Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing[a)]

Vijay Parthasarathy and Jerry L. Prince
Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

Maureen Stone[b)] and Emi Z. Murano
Department of Biomedical Sciences and Orthodontics, University of Maryland Dental School, Baltimore, Maryland 21201

Moriel NessAiver
Department of Radiology, University of Maryland Medical School, Baltimore, Maryland 21201

(Received 4 March 2005; revised 14 August 2006; accepted 20 September 2006)

A cine series of tagged magnetic resonance (MR) images of the tongue is used to measure tongue motion and its internal deformation during speech. Tagged images are collected in three slice orientations (sagittal, coronal, and axial) during repetitions of the utterance "disouk" (*/disuk/*). A new technique called harmonic phase MRI (HARP-MRI) is used to process the tagged MR images to measure the internal deformation of the tongue. The measurements include displacement and velocity of tissue points, principal strains, and strain in the line-of-action of specific muscles. These measurements are not restricted to tag intersections, but can be calculated at every pixel in the image. The different motion measurements complement each other in understanding the tongue kinematics and in hypothesizing the internal muscle activity of the tongue. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2363926]

PACS number(s): 43.70.Jt, 43.72.Ar [AL] Pages: 491–504

I. INTRODUCTION

Understanding the dynamics of tongue motion is important because the tongue plays a crucial role in speech, swallowing, and breathing. Measuring tongue motion is a challenging problem for a variety of reasons: the tongue's location deep within the vocal tract, high degrees of freedom during motion, and its rapidity of motion during speech and swallowing. A number of measurement techniques have been used to measure tongue motion during continuous speech. Imaging techniques measure the entire tongue surface contour noninvasively (cinefluoroscopy, ultrasound, and cine-MRI), but lack temporal resolution [Stone (1999)]. Point-tracking techniques have better temporal resolution (electromagnetic midsagital articulometer EMMA, electromagnetic articulograph EMA, x-ray microbeam), but sparsely represent the tongue surface.

Electromyography (EMG) is the theoretical ideal for interpreting muscle activity. Bipolar fine-wire electrodes have been used as the method of choice for measuring lingual muscle activity during speech and breathing [Eastwood *et al.* (2003); Honda (1996); Honda and Kusakawa (1997); Miyawaki *et al.* (1997); Sabisky *et al.* (2006); Sauerland and Harper (1976)]. EMG signals, however, are difficult to obtain and interpret, and the relation between action and function of the muscles becomes less clear as task complexity increases. In breathing, genioglossus muscle (GG) is basically considered a dilator muscle. In a slightly more complex movement such as vowel production, GG is subdivided into at least two portions, anterior and posterior, based on its fanlike fiber distribution. In the most complex motions, such as consonants, EMG signals are almost uninterpretable. In other words, the linkage between muscle activity and tongue behavior becomes increasingly ill-defined. Therefore, a method that considers fiber direction, tongue contour, and physiological data during voluntary movements is invaluable and in urgent need.

Measurement techniques that use magnetic resonance imaging (MRI) are proving to be important tools in the analysis of tongue behavior because of MRI's ability to measure not only the tongue's surface but its interior as well. The use of MRI in speech research began with the recording of steady-state vowels. Initial studies imaged one static image plane [cf. Baer *et al.* (1987)]. The applications of static MRI of the oral cavity have included anatomical studies [Lufkin *et al.* (1987); McKenna *et al.* (1990)], examination of cancers and neurological diseases [Cha and Patten (1988); Takashima *et al.* (1989)], language studies [Hoole *et al.* (2001, 2000)], and static vocal tract modeling [Engwall (1999)]. Static MRI has been compared to other tongue measurement techniques such as electropalatography, ultrasound, electromyography, and electromagnetic articulography [Engwall (2003a); Wein *et al.* (1990)]. Most commonly, static MRI of the oral cavity is used to measure vocal-tract boundaries, catalog vocal-tract shapes, and extract airway volumes

[a)]A portion of this work was presented in the proceedings of 15th International Congress of Phonetic Sciences in a paper titled, "Tracking tongue motion from tagged magnetic resonance images using Harmonic Phase imaging (HARP-MRI)."

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

for use in acoustic tube models [cf. Badin *et al.* (2002); Greenwood *et al.* (1992); Lakshminarayanan *et al.* (1991); Moore *et al.* (1992); Story *et al.* (1996, 1998); Sulter *et al.* (1992)].

Static MRI has limited capabilities in that it cannot be used to observe the changes in the tongue's position during speech. Hence, cine-MRI—acquisition of a time series of MR images—is used to observe the motion of the tongue. MR data are collected over multiple repetitions of a speech utterance, and an ensemble combination of the data produces a cine series of images. Multiple repetitions are necessary so that rapid tongue motions can be imaged with adequate spatial and temporal resolution [cf. Masaki *et al.* (1999)]. The images are collected in a specified orientation, usually sagittal, coronal, or axial [cf. Shadle *et al.* (1999)]. Multiple image orientations can be acquired with more repetitions [Stone *et al.* (2001b)]. A cine series of images can also be reconstructed using only a single repetition, but at the cost of reduced temporal and spatial resolution of the images. For example, using a single repetition per slice, Demolin *et al.* (2000) collected images at 4 Hz; Mady *et al.* (2002) studied glossectomized speech at 8 Hz; and Engwall (2003b) examined words at 9 Hz. Recently, Narayanan *et al.* (2004) have applied spiral MRI acquisition to do cine imaging collecting short sentences at 24 Hz with a single repetition.

A basic problem in speech research is to understand the relation between the dynamic forces of the muscles in the tongue's body and the kinematics of the tongue's surface. While cine-MRI has proven to be useful to study kinematics of the tongue surface, it does not give rich enough information about deformation of muscles in the tongue's body. The drawback of cine-MRI to provide information about the muscle deformation has been overcome through the use of MR tagging, which is a complementary method to EMG. MR tagging temporarily marks or "tags" a plane of tissue within the tongue. As the planes of tagged tissue move, the internal deformations can be observed. During early tagging experiments, reference tags were applied prior to a movement and deformed tags were captured either during or after the movement. Niitsu *et al.* (1994) separately captured vertical and horizontal tags of the head, first scanning the rest position followed by a vowel position held for 2 s. Niitsu's work first demonstrated the ability to visualize internal tongue deformation. Napadow *et al.* (1999a, b) extracted principal strains in the tongue's interior during nonspeech motions (protrusion and swallowing) by repeatedly measuring the rest and deformed positions. By progressively changing the time position of the deformed frame, several time frames were collected over multiple repetitions, and the data were combined into a pseudosequence. The tags could be hand-traced in the reference and deformed images, and subsequently used to hypothesize active muscle activity [Niitsu *et al.* (1992)].

Formation of a pseudosequence of tagged images and the subsequent hand-tracing of the tags is a manually intensive way of visualizing the progression of the internal deformation of the tongue. Naturally, the combination of cine-MRI with tagged MRI to create tagged-cine-MRI (tMRI) was the next substantial improvement in the measurement of the tongue's internal deformation. In tMRI, tags are applied prior to movement of the tissue, and cine-MRI is then collected throughout the entire motion. Tagged cine-MRI is a critical step forward in research on tongue behavior because it enhances the use of MRI in three ways. First, the tags can be tracked throughout the entire period of the speech utterance. Therefore, subtle deformations of the tongue can be visualized and quantified temporally. Second, while the existing tissue point tracking methods (x-ray microbeam, and electromagnetic articulography) track only the oral surface of the tongue, tMRI tracks the tongue surface completely from the tip to the root. Third, since the entire body of the tongue can be tracked, local tissue compressions can be calculated in the interior of the tongue. These local tissue compressions can be used to infer muscle activity, which can either be used in representational models of tongue control [Stone *et al.* (2000)] or as inputs to predictive models that map muscle contraction patterns to tongue surface deformations [Dang and Honda (1997); Iny *et al.* (2001)]. Local tissue compressions can also be used to test physiologically based models of tongue motion such as finite element models (FEM). Three-dimensional FEMs have been used to predict tongue behavior from muscle activity or vice versa [Dang and Honda (2004); Hashimoto and Suga (1986); Kakita *et al.* (1985); Kiritani *et al.* (1976); Levine *et al.* (2005); Wilhelms-Tricarico (1995)]. Local compressions from tMRI can also provide information about muscle activity that can serve as the bases for initial predictions for such models.

In order to quantify motion using tMRI images, tags need to be tracked and then combined to form motion estimates. A variety of image processing techniques have been proposed to estimate motion and deformation from tMRI images [Denney and McVeigh (1997); Guttman *et al.* (1994); Kerwin *et al.* (2000); Kerwin and Prince (2002); Ozturk and McVeigh (2000); Radeva *et al.* (1997); Young *et al.* (1995)]. These techniques include several semiautomatic tools to extract the tags from the tagged images and then combine the tag displacements into 2D or 3D motion measurements. All these methods were first developed to measure the motion of the heart from tagged MR images. In fact, harmonic phase MRI (HARP-MRI) was also first introduced as a fast automatic method to compute the deformation of the heart [Osman *et al.* (2000); Osman and Prince (2000)]. Even though there are considerable differences in the motion of the heart and the tongue, some of these methods could be adapted to measure tongue motion. Dick *et al.* (2000) adapted a 4D B-spline model to measure 3D strains and motion tracks in the tongue.

In the same vein, this work adapts HARP-MRI to measure tongue motion. In comparison to the existing methods of tagged image processing, HARP-MRI is faster and works with the least human intervention. Using the unique properties of HARP-MRI, we propose a fast MR imaging scheme that reduces the number of repetitions required by a factor of 2. We also develop algorithms to calculate motion measures that are unique to the tongue. These motion measures are subdivided into two categories: user-driven measures and data-driven measures. User-driven measurements include tracking of handpicked tissue points and calculating the

stretch between two hypothesized muscle endpoints. Data-driven measures, on the other hand, are calculated on all tissue points. These include velocities and principal strains. Data-driven measurements are objective measurements of internal tissue behavior and can be used to corroborate the user-driven measurements. They can also be used to develop and validate kinematic and dynamic models of the tongue.

II. THEORY

A. MR Imaging methods

MR tagging is traditionally done using the SPAMM (spatial modulation of magnetization) imaging protocol [Axel and Dougherty (1989a, b)]. A popular category of SPAMM tagging, which comes standard in many MR scanners, is called higher-order SPAMM (for example 1–4–6–4–1). This type of SPAMM tagging produces thin sharp black lines on the image. Even though these sharp tags are excellent for visualization, they have been found to be unsuitable for HARP-MRI processing [Kuijer *et al.* (2001)]. HARP-MRI performs better on smoothly varying tagging patterns produced by a technique called CSPAMM (complementary spatial modulation of magnetization) [Fischer *et al.* (1993)]. CSPAMM produces smoothly varying sinusoidal tags rather than sharp tags. Each CSPAMM tagged image is a combination of two tagged images, which are obtained separately using 1–1 SPAMM tagging. The 1–1 SPAMM tagging can be understood as the multiplication of the magnetization of the anatomy with a two-dimensional sinusoid. The two tagging sinusoids in a CSPAMM pair differ by a phase shift of π radians, which can be seen by comparing the two images in a vertically tagged CSPAMM pair in Figs. 1(a) and 1(b). Figures 1(d) and 1(e) show a horizontally tagged CSPAMM pair. Each of these images are acquired using a 1–1 SPAMM protocol. The combination of the images in each pair results in an improved tagged image that has twice the tag contrast than the individual 1–1 SPAMM images [Fischer *et al.* (1993)] and a tag persistence up to 1200 milliseconds.

The traditional way of combining CSPAMM images requires a complex subtraction of the two images in each pair. This subtraction is problematic since the phases of these images are corrupted by magnetic field inhomogeneity. In this work, we have used a recent method called MICSR (magnitude image CSPAMM reconstruction) to combine the pair of CSPAMM images [NessAiver and Prince (2003a)]. The MICSR method bypasses the use of complex data sets by using magnitude images only. If $A(x,y)$ and $B(x,y)$ represent two images in a CSPAMM pair, the MICSR reconstruction formula is given as

$$I_{\mathrm{MICSR}}(x,y) = |A(x,y)|^2 - |B(x,y)|^2. \quad (1)$$

The MICSR-reconstructed CSPAMM images are shown in Figs. 1(c) and 1(f). The MICSR method produces tagged images with pure sinusoidal tags without the constant bias that is characteristic of 1–1 SPAMM tagged images. To visualize this, notice that the magnitude images in Figs. 1(a), 1(b), 1(d), and 1(e) have a black background with crests of the rectified sinusoid seen as white. The sinusoids are rectified because only the magnitude is displayed, whereas in MICSR images the background is zero (gray), with both positive (white) and negative (black) excursions of the sinusoid. Since the output of MICSR reconstruction is pure sinusoidal tags, the use of MICSR is ideal for HARP-MRI processing. Also, improved tag contrast and tag persistence of these MICSR images improve our ability to acquire and analyze longer speech samples.

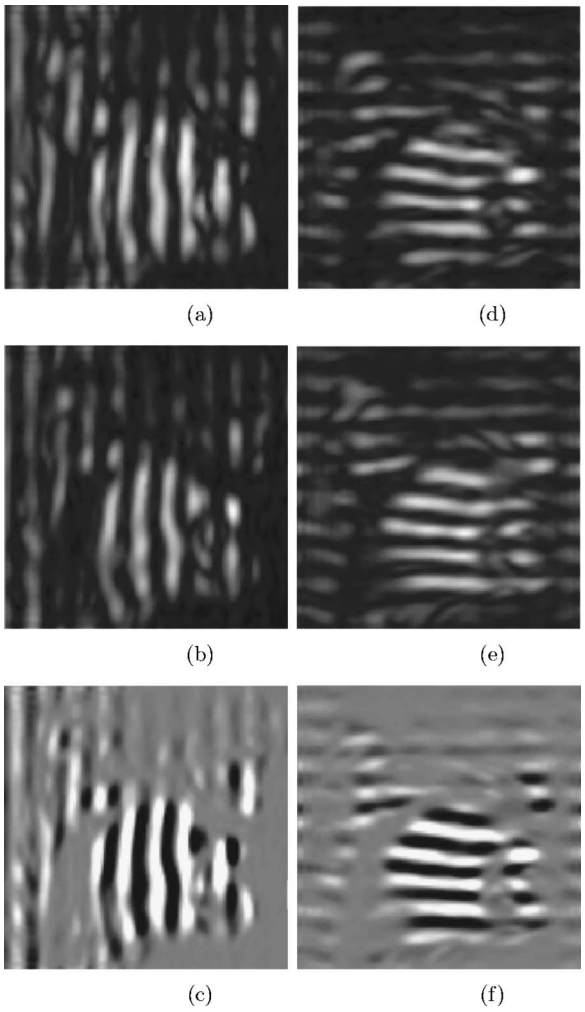


FIG. 1. Images (a) and (b) form a pair of vertically tagged CSPAMM images with phase shifted tags, and (c) is the MICSR combination of images (a) and (b). Images (d) and (e) form a pair of horizontally tagged CSPAMM images and (f) is the MICSR combination of images (d) and (e). Intensity of MICSR images (c) and (f) have been thresholded for better visualization of the tags.

B. Theory of HARP-MRI

Harmonic phase MRI (HARP-MRI) is used to estimate motion from sinusoidally tagged images [Osman *et al.* (1999)], an example of which is shown in Fig. 2(a). The tag lines, which were straight when initially applied, are now bent due to the local deformations in the tissue. The multiplication of the underlying magnetization of the anatomy with the tagging sinusoid produces two harmonic spectral peaks in the Fourier spectrum of the image [Fig. 2(b)]. The premise of HARP-MRI is that most of the spectral energy corresponding to the motion of the tissue is localized around

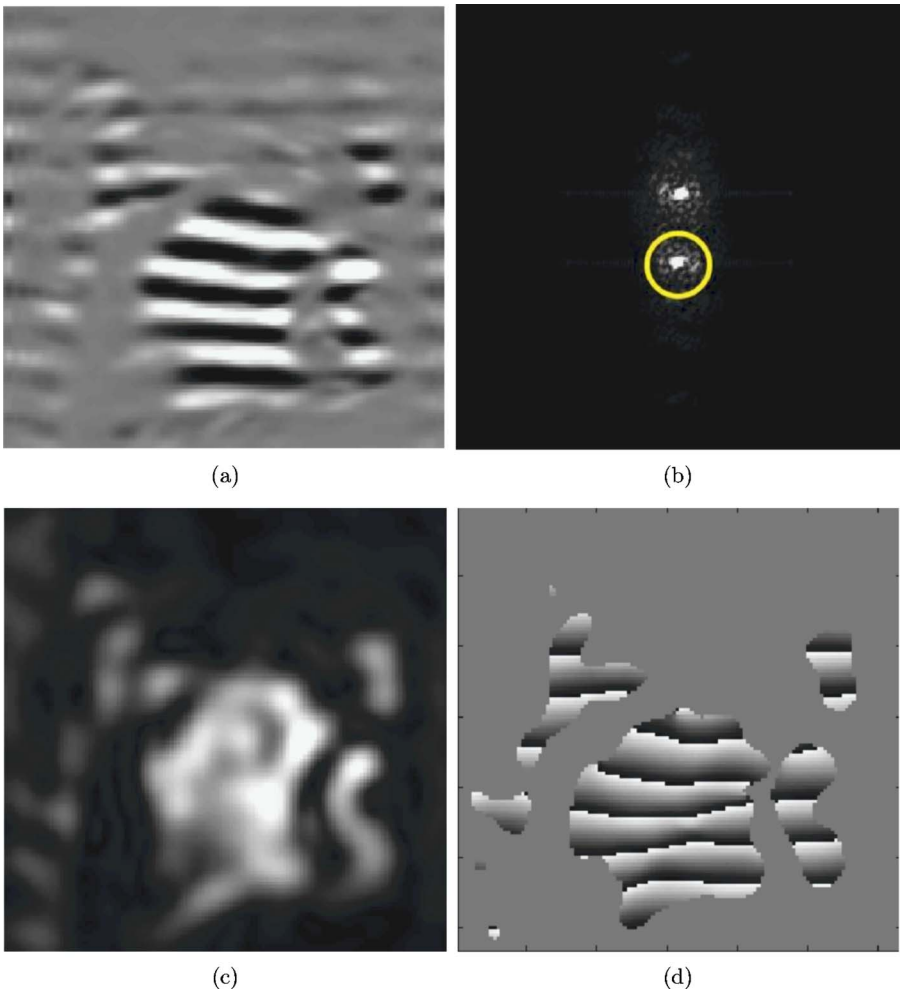


FIG. 2. (a) Horizontally tagged image produced by MICSR combination; (b) magnitude of its Fourier spectrum with the HARP bandpass filter; (c) harmonic magnitude image; and (d) harmonic phase image.

one of the harmonic spectral peaks. A bandpass filter [see the circle in Fig. 2(b)] is used to filter one of the harmonic peaks and the rest of the spectrum is zero padded. A *harmonic image* is constructed by calculating the inverse Fourier transform of the zero-padded spectrum. The resulting complex harmonic image can be expressed as

$$I(\mathbf{x},t) = D(\mathbf{x},t)e^{j\phi(\mathbf{x},t)}, \tag{2}$$

where $\mathbf{x}=[x_1\ x_2]^T$ is the position of the tissue point within an image and t is the time elapsed after the application of the tags [Osman *et al.* (2000)].

The *harmonic magnitude image* $D(\mathbf{x},t)$ [Fig. 2(c)] is proportional to the transverse magnetization at position $\mathbf{x}$ and time t. It represents an image of the anatomy, which is blurred by the bandpass filter. The basic concept of HARP-MRI is that the motion of the tissue is embedded in the *harmonic phase* (HARP) of the tagging sinusoid, which is measured as the phase $\phi(\mathbf{x},t)$ of the complex harmonic image. Figure 2(d) shows the image of the harmonic phase corresponding to the tagged image in Fig. 2(a). Since the phase of a complex image is constrained to lie between $-\pi$ and π, phase wraps occur in the HARP image. Note that these phase wraps exactly coincide with the troughs of the tag pattern, which indicates that the HARP phase has information about the motion of the tags. Mathematically, the harmonic phase $\phi(\mathbf{x},t)$ is linearly related to 2D tissue displacement $\mathbf{u}(\mathbf{x},t)$ as follows:

$$\phi(\mathbf{x},t) = -\boldsymbol{\omega}_0^{\mathrm{T}}\mathbf{u}(\mathbf{x},t) + \boldsymbol{\omega}_0^{\mathrm{T}}\mathbf{x}, \tag{3}$$

where $\boldsymbol{\omega}_0$ is the tagging vector, which corresponds to the orientation and frequency of the tag pattern in the image plane. In order to completely characterize displacement in two dimensions, we combine the harmonic phases from both the horizontal and vertical tags.

III. MATERIALS AND METHODS

A. Speaker and speech material

We demonstrate the application of HARP-MRI on one normal speaker. The speaker was a 27-year-old male, non-native speaker of Tamil accented English. The speaker did not have any dental fillings that might interfere with the MRI magnetic field. The speech material was the utterance */disuk/*, which was selected because it involves large movements of the tongue. The tongue makes an alveolar /d/, releases downward into /*is*/, moves backward for the /u/, upward into the velar stop /k/, and then returns forward for the next repetition.

B. Tagging and imaging parameters

Tagged MR images were collected on a Marconi 1.5 Tesla Eclipse scanner. The speaker laid supine in the MR scanner with the TMJ (temporomandibular joint) phased array coil positioned to image the area from the lower nasal

cavity to the upper trachea. Images in three different orientations were collected: sagittal, coronal, and axial. Eight sagittal slices, nine coronal slices, and nine axial slices were collected with a 5-mm slice thickness. On each slice, two sets of tagged images were obtained, one horizontal and the other vertical. Tagging is done using a 1–1 SPAMM tagging protocol with a tag spacing of 5 mm, which corresponds to a 10-mm period of the sinusoidal tag modulation. Imaging was done using a gradient echo imaging sequence with the following parameters: echo time of 2.2 milliseconds (ms), pulse repetition time of 4.13 ms, and a rectangular field of view of 200×200 mm. The data acquisition matrix was 64×22, zero filled to 128×128 sample points in Fourier space. Thus, the spatial resolution of the interpolated spatial voxel is $1.56 \times 1.56 \times 5$ mm. More details about the data acquisition matrix can be found in Appendix A.

C. Spatial resolution

The data acquisition matrix of size 64×22 in Fourier space corresponds to a tagged MR image with low spatial resolution (not tag resolution). This is unlike the other studies of the tongue that acquire MRI images with the primary objective of obtaining good spatial resolution. Therefore, it is natural to ask how does HARP-MRI work on such low resolution data. The key is to understand that HARP-MRI uses a bandpass filter to extract the complex harmonic image and needs only a small part of the Fourier spectrum to reconstruct motion. Therefore, it is sufficient to collect a small number of samples in Fourier space. Sampath *et al.* (2003) have validated this claim by showing that HARP-MRI estimates motion accurately even in cases of low resolution data. This unique property of HARP-MRI is critical in reducing the number of repetitions of the utterance required to reconstruct one tagged MR image.

D. Temporal resolution

The frame rate of the cine series is 18 Hz. Each cine series was collected over multiple repetitions, each repetition being 1 s long. The one second repetition time contained two parts: 667-ms recording time and 333-ms setup time for the next recording. With a frame rate of 18 Hz and 667-ms recording time, we collected 12 images for each cine series. In the rest of this paper we refer to each image in the cine tMRI series as one time frame.

E. Number of repetitions and repeatability

In order to use HARP-MRI on one 2D slice, we need four cine sequences of tagged images—two CSPAMM image sequences in two tag directions. In this research, we acquire each tagged cine series in four repetitions of the utterance. Therefore, we need 16 repetitions to track tongue motion on one slice. When compared to previous protocols that require 32 repetitions to acquire similar kind of data [Stone *et al.* (2001a)], this protocol represents a twofold improvement in the speed of data acquisition.

Since each tagged MR image is a combination of four repetitions, the speaker's ability to repeat the utterance precisely is critical to image quality. Variability across repetitions causes blurring of images and crossing of tags after combination. In order to ensure least variability across repetitions, each potential speaker is pretested and trained before the MRI experiment. Subjects are selected based on this pretest, which consists of repeating the proposed speech material (and others) to a metronome. The metronome is a combination of a repeating rhythmic tone [Shimada *et al.* (2002)] and an MR gradient sound that was recorded during a previous imaging session. This pretest is virtually identical acoustically to the experience in the scanner during tagged image acquisition. The speech material is repeated at 1-s intervals and the subject is asked to time his/her repetitions to this gradient sound cycle. The acoustic data are subsequently analyzed to measure the variability in repetition time. A variability threshold of 50 ms is set, above which the speaker is not used. The speaker used in this paper was very precise, with a standard deviation of 20 ms. Following the pretest, qualified subjects are trained by further synchronizing their utterance to the repeating rhythmic tone. The number of repetitions for the subject to get into rhythm is determined and later incorporated into the MR imaging protocol. The subject presented in this paper needed three repetitions to get into rhythm, and felt comfortable when the number of beats in the rhythmic tone was equal to the number of syllables in the utterance, which is two in this case.

F. Acquisition of cine-MRI and high-resolution static MRI

In addition to tagged MR images, untagged cine-MRI images were collected for the same speech task. The cine-MRI protocol requires eight repetitions per slice to reconstruct the cine image series. The data acquisition matrix is 66×64 (almost three times the spatial resolution of the tagged images) over a rectangular field of view of 200×200 mm, zero filled to 256×256 sample points in Fourier space. Thus, the spatial resolution of the interpolated spatial voxel is $0.75 \times 0.75 \times 5$ mm. Figures 3(a)–3(c) show images of one particular time frame of the coronal, sagittal, and axial slices, respectively. The tongue surface is seen better on the untagged images than on the tagged images because of its better spatial resolution. The untagged images are overlaid on the tagged images to visually correlate the changes in the tongue surface with the tongue's internal deformations. The untagged images are also used to create an articulation table (Table I) for */disuk/* by correlating the motion of the vocal-tract structures with the respective sounds. This table is later used to interpret the motion measurements from HARP-MRI.

Finally, in order to delineate the directions of specific muscles for this particular subject, a high-resolution static MRI is also acquired. An example of the high-resolution static MRI of a midsagittal slice is shown in Fig. 3(d). Anatomical landmarks delineated in the static MRI are used to outline the muscle directions in the tagged and untagged images. For example, in this study, the high-resolution sagittal image is used to find the origin and the insertions of the genioglossus muscle in order to find the strain along its line-of-action.

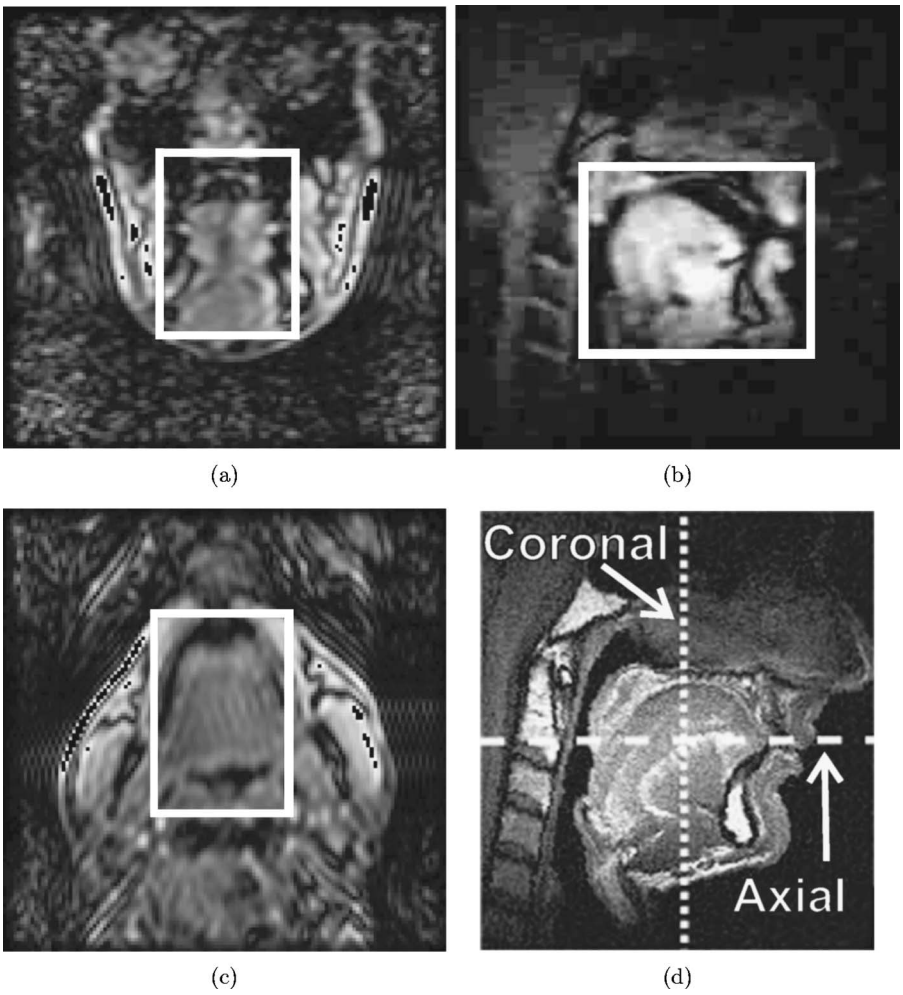


FIG. 3. Untagged image of one time frame of (a) coronal slice; (b) sagittal slice; (c) axial slice. White box shows region of interest around the tongue. (d) High-resolution static image of the anatomy to identify the muscle directions. White box shows a region of interest (ROI) around the tongue. The motion measures will be shown only for this ROI in order to avoid clutter of measurements in the surrounding tissues.

In this study, the motion is visualized on all three slice orientations. Of the eight sagittal slices, the results will be shown only on the midsagittal slice. The coronal and axial slices that are selected for visualization are marked in Fig. 3(d). While the selected coronal slice runs through the middle of the first upper molar, the axial slice runs parallel to the crown of the upper molars.

IV. RESULTS: VISUALIZATION AND DATA ANALYSIS

A. Checkerboard visualization

It is often helpful to visualize the local deformation of the tissue in addition to quantifying it. In order to better visualize the local tongue deformations, we introduce the checkerboard visualization [Fig. 4(a)], which is obtained by multiplying the horizontally and vertically tagged MICSR images and thresholding the resulting image [NessAiver and Prince (2003a), (2003b)]. Local elongation, compression, and shear of the tissue can be visualized as different shapes of the deformed squares. The images are upsampled using sine interpolation to increase tag crispness. The checkerboard display is purely a visualization tool and does not have any bearing either on the data analysis or on the resolution of HARP-MRI.

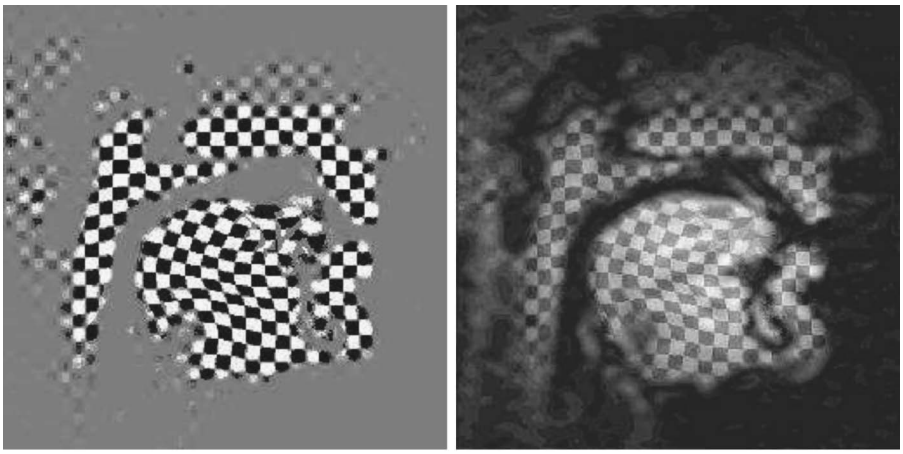
TABLE I. Articulation tabel for */disuk/* based on the shape of tongue and lips as seen in the untagged cine-MRI images.

Frames	Articulation
1–5	/*d*/
5–8	/*d*/−/*s*/
8–9	/*s*/−/u/
9–10	/u/−/k/
10–11	/k/
11–12	Preparation for next utterance

The checkerboard display in Fig. 4(a) shows the position of the tongue in the sixth time frame of the utterance */disuk/*. A considerable shearing and rotation can be seen in the body of the tongue. On the other hand, there is clear lack of deformation in the lips and the velum. The surface of the tongue is smeared due to reduced spatial resolution in the tagged images. In order to improve the visualization of the tongue surface, we overlay the checkerboard image onto the untagged image as shown in Fig. 4(b). This overlay visualization helps to better correlate the tongue's internal deformations with movements of the tongue surface. Note the marked improvement in the visualization of the tongue tip and tongue blade in Fig. 4(b) when compared to Fig. 4(a).

B. HARP-MRI analysis, visualization, and data interpretation

Equation (3) shows the linear relationship between the harmonic phase ϕ and tissue point displacement $\mathbf{u}$. From this



(a) (b)

FIG. 4. (a) MICSR trinary display. (b) Overlay of MICSR checkerboard and untagged image; note the better edge definition near the tongue tip in the overlay image.

estimate of displacement, the following motion measures can be calculated: (1) trajectories of tissue points; (2) velocities of tissue points; (3) principal strains; (4) strain between two points that are hypothesized to be the endpoints of a particular muscle. Even though two of these measures [(1) and (2)] have been reported in the context of the heart, in this research, we adapt the measures to the more complex movements of the tongue. In addition, our method to calculate principal strains and strains in the line-of-action of a muscle are unique to the analysis of the tongue.

The above quantities measured using HARP-MRI can be split into two categories: user-driven and data-driven. User-driven quantities require some manual intervention. The trajectories of tissue points and the line-of-action strains are user-driven quantities because they require the user to pick specific points in the tongue. On the other hand, the principal strains and the velocities are completely data driven and do not require any manual intervention. In the next few sections, we explain the calculation and interpretation of both kinds of measurements.

1. User-driven measurements

a. Trajectories of tissue points. The trajectory of a tissue point tracks its position at different times during a utterance. A tagged tissue point has two HARP values, one from the horizontal tags and another from the vertical tags. Let the horizontal HARP value be ϕ_H and the vertical HARP value be ϕ_V. Let the overall HARP vector be $\boldsymbol{\phi}=[\phi_H\ \phi_V]^T$. The HARP vector is a material property of the tissue. Therefore, as a tissue point moves, its HARP vector moves along with it. Hence, by tracking the HARP value through a cine series of images, the trajectory of a tissue point can be estimated [Osman *et al.* (1999)]. Consider a material point located at $\mathbf{x}_N$ at time t_N. If $\mathbf{x}_{N+1}$ is the position of this point at time t_{N+1}, then since HARP is a material property, we must have

$$\boldsymbol{\phi}(\mathbf{x}_{N+1},t_{N+1}) = \boldsymbol{\phi}(\mathbf{x}_N,t_N). \tag{4}$$

This relationship provides the basis for tracking $\mathbf{x}_N$ from time t_N to time t_{N+1}. Our goal is to find $\mathbf{x}$ that satisfies

$$\boldsymbol{\phi}(\mathbf{x},t_{N+1}) - \boldsymbol{\phi}(\mathbf{x}_N,t_N) = 0, \tag{5}$$

and then set $\mathbf{x}_{N+1}=\mathbf{x}$. This *HARP tracking* algorithm is repeated for consecutive time frames to track a particular point through the entire cine series and the trajectory is visualized as a path line.

b. Visualization. Local motion of the tongue is visualized by displaying trajectories of multiple handpicked tissue points (see Fig. 5). The small circle represents the positions of the tissue points in the first time frame, and the plus symbols represent the positions of the points in the last time frame. The trajectories are split into six (progressively darker) shades of gray. Each shade of gray corresponds to two of the 12 time frames. It is worth recalling that the locations of these tissue points need not be tag intersections. HARP tracking is capable of tracking arbitrary points on the image, even between pixels.

c. Interpretation. The timing and directions of the surface point trajectories corroborate the inhomogeneity of the

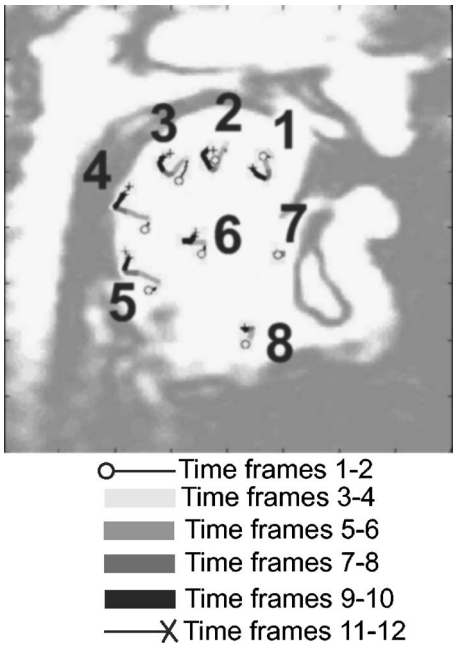
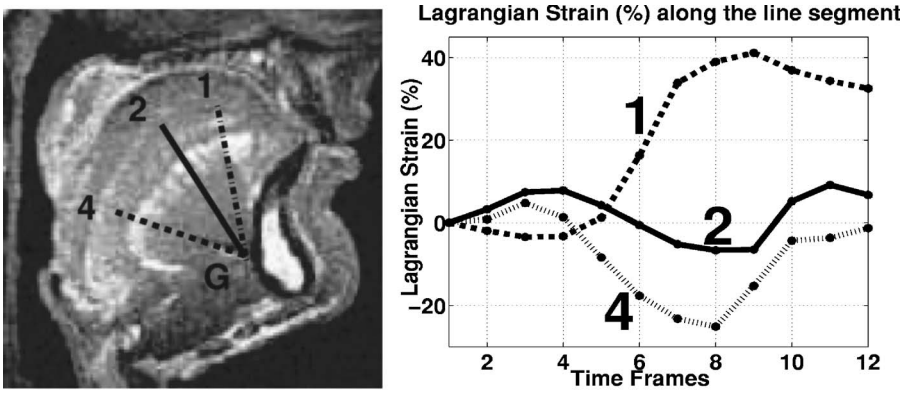


FIG. 5. Motion tracks for /*disuk*/ in a midsagittal slice. Black circles represent the position of the tissue point at time frame 1. The progression of shades from gray to black indicates the trajectory of the tissue point. In the electronic version, please zoom in more than 500% to see the trajectories clearly.



(a) (b)

FIG. 6. (a) Genioglossus muscle directions delineated on high-resolution static MRI. (b) Lagrangian strain in the three lines of action of genioglossus.

tongue surface movements (see Fig. 5). The most interesting differences occur in the middle of the motion. Point 1 first moves down and forward, and then begins to move back. Points 2 and 3 move down and back, and then begin to move back and up. Points 4 and 5 move back continuously, and then move slightly upward. These differences reflect at least three distinct regions of motion on the tongue surface. The internal points indicate further inhomogeneity. Point 6 moves straight back. Point 7, which is close to the origin of genioglossus, does not move at all, implying that the anterior lowering of point 1, 2, and 3 is due to verticalis, and not genioglossus. Point 8 shows an upward motion primarily, then a slight backward motion in the jaw muscles, despite the lack of jaw motion seen in point 7. This implies that jaw muscle contraction elevates the hyoid, and leads to thickening of the floor muscles, elevating this point and in turn the tongue.

The trajectories in Fig. 5 also reflect the coordination needed to execute the specific phonemes in the word. Note the differences in direction, extent, and timing of the trajectory motions. For points 4 and 5, the change from backward to upward motion occurs abruptly at time frame 9, when the tongue moves from /*u*/ to /*k*/. For point 1, the upward turn is slower (time frames 8–10) and is preceded by a lowering motion from the /*s*/ into the /*i*/. Points 2 and 3 turn upwards during time frames 8–10, ending with a forward slide during time frames 11 and 12 to get ready for the upcoming /*d*/ of the next repetition. Point 6 shows minimal upward/downward motion of the internal tongue.

d. Strain along the line-of-action of a muscle. The use of HARP tracking can be extended to calculate the change in length of a line segment by simply tracking the endpoints of the line segment. By choosing two endpoints on an image where we hypothesize a particular muscle to be, we can calculate the strain along the line-of-action of that muscle. Consider two points with reference positions $\mathbf{x}^1_{\text{ref}}$ and $\mathbf{x}^2_{\text{ref}}$ at time t_{ref}. At time t_N, we can track the positions of these points using HARP tracking as $\mathbf{x}^1_N$ and $\mathbf{x}^2_N$. The strain between these two points at time t_N with respect to the reference time can be calculated as

$$e = \frac{\|\mathbf{x}^2_N - \mathbf{x}^1_N\|}{\|\mathbf{x}^2_{\text{ref}} - \mathbf{x}^1_{\text{ref}}\|} - 1. \tag{6}$$

The strain is zero if the distance between the points remains unchanged, negative if there is shortening, and positive if there is lengthening. The reference time, t_{ref}, can be set at any time frame. In this work, all strain along the line-of-action are evaluated using the first time frame as the reference. Since this measure involves tracking of specified material points, the resulting strain is a Lagrangian measure.

e. Visualization. Figure 6(a) shows three lines hypothesized to be the lines of action of the genioglossus muscle. Figure 6(b) shows the strains calculated along these deforming lines. The origin and insertions of the genioglossus muscle were selected based on the high-resolution static image shown in Fig. 6(a). The origin of the genioglossus muscle (point *G*) was chosen as the midpoint of the flat inner surface of the mandibular bone in the midsagittal slice. Three insertion points were selected in anterior (point 1), middle (point 2), and posterior (point 4) compartments of genioglossus. The numbers correspond to the points that were tracked in the previous section. Note that the insertion of the genioglossus is picked a fraction below the tongue surface, even though the true insertion points are at the tongue surface. This is done because of the reduced spatial resolution at the tongue surface in the tagged image, which often leads to errors in tracking at the surface.

f. Interpretation. Since there is no movement at point 7 (as seen from Fig. 5), the Lagrangian strains are entirely due to the surface motion of tissue points 1, 2, and 4. As a result, these compressions and extensions reflect the subtle differences of adjacent tongue surface regions. Such subtle difference are not captured in existing surface measuring modalities, and are unique to HARP-MRI. The length of line 4 (dotted) increases from frame 1 to frame 3, and then starts to reduce by frames 4 and 5. These first five frames correspond to the utterance /*di*/. From frame 5 till frame 8, line 4 continues to compress. Frames 5 through 8 correspond to the start of /*s*/. The maximum compression takes place at frame 8, after which line 4 starts to expand until the last time frame. Line 2 (solid), although similar to line 4, has greater expansion and lesser compression than line 4. Moreover, the expansion into /*u*/ begins one time-frame later for line 2 and is in a more vertical direction, indicating substantial differences between the motion of these two surface regions. Line 1 (dashed) is quite different. Its length is relatively unchanged during the utterance /*d*/, but there is dramatic increase in length from frames 5 through 9, which corresponds

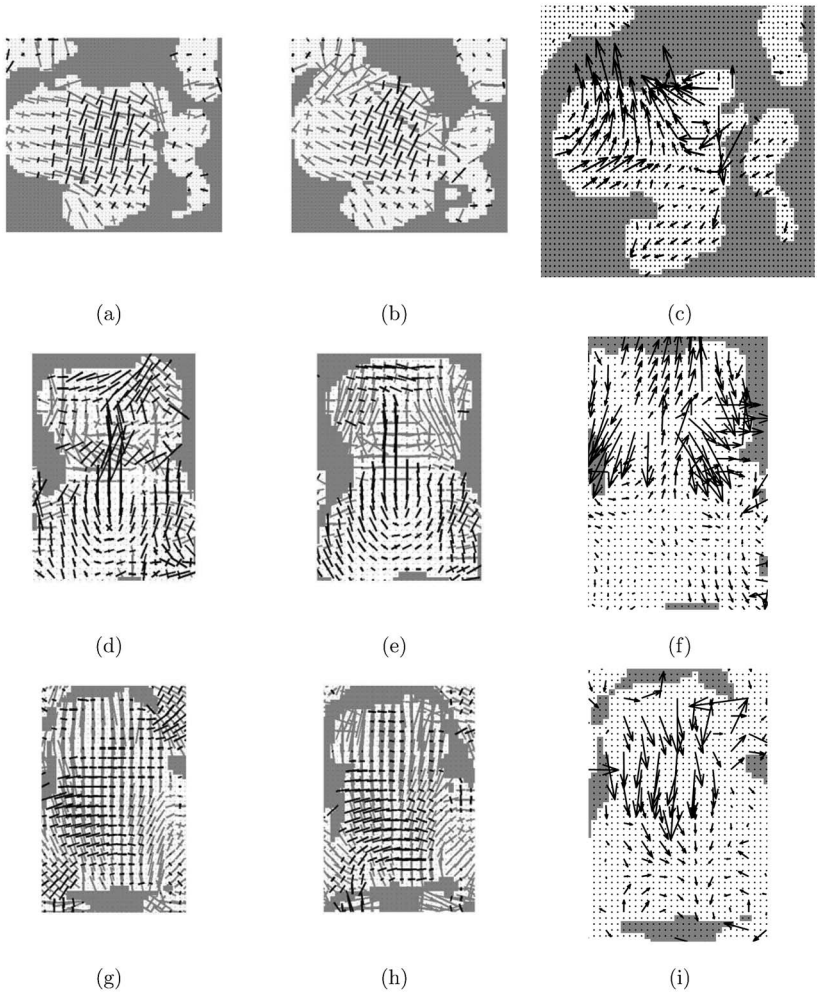


FIG. 7. The first two columns show the principal strains at time frames 9 and 10, respectively; black denotes compression and gray denotes expansion. The third column shows the velocity of tissue points calculated between time frames 9 and 10. While the first row shows the measurements for a midsagittal slice, the second and third rows show the measurements for an anterior coronal slice and a superior axial slice, respectively.

to /*su*/. This is consistent with the elevation of the entire tongue body, including the pharyngeal region. The maximum expansion occurs during /*u*/ at time frame 9, after which there is a decrease in length.

The user-defined measurements are useful to test hypotheses and to investigate relationships between the local surface and internal movements. For this data set the direction and rate of motion were consistent with three distinct surface regions containing independent spatiotemporal patterns of movement.

2. Data-driven measurements

Principal strains and tissue velocities are calculated from the data without any explicit modeling or user intervention. The data-driven measures help to corroborate the user-driven measures and to generate testable hypotheses.

a. Principal strains. Local stretch and compression in the interior of the tongue are objectively captured by the principal strains. The two principal directions are orthogonal and the strains along these directions capture the amount of maximal stretch or compression of a local tissue element. The strains are calculated on a regular grid on each pixel in the deformed tissue configuration. The principal strains can be derived from the Eulerian strain tensor E^* [Lai *et al.* (1996)], which is defined as follows:

$$E^* = \frac{1}{2}(I - B^{-1}), \tag{7}$$

where $B = FF^{\mathrm{T}}$ is the left Cauchy-Green strain tensor, $F = (\nabla \boldsymbol{\phi}(\mathbf{x},t))^{-1}\Omega^{\mathrm{T}}$ is the deformation gradient tensor, and $\Omega([\boldsymbol{\omega}_1 | \boldsymbol{\omega}_2]^{\mathrm{T}})$ is a matrix whose columns are two linearly independent tag direction vectors, namely horizontal and vertical. The matrix $\nabla \boldsymbol{\phi}(\mathbf{x},t)$ can be proven to be always invertible and hence E^* always exists. The eigenvectors of E^* represent the principal directions of strain. Using the definition of E^* [Lai *et al.* (1996)], we define strain in the principal direction as

$$S_i = 1 - \sqrt{\beta_i}, \quad i = [1,2], \tag{8}$$

where β_i is the ith eigenvalue of B^{-1}. Therefore, S_i represents the strains in the principal directions. More details about Eq. (8) can be found in Appendix B.

b. Visualization. The first and second columns in Fig. 7 show the principal strains corresponding to the motion from /u/ to /k/ and during /k/, respectively. The three rows correspond to the sagittal, coronal, and axial orientations. The principal strains are plotted in the form of a cross, with black denoting compression and gray denoting expansion. The length of the lines in the cross denotes the amount of strain, and the direction of the lines point to the directions of maxi-

mal strain. Although the strains are calculated at all pixels, strain is downsampled in the display to improve clarity.

c. Interpretation. In the midsagittal strains shown in the top row, there is a distinct difference in the local strain pattern in the anterior and posterior parts of the tongue. In the anterior and central parts, vertical compression is accompanied by either horizontal or oblique expansion. In contrast, in the posterior and the velar parts of the tongue there is expansion in both principal directions, though primarily horizontal in time frame 9. An interesting observation is that, even though the image is in the midsagittal plane, the direction of compression lines does not correspond to the direction of the genioglossus muscle. To better interpret the compressions an anterior coronal slice is considered, the position of which is shown in Fig. 3(d). The middle row of Fig. 7 shows the principal strain in the coronal slice. In both the coronal images, especially in time frame 10, one can see a midline vertical compression surrounded by a bilateral expansion. This pattern follows the anatomical distribution of the genioglossus muscle surrounded by the bilateral intrinsic muscles. This is consistent with genioglossus activation. There is a reduction in genioglossus compression between frames 9 and 10; the vertical lines are shorter and fewer in number. At the same time, expansion is seen laterally in the direction of verticalis. Both of these are consistent with the tongue elevation into /*k*/. The oblique lines below the tongue are consistent with the effects of mylohyoid contraction, which would further elevate the tongue into the /*k*/.

Figures 7(g) and 7(h) show the principal strains in time frames 9 and 10 for the axial slice, which is marked in Fig. 3(d). The lateral-to-medial compression in the axial images can be hypothesized to be because of the contraction of the transverse muscle, which can in turn explain the horizontal expansion in the sagittal plane and the bilateral vertical expansion in the coronal plane. Also, compression patterns around the pharynx resemble the activity of the styloglossus and the palatoglossus muscles. The combined visualization of the principal strains in the three orthogonal planes clearly corroborates the idea that the tongue is a muscular hydrostat, where a compression in one direction will result in an expansion in the another direction.

d. Velocity fields. The velocity of a tissue point at time t can be calculated from the HARP vector as follows:

$$\mathbf{v}(\mathbf{x},t) = -\left[\nabla \boldsymbol{\phi}(\mathbf{x},t)\right]^{-1} \frac{\partial \boldsymbol{\phi}}{\partial t}. \tag{9}$$

The calculation of velocity involves both the spatial gradient of phase within an image, and also the temporal gradient of phase across two time frames. In both cases, the forward difference operator is used.

e. Visualization. Figures 7(c), 7(f), and 7(i) show the velocity vectors between time frame 9 and 10, in the sagittal slice, coronal slice, and axial slice, respectively. These are the same slices that were considered for the strain computations. The length and direction of the arrow heads denote the magnitude and the direction of the velocity. Like the display of the principal strains, the vectors are displayed only on every other pixel for the sake of clarity.

f. Interpretation. The principal strains and velocity fields are complementary to each other. The principal strains indicate expansion and compression without indicating direction of motion, since the stretch may occur in either or both directions. On the other hand, velocity fields indicate direction and extent of motion, but do not indicate whether the tissue point motion was due to translation, expansion, or both.

Between time frames 9 and 10, the tongue was moving from /*u*/ to /*k*/. Examination of the sagittal velocity field in Fig. 7(c) shows the most salient features to be large upward motion of the upper tongue surface. The coronal image in Fig. 7(f) indicates that the upward motion occurs medially and that the sides of the tongue move laterally and downward. These patterns create an explicit map of the local tongue behaviors that comprise the surface elevation. For example, the motion from /*u*/ to /*k*/ requires tongue elevation and contact of the anterior-dorsum with the velum. There are a number of ways this could be executed. The tongue is a volume-preserving muscular hydrostat with interdigitated three-dimensional muscle architecture that allows complex deformations. Therefore, although the tongue can move rigidly by using the extrinsic muscles, local motion is another feasible option. Nonrigid local motion is an advantageous way to accomplish the constriction requirements of the target sound, while incorporating coarticulatory constraints on noncritical regions. The present data set indicates that the /*u*/ to /*k*/ movements actively incorporate features of nonrigid motion. First, the sagittal image shows that the tongue elevation is not uniform in direction between /*u*/ and /*k*/. Each region moves optimally to elevate the anterior-dorsum, which results in different directions of motion throughout the tongue. The posterior tongue actually, moves inward, an excellent volume-preserving mechanism to further facilitate anterior-dorsal elevation.

In the coronal slice, the upward motion is revealed to be primarily in the medial regions of the tongue, presumably because the palate is arched in the center and lower laterally, thus requiring less upward motion from the lateral margins of the tongue. Further, Fig. 7(c) shows that as the tongue contacts the palate, the tongue edges lower and spread out, suggesting recoil, or features of the next sound.

3. Complementary analysis between tissue motion and tissue compression

Comparison of the velocity fields with the principal strains reveals interesting relationships between translation and stretch of specific muscles. The two sagittal images [Figs. 7(a) and 7(b)] indicate homogeneous regions of strain in the anterior and posterior tongue from /*u*/ to /*k*/, with compression occurring only in the anterior tongue. The strain directions rotate from vertical/horizontal in direction [Fig. 7(a)] to oblique [Fig. 7(b)], and the upper surface of the tongue expands considerably between Figs. 7(a) and 7(b). These two regions translate into a velocity pattern [Fig. 7(c)] with at least three regions of surface motion (from front to back: backward, upward, and inward), as well as two horizontal divisions, the upper moving upward and the lower moving forward.

Frames 9 and 10 in the axial slice [Figs. 7(g)–7(i)] show patterns that are consistent with the sagittal data. Lateral-to-

medial shortening is consistent with both active contraction of transverse muscle and passive compression due to contact with the palate. The vector field in the axial plane [Fig. 7(i)] shows little in-plane motion in the posterior tongue between frames 9 and 10, consistent with motion orthogonal to the axial plane.

The coronal images add more information. The principal strains [Fig. 7(d) and 7(e)] show compression at the location of genioglossus (midline) and at the location of mylohyoid (bottom). Active contraction of genioglossus would oppose the upward motion seen in Fig. 7(f). The velocity field shows a reduction, or checking, in upward motion at the location of the GG compression. The tongue surface moves upward medially [Fig. 7(f)] and compresses left-to-right. This compression may be the passive response to the large lengthwise expansion of superior longitudinal required for the tongue surface to reach the palate. Compression seen in the mylohyoid region beneath the tongue [Figs. 7(d) and 7(e)] could reflect active contraction; mylohyoid has been found to contract during production of /k/ [Hirose (1971)]. The lateral tongue lowering in the velocity field [Fig. 7(f)] maps to the regions of expansion in the lateral tongue [Fig. 7(e)].

V. DISCUSSION

A. Advantages of using HARP-MRI

There are several technical advantages of analyzing tagged images using HARP-MRI. First, the motion measurements are not restricted to the points where tag lines intersect. Motion can be measured at the same resolution as the image resolution at every pixel. Second, since HARP-MRI uses only a small part of the Fourier spectrum of the tagged image to reconstruct motion. The ability to reconstruct motion from such small amount of Fourier spectral data is key in reducing the number of repetitions and also to improve the frame rate of the cine series. Third, since all these different measures are obtained from the same modality in the same scanning session, registration between modalities becomes unnecessary. Finally, HARP-MRI requires minimal human intervention and is computationally very fast. Tracking of one tissue point can be done within a second, and the calculation of strains and velocities for one tagged image can be measured within one minute.

Apart from the technical advantages, HARP-MRI brings a new perspective in the understanding of workings of the tongue. The combined analysis of the different motion measurements is a powerful tool in understanding tongue dynamics, and we clearly see three possible uses. First, plausible muscle activity can be inferred for testing with models. Second, models of motor control can be tested against the strain and velocity patterns. Third, multiplanar corroboration of muscle behavior can be observed.

B. Temporal resolution

The time resolution of 18 Hz, which is used in this study, is low when compared to the rate of normal speech. Even though we had the potential to collect data at nearly twice the rate (30 Hz), the speaker's temporal imprecision resulted in poor image quality and faded tags. At higher rates of data collection, even the slightest speaker imprecision between repetitions tends to cause improper ensemble combination. By slowing down the frame rate, we were effectively averaging over larger time units, thereby reducing the variability.

Reduction in frame rate is only a temporary solution to the speaker imprecision problem. We believe that data collected at higher frame rates have numerous advantages. Therefore, in spite of its current drawback with regards to speaker variability, we believe tMRI should be pursued. These drawbacks could be approached from three different directions. First, an image processing approach where a spatiotemporal registration of the four tagged data sets could be used. Second, an MR imaging approach could be used, where either the imaging sequences are designed to collect tagging data in one repetition (for example, grid tags with spiral acquisition) or the data from separate repetitions could be combined using retrospective gating methods. Third, an MR imaging approach where prospective gating could be done using the onset of speech.

C. Selection of speech materials

Temporal resolution is a key factor in deciding the speech materials that can be used for this study. HARP tracking makes a small motion assumption, which means that HARP tracking will be unable to track very large interframe motion. In order for HARP-MRI to track motion of a tissue point, the motion should be less than one half the tag separation between two consecutive frames. Therefore the temporal resolution, which defines the time interval between two consecutive frames, should be adequate for the particular utterance. Our experience has shown that tracking error is also likely to occur at the surface of the tongue because these points are not between two tags.

D. Future work

In this work, we have used 16 repetitions to fully track the motion of the tongue. The number of repetitions, however, needs to be further reduced in order to extend the scope of this study to nonexpert speakers and patients. In order to do this, research into the theoretical limits of HARP is ongoing [Parthasarathy and Prince (2004)]. Understanding of these issues will help predict whether longer and more complex speech patterns can be measured using HARP-MRI.

Measuring the motion and deformation is only the first part in understanding tongue dynamics. In this work, tissue compressions were used to hypothesize muscle activity. We plan to use these compressions in representational models of tongue control. It is important, however, to understand the assumptions made in moving from a pure kinematic measurement to a dynamic model. The observed compressions were assumed to be contractions and the effects of muscle contractions were also assumed to be direct. We acknowledge that there are many possible strategies of motor equivalence. Future work on representational and predictive models should include multiple synergistic patterns to achieve the compressions and surface shapes seen in the data.

VI. CONCLUSION

In this paper, we have presented a new method to measure the motion of the tongue during speech. We have demonstrated the use of HARP-MRI on tagged MRI images. HARP-MRI is an automatic tag tracking technique that estimates motion and deformation at every pixel in a tagged image. The motion measurements include trajectories of tissue points, principal strains at every pixel, strains in the line-of-action of a particular muscle, and tissue velocities. The different motion measurements can be visualized in a variety of ways and are complementary to each other. These measures were calculated on a 3D data set with images in the sagittal, axial, and coronal planes. Even though our data were collected in 3D, the measurements are still on 2D planes. We are actively pursuing 3D imaging and image processing approaches to measure 3D motion using tagging. The final goal is to be able to input these 3D motion measures into representational models in order to better understand the overall muscular compressional patterns in the production of speech.

ACKNOWLEDGMENTS

This research was funded in part by National Institute on Deafness and Other Communication Disorders (NIDCD) Grant R01 DC01758. We also acknowledge Rao Gullapalli, Jiachen Zhuo, and Harsh Agarwal for their contribution in collecting the tagged MR data.

APPENDIX A: DATA ACQUISITION MATRIX

The field of view is square with 200 mm along the tagging direction and 200 mm in the perpendicular direction. The period of each tag cycle is 10 mm. Therefore, there are 20 (200/10) tag cycles across the field of view. This implies that in the tagging direction the harmonic peaks will be centered at ±20 in Fourier space. If it is assumed that HARP processing needs only ±10 Fourier space samples on either side of the harmonic peak, then all we need to sample is 60 points, from −30 to +30. We sample −32 to +31, a total of 64 points (a power of 2) to make the inverse fast Fourier transform (FFT) algorithm efficient. Essentially, we divide Fourier space into three parts along the tagging direction. While the central third is centered around the center of Fourier space, the negative and positive thirds are centered around the negative and positive harmonic peaks respectively. Along the direction perpendicular to the tags, 23 (−11 to 11) Fourier points are sampled.

APPENDIX B: STRAIN IN PRINCIPAL DIRECTION

In strict continum mechanical sense, the term principal strains are not defined for finite deformation. In this paper, we use the term principal strain to mean strains in the directions that have no shear, i.e., principal directions. The terms "principal strain" and "strain in the principal directions" are used interchangeably.

Let F be the deformation gradient tensor. Using polar decomposition, $F=VR$, where V is a symmetric tensor called the right stretch and R is an orthogonal tensor that characterizes the rotational component. The left Cauchy-Green strain tensor B is defined as $B=V^2$.

Let λ_i, $i=[1,2]$ be the ith eigenvalue of V. Then, the eigenvalues of B^{-1} are

$$\beta_i = \frac{1}{\lambda_i^2}. \tag{B1}$$

The Eulerian strain tensor, E^* is defined as, $E^*=\frac{1}{2}(I-B^{-1})$. Therefore, the eigenvalues of E^* are given as $\frac{1}{2}(1-\beta_i)$, $i=[1,2]$. Since B is symmetric, E^* is also symmetric and hence can be diagonalized when E^* is expressed along its eigenvectors e_1 and e_2 as its basis. Given the definition of the diagonal terms in the E^* [Lai *et al.* (1996)], we get

$$\frac{\|ds_i\|^2 - \|dS_i\|^2}{\|ds_i\|^2} = (1-\beta_i), \tag{B2}$$

where dS_i is an infinitesimally small piece of tissue in the ith direction when the tags were applied and ds_i is the same tissue at a later time after deformation. From the above relation, the unit Eulerian strain along the principal direction, S_i, can be derived as

$$S_i = \frac{\|ds_i\| - \|dS_i\|}{\|ds_i\|} = 1 - \sqrt{\beta_i}, \quad i=[1,2]. \tag{B3}$$

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

Axel, L., and Dougherty, L. (**1989b**). "MR imaging of motion with spatial modulation of magnetization," Radiology **171**, 841–845.

Badin, P., Bailly, G., Reveret, L., Baciu, M., and Segebarth, C. (**2002**). "Three-dimensional linear articulatory modeling of tongue lips and face, based on MRI and video images," J. Phonetics **30**(3), 533–554.

Baer, T., Gore, J., Boyce, S., and Nye, P. (**1987**). "Application of MRI to the analysis of speech production," Magn. Reson. Imaging **5**(5), 1–7.

Cha, H., and Patten, B. M. (**1988**). "Amyotrophic lateral sclerosis abnormalities of the tongue on magnetic resonance imaging," Am. Neurological Association **25**, 468–472.

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).

Dang, J., and Honda, K. (**2004**). "Construction and control of a physiological articulatory model," J. Acoust. Soc. Am. **115**, 853–870.

Demolin, D., Metens, T., and Soquet, A. (**2000**). "Real time MRI and articulatory coordinations in vowels," in Proceedings of International Speech Production Seminar.

Denney, T. S. Jr., and McVeigh, E. R. (**1997**). "Model-free reconstruction of three-dimensional myocardial strain from planar tagged MR images," J. Magn. Reson Imaging **7**, 799–810.

Dick, D., Ozturk, C., Douglas, A., McVeigh, E., and Stone, M. (**2000**). "Three-dimensional tracking of tongue motion using tagged-MRI," in International Society for Magnetic Resonance in Medicine, 8th Scientific Meeting and Exhibition, Denver, Co.

Eastwood, P. R., Allison, G. T., Shepherd, K. L., Szollosi, I., and Hillman, D. R. (**2003**). "Heterogenous activity of the human genioglossus muscle assessed by multiple bipolar fine-wire electrodes," J. Appl. Physiol. **94**, 1849–1858.

Engwall, O. (**1999**). "Modeling of the vocal tract in three dimensions," in Proceedings of the Eurospeech Meeting, pp. 113–116.

Engwall, O. (**2003a**). "Combining MRI, EMA and EPG measurements in a three-dimensional tongue model," Speech Commun. **41**(2–3), 303–329.

Engwall, O. (**2003b**). "A revisit to the application of MRI to the analysis of speech production—testing our assumptions," in 6th International Seminar on Speech Production.

Fischer, S. E., McKinnon, G. C., Maier, S. E., and Boesiger, P. (**1993**).

"Improved myocardial tagging contrast," Magn. Reson. Med. **30**, 191–200.

Greenwood, A. R., Goodyear, C. C., and Martin, P. A. (**1992**). "Measurements of vocal tract shapes using magnetic resonance imaging," in IEE Proceedings, pp. 5553–5560.

Guttman, M. A., Prince, J. L., and McVeigh, E. R. (**1994**). "Tag and contour detection in tagged MR images of the left ventricle," IEEE Trans. Med. Imaging **13**(1), 74–88.

Hashimoto, K., and Suga, S. (**1986**). "Estimation of the muscular tensions of the human tongue by using a three-dimensional model of the tongue," J. Acoust. Soc. Jpn. **7**, 39–46.

Hirose, H. (**1971**). "Electromyography of the articulatory muscles: Current instrumentation and technique," Haskins Lab Status Rep Speech Res. **25–26**, 73–86.

Honda, K. (**1996**). "Organization of tongue articulation for vowels," J. Phonetics **24**, 39–52.

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

Hoole, P., Geng, C., and Winkler, R. (**2001**). "Towards a speaker-independent representation of tongue-posturing for speech," in 4th International Speech Motor Conference, edited by B. Maassen, W. Hulstijn, R. Kent, H. Peters, and P. Van Lieshout, pp. 138–141.

Hoole, P., Wismuller, A., Leinsinger, G., Kroos, C., Geumann, A., and Inoue, M. (**2000**). "Analysis of tongue configuration in multi-speaker, multi-volume MRI data," in Proceedings of 5th Speech Production Seminar, pp. 157–160.

Iny, D., Lundberg, A., Levine, W. S., and Stone, M. (**2001**). "The dynamics and control of the tongue during speech," in Proceedings of the American Control Conference, pp. 246–251.

Kakita, Y., Fujimura, O., and Honda, K. (**1985**). "Computation of mapping from the muscular contraction pattern to formant in vowel space," in *Phonetic Linguistics; Essays in honor of Peter Ladefoged*, edited by V. Fromkin (Academic Press, NY).

Kerwin, W. S., and Prince, J. L. (**2002**). "Kriging filter for space-time interpolation," Adv. Imaging Electron Phys. **124**, 139–193.

Kerwin, W. S., Osman, N. F., and Prince, J. L. (**2000**). "Image processing and analysis in tagged cardiac MRI," in *Handbook of Medical Image Processing and Analysis*, edited by I. N. Bankman (Academic Press), pp. 375–392.

Kiritani, S., Miyawaki, K., Fujimura, O., and Miller, J. (**1976**). "A computational model of the tongue," Ann. Bull. RILP Univ. Tokyo **10**, 243–251; URL⟨http://www.umin.ac.jp/memorial/rilp-tokyo/ ⟩. Last accessed 12/6/06.

Kuijer, J. P., Jansen, E., Marcus, J. T., Rossum, A. C. V., and Heethaar, R. M. (**2001**). "Improved harmonic phase myocardial strain maps," Magn. Reson. Med. **46**(5), 993–999.

Lai, W. M., Rubin, D., and Krempl, E. (**1996**). *Introduction to Continuum Mechanics* (Butterworth-Heinemann, London).

Lakshminarayanan, A. V., Lee, S., and McCutcheon, M. J. (**1991**). "MR Imaging of the vocal tract during Vowel Production," J. Magn. Reson Imaging **1**, 71–76.

Levine, W. S., Torcaso, C. E., and Stone, M. (**2005**). "Controlling the shape of a muscular hydrostat: A tongue or tentacle," *New Directions in Control Theory and Applications*, edited by W. P. Dayawansa, A. Lindquist, and Y. Zhou, Lecture Notes in Control and Information Sciences, Vol. 321 (Springer Publishing Co.), pp. 207–222.

Lufkin, R., Christianson, R., and Hanafee, W. (**1987**). "Normal magnetic resonance imaging anatomy of the tongue, oropharynx, hypopharnyx and larnyx," Dysphagia **1**, 119–l27.

Mady, K., Sader, R., Zimmermann, A., Hoole, P., Beer, A., Zeilhofer, H., and Hanning, C. (**2002**). "Assesment of consonant articulation in glossectomee speech by dynamic MRI," in Proceedings of ICSLP 2002.

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. **20**, 375–379.

McKenna, K. M., Jabour, B. A., Luftkin, R., and Hanafee, W. (**1990**). "Magnetic resonance imaging of the tongue and oropharynx," Top Magn. Reson Imaging **2**, 49–59.

Miyawaki, K., Hirose, H., Ushijima, T., and Sawashima, M. (**1997**). "A preliminary report on the electromyographic study of the activity of lingual muscles," Ann. Bull. RILP Univ. Tokyo **9**, 91–1066; URL⟨http://www.umin.ac.jp/memorial/rilp-tokyo/ ⟩. Last accessed 12/6/06.

Moore, C. C., O'Dell, W. G., McVeigh, E. R., and Zerhouni, E. A. (**1992**). "Calculation of three-dimensional left ventricular strains from biplanar tagged MR images," J. Magn. Reson Imaging **2**, 165–175.

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. **322**, 1–12.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., and Byrd, D. (**2004**). "An approach to real-time magnetic resonance imaging for speech production," J. Acoust. Soc. Am. **115**, 1771–1776.

NessAiver, M., and J. L. Prince (**2003a**). "Magnitude image CSPAMM reconstruction (MICSR)," Magn. Reson. Med. **50**(2), 331–342.

NessAiver, M., and Prince, J. L. (**2003b**). "Visualization of myocardial motion using MICSR trinary checkerboard display," in Inf. Processing in Medical Imaging **18**, 573–585.

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 of Logopedics and Phoniatrics, University of Tokyo, **26**, 149–156; ⟨URLhttp://www.umin.ac.jp/memorial/rilp-tokyo ⟩. Last accessed 12/6/06.

Niitsu, M., Kumada, M., Campeau, G., Niimi, S., Riederer, S. J., and Itai, Y. (**1994**). "Tongue displacement: Visualization with rapid tagged magnetization-prepared MR imaging," Radiology **191**, 578–580.

Osman, N. F., and Prince, J. L. (**2000**). "Visualizing myocardial function using HARP MRI," Phys. Med. Biol. **45**(6), 1665–1682.

Osman, N. F., McVeigh, E. R., and Prince, J. L. (**2000**). "Imaging heart motion using harmonic phase MRI," IEEE Trans. Med. Imaging **19**(3), 186–202.

Osman, N. F., Kerwin, W. S., McVeigh, E. R., and Prince, J. L. (**1999**). "Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging," Magn. Reson. Med. **42**, 1048–1060.

Ozturk, C., and McVeigh, E. R. (**2000**). "Four-dimensional B-spline based motion analysis of tagged MR images: Introduction and *in vivo* validation," Phys. Med. Biol. **45**, 1683–1702.

Parthasarathy, V., and Prince, J. L. (**2004**). "Dynamic range of harmonic phase magnetic resonance imaging (HARP-MRI)," in Proceedings of the International Society of Biomedical Imaging, pp. 884–887.

Radeva, P., Amini, A., and Huang, J. (**1997**). "Deformable B-solids and implicit snakes for 3D localization and tracking of MRI-SPAMM data," Comput. Vis. Image Underst. **66**(2), 163–178.

Sabisky, J. P., Butler, J. E., Fogel, R. B., Taylor, J. L., Trinder, J. A., Davis, D. P., and Gandevia, S. C. (**2006**). "Tonic and phasic respiratory drives to human genioglossus motoneurons during breathing," J. Neurophysiol. **95**(4), 2213–2221.

Sampath, S., Derbyshire, A., Osman, N. F., Atalar, E., and Prince, J. L. (**2003**). "Real-time imaging of two-dimensional cardiac strain using a fast HARP pulse sequence," Magn. Reson. Med. **50**(1), 154–163.

Sauerland, E. K., and Harper, R. M. (**1976**). "The human tongue during sleep: Electromyographic activity of the genioglossus muscle," Exp. Neurol. **51**, 160–170.

Shadle, C., Mohammad, M., Carter, J., and Jackson, P. (**1999**). "Dynamic magnetic resonance imaging: New tools for speech research," in *proceedings of the 14th Int. Cong. Phon. Sci.*, pp. 623–626.

Shimada, Y., Fujimoto, I., Takemoto, H., Takano, S., Masaki, S., Honda, K., and Takeo, K. (**2002**). "4D-MRI using the synchronized sampling method," Jpn. J. Radiol. Technol. **53**(12), 1592–1598.

Stone, M. (**1999**). "Laboratory techniques in speech science," in *A Handbook of Phonetic Science*, edited by W. Hardcastle, A. Marchal, and N. Hewlett, Blackwell Handbooks in Linguistic Series (Blackwell, Ltd., Oxford), pp. 11–32.

Stone, M., Dick, D., Davis, E., Douglas, A., and Ozturk, C. (**2000**). "Modelling the internal tongue using principal strains," in *Proceedings of the 5th Speech Production Seminar, Kloster-Seeon*, pp. 133–136, Germany.

Stone, M., Davis, E. P., Douglas, A. S., NessAiver, M., Gullapalli, R., Levine, W. S., and Lundberg, A. (**2001a**). "Modeling the motion of the internal tongue from tagged cine-MR images," J. Acoust. Soc. Am. **109**(6), 2974–2982.

Stone, M., Davis, E. P., Douglas, A. S., NessAiver, M., Gullapalli, R., Levine, W., and Lundberg, A. J. (**2001b**). "Modeling tongue surface contours from cine-MRI images," J. Speech Lang. Hear. Res. **44**(5), 1026–1040.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **10**(3),

537–554.
Story, B. H., Titze, I. R., and Hoffman, E. A. (**1998**). "Vocal tract area function for an adult female speak based on volumetric imaging," J. Acoust. Soc. Am. **104**(1), 471–487.
Sulter, A. M., Miller, D. G., Wolf, R. F., Schutte, H. K., Wit, H. P., and Mooyaart, E. L. (**1992**). "On the relation between the dimensions and resonance characteristics of the vocal-tract—a study with MRI," Magn. Reson. Med. **10**(3), 365–373.
Takashima, S., Ikezoe, J., Harada, K., Akai, Y., Hamada, S., Arisawa, J., Morimoto, S., Masaki, N., Kozuka, T., and Maeda, H. (**1989**). "Tongue cancer: Correlation of imaging and sonography with pathology," American Journal of Neuroradiology **10**, 419–424.
Wein, B., Drobnitzky, M., and Klajman, S. (**1990**). "Magnetresonanztomographie and Sonographie bei der Lutbildung (Magnetic Resonance Imaging and Sonography During Phonation)," Fortschr Roentgensr **153**, 408–412.
Wilhelms-Tricarico, R. (**1995**). "Physiological modeling of speech production: Methods for modeling soft-tissue articulators," J. Acoust. Soc. Am. **97**, 3805–3898.
Young, A. A., Kraitchman, D. L., Dougherty, L., and Axel, L. (**1995**). "Tracking and finite element analysis of stripe deformation in magnetic resonance tagging," IEEE Trans. Med. Imaging **14**(3), 413–421.