

A CONTINUUM MECHANICS REPRESENTATION OF TONGUE DEFORMATION

*Edward P. Davis, Andrew S. Douglas and Maureen Stone**

*Department of Mechanical Engineering
The Johns Hopkins University
Baltimore MD 21218–2686 USA*

**Division of Otolaryngology
University of Maryland Medical School
Baltimore, MD 21210 USA*

ABSTRACT

Understanding the kinematics of the tongue during normal speech will provide important information both for accurate modeling of the acoustics of the vocal tract and for clinical diagnosis and enhanced treatment of persons with abnormal speech due to tongue motion impediments. Measuring tongue motion can be done using magnetic resonance imaging (MRI) which cannot provide the temporal resolution required for normal speech, but can create pseudomovement by imaging a single moment in a vocalization from multiple repetitions of a speech gesture. Finding the relevant biomechanical information from these measurements requires a global kinematic model of tongue motion. Preliminary results from modeling the tongue show that the motion is not that of a rigid body. Indeed, a kinematic model can reveal important biomechanical features, such as regional muscle stretch and velocity requirements as well as the strain distribution. The parameters which describe the kinematic model also provide a quantitative means of comparing the tongue motion, and hence the acoustic control of different subjects.

1. INTRODUCTION

The motion of the tongue during normal speech is both rapid and complex. To understand both normal and abnormal speech and to provide part of a scientific basis for a comprehensive model of vocal tract acoustics, a rigorous mechanical model of the motion of the tongue during speech needs to be developed. A realistic model needs to account for the fact that the tongue musculature is fluid saturated and that the deformations which accompany motion must be locally volume preserving. These deformations are important because they reveal the underlying nature of the tongue kinematics and are key to the tongue kinetics. In three dimensions, the tongue motion must reflect the interaction of the muscles which are actively contracting and also those which remain passive during any given motion. The tongue motion is the result of the interaction of both lengthwise and crosswise muscle deformations, which delineate the degrees of freedom of tongue behavior.

2. IMAGING SOFT TISSUE

There are several ways in which the motion of soft tissue can be imaged. The first is to track points, such as using bi-planar cine x-ray imaging systems to identify the 3-D motion of radiopaque

markers. Although this approach has been used extensively [1] to quantify the motion of canine hearts, it is unsuitable for speech due to the amount of x-ray exposure.

Similarly, 2-D marker tracking can be done using x-ray microbeam or electromagnetic articulometry [2,3]. However, these methods do not provide motion information on the entire tongue surface or of points within the tongue. Another method with which to analyze the motion of the tongue during normal speech is ultrasound [4] which detects sharp boundaries in acoustic impedance at the tongue's upper surface. Ultrasound has the temporal resolution required to image the rapid motion of the tongue during speech. However, it does not provide the motion of the entire tongue, only the upper surface.

An excellent method for imaging the entire tongue is magnetic resonance imaging (MRI). An MRI image consists of pixels each with a defined gray-scale intensity. These are obtained by quantifying the characteristic magnetic dipole relaxation times in polar molecules to differentiate between tissue types in static images. Additional information and "pseudomovement" can be obtained from MRI if the tissue is "tagged." This involves both a non-uniform excitation of the soft tissue and the subsequent imaging of this tissue. Essentially, magnetic markers, the tags, are placed in regular patterns and then imaged. Since the magnetic fields decay in time, MRI provides about a 500 ms "window" during which meaningful movement data can be gathered. The resolution of the tags in a single moving image is poor, so multiple repetitions of the behavior are performed at a single moment in time across all repetitions. This creates a "pseudomovement," that is, the average of multiple, nearly identical, movements [5].

3. IMAGE PROCESSING

Ultrasound can measure tongue motion during normal speech. However, current ultrasound imaging has the transmitting/receiving transducer placed below the jaw and images the upper tongue surface only. Often, the tip of the tongue is lost because of the air space below it. Mid-line (mid-sagittal) data is easily acquired using common ultrasound devices. Using more advanced instruments, it is now possible to reconstruct the full three dimensional tongue surface [6].

MRI imaging has been used to reconstruct 3-D airway volumes for static sounds—vowels and fricatives [7,8,9]—but not to extract tongue features. The current MRI dataset consists of four midsagittal test images of the tongue: at rest, and saying /a/, /i/ and

/u/. As an example, figure 1 shows an MRI image of the tongue while vocalizing the /a/ sound. (This image had the most complex surface profile). Image processing techniques are used on each image to identify a discrete number of pixels, which represents the tongue surface. The pixel locations are shown as black circles with white interiors on the image in figure 1.

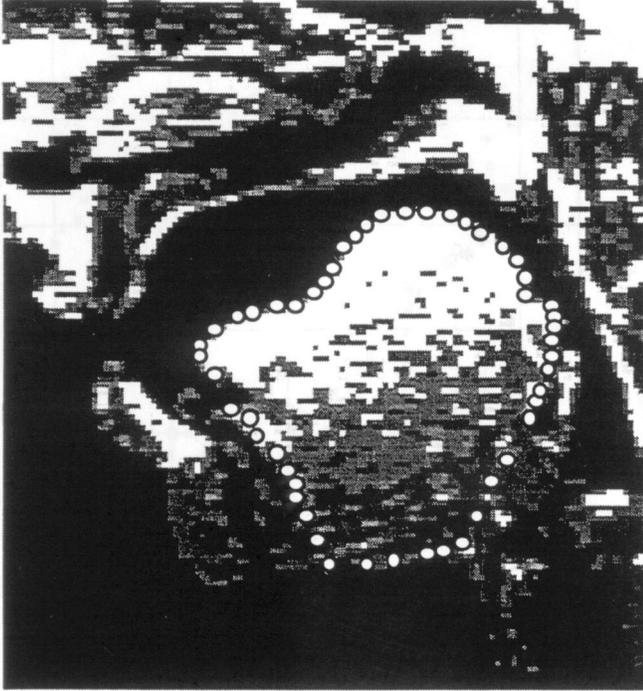


Figure 1: MRI image of the tongue in the /a/ position. (The lips are to the left and the airway is the black region above the tongue.) The tongue surface has been identified and marked by the large white circles.

4. MODELING TONGUE KINEMATICS

In developing a model of global tongue motion, the kinematic constraints associated with the volume preserving nature of the tongue and the physiological constraints of the musculature need to be accommodated. This procedure is similar to a kinematic model developed to describe cardiac motion [10,11,12]. The preliminary model presented here, examines only the mid-sagittal plane and does not limit the muscle stretches of the inter-fiber shear. It accounts for global rigid-body motions, homogeneous stretches in the longitudinal (x) and vertical (y) directions, homogeneous shear in the mid-sagittal plane, as well as simple bending in the vertical and horizontal directions. For example, if the position of a material point in the reference (rest) configuration is (X,Y) , then its displacement as a result of the horizontal, or x -bend deformation, is given by $u = -2a_4 XY$ in the x direction and $v = a_4 X^2$ in the y direction. These deformations would be sufficient to model the motion from rest to /i/ and from rest to /u/, but cannot account for the sharp depression in the /a/ image (which is a result of strong contraction in the frontal section of genioglossus). An additional deformation mode was therefore added, which accommodated this surface depression. This shape took the general form

$$u(r, s) = a_6 r^2 \exp(-bs^2)$$

where (r, s) are local orthogonal coordinates chosen such that r corresponds to the direction of the depression and s is tangent to the tongue surface, a_6 and b are constants set by the optimization of the model.

The next step was to choose the magnitudes of each deformation mode (the values of the model parameters) such that the position of the tongue surface as predicted by the model was as close as possible to the surface as observed using MRI imaging.

When optimizing the choice of model parameters using only surface information from MRI images, account has to be taken of the lack of a one-to-one correspondence between points representing the surfaces in the reference (rest) position and in the deformed (/a/ , /i/ and /u/) positions. One procedure used to find model parameters is to minimize the distance between the positions of the points in the reference configuration, as mapped by the model, to the line which joins the two nearest neighbor points defining the observed deformed tongue surface. Actual implementation of such an algorithm is enhanced when one first minimizes the distance from the position to which a pixel in the reference image is mapped (by the kinematic model) to the nearest identified surface pixel in the deformed image and then minimizes the distance to a line joining the two nearest pixels. In order to make the cost function a complete metric, symmetry requires minimizing with respect to both the forward and inverse mapping of the points; also the triangle inequality must hold and the cost function must be positive definite (both of which are satisfied by making the cost function the sum of distances). Optimization was done using the downhill simplex method (in 11 dimensions).

With such sparse data defining the motion of the tongue, for some models of deformation it is possible to find a number of physiologically unreasonable choices of model parameters, each of which describe deformations of the surface which result in shapes close to that which is observed. This is because only surface profile information is available. Examples of such deformations would include deformations akin to the hour-glassing seen in some finite element computations and strains which are regionally highly inhomogeneous or beyond the physiological limits of muscle contraction or extension. To counteract this possibility, we intend, in future, to include sarcomere stretches using the anatomy of the tongue as kinematic constraints.

With MRI tagging, the same discrete material points (identifying a single piece of muscle tissue) are imaged in both the reference and the deformed images. Tags are also distributed throughout the tongue musculature, so data from the tongue interior is also available. Again, an optimal choice of model parameters is one which minimizes the distance between the tag positions in the deformed position as observed and as computed by the model. However, in this case, the one-to-one correspondence between the tags in the reference and deformed images, makes the optimal choice of model parameters easier to find. The present dataset was untagged. This dataset was collected for other purposes, but has been used in this project as test data for developing the algorithm for identifying tongue kinematics from surface motion.

5. RESULTS

The measured surface points in the rest position were mapped (deformed by the model) into each of the other observed tongue shapes (/a/, /i/ and /u/) using various amounts of stretch, shear, bending and depression. Table 1 shows the value of each deformation parameter. The model motions do not fully account for all of the tongue movement, and figure 2, 3 and 4 show the difference between the actual tongue surface profile and the modeled surface for each shape (/a/, /i/ and /u/).

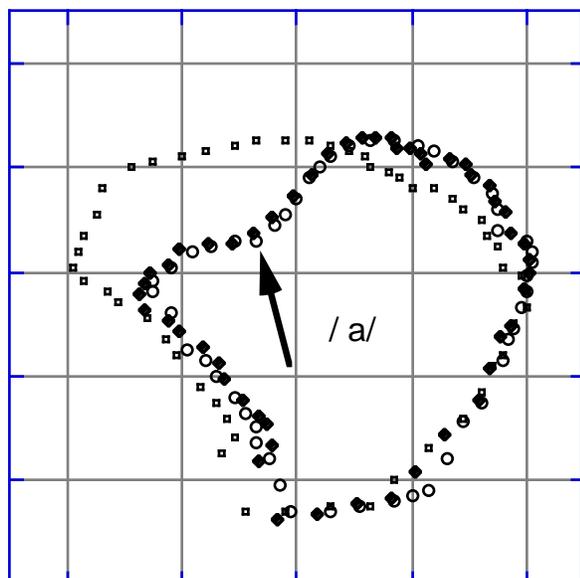


Figure 2: Comparison of the tongue surface profiles as modeled (black diamonds) and as observed (unfilled circles) for /a/. The small open squares are the rest position. The arrow points to the depression in the /a/ shape and is in the local positive r direction.

Notice that the primary motion in figure 2 is associated with the formation of the depression in the tongue upper surface. The areas of the rest profile and the /a/ profile are different, this means that the planar motion of the musculature in the sagittal plane must be accompanied by motion normal to that plane to accommodate the volume preserving deformation of the tongue. The values in table 1 show that the amplitude of the depression mode in the /i/ and /u/ sounds are less than 20% of that in the /a/ sound.

The primary motion in figure 3 is a contraction of the longitudinal dimension consistent with the dominant negative x -stretch and positive x -bend modes of deformation required to form /i/.

The primary motion in figure 4 is a combination of a contraction of the longitudinal dimension and a planar shear which lifts the rear of the tongue and depresses the front of the upper surface. These motions are shown by the dominant negative x -stretch and xy -shear modes of deformation required to form /u/ (see table 1).

The shapes shown in figures 2, 3 and 4 show that the tongue profile, as defined by the model, accurately represented the observed profile on the upper surface of the tongue. Near the back of the tongue, the MRI images did not clearly define the tongue

surface (see figure 1) and the "fit" in each case was weaker in this region. Also, in the observed /a/ image (see figures 1 and 2), there was a region of sharp shape change just under the tip that was not resolved by the model.

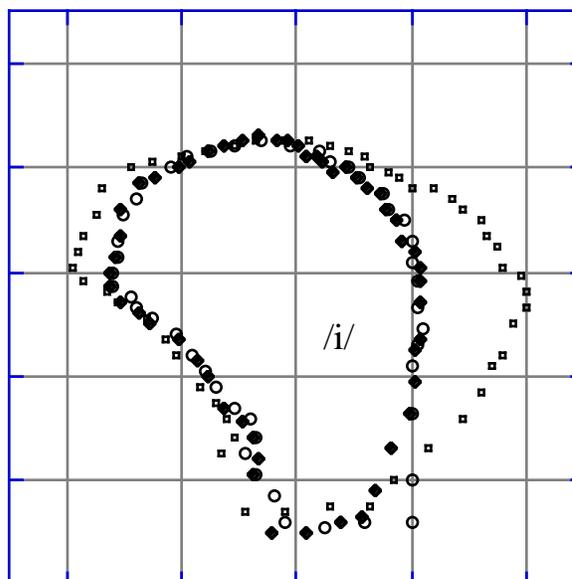


Figure 3: Comparison of the tongue surface profiles as modeled (black diamonds) and as observed (unfilled circles) for /i/. The small open squares are the rest position.

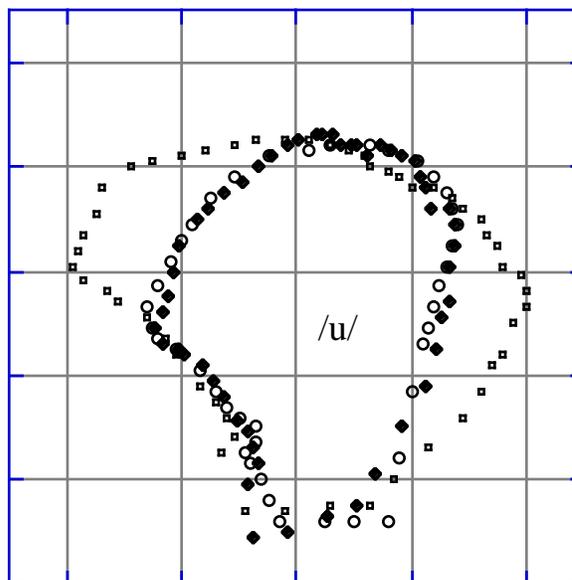


Figure 4: Comparison of the tongue surface profiles as modeled (black diamonds) and as observed (unfilled circles) for /u/. The small open squares are the rest position.

Table 1 shows the numerical values of the model deformation parameters (rigid-body motion parameters are not shown). These are consistent with our general notions of the required tongue

motions. For example, from rest to /i/ requires an overall contraction in the (longitudinal) front-to-back direction. The model quantifies this contraction since we can compute the overall stretch ratio in the x direction, $\lambda_x = 1 + a_1 = 1 - 0.2559 = 0.74$ (or 74%) which is effectively a Lagrangian strain $E_{xx} = -0.22$. However, this is just the strain associated with one component of the model deformation. The actual muscle fiber strains are given by all the deformation modes in the model. In particular, the x-bend will also dramatically affect the muscle strains E_{xx} .

Model parameter		/a/	/i/	/u/
x-stretch	(a_1)	-0.0915	-0.2559	-0.1864
y-stretch	(a_2)	0.0486	0.0596	0.0020
xy-shear	(a_3)	-0.0624	-0.0258	-0.1992
x-bend	(a_4)	0.0353	0.2167	0.2047
y-bend	(a_5)	0.0151	0.1349	0.1600
Depression	(a_6)	0.4129	0.0875	-0.0845

Table 1: Model parameters for each of the sounds /a/, /i/, and /u/. The deformation mode magnitudes do not have the same units. Comparison of mode magnitudes (a 's) from sound to sound is meaningful.

As an example of the strain distribution, figure 5 shows a gray-scale image of the contours of Lagrangian shearing strain in the x - y plane (E_{xy}). The line of action of the depression is indicated by a sharp transition from positive to negative in the shear strain (figure 5). This sharp change in shear direction may be due to the localized contraction of genioglossus (see figure 2).

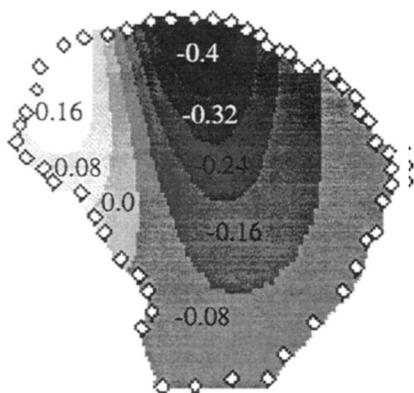


Figure 5: The rest tongue profile showing a gray-scale image of the contours of Lagrangian shearing strain in the x - y plane (E_{xy}) for /a/.

Other strain profiles (for /i/ and /u/) are much less heterogeneous, showing a gradual variation in strain associated with more uniform muscle activation. As an example, the contours of Lagrangian

shearing strain in the x - y plane (E_{xy}) for /i/ are shown in figure 6. The large positive values for shear near the forward region of the tongue show that this area has to be sheared downwards. The negative shearing at the back of the tongue shows that this region is to be sheared upwards to form /i/.

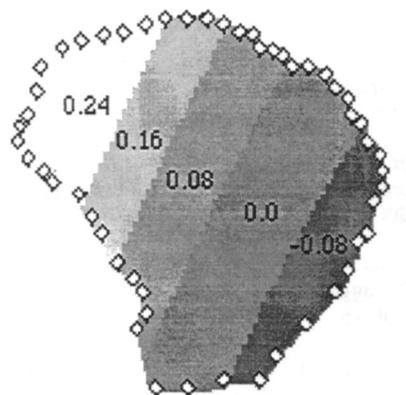


Figure 6: The rest tongue profile showing a gray-scale image of the contours of Lagrangian shearing strain in the x - y plane (E_{xy}) for /i/.

6. CONCLUSIONS

Models of tongue motion can identify the specific modes of deformation required to produce the tongue shapes associated with various sounds. When the only information available is the tongue surface profile (such as in untagged MRI images or from ultrasound imaging of the tongue during speech) a global kinematic model has to be developed for a description of tongue motion.

The tongue deformation model can then be used to identify the magnitude of its different deformation modes and can be used to quantify physiologically important features, such as strains and muscle stretches.

The absence of discrete point motion and data in the tongue interior dramatically increases the complexity required to resolve motion and may be a source of significant uncertainty in the results.

Future models need to include the appropriate penalties for biologically unlikely motion (excessive sarcomere stretch, for example). Three dimensional models will have to be formulated with the kinematic constraint of locally volume preserving deformation modes.

7. ACKNOWLEDGMENTS

Supported (in part) by research grant number R01 DC 01758 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

8. REFERENCES

1. Hunter, W.C., & Zerhouni, E.A. "Imaging distinct points in left ventricular myocardium to study regional wall deformation."

In: J.A. Anderson (ed.) *Innovations in Diagnostic Radiology*, Springer-Verlag, New York, 1989, pp. 169–190.

2. Westbury, J. R. (1991) The significance and measurement of head position during speech production experiments using the x-ray microbeam system. *J. Acoust. Soc. Amer.*, 89 (6), 1782-91.
3. Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (1993) Electro-magnetic midsagittal articulometer systems for transducing speech articulatory movements. *J. Acoust. Soc. Amer.*, 92, 3078-96.
4. Unser, M., and Stone, M. "Automated detection of the tongue surface in sequences of ultrasound images." *J. Acoust. Soc. Amer.*, 91: 3001–3007, 1992.
5. Zerhouni, E.A., Parish, M.S., Rogers, W., Yang, A., and Shapiro, E.P. Tagging of the human heart by magnetic resonance. *Radiology*, Vol. 16, 59–63, 1989.
6. Stone, M., and Lundberg, A. (1996) Three-dimensional tongue surface shapes of English consonants and vowels. *J. Acoust. Soc. Amer.* (in press-June).
7. Baer, T., Gore, J., Gracco, C., and Nye, P. (1991) Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. *J. Acoust. Soc. Amer.*, 90, 799-828.
8. Narayanan, S., Alwan, A. and Haker, K. (1995) An articulatory study of fricative consonants using magnetic resonance imaging. *J. Acoust. Soc. Amer.*, 98 (3), 1325-47.
9. Moore, C. (1992) The correspondance of vocal tract resonance with volumes obtained from magnetic resonance images. *J. Speech and Hearing Res.*, 35, 1009-1023.
10. Arts, T., Hunter, W.C., Douglas, A.S., Muijtjens, A.M.M., and Reneman, R.S. Description of the deformation of the left ventricle by a kinematic model. *J. Biomechanics*, Vol. 25, 1119–1127, 1991.
11. Waks, E., Prince, J.L. and Douglas, A.S. Cardiac motion simulator for tagged MRI, to appear in Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, San Francisco, June 21-22, 1996.
12. Arts, T., Hunter, W.C., Douglas, A.S., Muijtjens, A.M.M., Corsel, J.W. and Reneman, R. S. Macroscopic three-dimensional motion patterns of the left ventricle. In *Interactive Phenomena in the Cardiac System*. Plenum Press, New York, 383–392, 1993.