Modeling the motion of the internal tongue from tagged cine-MRI images

Maureen Stone[a)]
Division of Otolaryngology, University of Maryland School of Medicine, 16 South Eutaw Street, Room 525, Baltimore, Maryland 21201

Edward P. Davis and Andrew S. Douglas
Department of Mechanical Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218

Moriel NessAiver and Rao Gullapalli
Department of Radiology, University of Maryland School of Medicine, 22 North Greene Street, Baltimore, Maryland 21201

William S. Levine
Department of Electrical Engineering, University of Maryland, A. V. Williams Building, College Park, Maryland 20740

Andrew Lundberg
Department of Computer Science, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218

(Received 19 April 1999; accepted for publication 30 November 2000)

A new technique, tagged Cine-Magnetic Resonance Imaging (tMRI), was used to develop a mechanical model that represented local, homogeneous, internal tongue deformation during speech. The goal was to infer muscle activity within the tongue from tissue deformations seen on tMRI. Measurements were made in three sagittal slices (left, middle, right) during production of the syllable /kɑ/. Each slice was superimposed with a grid of tag lines, and the approximately 40 tag line intersections were tracked at 7 time-phases during the syllable. A local model, similar to a finite element analysis, represented planar stretch and shear between the consonant and vowel at 110 probed locations within the tongue. Principal strains were calculated at these locations and revealed internal compression and extension patterns from which inferences could be drawn about the activities of the Verticalis, Hyoglossus, and Superior Longitudinal muscles, among others. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1344163]

PACS numbers: 43.70.Aj, 43.70.Bk [AL]

I. INTRODUCTION

It is difficult to measure the movements of local regions within the human tongue during speech and determine their effect on the tongue surface. It is useful to do so, however, because the human tongue has a complicated musculature that is not easily reconciled with the deformation patterns seen on the tongue surface during speech. Moreover, the relationship between muscle activity and surface deformation is an important component in understanding how speech is controlled and how it is disrupted in various disorders. As a first step, the model and data presented here are intended to add details about the local kinematics of the tongue during speech. By interpolating the deformation of the tongue between observed MRI tag-points in a reference (consonant) and deformed (vowel) state, detailed information on the (2D) strain and in-plane muscle contraction was obtained. This allowed us to investigate tissue deformation patterns and infer related muscle activity.

There is a small body of material detailing the activity of the internal tongue using tagged Magnetic Resonance Imaging. These data have been used to predict tongue muscle contractions for rest-to-vowel motion (Niitsu *et al.*, 1992; Kumada *et al.*, 1992; Dang and Honda, 1997) and nonspeech movements (Napadow *et al.*, 1999a, 1999b). Some anatomical information on tongue muscles is also available from dissections and histological studies (Abd-El Malek, 1939; Carpentier and Pajoni, 1989; Doran, 1975; Miyawaki, 1974; Shawker *et al.*, 1984; Takemoto, 2001).

Electromyography (EMG) data are another source of muscle information (Miyawaki, 1975; MacNeilage and Sholes, 1964; Sauerland and Mitchell, 1975). Within the tongue, however, EMG data are difficult to collect and interpret due to substantial muscle interdigitation, cross-talk, diffuse fiber distribution of the muscle of interest (Perlman *et al.*, 1989) and fatigue (Faber and Raphael, 1989). Promising work has used EMG to predict articulator movements

[a)]Electronic mail: mstone@umaryland.edu

from speech motor commands (Honda and Kusakawa, 1997), and has compared EMG to tongue configurations (MacNeilage and Sholes, 1964; Maeda and Honda, 1994; Niimi *et al.*, 1994), and acoustic spectra in vowels (Maeda and Honda, 1994; Baer *et al.*, 1988). Nonetheless, EMG is not a commonly used instrument, because it is unpleasant, invasive, and difficult to interpret. Tagged MRI could supplement EMG if it successfully represented tongue muscle contraction.

Soft tissue such as tongue musculature is generally regarded as volume preserving (Fung, 1993). Therefore, tongue motion is the result of local tissue distortion (without volume change). Since muscle activation is contractile in nature, there is an associated elongation of tissue in the direction transverse to the primary contraction. By measuring the local tongue deformation, principal stretches and their associated directions can be determined. Using our knowledge of tongue anatomy, these directions can be compared with the orientation of the musculature to deduce which muscles locally are active during a particular phase of tongue motion. These inferences can provide important information about the neural control and muscle activation in both normal and disordered speakers. In the present study, a simple model of one plane of the tongue was developed based on a finite-element type mesh derived from the movement of tagged Cine-Magnetic Resonance Imaging (tMRI) tag-points. The modeled plane could be probed at regions of any size and principal strains could be calculated for that region.

Four concepts are necessary to keep in mind when examining how MRI tagging is accomplished. First, only the spins aligned along the Z axis can become excited by the applied rf energy. Second, only the spins on the XY or transverse plane can produce an MRI signal. Third, the signal that corresponds to a single pixel in the image is actually the vector sum of more than 1015 spins. That signal can be destroyed by causing the spins in a single pixel to lose phase coherency. Fourth, the time necessary for the spins to realign along the Z axis is measured in hundreds of milliseconds (ms). Tag lines are stripes of tissue that have been tipped onto the XY plane and then forced to become dephased, temporarily destroying their ability to produce a coherent signal. A simplified description of how this is accomplished follows.

First, apply an rf pulse that tips all spins in a plane by 45°. Next apply a linear magnetic gradient along one direction in the imaging plane, temporarily changing the resonance frequency of these excited spins, resulting in an accumulation of phase along that direction, say ten complete cycles across the field of view (FOV). There are now ten lines that have the same phase (call it 0°) immediately after the 45° rf pulse and ten lines in between that are 180° out of phase. Now apply a second rf pulse of $-45°$. Those spins with 0° phase will be tipped back up to the Z axis, those with 180° phase will be tipped a total of $-90°$ down onto the XY plane. Finally, another large gradient is applied that totally dephases all of the spins on the XY plane. These spins are no longer producing a coherent signal and cannot be re-excited until they have recovered a significant portion of their Z magnetization. The process lasts on the order of a few ms, producing alternating regions of tissue that produce no signal or full signal. The number of bands, and hence the distance between dark stripes, is determined by the strength and duration of the gradient between the two rf pulses. A dark grid is applied by first laying down horizontal lines (x axis) followed by vertical lines (y axis). Once these tag lines of little or no signal have been created, a standard cine (motion) imaging sequence can be used to track the motion and deformation of the tag lines.

Deformation models have been developed for the heart, which include circumferential shortening and stretching of the ventricles, allowing tagged Cine-MRI (tMRI) to study myocardial deformation (cf. McVeigh, 1996; McVeigh and Atalar, 1992). tMRI of the heart provides displacement and motion of discrete points, allowing computation of principal strains and directions. The heart differs from the tongue, however, in important ways. The heart has relatively simple, repetitive, kinematics that do not vary much with cardiac contractility, stroke volume, preload, or afterload. Consequently there is no regional variation (except in the case of infarction). Heart beats are highly repeatable and are independent of the rate of contraction (Douglas *et al.*, 1991). Heart beat cycles are easily synchronized temporally to the electrocardiogram (ECG) because the heart produces a large electrical depolarization (R-wave peak) immediately prior to contraction.

What is different about this study of the tongue is that it examines a soft tissue structure with highly variable and complex deformation caused by the differential stimulation of an anatomically highly complex (architecturally) musculature. The small rapid internal movements of the tongue used in speech provide a challenge for the tMRI procedure, since it has a fairly poor temporal (18 Hz) and spatial (1.9 mm$\times$1.9 mm$\times$7.0 mm) resolution. This study, therefore, was an experimental application of tMRI to speech, to determine the extent to which small speech movements could be resolved. Measurements were made of the small deformations internal to the tongue during CV syllables for evidence of regional activity that reflected specific muscle contractions.

The goal of this study was to show the application of tMRI and a fairly simple model of tongue deformation as a means of illustrating internal tongue stretch and inferring muscle activity during speech. The tMRI procedure and local model presented here identified the direction and extent of stretch at local tissue points. From the stretch and shear data local principal strains were calculated and underlying muscle contraction patterns were inferred. Because of the experimental nature of this method the model was applied only to a limited, preliminary data set.

II. METHODS

A. Subject and speech materials

A 19-year-old male, native speaker of English demonstrated the application of the model. The subject had no den-

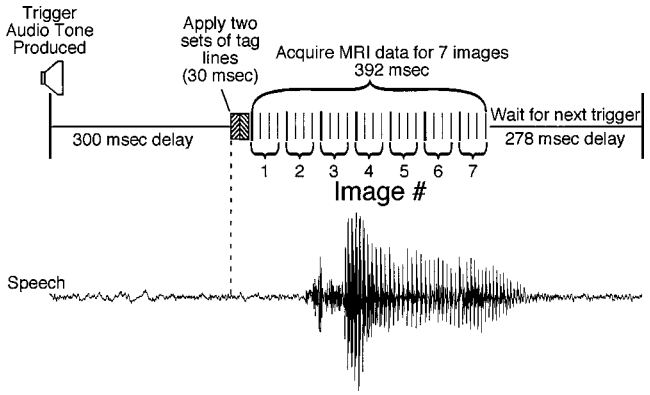

FIG. 1. Procedure for producing a tag and measuring seven time-phases during each repetition.

tal fillings that might interfere with the MRI magnetic field. The speech materials were the syllable /kɑ/. A standard MRI head coil was used to record the MRI data and also position the head. There appeared to be no extraneous head motion (see Sec. I D below). Three tMRI slice sequences were collected in the left, middle, and right sagittal planes. The syllable /kɑ/ was repeated 32 times per slice, or 96 times in total. The spatial quality of the image depended on the ability of the subject to repeat the syllable accurately 96 times on cue, at one per second. An ECG simulator triggered a tonal cue at the *R*-wave peak of the ECG simulator and created the tags 300 ms later (Fig. 1). At a later date, an audio recording was made as the subject repeated the syllable 16 times to the beat of a metronome. Acoustic analysis showed that the /k/ release occurred a mean of 534 ms (s.d. 37) after the acoustic cue, suggesting that the first data frame was collected within or before the stop, with the tongue pre-positioned in consonantal position. Visual inspection of the tMRI data during collection indicated that the tongue began moving downward in the second or third time-phase, with the maximum tongue position always in the first time-phase.

B. Instrumentation and data collection

A Picker 1.5 Tesla Edge System collected the data using tagged Cine-MRI. The subject lay supine in the MRI scanner with the neck coil positioned to image the area from the lower nasal cavity to the upper trachea. Three sagittal slices, left (L), middle (M), and right (R), were collected. During each syllable repetition, 7 time-phases were recorded at 56 ms per time-phase, or the equivalent of 18 samples per second (Fig. 1). The signal emitted by the protons, in a single time-phase in a single slice, was summed across multiple repetitions of the same syllable producing a single image-sequence of 7 time-phases per syllable, i.e., a cyclic sequence of the C-to-V motion. The first four time-phases are displayed in Fig. 2, showing the movement from /k/ to /ɑ/.

A SPAMM-like procedure (Spatial Modulation of Magnetization prior to imaging) (Axel and Dougherty, 1989a, 1989b) was used with the following parameters: a gradient echo with an echo time of 4.6 ms and a repetition time of 14 ms, a phase encode group size of 4, a field of view of 24 cm, and a slice thickness of 7 mm. A series of five rf pulses was used, with relative amplitudes of $[+1-4+6-4+1]$ to produce sharp dark tag lines separated by a distance of 11 mm. This complex rf pulse was applied sequentially to the *x* and *y* axis taking a total time of 30 ms, and resulting in a grid of horizontal and vertical tag lines. Seven time-phases (56 ms each) were recorded during the succeeding 392 ms as the tongue moved into the vowel (see Fig. 1). Therefore, the one second repetition time is allotted to a time lag of 300 ms between the trigger and the tag, a period of tag onset and decay lasting about 420 ms (which includes creating the tags and the 7 time-phases), and a final period (about 280 ms) devoted to looking for the next trigger.

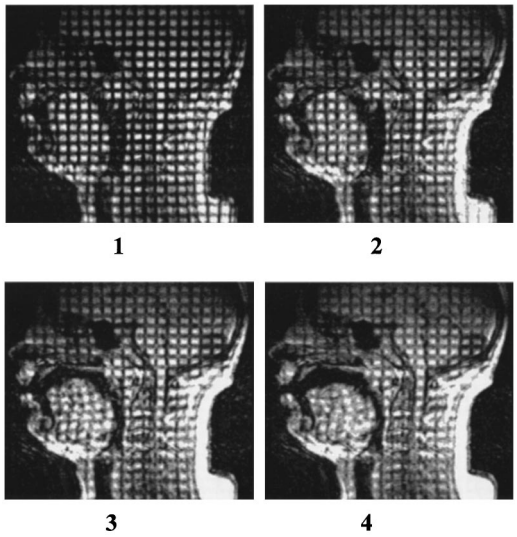

FIG. 2. MRI image sequence of the first four frames in the /k/ to /ɑ/ motion. (a) is the /k/ and (d) is the /ɑ/. (b) and (c) are the transitional frames.

Thirty-two repetitions were needed for reconstruction of the 7 time-phases in a single plane, because only 4 of the 128 lines of *k*-space were acquired per repetition $(128/4=32)$. It should be noted that *k*-space is Fourier or frequency-based space. The third line in each group of four contained the high frequency or edge definition material. By not collecting fewer lines of *k*-space at each sweep we sacrificed some edge definition and temporal resolution. The gain was reduced syllable repetitions and subject fatigue. A similar collection process was done without tagging by Masaki *et al.* (1999). That study differed from the present one in three ways. First, their temporal resolution was 25 ms and ours 56 ms. Second, their acquisition required 128 repetitions for each slice while ours used 96 repetitions for three slices. Third, they were only able to look at gross anatomical features; the use of tagging lines in this paper enabled us to look at deformation of the tongue tissue with a resolution of roughly 2 mm.

In tMRI images, the intersection of the tag lines are material points that move with the tissue. Based on the motion of these discrete points, the motion of the entire tongue was interpolated using the same low-order polynomials that are employed in finite element analysis (Cook *et al.*, 1989). Movement of the tongue normal to the sagittal plane was assumed to be insignificant at midline, and indeterminate in the parasagittal slices without 3D data.

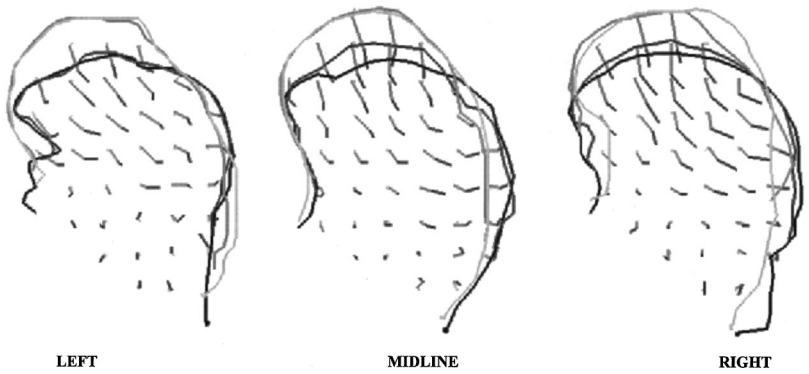

FIG. 3. Time course of tag movement for first four time-phases of the /k/ to /ɑ/ movement. Line color progresses from light gray (/k/) to black (/ɑ/) over time. All MRI based figures are oriented with the tongue tip on the left.

C. Internal measurement and modeling

The intersections of tag-lines with each other and with the tongue surface were measured for all time-phases between the consonant and vowel and stored as *xy* coordinates. There were about 40 intersections or ''tag-points'' in each image. The lowest anterior point of the tongue was defined as the intersection of the inferior-posterior edge of the jaw (black) and the nearest horizontal tag line in the tongue [Fig. 2(a)]. The bone of the jaw is black and the marrow within the jaw is light gray. It is possible that the black region anterior to the tongue included air from the sublingual cavity. If the black area were air (i.e., soft tissue unattached) more motion would be expected in the adjacent tongue region than if the black were bone (soft tissue attached). Air cannot be distinguished from bone on MRI images, so the black area could not be better defined. Movement of the lower tongue, especially the anterior region, was quite small (Fig. 3) suggesting a bony attachment. The lowest posterior point was defined similarly as the intersection of the inferior–posterior surface of the epiglottis with the nearest horizontal tag line.

Once the anterior and posterior tag-points were specified at the base of the tongue, all the tags were tracked across the sequence. We were consistent in the points chosen across time-phases. The experimenter was assisted in tracking the points by (1) observing tag position in the preceding and following time-phases and (2) digitally superimposing the reference tag-points on the current image. Measured points were within the tongue proper, that is, above the jaw-opening muscles, which were brighter than the tongue. As the tags decayed, muscle distinctions became clearer. Therefore, after all the time-phases were measured, the base of the tongue was determined on the vowel images [Fig. 2(d)]. Tags on previous time-phases were then added or deleted to equalize the number of points on all images within a single image sequence. The vowel frame was defined as the last time-phase before the posterior tongue surface changed direction back toward the consonant, and was typically the third, fourth, or fifth time-phase. 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, the entire tongue usually appeared to reach maximum in a single frame.

As the tongue moved between two sounds the tags deformed, reflecting the internal tissue deformation. Principal strains were calculated on the extreme consonant and vowel shapes, because the intervening movements were too small to rule out measurement error.

1. Deformation model

The extreme consonant (reference) and vowel (deformed) tag-points were modeled using principal strains. While these intersection points were treated as tissue points in the plane, they actually represent the motion of a volume with a depth of 7 mm perpendicular to the image plane. Nodal motion therefore represented an average of the tongue motion across that region. In the reference state the nodes had coordinates (X_j, Y_j). In the deformed state k (where $k=1,...,7$ indicating which of the time phases was being examined) the nodes had coordinates $(x_j^{(k)}, y_j^{(k)})$.

The two in-plane displacement components, $u_j = u(X_j, Y_j) = x_j^{(k)} - X_j$ and $v_j = v(X_j, Y_j) = y_j^{(k)} - Y_j$, were calculated at each node. A finite-element type mesh, made up of triangular and quadrilateral elements, was created from these nodes. The details of the node point identification algorithms and point-tracking procedures can be found in Davis (1999). Displacement of any arbitrary points within the tongue was calculated from the finite-element type basis functions, $\phi_j(X,Y)$ and the displacements measured (u_j, v_j) at each associated node. These basis functions have the characteristic that $\phi_i(X_j, Y_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta defined by $\delta_{ij}=1$ if $i=j$ and $\delta_{ij}=0$ if $i \neq j$.

First the number of closest nodes, N, was found which defined the quadrilateral $(N=4)$ or triangular $(N=3)$ element in which any generic point was located. Second, the N closest nodes, (X_j, Y_j), $j=1,...,N$, were used to define the basis functions, $\phi_j(X,Y)$ in that element (Cook *et al.*, 1989). The displacement of any point in an element could then be found using the known displacements, u_j and v_j, and the basis functions $\phi_j(X,Y)$ from

$$u(X,Y) = \sum_{j=1}^{N} u_j \phi_j(X,Y)$$

and (1)

$$v(X,Y) = \sum_{j=1}^{N} v_j \phi_j(X,Y).$$

2. Computing deformation, stretch and strain

This principal strain model gives the simplest possible kinematic interpretation to local tongue deformations. The tag displacements can be computed from Eq. (1), and we can find the in-plane components of the displacement gradient tensor, **H**, namely

$$
\begin{aligned}
H_{11} &= \frac{\partial u(X,Y)}{\partial X} = \sum_{j=1}^{N} u_j \frac{\partial \phi_j(X,Y)}{\partial X}, \\
H_{22} &= \frac{\partial v(X,Y)}{\partial Y} = \sum_{j=1}^{N} v_j \frac{\partial \phi_j(X,Y)}{\partial Y}, \\
H_{12} &= \frac{\partial u(X,Y)}{\partial Y} = \sum_{j=1}^{N} u_j \frac{\partial \phi_j(X,Y)}{\partial Y}, \\
H_{21} &= \frac{\partial v(X,Y)}{\partial X} = \sum_{j=1}^{N} v_j \frac{\partial \phi_j(X,Y)}{\partial X},
\end{aligned}
\tag{2}
$$

where $\phi_j(X,Y)$ and the derivatives of $\phi_j(X,Y)$ are known everywhere in any element.

Principal strains and their directions can be computed from the two eigenvalues, E_i, and their associated eigenvectors, $\mathbf{N}_i$, of the in-plane Lagrangian strain tensor, $\mathbf{E} = \frac{1}{2}[\mathbf{H}\mathbf{H}^T + \mathbf{H}^T\mathbf{H}]$. The stretch, $\lambda_{\mathbf{M}}$, in any direction, **M**, can be computed from the relationship $\lambda_{\mathbf{M}} = \sqrt{2\mathbf{M}\cdot\mathbf{E}\mathbf{M}+1}$ (Lai *et al.*, 1993, p. 134).

D. Measurement error and repeatability

Three possible sources of error affect the MRI measurements. The first source is image quality, and includes spatial and temporal resolution of the image, as well as magnetic distortion due to tagging. The second is human measurement error, and the third is speaker variation, specifically speech precision and head movement.

Spatial resolution of the acquired MRI image, was based on a 128×128 pixel grid. Resolution was increased to 256 ×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 0.95 mm. Thus the minimum resolvable movement, absent gray scale, was 2 pixels or 1.9 mm.

The temporal resolution was reduced by the large time window. The time window (56 ms/time-phase) meant that tongue movement might cause blurring of the surface in an image. The magnetic disturbance of the tags caused frames with newly laid tags (consonants) to have better tag quality, and poorer surface clarity, than later frames (vowels) where the tags have decayed (see Fig. 2). Poor image clarity increases human measurement error. Human measurement error was assessed by having a second judge re-measure the tag-points in the midline /k/ and /ɑ/ time-phases. For the consonant frame, the average (maximum) error value was $x = 0.5$ (0.95) mm, and $y = 0.5$ (1.9) mm. For the vowel frame the average and maximum errors were larger: $x = 1.4$ (4.8) and $y = 1.8$ (6.6) mm. These numbers reflect human measurement error, image resolution and speaker precision.

Speaker contributions to error included temporal precision (discussed in Sec. II A above) and extraneous head motion. The latter was considered, because the head coil limited but did not preclude head motion. Six specific tag-points, three vertical and three horizontal, were measured in the brain portion of the image of the consonant and vowel time-phases. Five of the six points differed by one pixel, the sixth differed by one in each direction. These differences were within the measurement error indicating no evidence of C-to-V head motion.

III. DATA APPLICATION

A natural use of the principal strain model is to infer activity of tongue muscles from regions of compression. This is done by measuring the tag-points consistent with the muscles' lines of action in the consonant (reference) and vowel (deformed) time-phases. A wire mesh of both data sets is created by applying the deformation model and calculating in-plane principal strains throughout the tongue. The patterns of compression are then examined for patterns reflective of muscle contraction.

A. Tag-point measurements

Tag-points were measured in all four time-phases from /k/ to /ɑ/ (Fig. 3). In all images the tongue tip is on the left. The figures show time-lapse movement of the individual tags with the first time-phase (/k/) in gray and the last (/ɑ/) in black. The tag trajectories depict local differences in motion. For example, the midsagittal slice shows directional differences; the top third moves downward and the central third backward. The local motion also differed in range; the top moved farther than the center.

B. Wire mesh deformation grids

Tag-point wire mesh grids for left, middle and right slices of the /k/-to-/ɑ/ deformation appear below (Fig. 4). This figure, and the derived strains (Fig. 5), cannot be directly compared to Fig. 3. These figures present deformations of tag grid cells (four points), not single tag-points, and are observed only at the extreme consonant and vowel position. The intervening time-phases are not considered. The grids are made by connecting with straight lines all the tag-points, including surface tag-points, which are the intersections of the tag lines with the surface. This is why the surfaces of Fig. 4 look angular. The in-plane grids for the /k/ and /ɑ/ contain the same tissue points; the differences are due to changes in the shape of the tongue. For viewing convenience, four cells have been shaded on each grid, indicating local deformation at the top, central, back, and bottom regions. An uncompressed cell is 1.1 cm per side.

In Fig. 4 the top lowered substantially in all three slices (gray to black). The posterior tongue moved backward about half that amount. The upper-back was lower for /k/-L,M than /k/-R (gray). The shaded cells in the center exhibit horizontal expansion, shear, and vertical compression, especially on the right (black square). The topmost cells compressed and lowered vertically, while the upper-back rotated. The bottom showed some horizontal expansion.

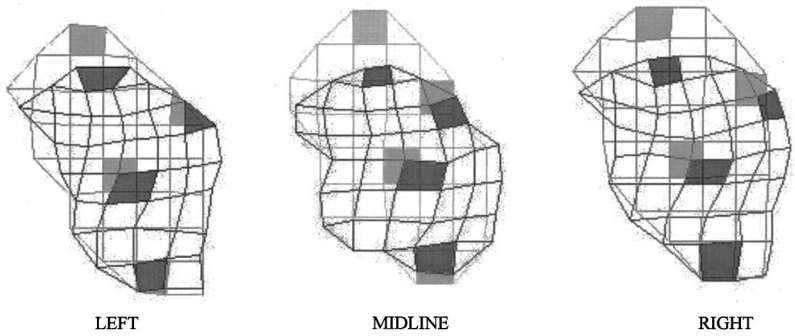

FIG. 4. Wire mesh representations of /k/ (gray) and /ɑ/ (black) for left, middle, and right slices.

C. Principal strains

Once the model was developed, principal strains could be calculated for any portion of the tongue, cover any size area, and reflect the strain at that location. Principal strains were computed at 110 locations in the tongue. The strain at each location was derived from the tag-based model. Principal strains for /kɑ/ appear in Fig. 5. Gray indicates extension and black compression.

The lower third of the tongue compressed primarily downward and backward accompanied by upward and backward extension. The middle tongue compressed vertically and extended horizontally a small amount; the top compressed inwardly. In all three slices the tip extended: forward at midline, forward/upward at left and right. On the left a band of tags just below the tongue surface compressed lengthwise, parallel to the tongue surface (black lines). On the right the two surface tag rows also were distinct from the deeper tags by being more greatly compressed (see Fig. 5). The location and direction of local compression could be compared to the anatomy seen in Fig. 6.

IV. DISCUSSION

The present study is the first to use tagged Cine-MRI to mathematically model tongue muscle activity during speech, although Napadow *et al.* (1999a, b) used a similar procedure to study the midsagittal tongue during swallowing and nonspeech movements. Several features of the present study differ from other speech studies. First, previous studies have examined the midsagittal plane and inferred muscle action through visual inspection. Niimi *et al.* (1994) visually inspected midsagittal MRI tag line deformations and hypothesized muscle contraction. Kumada *et al.* (1992) and Niitsu *et al.* (1994) drew lines manually on midsagittal slices to connect deformed tag-points. From these they inferred active muscles. In the present study, principal strains modeled regional homogeneities and allowed inference of underlying muscle contraction patterns. The second feature was the choice of a consonant as the reference shape. The above studies used rest as the reference position, but rest position is variable, because it is not constrained by any acoustic requirements, and the starting position of the tongue affects what muscles are used to achieve the end position. The present study used /k/ because it is an extreme shape and requires a specific downward movement into the vowel. The third feature is that the present study used Cine-MRI, which allowed us to choose the target frame from an actual motion sequence, rather than examine the deformation from one sustained position to another.

A. Internal segments and muscle activity

This study extracted principal strains for small local tongue regions in an effort to expose compression patterns consistent with lines of action of possible tongue muscle contractions. The wire mesh grids (Fig. 4) depict local deformations; the principal strains (Fig. 5) reflect local tissue compression and expansion. Muscle lengthening (expansion) must always be passive, because active muscle contraction causes shortening. Muscle shortening, however, can be due to active muscle contraction, or to passive compression of the tissue. In addition, the tongue is globally and locally volume preserving, i.e., tissue cannot be increased or decreased, only relocated. Therefore, when both in-plane strains at a single point indicate compression, extension must have occurred orthogonally to the slice, i.e., in the cross-sectional, or lateral plane.

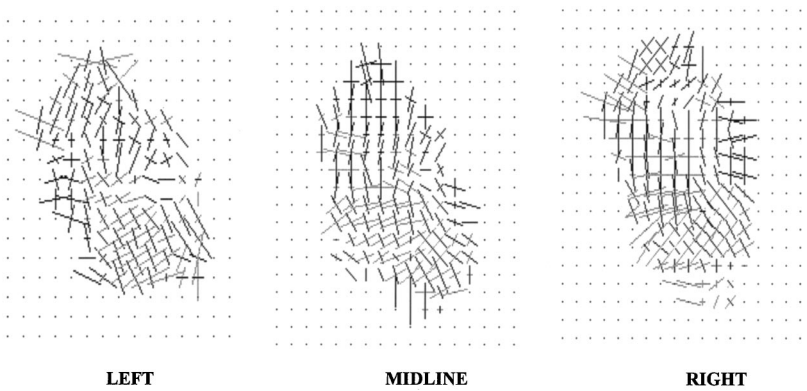

FIG. 5. Principal strains for the /k/ to /ɑ/ deformation for left middle and right slices. Black lines indicate local tissue compression and gray indicate local expansion.

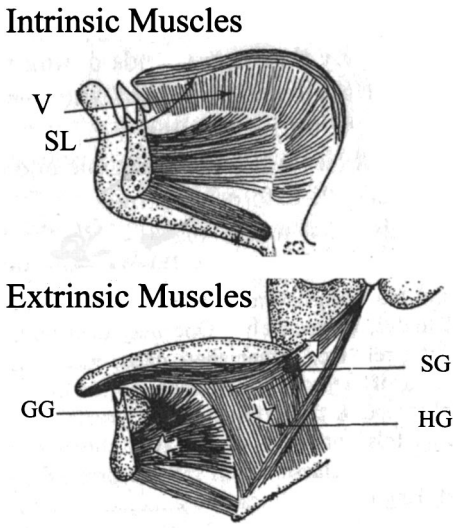

FIG. 6. Muscles of the tongue: Superior Longitudinal (SL), Verticalis (V), Genioglossus (GG), Styloglossus (SG), and Hyoglossus (HG). From Dew and Jenson, *Phonetic Processing*, 1977, pp. 92–93.

In Fig. 5, the principal strains isolated several local regions: the lower third, the upper tongue and the circumference. These regions were used to predict the role of specific muscles in the deformation. First, consider the lower third of the tongue. The downward and backward compression seen in the lower third of the tongue (Fig. 5) is consistent with Hyoglossus (HG) muscle contraction (Fig. 6). Next consider the upper tongue in Fig. 5. Compression here was almost uniformly vertical, particularly at midline and right, consistent with Verticalis contraction (Fig. 6). Verticalis fibers interdigitate completely with several other muscles and so cannot be isolated using EMG. As a result, the muscle does not always appear in tongue muscle models (cf. Dang and Honda, 1997; Honda and Kusakawa, 1997), or its role is based on anatomy, not physiology (Perkell, 1974; Wilhelms-Tricarico, 1995). The vertical compression in the middle and upper tongue was sharply separated spatially from the backward compression of the lower third. This spatial distinction greatly reduces the likelihood that HG alone lowered and backed the tongue, and strongly supports Verticalis contraction as a part of a rapid tongue lowering gesture for /kɑ/. Contracting both muscles would increase the rate and distance of surface lowering compared to HG alone.

The third region that acted as a unit was the circumference of the tongue—the outer two rows of tags. On the left particularly, circumferential shortening of the tongue surface was consistent with contraction of the Superior Longitudinal muscle (SL), as can be seen in the strains of the left side (Fig. 5). SL (see Fig. 6) may have as many as three roles in tongue movement. First, SL contraction shortens the tongue, because it is a lengthwise muscle (Kier and Smith, 1985). Since the tongue surface is curved in the sagittal plane, SL contraction should shorten tongue length circumferentially from the tongue tip to the hyoid. Second, SL, in combination with Genioglossus Anterior, can elevate the tongue tip (Napadow *et al.*, 1999b, p. 4; Smith and Kier, 1989; Thexton *et al.*, 1998). To cause upward curling, GGA contraction would locally stiffen the tongue behind the tip, which is consistent with the small concavity often seen in /ɑ/. At the same time, SL would compress the upper surface longitudinally. This must be done without simultaneous compression of the lower surface (Inferior Longitudinal) or the tongue will merely shorten (Kier and Smith, 1986). In the present data, the tongue tip expanded horizontally at midline, and obliquely on the sides, orthogonal to and consistent with SL activity. This expansion, however, did not result in tongue tip protrusion or elevation (Fig. 4), because other muscles pulled the entire tongue backward and downward. As a third role for SL, we propose that contraction of the orthogonal SL and Verticalis fibers might create a ''belt'' that stiffens the tongue circumference and reduces its degrees of freedom. A belt could be used to resist outward expansion. Resistance to expansion was inferred from the strains in the back two rows of the rear tongue, which rarely extended horizontally despite the backward movement of /ɑ/ (Fig. 5). One cannot completely discount, however, the possibility that this, or any of the other strains, is due to passive compression created indirectly by the contraction of other muscles or to the supine position.

Genioglossus anterior (GGA) has long been associated with local depressions of the anterior tongue. For example, in tMRI images, Niimi *et al.* (1994) attributed downward motion from rest-to-/ɑ/ to GGA. Their data exhibited a depression posterior to the tip. Considerable ultrasound data has shown the same depression in /ɑ/ (cf. Stone and Lundberg, 1996). More importantly, Davis *et al.* (1996) found that for steady state /ɑ/ the local depression behind the tip could be modeled as a shear deformation consistent with GGA contraction. Anatomy studies also have found that GGA fibers do not curve forward or extend into the tip (Takemoto, 2001). Contraction of fibers oriented vertically and inserting behind the tip will lower the surface immediately posterior to the tip, facilitating tip elevation. Figure 4, in the current data, showed vertical compression behind the tongue tip that was in agreement with GGA fiber direction. Inspection of the midsagittal compression lines (Fig. 5), however, did not disclose trajectories directly following the fibers of GGA. The strains did not distinguish the local depression from the general lowering of the upper surface, perhaps because the direction of action of Verticalis and GGA are virtually identical in the midline upper tongue. In the lower tongue, some support for GGA involvement was seen. Backward compression, due to HG contraction, was less pronounced in the anterior two-thirds of the lower midline tongue than in the left and right slices. GGA activity may have caused this attenuation. Measurement of additional tasks and subjects, and increased tMRI resolution, will create a database of internal tongue strains that can be interpolated and interpreted with more confidence.

B. Does the tongue surface reflect internal tag deformation patterns?

The surface and tag movements of Fig. 3 were compared. The tags tracked movement of specific tissue points at the surface and within the tongue. The upper surface of the tongue moved downward and the posterior surface moved backward. The internal tag movements of the upper tongue

were consistent with the movement of the upper surface and the tag movements of the middle tongue with the back surface. With a larger quantity of data and consistent associations, surface and internal relationships might be compiled in a standard reference table. Such a table would be useful in mapping normal muscle-to-surface relationships, and as baseline data for comparison with articulatory disorders of various origins.

Two characteristics of the internal tissue point motion supplement the knowledge gained from 2D models. The first characteristic was left-to-right asymmetry in the tag trajectories (Fig. 3). On the left, the upper tongue moved in a fairly linear path downward and backward. On the right, it moved downward and then backward. Because the data did not include the coronal and axial planes, asymmetries like this were difficult to interpret. Some asymmetries are found in normal human anatomy. Additionally, functional asymmetries could result if a muscle on one side of the tongue contracted more forcefully, or a passive pull on the tissue modified its path asymmetrically. The second characteristic of note was that the surface and tag motions were not always simultaneous. For example, in Fig. 3 the movement of the left and right tongue surface occurred primarily between time-phases 2 and 3. The upper tags, however, moved quite a bit between time-phases 3 and 4 (black). This movement may have been a passive response to the rapid surface lowering at phase 3 or a reflection of activity in other planes. True 3D volumetric data are needed to get a fully accurate representation of the internal-to-surface relationships.

V. CONCLUSIONS

This preliminary study supports the continued use of tagged Cine-MRI as a method of examining muscle activity between phonemes during speech. Data such as these can supplement EMG data, and the procedure is noninvasive. From tags measured at the intersection of horizontal and vertical tag lines, the deformation of tissue-points within the tongue were measured with a resolution of about 0.95 mm. Principal strains detailed regions of compression and extension in the tongue and were interpreted in relation to internal tongue muscle activity. Two muscles, from which it is difficult (Superior Longitudinal) or impossible (Verticalis) to acquire accurate EMG data, appeared to contract in /k/-to-/ɑ/ tongue lowering. Some evidence for GGA contraction was seen in the anterior tongue consistent with surface lowering and depression formation.

This local homogeneous model for deriving principal strains from tMRI tongue images appears promising for future examination of the tongue musculature. It enabled us to examine stretch independently from translation, to infer contraction of muscles that are difficult to access with EMG, and to hypothesize new roles for muscles from the data. We expect predictive models and additional data to improve the accuracy of these inferences.

Since this data set was collected on a single subject, and contains simplified speech material, these results cannot be generalized. Rather, they are presented to illustrate the valuable information that can be derived from a principal strain model.

ACKNOWLEDGMENTS

The authors would like to thank Michael Cohen, Rafael Laboissiere, Heni Yehia, Anders Lofqvist, and one anonymous reviewer for their helpful comments on this paper. This project was supported in part by Grant No. DC025681 from the National Institute of Deafness and Other Disorders, and Northrop Grumman, Inc.

Abd-El Malek, S. (**1939**). ''Observations on the morphology of the human tongue.'' J. Anatomy **73**, 201–210.

Axel, L., and Dougherty, L. (**1989**). ''MR imaging of motion with spatial modulation of magnetization.'' Radiology **171**, 841–849.

Axel, L., and Dougherty, L. (**1989**). ''Heart wall motion: Improved method of spatial modulation of magnetization for MR imaging.'' Radiology **172**, 349–350.

Baer, T., Alfonso, P., and Honda, K. (**1988**). ''Electromyography of the tongue muscles during vowels in /əpVp/ environment,'' Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, U. of Tokyo, Vol. 22, pp. 7–19.

Carpentier, P., and Pajoni, D. (**1989**). ''The tongue: A complex muscular array,'' Rev. Orthop. Dento. Faciale **23**, 19–28 (in French).

Cook, R., Malkus, D., and Plesha, M. (**1989**). *Concepts and Applications of Finite Element Analysis* (Wiley, New York).

Dang, J., and Honda, K. (**1997**). ''A physiological model of the tongue and jaw for simulating deformation in the midsagittal and parasagittal planes,'' J. Acoust. Soc. Am. **102**, 3167(A).

Davis, E. (**1999**). ''Measurement and Kinematic Modeling of the Human Tongue,'' Ph.D. Dissertation, Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD.

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

Dew, D., and Jensen, P. (**1977**). *Phonetic Processing* (Charles E. Merrill, Columbus, OH), pp. 92–93.

Doran, G. A. (**1975**). ''Review of the evolution and phylogeny of the mammalian tongue,'' Acta Anat. **91**, 118–129.

Douglas, A., Rodriguez, E., O'Dell, W., and Hunter, W. (**1991**). ''Unique strain history during ejection in canine left ventricles,'' Am. J. Physiol. **260**, H1596–H1611.

Faber, A., and Raphael, L. (**1989**). ''Relationship of recorded EMG signals to within- and cross-utterance acoustic variation,'' J. Acoust. Soc. Am. Suppl. 1 **85**, S58.

Fung, Y. C. (**1993**). *Biomechanics, Mechanical Properties of Living Tissues* (Springer Verlag, New York).

Honda, K., and Kusakawa, N. (**1997**). ''Compatibility between auditory and articulatory representations of vowels,'' Acta Oto-Laryngol., Suppl. **532**, 103–105.

Kier, W., and Smith, K. (**1985**). ''Tongues, tentacles and trunks: The biomechanics of movement in muscular-hydrostats,'' Zoological Journal of the Linnean Society **83**, 307–324.

Kumada, M., Niitsu, M., Niimi, S., and Hirose, H. (**1992**). ''A study on the inner structure of the tongue in the production of the 5 japanese vowels by tagging snapshot MRI,'' Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, U. of Tokyo, Vol. 26, pp. 1–13.

Lai, W., Rubin, D., and Krempl, E. (**1993**). *Introduction to Continuum Mechanics*, 3rd ed. (Pergamon, New York).

MacNeilage, P., and Sholes, G. (**1964**). ''An electromyographic study of the tongue during vowel production,'' J. Speech Hear. Res. **7**, 209–232.

Maeda, S., and Honda, K. (**1994**). ''From EMG to formant patterns of vowels: The implications of vowel spaces,'' Phonetica **51**, 17–29.

Masaki, S., Tiede, M., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., and Ninomiya, N. (**1999**). ''MRI based speech production study using a synchronized sampling method,'' J. Acoust. Soc. Jpn. (E) **20**, 375–379.

McVeigh, E. (**1996**). ''MRI of myocardial function: Motion tracking techniques,'' Magn. Reson. Imaging **14**, 137–150.

McVeigh, E., and Atalar, E. (**1992**). ''Cardiac tagging with breath-hold Cine-MRI,'' Magn. Reson. Imaging **28**, 318–327.

Miyawaki, K. (**1974**). ''A study on the musculature of the human tongue,'' Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, U. of Tokyo, Vol. 8, pp. 23–49.
Miyawaki, K. (**1975**). ''A preliminary report on the electromyographic study of the activity of lingual muscles,'' Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, U. of Tokyo, Vol. 9, pp. 91–106.
Napadow, V. J., Chen, Q., Wedeen, V. J., and Gilbert, R. J. (**1999a**). ''Biomechanical basis for lingual muscular deformation during swallowing.'' Am. J. Physiol. **277**, G695–G701.
Napadow, V. J., Chen, Q., Wedeen, V. J., and Gilbert, R. J. (**1999b**). ''Intramural mechanics of the human tongue in association with physiological deformations,'' J. Biomech. **32**, 1–12.
Niimi, S., Kumada, M., and Niitsu, M. (**1994**). ''Functions of tongue-related muscles during production of the five Japanese vowels,'' Annual Bulletin of the Research Institute Logopedics and Phoniatrics, U. of Tokyo, Vol. 28, pp. 33–40.
Niitsu, M., Kumada, M., Niimi, S., and Itai, Y. (**1992**). ''Tongue movement during phonation: A rapid quantitative visualization using tagging snapshot MRI imaging,'' Annual Bulletin of the Research Institute Logopedics and Phoniatrics, U. of Tokyo, Vol. 26, pp. 149–156.
Perkell, J. (**1974**). ''A physiologically oriented model of tongue activity in speech production,'' Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Perlman, A., Luschei, E., and Dumond, C. (**1989**). ''Electrical activity from the superior pharyngeal constrictor during reflexive and nonreflexive tasks,'' J. Speech Hear. Res. **32**, 749–754.
Sauerland, E. K., and Mitchell, P. (**1975**). ''Electromyographic activity of intrinsic and extrinsic muscles of the human tongue,'' Texas Reports on Biology and Medicine **33**, 445–455.
Shawker, T., Sonies, B., and Stone, M. (**1984**). ''Sonography of speech and swallowing,'' in *Ultrasound Annual* (Raven, New York), pp. 237–260.
Smith, K. K., and Kier, W. M. (**1989**). ''Trunks, tongues, and tentacles: Moving with skeletons of muscle,'' Am. Sci. **77**, 29–35.
Stone, M., and Lundberg, A. (**1996**). ''Three-dimensional tongue surface shapes of English consonants and vowels,'' J. Acoust. Soc. Am. **99**, 3728–3737.
Takemoto, H. (**2001**). ''Morphological analyses of the human tongue musculature for three-dimensional modeling,'' J. Speech Hear. Res. (in press).
Thexton, A. J., Crompton, A. W., and German, R. Z. (**1998**). ''Transition from suckling to drinking at weaning: a kinematic and electromyographic study in miniature pigs,'' J. Exp. Zool. **280**, 327–343.
Wilhelms-Tricarico, R. (**1995**). ''Physiological modeling of speech production: Methods for modeling soft-tissue articulators,'' J. Acoust. Soc. Am. **97**, 3085–3098.