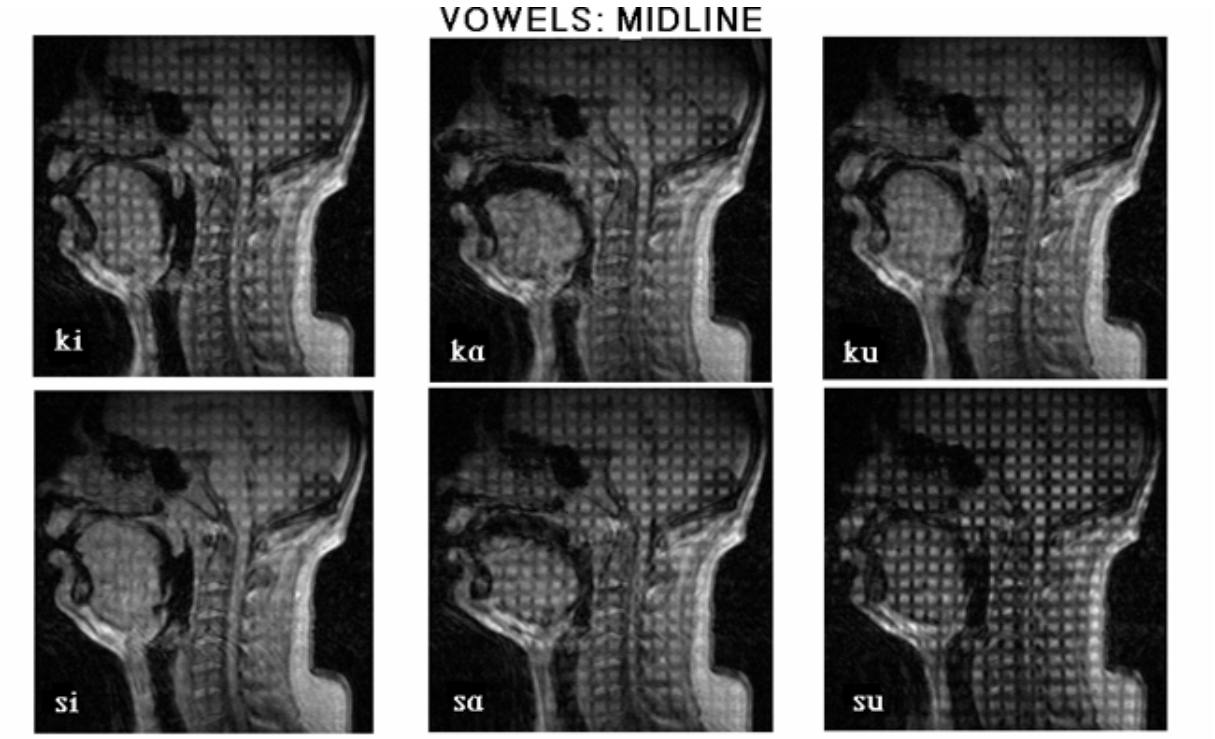

Figure 2. Raw MRI images of extreme vowels at midline.

Visual inspection of the MRI data during collection indicated that the tongue began moving from the consonant to the vowels in the second or third time-phase, (see Figure 3). Subsequent acoustic analysis of 16 repetitions showed vowel onset to be more than 300 ms after the cue. The data also indicated good precision across repetitions (small std. dev.'s). For the /k/ data, mean (std. dev.) vowel onset occurred at: /ki/-476(20) ms; /kA/-534(37) ms; and /ku/ 462(28) ms. For the /s/, vowel onset was at /si/-568(38) ms; /sA/-643(38) ms; /su/-544(34) ms. The /s/ onset was at: /si/-374(35) ms; /sA/-461(41) ms; /su/-331(46) ms. The /s/ duration was /si/-193(17) ms, /sA/-183(24) ms, and /su/-213(34) ms. The /sV/ syllable durations were /si/-483(20) ms, /sA/-465(22) ms, and /s/-573(21) ms. Therefore, the 300 ms delay between presenting the acoustic cue and triggering the tags placed the tag onset prior to the consonants. The lack of motion between the first and second time-phase indicates that the tongue was in consonant position before the syllable

began.

Instrumentation and Data Collection

A Picker 1.5 Tesla Edge System was used to collect the data. The data, available to us from another study (Stone, et al., to appear), were acquired using tagged Cine-MRI. This procedure not only "tags" the tissue, but also allows the collection of time-phases (frames) throughout a cyclic motion. Thus we were able to collect seven moments in time during the CV syllable. The present study did not make use of the tags, but the tongue surface contours, which appeared with excellent resolution, were extracted at each time-phase (Figures 1 & 2).

The subject was positioned in a supine position in the MRI scanner with the neck coil positioned to image the area from the lower nasal cavity to the upper trachea. Comparisons of tongue measurements made in upright and supine position during speech have found postural changes in the non-critical regions of the tongue as a function of position (Tiede, Masaki, Wakumoto & Vatikiotis-Bateson, 1997, Tiede, Masaki and Vatikiotis-Bateson, 2000). The present tongue shapes, therefore, as with all MRI studies, will likely be colored by the location of gravity.

Spatial resolution of the acquired MRI image was a 128 x 128 pixel grid. Resolution was increased to 256 x 256 in the raw image by zero filled Fourier interpolation. Pixel to mm conversion was calculated by dividing the field of view (240 mm) by the number of pixels (256). One displayed pixel, therefore, equaled .95 mm. Thus the minimum resolvable movement, absent gray scale, was 2 pixels or 1.9 mm.

Tagged Cine-MRI involves tagging tissue and then recording multiple frame, or cine, data. An abbreviated explanation follows of the cine recording procedure and its resultant frames, or time-phases. The tagging component of the recording procedure is not discussed here, since it is not germane to this paper (only surface contours and multiple time-phases are used). For a

detailed explanation of the tagging, see Stone et al., (to appear). The syllable is repeated by the speaker 96 times at a rate of one per second. Each second is divided as follows. A pulse is generated by the ECG signal generator used in cardiac studies. The pulse triggers the acoustic cue to the subject, and 300 ms later, the application of the tags to the tissue. The tags decay throughout the next 500-600 ms. During this time, 7 time-phases are sampled, each collected over a 56 ms time period, or 392 ms total (this is equivalent to 18 samples per second: 56 x 18 = 1.008s). In the remaining period the machine looks for the next trigger signal.

Data sets of images were collected in three sagittal planes, left (L), middle (M), and right (R). The image planes were 7-mm thick with a 4-mm separation (11-mm center to center). Although the tags were not used in the study, the time-phases used in the tagging protocol were. Capture of the 56 ms time-phases was done by the ensemble summation technique mentioned above. For each syllable and each slice, the seven time-phases were summed to produce a single sequence of C-to-V motion. The seven time-phases covered the C-to-V transition.

The tongue shape in the first time-phase (or MRI frame) is referred to as the "reference" configuration. This was always the consonant in the consonant to vowel syllable (CV). About 40 points ($N_R \approx 40$), identified by the horizontal and vertical coordinates, (X_k, Y_k) ($k = 1$ to N_R), were required to densely represent the surface of the tongue. The points were chosen at approximately four pixels apart. Image processing software was developed to allow hand selection of points (Davis, 1999). Since these data are two-dimensional, X and Y were used to describe the horizontal and vertical positions in the reference.

The tongue shape in the vowel frame was the "deformed" configuration and was used to define a similar number ($N_O \approx 40$) of points on the surface of the tongue, $(x_k^{(O)}, y_k^{(O)})$ ($k = 1$ to N_O). Here N_O defines the number of surface points selected in the time-

phase of the observed (O) vowel. Note that since we are not measuring the motion of specific tissue points on the surface of the tongue, the generic reference point (X_k, Y_k) does not move necessarily to $(x_k^{(O)}, y_k^{(O)})$. In addition movement of the tongue normal to the sagittal plane was assumed to be insignificant at midline and was indeterminate in the parasagittal slices without 3D data.

Vowel frames were defined as the last frame before the tongue changed direction back toward the consonant, and were typically the third, fourth or fifth time-phase of the seven frame sequence (see Figure 3). Different portions of the tongue do not reach maximum vowel position at the same time. Our criterion was for the back of the tongue to reach maximum. Due to the long time-phase, 56 ms, usually the entire tongue appeared to reach maximum in a single frame.

Deformations were computed between: (1) the extreme consonant and vowel (C/V) frames within a syllable, (2) the identical consonant (C/C) and (3) the identical vowel (V/V) in different contexts. As the tongue moved from consonant to vowel, points along the tongue surface were identified by hand on all frames. In the initial consonant frame, the strong effect of MRI tagging made it difficult to identify the lower edge of the tongue, (though these edges could be distinguished in the subsequent vowel frames, Figures 1 and 2). The lower edge of the tongue, therefore, was not measured. Omitting the lower tongue edge gives movements of the upper edge twice the weight of either side in the calculation of the vertical deformations. To offset this effect, the lowest points in the front and back were oversampled. This oversampling was done by taking a grid of four adjacent points at the lowest point location. This gave the lowest points four times the weight of the other points, to compensate for the missing lower edge (see Figure 4 below).

The lowest points on the tongue had to be chosen consistently in all images to define the vertical extent of the tongue. In the front of the tongue, the lowest point was the intersection of a

horizontal tag with the lowest, most posterior corner of the mandible. In the back, the lowest point was chosen to be the inflection point at the base of the tongue and the top of the airway. Here also, the lowest point on the tongue were chosen to be intersection between a tag line and the surface to assure that the same tissue point were used in the reference and deformed images. These points help determine the amount of vertical translation and stretch present in the tongue. The epiglottis was omitted in the /ki/, /si/, and /su/ data, because of its distance from the tongue body. A vertical line connected the valleculae (cavity) above the epiglottis to the pharyngeal space below it. In the /kA/, /sA/, and /ku/ data the epiglottis abutted the tongue and was included in the measurement. In the V-to-V and C-to-C comparisons, some frames were remeasured to assure that each comparison either did or did not contain the epiglottis.

Measurement Error and Repeatability

Several sources of error affected the MRI measurements. They were: temporal resolution of the MRI image; magnetic distortion due to tagging; measurement error; speaker precision (discussed in Methods); and head movement.

The slow temporal resolution (56 ms/time-phase) and the magnetic disturbance from the tags reduce tongue edge clarity. Because of the large time window, each slice includes some tongue movement, which can cause blurring of the surface. In addition, the magnetic disturbance of the newly laid tags causes consonants to have poorer edge clarity than vowels (Figure 1 vs Figure 2). Even so, the consonants were no more difficult to measure than the vowels. The worst measurement error occurred for /A/. Measurement error was assessed by having the initial reviewer remeasure 4 of the 36 contours, about 10% of the data. L_2 norm comparisons of the two measurements of each curve showed the following error: /s/ in /su/-L = 0.67 mm; /k/ in /kA/-M =

0.79 mm; /i/ in /si/-M = 0.51 mm; and /A/ in /sA/-R = 1.11 mm. These numbers are within the measurement error, except for /sA/-R. The error values include errors due to pixel resolution (.95mm); temporal resolution (.56 sec); magnetic disturbance; subject repetition imprecision (up to 46 ms); and true measurement error.

The effect of head motion was also considered. Six tag points in the brain (3 vertical, 3 horizontal) were measured in the consonant and vowel frames of the first and last data sets. This allowed us to measure head motion between consonant and vowel, and also between the beginning and end of the recording session. Measurements of /kA/ and /sA/ found that in no comparison was there a difference of more than one pixel in x or y. Moreover, the errors were not systematic. The overall pixel resolution difference is less than the accepted measurement error indicating no evidence of C-to-V head motion or long-term drift.

Model of Deformation

The objective in modeling tongue deformation was to provide parameters to describe the differences between the C and V tongue contours so that quantitative comparisons could be made of the deformations: translation, rotation, stretch and shear. Rigid body motion consisted here of 2D horizontal (x) and vertical (y) translation and in-plane rotation. Homogeneous deformations were described by the two, in-plane (orthogonal) principal strains (PS) and their directions. The two principal strains are the maximum and minimum axial strain, which quantify the local extension per unit reference length. Goodness of fit is estimated by calculating L_2 norms between the modeled and observed vowel shapes (see below).

Strains, principal strains, and shear are explained as follows. When a solid body stretches, at each point in that body neighboring points will change their relative *distance* and *angle* during

the deformation. For rigid bodies, the stretch in every direction is unity (no stretch). For a non-rigid, two-dimensional body, or a single plane of a three-dimensional body, the stretch of any material element is defined as the ratio of the new length to its original length ($\Delta\ell/L$) in a specific direction. The maximum and minimum (or principal) stretches will be in orthogonal directions and are the eigenvalues of the stretch tensor. Stretch and strain are different expressions of the same deformation ($\Delta\ell/L$). Stretch expresses the deformation distance in units of measurement (e.g., mm); strain expresses the deformation as percentage of original length. Hereafter, we consider these deformations in units of strain.

For a two-dimensional body or a single plane of a three-dimensional body, the shear, or shear strain, is a measure of the change in angle, as a result of deformation, between any two material elements that were orthogonal in the original configuration. For example, if one has a vertical stack of pennies and each is moved slightly to the right with respect to the penny below, there has been no area change, but a shape change is quantified by the shear, or change in angle, between the left edges of the pennies and the horizontal.

The homogeneous deformations of strain and shear were used to calculate principal strains for the CV movements. Principal strain is the direction of strain in which shear is zero. Because there is no shear component, the strains are maximal. Principal strains express the in-plane maximum and minimum strain, which may be compression or expansion.

The model presented here examines only single planes individually. An origin was chosen at the center of the tongue in the reference image. The model of deformation mapped any point, say P, in the reference shape occupying position (X, Y) with a new model position $(x^{(M)}, y^{(M)})$ in the deformed shape. We chose a coordinate system such that X (and x) defined the horizontal position, and Y (and y) the vertical position of any point along the tongue contours.

The motion of the generic point, *P*, was then described by two functions, *f* and *g*, which depend on its original position and on the parameters c_i ($i = 1$ to M, where M is the number of parameters used to describe the motion). A six parameter model is required to capture planar rigid body motion and homogeneous deformation, thus $M = 6$, and

$$x^{(M)} = f(X, Y; c_1, c_2, \ldots, c_6)$$
$$y^{(M)} = g(X, Y; c_1, c_2, \ldots, c_6).$$

The parameters, seen in Table 1, are (c_1) horizontal translation, (c_2) vertical translation, (c_3) rotation about the center of the image, (c_4) homogeneous horizontal stretch, (c_5) homogeneous vertical stretch, and (c_6) uniform simple shear. Previous work (Davis, Douglas & Stone,1996) used the present 6 parameters plus three more: x-bend, y-bend, local depression. These higher order deformations add more detail to the model. They were not used here, however, in order to determine if a global average of the first order components could describe tongue motion with sufficient accuracy. The averaged global principal strains were used to examine coarticulatory effects on tongue shape and LR asymmetry.

The final motion was represented in terms of the six motion parameters c_i. For any specific point P_k, with reference position (X_k, Y_k), the deformed position was given by

$$x_k^{(M)} = c_1 + (\cos c_3 - c_6 \sin c_3)(1 + c_4)X_k + (c_6 \cos c_3 - \sin c_3)(1 + c_5)Y_k \qquad (1a)$$

$$y_k^{(M)} = c_2 + (\sin c_3 + c_6 \cos c_3)(1 + c_4)X_k - (c_6 \sin c_3 - \cos c_3)(1 + c_5)Y_k \qquad (1b)$$

Table 1. Homogeneous deformations applied globally to the surface MRI data.

Parameter	Deformation	Formula
c_1	x translation	$x^{(M)} = X + c_1$
c_2	y translation	$y^{(M)} = Y + c_2$,
c_3	rotation	$x^{(M)} = X \cos c_3 - Y \sin c_3$ and $y^{(M)} = X \sin c_3 + Y \cos c_3$.
c_4	homogeneous horizontal stretch	$x^{(M)} = (1 + c_4)X$ and $y^{(M)} = Y$.
c_5	homogeneous vertical stretch	$y^{(M)} = (1 + c_5)Y$ and $x^{(M)} = X$..
c_6	uniform simple shear	$x^{(M)} = X + c_6 Y$ and $y^{(M)} = c_6 X + Y$.

L_2 Norm Analyses.

The L_2 norm (Equation 3) was used as a measure of the difference between two tongue contours. Given two sets of points, the L_2 norm compares this difference by measuring the squared distance between the two contours. The L_2 norm was used to (1) determine whether a significant tongue motion had occurred between two time-phases and (2) find the optimal values of the motion parameters, c_j.

The first use of the L_2 norm was to compare the reference and deformed shapes of two measured contours: C/V, C/C or V/V. Ideally, when quantifying tongue motion, identical tissue points on the deforming tongue contour are measured in both the reference and the (observed) deformed configurations. In that case there is a one-to-one correspondence between reference points (X_k, Y_k) $(k = 1 \text{ to } N^{(R)})$ and deformed points $(x_j^{(O)}, y_j^{(O)})$ $(j = 1 \text{ to } N^{(O)} = N^{(R)})$. However, in the MRI images, we are unable to identify distinct points along the tongue contour. The points on the reference contour do not necessarily correspond to the points on the observed deformed contour. Thus we need to generalize the L_2 norm measure, Eq. (3).

To generalize the L_2 norm we considered two tongue contours denoted as A (consonant) and B (vowel) respectively. The first was defined by the linear splines connecting the points $(x_k^{(A)}, y_k^{(A)})$. The second contour was defined by the linear splines connecting the points $(x_j^{(B)}, y_j^{(B)})$. We identified the points on the second contour, B, closest to each of the $N^{(A)}$ points $(x^{(1)}, y_k^{(A)})$ and called them $(\hat{x}_k^{(B)}, \hat{y}_k^{(B)})$, $(k = 1 \text{ to } N^{(A)})$. We also identified the points on the first contour closest to each of the $N^{(B)}$ points $(x_j^{(B)}, y_j^{(B)})$ and called them $(\hat{x}_j^{(A)}, \hat{y}_j^{(A)})$, $(j = 1 \text{ to } N^{(B)})$. We define the measures

$$L_2^{(A-B)} = \frac{1}{\sqrt{N^{(A)}}} \sqrt{\sum_{j=1}^{N^{(A)}} \left[\left(x_k^{(A)} - \hat{x}_k^{(B)} \right)^2 + \left(y_k^{(A)} - \hat{y}_k^{(B)} \right)^2 \right]}, \qquad (2a)$$

$$L_2^{(B-A)} = \frac{1}{\sqrt{N^{(B)}}} \sqrt{\sum_{j=1}^{N^{(B)}} \left[\left(x_j^{(B)} - \hat{x}_j^{(A)} \right)^2 + \left(y_j^{(B)} - \hat{y}_j^{(A)} \right)^2 \right]}, \qquad (2b)$$

such that the total L_2 norm is given by

$$L_2(A,B) = \tfrac{1}{2} \left[L_2^{(A-B)} + L_2^{(B-A)} \right]. \qquad (3)$$

The second use of the L_2 norm in these data was to measure modeling error or goodness of fit of the modeled parameters, c_i. The Observed-to Modeled (O/M) L_2 norm was the measure of the distance between the points on the deformed tongue contour as mapped by the model $(x_j^{(M)}, y_j^{(M)})$, (equation 1), and the observed points $(x_j^{(O)}, y_j^{(O)})$. Thus,

$$L_2(c_i) = \frac{1}{\sqrt{N}} \sqrt{\sum_{j=1}^{N} \left[\left(x_j^{(M)} - x_j^{(O)} \right)^2 + \left(y_j^{(M)} - y_j^{(O)} \right)^2 \right]}. \qquad (4)$$

If the model had sufficient fidelity to capture the contour's motion perfectly, and there was no measurement error, then the six motion parameters, c_i, could be chosen such that $x_j^{(M)} = x_j^{(O)}$ and $y_j^{(M)} = y_j^{(O)}$ $j = 1$ to N gives $L_2(c_i) = 0$. This would be the O/M L_2 norm.

Deformation Components

The optimum values of the six motion parameters, c_j, minimize the O/M L_2 norm. Minimization was accomplished using a downhill simplex method (Press, Flannery, Teukolsky, & Vetterling, 1989, p. 326-330). By minimizing the O/M L_2 norm, defined in equation 4, we could compute each of the deformation components as quantified by its magnitude, c_j. For example, horizontal strain, c_4, is in percentages; horizontal translation, c_1, is in mm.

Model Fidelity

Interpretation of the reference-to-deformed L_2 norms (C/V, C/C, V/V), depends on the fidelity of the model. The present model is homogeneous in nature. Therefore, the degree to which it represents the shape difference between the reference contour and the observed deformed contour depends upon the extent to which that difference is homogeneous. If the difference is homogeneous, the model will have high fidelity and the L_2 norm will be small between the observed (O) deformed contour and the modeled (M) deformed contour (O/M). If the O/M L_2 norm is too large, the homogeneous model lacks sufficient degrees of freedom and cannot simulate tongue motion with adequate fidelity. O/M L_2 norm differences of 1.90 mm or less were within 1 pixel of movement and were attributed to error (low image resolution). O/M L_2 norm differences above 1.90 mm reflect components of the deformation that were not captured by the model. In other words, C/V, V/V, and C/C L_2 norm values greater than 1.90 mm reflect true differences in deformation, while O/M L_2 norm values greater than 1.90 mm reflect reduced model fidelity. The averaged O/M L_2 norm values represent global goodness-of-fit measures for the model's ability to capture C-to-V motion (Table 2), and context accommodation (Table 3).

Results for the Data Application

Representing Consonant-to-Vowel Deformation

The first five tongue surface contours in the /k/ to /A/ movement are drawn in Figure 3. The first contour is black and the last is light gray. The tongue tip is on the left. Deformation between the consonant and vowel (first and last contours) was examined in all syllables in two ways: by mathematically calculating the shape differences between each consonant and vowel (L_2 norms), and by decomposing the movement into mechanical deformations (c_i).

Figure 3. Left, middle and right contours for the sequence /kA/. The first five time-phases are shown in sequence from black to light gray.

First consider the L_2 norm comparisons seen in the first two columns of Table 2. The C/V L_2 norm reflects the motion between the C and V contours. The syllable with the largest deformation was /kA/, with L_2 norm differences of: Left: 5.8 mm, Midline: 4.9 mm, Right: 6.2 mm. For /ki/ and /ku/, the L_2 differences were smaller since those vowels were more similar to /k/ in shape and position. For /s/, L_2 norms were quite small. There was less motion (deformation) from /s/-to-V than /k/-to-V.

Second consider the L_2 norm as a measure of goodness-of-fit for the model. The O/M L_2 norm (Table 2, col 2) represents the difference between the observed vowel (O) and the modeled vowel (M), with 1.9 mm being the threshold.. For the /s/ most O/M L_2 norms were below 1.9 mm but the movements are also quite small. For the /k/ /kA/-L had the worst fit of the model (3.5mm). Of the 18 deformations, 12 were below 1.9 mm and the average O/M L_2 norm was 1.8, indicating that in general the model represented the deformations well.

Table 2. Consonant to Vowel Deformations for left (L), middle (M) and right (R) sagittal slices. L_2 Norms compare contours of vowels with consonants (C/V) and to models (O/M). Each comparison is represented by the deformation components of translation (trans), rotation and principal strain (PS). Deformation direction is C-to-V.

		C/V L2 (mm)	O/M L2 (mm)	X-trans (mm)	Y-trans (mm)	Rotation (degrees)	PS 1 (%)	PS 2 (%)	PS Direction (degrees)	L,M,R edge Arrival phase
ki	L	3.4	2.2	-0.6	-0.1	3	5%	-14%	17	3
	M	2.5	1.6	-1.2	-1.2	-2	3%	-6%	45	3
	R	2.8	2.0	2.2	0.7	1	7%	2%	-23	4
kA	L	5.8	3.5	-2.3	-7.6	8	2%	-10%	-38	3
	M	4.8	1.5	-3.6	-6.5	5	8%	-15%	-7	4
	R	6.2	1.7	-1.2	-7.9	3	18%	-12%	-4	4
ku	L	4.4	2.0	-1.9	-0.7	4	0%	-24%	29	4
	M	3.2	1.9	1.0	-1.9	3	0%	-12%	18	3
	R	3.1	1.8	1.2	-4.8	4	2%	-4%	13	4
si	L	1.7	1.1	-0.2	-0.3	-1	0%	-7%	-85	4
	M	1.9	1.6	-0.7	0.3	2	0%	-6%	-2	5
	R	1.6	1.4	-0.6	0.9	-1	2%	-3%	33	2
sA	L	2.4	1.5	0.2	0.1	2	10%	-3%	-17	2
	M	2.0	1.5	0.3	0.1	1	8%	1%	-32	3
	R	2.7	1.6	0.3	-2.0	0	11%	6%	79	4
su	L	3.3	1.7	-1.4	-1.6	1	8%	-11%	-74	1
	M	2.4	1.6	-1.3	1.0	-3	6%	-6%	-85	2
	R	3.3	2.2	-1.9	6.9	-4	-4%	-12%	29	1
Average L2		3.2	1.8							
Std. Dev.		1.3	0.5							

Of the several deformation components extracted by the model (Table 1 and Equation 1), consider first the rigid body deformations. Even non-rigid bodies have a rigid, or zeroth order, component to their movement, which can be represented as translation and rotation. Rigid body translation (Table 2, col 3 and 4) is in mm with negative numbers being backward (x) or downward (y); rotation is in degrees (col 5) with positive numbers indicating clockwise rotation. The data indicated that the x and y translations for the tongue were small in most syllable movements. The exception, /k/-to-/A/ motion, contained large downward translations for all three slices (-7.6, -6.5, -7.9 mm) consistent with its large opening movement. Translation asymmetry was seen for /ku/ and /su/, which only had large motions on the right (downward -4.8mm and upward 6.9 mm respectively).

Next consider the homogeneous deformations, expressed as principal strains, i.e., the direction of maximal strain. Principal strain (PS1 and PS2) values are expressed as percent change with positive numbers referring to extension and negative numbers to compression (col 6 and 7). The PS direction (col 8) indicates the direction of the principal compression or extension. In the images, the subject faces toward the left. Positive numbers rotate the strain direction clockwise from the horizontal (PS 1) or vertical (PS 2) plane.

Compression and extension appeared to be the major component of these deformations, as seen when comparing the translations and the principal strains. For /kA/, compression was almost entirely downward (PS angle = -7^0, -4^0), or downward and backward (-38^0). Expansion was backward or backward and upward. Figure 4 shows the reference, deformed and modeled contours for /kA/-LMR. The gray squares are the /k/; the gray circles are the observed /A/; and the black diamonds are the modeled /A/. For /ku/ (not shown), compression also was downward (PS 2). There was no backward expansion, however, which indicates lateral expansion (widening) in

response to the vertical compression.

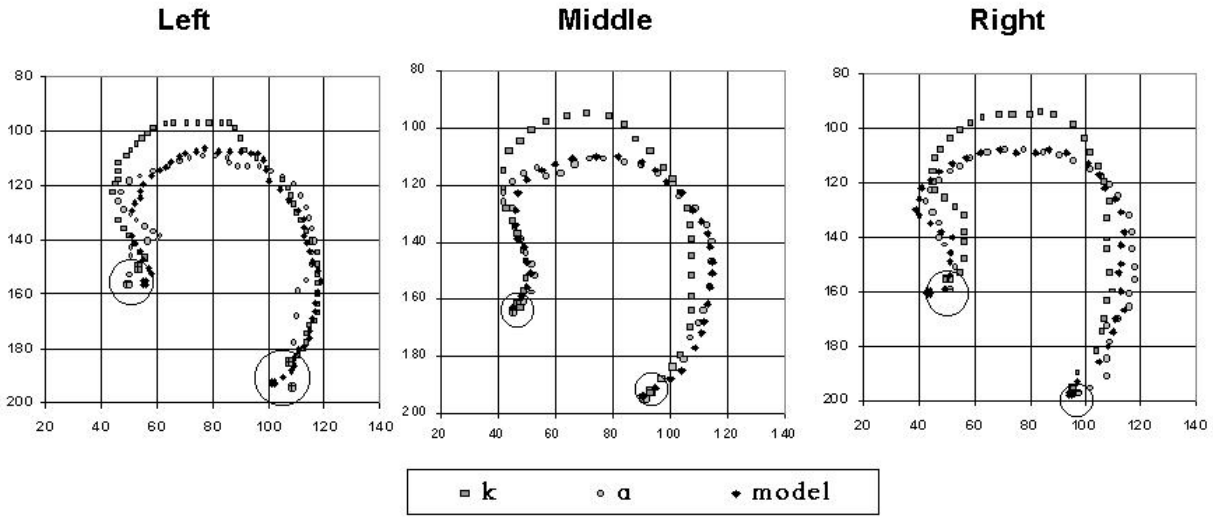

Figure 4. Measured consonant (gray squares) and vowel (gray circles) contours for the syllable /kA/, and the model surface for the vowel (black diamonds). Principal strains drawn in the middle are extension (PS1-solid line) and compression (PS2-dashed line). Large symbols indicate oversampling of lowest points (circled).

Tongue tip representation is a challenge for a homogeneous model because its deformation may be uncorrelated with the rest of the tongue. This can be seen in /ka/ - L, where the large O/M L_2 norm indicated poor goodness-of-fit. Figure 4 pinpoints this error at the tongue tip. The horizontal extension was underrepresented (also shown in PS1 = 2%). This contrasted with /ka/-R where the model captured the tip well, as seen by the small O/M L_2 norm (1.7 mm) and an appropriately large extension (PS1 = 18%). In previous work, bend and local deformations were represented using this model (Davis, Douglas, and Stone, 1996). The calculation of principal strains, however, precludes the inclusion of higher order deformations. This is not necessarily a disadvantage. The addition and deletion of higher order deformations is a feature that allows the model to adapt to different goals, such as representing homogeneous motion strains (as in this study) or identifying higher order deformations.

Context Effects

The contour shapes of each consonant and vowel were compared across phonetic contexts. Figure 5 depicts overlaid midsagittal contours of /k/ and /s/ in the three vowel contexts, and Figure 6 depicts /i/, /A/, and /u/ in the two consonant contexts. Table 3 presents L_2 norms and modeled deformations comparing the identical phoneme in different contexts. Positive and negative strains in Table 3 indicate differences in tongue body length, not compression and extension, since the two sounds are not contiguous in time. Translation reflects position differences, not movement. The V/V L_2 and C/C L_2 norms are, therefore, a measure of coarticulation, as described in equation 4. Larger numbers indicate greater intra-phoneme differences due to context and the O/M L_2 norms indicate their homogeneity.

Table 3. Context Effects. L_2 Norms compare contours of the same phonemes in different contexts (C/C, V/V), and to models (O/M). The L_2 Norms show divergence in shape. The components show divergence in position, rotation, and length due to context differences.

C effect on V (deformation direction k-to-s)

	V/V L2 (mm)	O/M L2 (mm)	horizontal position (mm)	vertical position (mm)	Rotation (degrees)	horizontal length %	vertical length %
i - i L	2.3	1.7	1.2	-0.9	1	-2	-1
A - A L	3.0	2.0	2.7	4.0	-4	2	-8
u - u L	2.7	1.9	0.8	-1.1	3	2	-5
i - i M	2.4	1.5	-0.1	-0.4	5	4	-6
A - A M	4.6	1.9	4.2	2.3	1	1	-2
u - u M	4.0	2.6	-0.8	1.5	-1	6	6
i - i R	3.8	2.7	-1.6	-3.3	0	-4	-3
A - A R	4.4	2.6	2.0	3.5	3	-2	1
u - u R	3.7	2.0	-3.4	2.2	-2	-9	-16
Average O/M	3.4	2.1					
Std. Dev. O/M	0.9	0.4					

V effect on C (deformation direction in parenthesis)

	C/C L2	O/M L2	horizontal position (mm)	vertical position (mm)	Rotation (degrees)	horizontal length %	vertical length %
k(i)-k(A) L	5.9	2.1	-5.1	-4.5	-2	**-5**	**-5**
k(i)-k(u) L	6.1	4.0	-4.3	-1.7	2	8	9
k(A)-k(u) L	5.3	3.0	2.6	1.8	-1	-3	7
s(i)-s(A) L	5.6	2.0	-4.8	-0.5	-2	-9	-3
s(i)-s(u) L	3.9	1.6	-2.8	-0.8	-2	0	2
s(A)-s(u) L	3.5	1.9	1.8	-0.3	1	11	3
k(i)-k(A) M	6.4	2.0	-5.3	-1.4	2	10	-4
k(i)-k(u) M	5.3	2.8	-4.1	-1.1	1	5	0
k(A)-k(u) M	2.2	1.4	1.2	0.2	3	-2	1
s(i)-s(A) M	6.1	1.5	-3.7	-4.0	9	20	-14
s(i)-s(u) M	4.4	2.0	-2.4	0.0	4	15	-8
s(A)-s(u) M	3.8	2.0	1.1	4.4	-4	-3	5
k(i)-k(A) R	3.2	1.2	-1.6	-1.6	-1	2	10
k(i)-k(u) R	2.5	1.4	-0.2	-0.4	1	4	8
k(A)-k(u) R	1.9	0.9	1.3	0.9	-1	1	-2
s(i)-s(A) R	3.2	1.8	-1.3	1.7	1	7	-15
s(i)-s(u) R	3.5	3.0	-1.0	0.4	2	7	-5
s(A)-s(u) R	3.2	2.3	0.8	-0.8	2	-1	13
Average L2	4.2	2.0					
Std. Dev. L2	1.5	0.8					

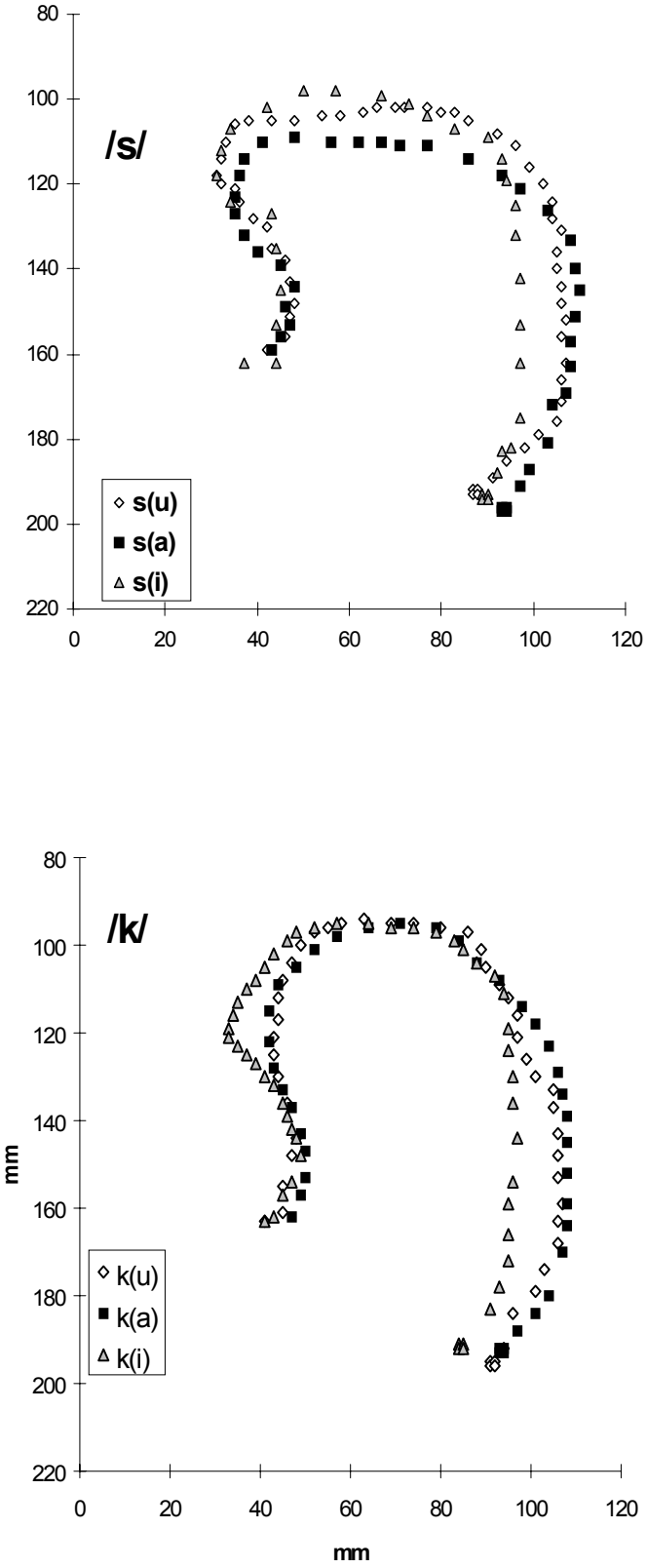

Figure 5. Contours of midline /k/ and /s/ in the three vowel contexts.

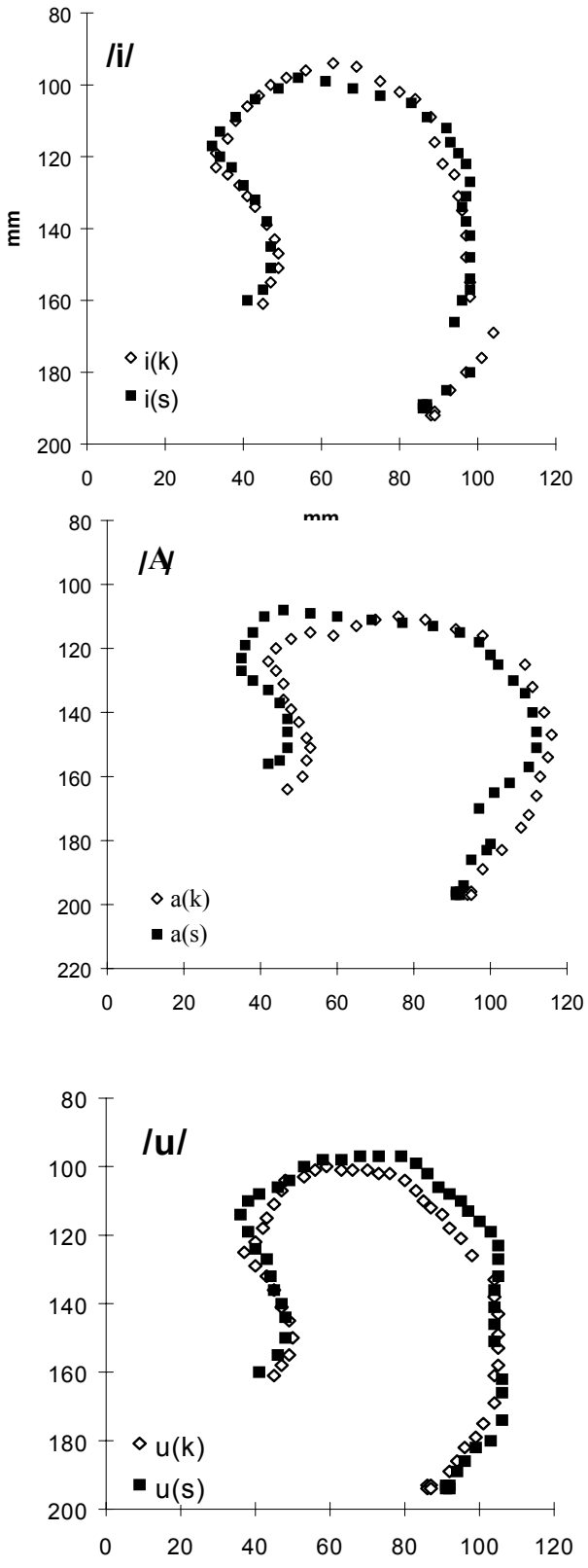

Figure 6. Contours of /i/, /A/, and /u/ in the two consonant contexts.

Figure 5 shows that there was considerable midline coarticulation for the consonants; only the constriction in /k/ was invariant. Figure 6 shows little coarticulation for the vowels, only /A/ had notable coarticulation; the tongue was farther forward and longer in /s/ than /k/ context. Table 3 offers numerical confirmation of these differences plus data for the parasagittal slices. Examining the midline, the average L_2 norms were larger for consonants (C/C=4.4mm) than vowels (V/V=3.4mm), indicating larger context effects on consonants than vowels. Interestingly, these averages were as large or larger than the average differences in C-to-V shape (Table 2, Average C/V L_2 =3.2mm). The exception was the /kA/-/ku/ comparison, which had similar tongue shapes for the /k/. Figure 5 depicts this similarity as well as the much more anterior tongue position for /k/ in /i/ context. Most of the C/C and V/V shape differences were in the horizontal plane, the opposite direction from the C/V deformations.

Left-to-Right (LR) Asymmetry

LR differences in tongue position and length were found for both consonants and vowels. For the consonants, contour differences (C/C L2 norms) were larger on the left side than the right indicating greater accommodation to the vowels on the left side. Although the values are smaller, the vowels appeared to have slightly more context effect on the right.

The horizontal deformations particularly reflected this asymmetrical coarticulation. Mean values for consonants reflect the greater coarticulation on the left (position L=3.8mm, R=1.0mm, length L=6.5%, R=3.7%). Vowel coarticulation was smaller than consonant, and greater differences were seen on the right (position L=1.6, R=2.3mm, length: (L=2%, R=5%). In the C/C and V/V comparisons, differences in position are more important than length, because the L_2 norm differences are small and length differences are expressed as percentages.

A second type of LR difference was the leading-lagging edge. Often a leading or lagging edge was seen, in which one contour (L, M, R) arrived at maximum a frame early or late. Table 2 (last column) lists the time-phase at which each edge arrived at the vowel (1-7). Lower numbers indicate earlier arrival. For /kA/, the left edge reached maximum at phase 3, the middle and right at phase 4. There was then no movement through phase 5 for any slice (see Figure 3). For /ku/ the middle contour arrived earlier (phase 3) than the others.

In summary, tongue motions between consonants and vowels were compared and modeled. The OM L_2 norm showed that two-thirds of the motions had small error and thus were well represented by a homogeneous model. CV L_2 norm comparisons showed that the /k/-to-/A/ motion was the largest of the six syllables and the /k/-to-V movements were larger than the /s/-to-V movements in general. The component deformations described the C-to-V motions as containing mostly vertical compression, horizontal expansion and minimal translation. Considerable coarticulation was seen in this data set, with generally more shape accommodation on the left for consonants and the right for vowels. At times coarticulation caused larger shape differences to occur between the same phoneme in different contexts than between adjacent consonants and vowels.

Discussion

This study is a new application of continuum mechanics to modeling tongue motion. It decomposes tongue surface motion into a compact set of global, homogeneous deformations and determines statistically the model's success. A model of this sort is valuable because most tongue measurements are of the surface not the internal tongue. Moreover, simplification of tongue behavior is helpful in representing and comparing group differences. The model should prove useful in studying speech motor control because it captures C-to-V movement, coarticulation and asymmetry.

Modeling Consonant-to-Vowel Motion

Despite the fairly coarse spatial and temporal resolution of the MRI data, the global model was able to describe deformations between maximal consonant and vowel positions, especially for the larger movements, such as /kA/. For example, the model indicated that this subject used compression and extension more than translation in C-to-V tongue movement, especially in the horizontal direction (Table 2). Traditional models of tongue motion assume the tongue can translate without deformation, ie rigid body motion, usually due to extrinsic muscle activity (cf. Perkell, 1974, Dang and Honda,). The behavior exhibited by this subject, however, is more consistent with the muscular hydrostat model of soft tissue motion, in which muscle contraction causes motion by creating deformations that compress the structure in one direction or location and expand it in another (Kier and Smith, 1985). The tongue exhibited just this pattern in the C-to-V tongue motions. Principal strains showed that compression and extension accounted for much more of the movement than translation and rotation (Table 2).

Because the principal strains quantify the direction and amount of tissue compression, one

can infer muscle contraction patterns. One must be very careful about inferring muscles from compression data since not all compression is active. However, Table 2 and Figure 4 provide compression data (negative principal strains) that hint at muscle involvement. In Figure 4, the PS lines are drawn in the middle of each tongue slice indicating the direction and amount of maximum compression (dotted) and expansion (solid) as the tongue moved from the /k/ to /A/. The right slice had only downward compression; the left had downward and backward. Both were consistent with hyoglossus (HG) contraction, since HG has purely vertical fibers and also oblique fibers running toward the tongue tip (AbD El-Malek, 1939; DuBrul, 1976; Sokoloff, 1989). It is not known how independent is the control of these fiber bundles, but even slight differences in their bilateral activation would cause PS angle difference. Results of this sort can be tested in predictive models, eventhough the quantities themselves do not definitively define muscular events.

Coarticulation

More similar C and V phoneme shapes require less C-to-V motion. Less C-to-V motion requires less work and so is physiologically more cost effective. The present speech task consisted of repeating a CV syllable 96 times, thus alternating C and V gestures. This alternating repetition is essentially diadochokinetic, which is the "function of arresting one motor impulse and substituting another that is diametrically opposite" (Dorland, 1968). While fatigue could be a factor in diadochokinetic motions, the primary difficulty is considered to be the rapid reversal of motor commands. The present subject reduced the difficulty of the alternating motions in two ways. First, he coarticulated more in the horizontal or anterior-to-posterior (AP) plane and moved more in the vertical or superior-to-inferior (SI) plane. Second by using consonant accommodation

more on the left side, and vowel accommodation more on the right he created left-right movement asymmetry.

In the first case, the SI/AP dichotomy for consonants can be explained by the complete linguopalatal contact require for /k/ (and /s/ laterally), which constrains SI tongue position during the consonant. In the AP direction the consonantal tongue is less constrained, allowing contextual accommodation horizonally. This phenomenon, is well known for /k/ whose palatal constriction moves horizontally in front vs. back vowel context. Tongue accommodation to vowels has also been shown for /s/ (Stone and Vatikiotis-Bateson, 1995). Such accommodation allows more rapid CV motion by reducing the distance in the AP direction.

In the second case coarticulation occurs more on one side for consonants and the other for vowels, which may allow the SI constraint to be eased, as seen in Figure 4. For /kA/, the superior limits of /k/ were similar for all three slices, but there was greater posterior contact for /k/-R. The lesser palatal contact at /k/-L,M, an accommodation to the vowel, reduced the distance the tongue must travel from the C to the V, allowing a faster motion. Figure 3 and Table 2 last col., show that the left-tongue reached the /A/ by the third time-phase; the middle and right reached it by the fourth. We believe that there is no significance to which side reaches the vowel first.

It is interesting to note that the consonants experienced greater contextual accommodation than the vowels. Consonants, which are constrained by their constriction/occlusion locations, have historically been shown to accommodate less than vowels and to cause less accommodation in vowels than other vowels do (MacNeilage and DeClerk, 1969, Ohman, 1966, Fowler, 1980, Recasens 1987). The present results examine the tongue from a different perspective, as a single unsegmented structure. When the focus is on the entire tongue, consonant articulation is considerably greater than when considering the constriction width alone (see also, Recasens 1987).

Despite constriction constraints the tongue clearly can use asymmetry and AP motion to allow more and previously unexplored types of consonant accommodation.

A final consideration in these CV motions is gravity. Since the subject is supine, gravity is located at the back of the head, as discussed in Methods. AP accommodations, therefore, probably include a non-homogeneous resistance to gravity (Tiede et al., 2000). The effects of supine position on MRI data cannot yet be filtered out. Therefore, the present data, and MRI data in general, may not be entirely consistent with upright tongue motion.

Lingual Asymmetry

Lingual asymmetry is common in static postures (cf., Story et al., 1996, Stone and Lundberg, 1996) and in movements (Stone, 1990). LR asymmetry in tongue movement is especially apparent in repetitive tasks such as syllable repetition. Two reasons for LR asymmetry are biomechanical imprecision and purposeful rotation. We believe both effects occurred here as a result of rapid repetitive motion and coarticulatory accommodation.

The LR asymmetry in the C-to-V motion may have been enhanced by the use of alternating C and V gestures, but even in ordinary speech subjects often move the tongue asymmetrically. Stone (1990) observed LR rotation with a lateral (side) pivot and a central pivot, the lateral being much more common. The observed LR asymmetry is consistent with mechanical rotations that decrease the mass to be moved thus reducing effort and increasing speed. There is no information at the present time as to whether asymmetry is acoustically important. Its non-systematic nature (Stone, 1990) suggests that it may not be.

True rotation involves active control, so it is of interest to consider the reasons for its occurrence (reduction of biological cost), when it occurs (speaker and task effects), and how it

might be accomplished. A simple strategy for rotation would be to differentially contract bilateral pairs of muscle, thus reducing motion on one side. Alternatively, it could result from asymmetrical tongue shape for the starting or ending phoneme, tissue inertia or other passive biomechanical effects. The present data contained some deformations that were consistent with LR rotation (Table 2, C/V L_2 , PS2), note especially /ku/, but are too few to speculate extensively on active rotation and passive asymmetry. However, rotation as a cost effective movement strategy should be studied further.

In summary, the model represented fairly well tongue deformations in consonant-to-vowel motion. Translation and rotation indicated direction and extent of rigid-body movement; the results were consistent with phonetic expectations. Global principal strains indicated the direction and amount of maximum tongue compression and extension in each plane for each syllable. The goodness-of fit measure indicated that some deformations contained higher order components. The model also captured coarticulatory strategies, such as trade-offs in SI and AP dimensions and LR asymmetries, which reflect strategies for accommodation to context, gravity and palatal constraints. From these data one can make preliminary inferences about possible muscle activity, which can be tested in predictive models. This is a methods paper, whose goal is to lay the groundwork for applications that can compare, individuals, dialects and disorders, and the intermediate steps in the motion.

Conclusions

A homogeneous global model of tongue surface motion can be used to interpret dynamic aspects of speech even when data are interpreted conservatively. The spatial resolution of Cine-MRI images allowed adequate measurement of tongue features such as deformation patterns, context effects, and left-right differences. Thus the model captured the global homogeneous features of the C-to-V movement that it was designed to measure, though it missed higher order deformations such as tip extension. The modeled deformations, though exclusively homogeneous in nature, were sensitive enough to capture three important aspects of this subject's tongue motion. The first was the importance of stretch relative to translation. The second was the use of AP consonantal shape accommodation to compensate for the relative invariance required in the SI dimension. The third was the use of LR rotation to reduce the distance traveled (biological cost) in the CV movements. All three of these features are useful for understanding tongue control and representing speech disorders.

Since this data set was collected on a single subject, and contains simplified speech material, the results cannot be generalized widely. However, these are useful preliminary data that suggest many ideas for future investigation. They are presented to illustrate ways in which speech behaviors can be represented using a global homogeneous model with few control parameters.

Acknowledgements

This project was supported in part by grant DC025681 from the National Institute of Deafness and Other Disorders, and Northrop Grumman Inc. Parts of this work were presented at the Fifth European Conference on Speech Communication and Technology (Eurospeech), Rhodes, Greece,

September 1997.

References

Abd-El Malek, S. (1939). A contribution to the study of the movements of the tongue in animals with special reference to the cat. *J. Anatomy* 73:15-33, 1938.

Alwan, A., Narayanan, S., & Haker, K. (1997). Towards articulatory-acoustic models for liquid consonants based on MRI and EPG data. Part II: The rhotics. *Journal of the Acoustical Society of America*, *101* (2), 1078-1089.

Baer, T., Gore, J., Boyce, S., & Nye, P. (1987). Application of MRI to the analysis of speech production. *Magnetic Resonance Imaging, 5*, 1-7.

Baer, T., Gore, J., Gracco, C., & Nye, P. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *Journal of the Acoustical Society of America, 90*, 799-828.

Crary, M.A., Kotzur, I.M., Gauger, J., Gorham, M., & Burton, S. (1996). Dynamic magnetic resonance imaging in the study of vocal tract configuration. *Journal of Voice*, *10* (4), 378-388.

Dang, J. & Honda, K. (1997). A Physiological Model of the Tongue and Jaw for Simulating Deformation in the Midsagittal and Parasagittal Planes. *Journal of the Acoustical Society of America, 102* (5), 3167A.

Davis, E. (1999). Measurement and kinematic modeling of the human tongue. Doctoral dissertation, Johns Hopkins University Department of Mechanical Engineering.

Davis, E., Douglas, A., & Stone, M. (1996). A continuum mechanics representation of tongue motion in speech. *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, PA, *2*, 788-792.

Demoulin, D., George, M., Lecuit, V., Metens, T., Soquet, A., & Raeymaekers, H. (1997). Coarticulation and Articulatory Compensations Studied by Dynamic MRI. *Proceedings of the 5th European Conference on Speech Communication and Technology, EuroSpeech,* Rhodes – Greece. *1*, 43-46.

Dorland's Pocket Medical Dictionary. 21st edition (1968) Philadelphia, PA., W.B. Saunders Co.

DuBrul, E. L. (1976). Biomechanics of speech sounds. *Ann. NY Acad. Sci*. 280: 631-642.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing control. *Journal of Phonetics, 8,* 113-133.

Honda, K., Himi, H., and Dang, J. (1994). A Physiological Model of Speech Production and the Implication of Tongue-Larynx Interaction. *Proceedings of the International Conference on Spoken Language Processing.* Yokohama, Japan, 175-178.

Kier, W., and Smith, K. (1985) Tongues, tentacles, and trunks: the biomechanics of movement in

muscular-hydrostats. *Zoological Journal of the Linnaean Society, 83*, 307-324.

Kroger, B. (1992). Minimal rules for articulatory synthesis. [*Signal Processing VI: Theories and Applications* Ed. by. J. Vandewalle, R. Boite, M. Moonen, A. Oosterlinck. Elsevier Science Publishers.] In J.Vandewalle, R. Borte, M. Moonen & Oosterlinck (Eds.), Signal Processing VI: Theories and Applications (pp.xx). Publication location: Elsevier Science Publishers.

Lundberg, A., and Stone, M. (1999) Three-dimensional Tongue Surface Reconstruction: Practical Considerations for Ultrasound Data. *Journal of the Acoustical Society of America*, *106*, 2858-2867.

MacNeilage, P., and DeClerk, J. (1969) On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America, 45, 1217-1233.*

Maeda, S. (1991). On articulatory and acoustic variabilities. *Journal of Phonetics*, *19*, 321-331.

Masaki, S., Tiede, M.K., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., and Ninomiya, N. (1997). MRI observation of dynamic articulatory movements using a synchronized sampling method. *Journal of the Acoustical Society of America, 102,* 3166 (A).

McGowan, R & Cushing, S. (1999). Vocal tract normalization for midsagittal articulatory recovery with analysis-by-synthesis. *Journal of the Acoustical Society of America*, *106,* 1090-1105.

Mermelstein, P. (1973). Articulatory model for the study of speech production. *Journal of the Acoustical Society of America, 53*, 1070-1080.

Moore, C.A. (1992). The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images. *Journal of Speech, Language, Hearing Research*, *35* (5), 1009-1023.

Narayanan, S., Alwan, A., and Haker, K. (1997). Towards articulatory-acoustic models for liquid consonants based on MRI and EPG data. Part 1: The laterals. *Journal of the Acoustical Society of America, 101* (2), 1064-1077.

Ohman, S. (1966) Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, **39**, 151-168.

Ong, D., and Stone, M., (1998). Three Dimensional Vocal Tract Shapes in /r/ and /l/: A Study of MRI, Ultrasound, Electropalatography, and Acoustics. *Phonoscope, 1*, 1-13.

Perkell, J. (1974). A physiologically oriented model of tongue activity in speech production. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Press, W., Flannery, B., Teukolsky, S. and Vetterling, W. (1989) *Numerical Recipes in Pascal, the Art of Scientific Computing.* Cambridge, England: Cambridge University Press.

Recasens, D. (1987) Long range coarticulation effects for tongue dorsum contact in VCVCV

sequences. *Speech Communication, 8*, 293-307.

Sanguineti, V., Laboissiere, R., and Payan, Y. (1997) A control model of human tongue movements in speech. *Biol.Cybern.* 77(1):11-22, 1997.

Sanguineti, V., Laboissiere, R., and Ostry, D.J. (1998) A dynamic biomechanical model for neural control of speech production. *J.Acoust.Soc.Am.* 103(3):1615-1627, 1998.

Sokoloff, A. J. (1989). The organization of the hypoglossal nucleus: an experimental neuroanatomical investigation of hypoglossal-lingual and cortico-hypoglossal projections in *Macaca fasicularis* and other vertebrate species. Ph.D. thesis, Harvard University.

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. & 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., Davis, E., Douglas, A.S., Ness Aiver, M., Gullapalli, R., Levine, W., Lundberg, A. (to appear, 2001). Modeling the motion of the internal tongue from tagged cine-MRI images. *Journal of the Acoustical Society of America.*

Stone, M. and Vatikiotis-Bateson, E. (1995): Trade-offs in tongue, jaw and palate contributions to

speech production. *J. Phonetics*, *23*, 81-100.

Story, B.H., Titze, I.R., and Hoffman, E. (1996). Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America, 100* (1), 537-554.

Tiede, M.K., Masaki, S., Wakumoto, M. and Vatikiotis-Bateson, E. **(**1997). Magnetometer observation of articulation in sitting and supine positions. *Journal of the Acoustical Society of America*, *102* (2*),* p. 3166.

Tiede, M.K., Masaki, S. and Vatikiotis-Bateson, E. (2000). Contrasts in speech articulations observed in sitting and supine conditions. *Proceedings of the Fifth Seminar on Speech Production: Models and Data. Bavaria, Germany,* May 1-4. 25-28.

Tiede, M.K. and Yehia, H. (1996). A shape-based approach to vocal tract area function estimation. *Journal of the Acoustical Society of America, 100* (4), 2658.

Wilhelms-Tricarico, R. (1995). Physiological modeling of speech production: Methods for modeling soft-tissue articulators *Journal of the Acoustical Society of America, 97* (5), 3085-3098.