JSLHR

Research Article

Analysis of 3-D Tongue Motion From Tagged and Cine Magnetic Resonance Images

Fangxu Xing,[a] Jonghye Woo,[b] Junghoon Lee,[a,c] Emi Z. Murano,[c] Maureen Stone,[d] and Jerry L. Prince[a]

Purpose: Measuring tongue deformation and internal muscle motion during speech has been a challenging task because the tongue deforms in 3 dimensions, contains interdigitated muscles, and is largely hidden within the vocal tract. In this article, a new method is proposed to analyze tagged and cine magnetic resonance images of the tongue during speech in order to estimate 3-dimensional tissue displacement and deformation over time.
Method: The method involves computing 2-dimensional motion components using a standard tag-processing method called harmonic phase, constructing superresolution tongue volumes using cine magnetic resonance images, segmenting the tongue region using a random-walker algorithm, and estimating 3-dimensional tongue motion using an incompressible deformation estimation algorithm.
Results: Evaluation of the method is presented with a control group and a group of people who had received a glossectomy carrying out a speech task. A 2-step principal-components analysis is then used to reveal the unique motion patterns of the subjects. Azimuth motion angles and motion on the mirrored hemi-tongues are analyzed.
Conclusion: Tests of the method with a various collection of subjects show its capability of capturing patient motion patterns and indicate its potential value in future speech studies.

The human tongue is highly deformable during speech and is able to perform precise motion over short periods of time because of its complex internal muscle architecture (Abd-el-Malek, 1955; Takemoto, 2001). Studying the motor control of the tongue muscles, including their interaction and cooperation for producing speech, has always been an interesting topic for oral surgeons, linguists, and speech-language pathologists. This is because the tongue, which is a volume-preserving structure devoid of bones and joints, moves entirely by deformation to create the shapes critical to speech, chewing, and swallowing (Maton, 1997). Therefore, developing a fast and accurate method for quantitatively analyzing tongue motion from medical imaging has always been an important topic to aid in speech studies.

The goal of the present research is to create and test a method to extract three-dimensional (3-D) tissue-point motion from within the tongue and use it to distinguish between glossectomy and control motion patterns in an interpretable way. Magnetic resonance imaging (MRI) is capable of revealing both anatomical structures and tissue motion. The internal motion of the tongue muscles has been captured by tagged MRI (Parthasarathy, Prince, Stone, Murano, & NessAiver, 2007). Through spatial modulation of the electromagnetic field, magnetic tags can be placed within the tissue and then deform with the tongue as it moves. Such deformed tag patterns contain motion information that is reconstructed by processing a sequence of tagged MRI acquired over time. Although tagged MRI captures motion well, it provides only low-resolution anatomical information at tissue–air boundaries. Therefore, in the present imaging protocol, an additional image sequence—a cine-MRI sequence without tags—is collected at the same positions and same time instances during additional repetitions of the motion cycle. These images can be used to segment the tongue region.

A few immediate difficulties follow the data acquisition. First, tagged MRI contains motion information only in two dimensions, and multiple orientations must be combined to produce 3-D motion. Second, manual segmentation of multiple cine-MRI slices in multiple time frames (a typical number of images is ~800 slices per subject) is a time-consuming task. Third, due to the limitations

[a]Johns Hopkins University, Baltimore, MD
[b]Harvard Medical School, Boston, MA
[c]Johns Hopkins University School of Medicine, Baltimore, MD
[d]University of Maryland School of Dentistry, Baltimore, MD
Correspondence to Fangxu Xing: fxing1@jhu.edu

Editor: Jody Kreiman
Associate Editor: Kate Bunton
Received June 6, 2014
Revision received November 29, 2014
Accepted December 2, 2015
DOI: 10.1044/2016_JSLHR-S-14-0155

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

in the acquisition process and the nature of tagged MRI, the tongue's motion can only be sampled at sparse spatial locations and relatively low resolution (Zerhouni, Parish, Rogers, Yang, & Shapiro, 1988). Therefore, for consistent scientific study and potential clinical use, there is a need for automated methods to quantify the tongue's motion in an efficient way. Moreover, after obtaining estimates of the 3-D tongue motion from multiple subjects in the same speech task, it remains challenging to analyze the resulting data in order to reveal population similarities and differences. The inconsistent speaking rate between subjects, for example, is a major obstacle. Therefore, a method is needed to provide a time alignment of the displacement data. The high dimensionality of the time sequences of 3-D displacements is another obstacle. Therefore, another method is needed to perform dimensionality reduction so that population analysis can be readily carried out.

Among various subjects, the motion of people who have received a glossectomy—a surgery to remove a malignant section of the tongue—is of special interest, because these people have altered tongue morphology and may combine unusual motor control strategies for speech (cf. Nicoletti et al., 2004). The National Cancer Institute estimates that 48,330 people in the United States develop oral or oropharyngeal cancer per year and that 9,570 of these people will die from it (American Cancer Society, 2016). Tumor size and location are the two most important factors that affect speech (Heller, Levy, & Sciubba, 1991; Nicoletti et al., 2004; Pauloski et al., 1998). The data set used in this article controls for tumor location by including only subjects with a unilateral tumor occurring just behind the tongue tip. People with both small and moderate-sized tumors, up to 4 cm long, are included, and the effects of tumor size on motion pattern are considered. Moreover, reduced control in the tip on the resected side may cause additional motion differences between subjects in the glossectomy group and the control group. Tongue-tip fricatives such as /s/ are more challenging for these people (Heller et al., 1991), so the speech task includes the sound /s/. In a word, understanding the motion differences made by people who have and have not received a glossectomy will assist surgical decisions as well as speech and swallowing remediation. Therefore, the present article aims to provide a new and useful tool for learning typical motor control and adaptive behaviors used by people who have received a glossectomy to compensate for morphological changes.

Previous methods have been proposed for the computation and analysis of 3-D motion from tagged MRI. However, most have been developed with the heart as the target organ, and therefore processing tongue data of this type requires novel techniques. There have been some well-established algorithms to handle some steps toward the final goal: the harmonic phase (HARP) algorithm (Osman, McVeigh, & Prince, 2000) for computing 2-D motion from tagged data; the incompressible deformation estimation algorithm (IDEA; Liu et al., 2012) for estimating 3-D motion from the HARP result; the superresolution methods (Woo, Murano, Stone, & Prince, 2012) for building 3-D tongue volumes; segmentation algorithms for providing automatic or semiautomatic labeling of the tongue region (Grady, 2006; Han, Xu, & Prince, 2003); and principal-components analysis for dimensionality reduction of populations of vector fields (Stone, Liu, Chen, & Prince, 2010). The proposed method, called TMAP for tongue motion analysis pipeline, incorporates all of these steps in a carefully optimized semiautomatic pipeline that also includes steps for time alignment and a population analysis. TMAP is evaluated with a data set from 16 people in a control group and five people who have received a glossectomy completing a specific speech task. The efficacy of TMAP is demonstrated and its potential for quantitative motion analysis of the tongue is revealed.

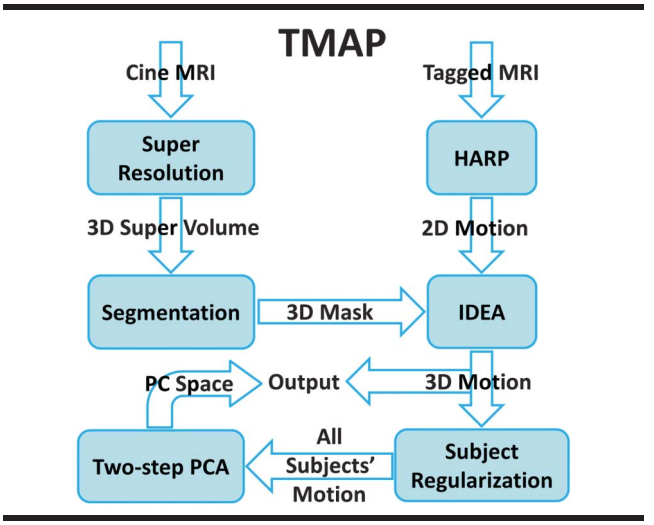
Method

As illustrated in Figure 1, the input to TMAP was tagged and cine MRI data. The output was the 3-D motion field and multisubject statistical-analysis result. The entire TMAP was implemented within in-house HARP 5 software with user interfaces that were based on MATLAB (MathWorks, Natick, MA). Each method in the pipeline is described in the following sections.

Subjects and Speech Material

Subjects in this study were five people who have received a glossectomy and 16 who have not. Four of the subjects in the glossectomy group had small (T1) tumors, and the fifth had a midsize (T2) tumor. The T1 tumors were removed with a glossectomy and the wound closed by sutures. The wound from the T2 tumor was closed by adding external tissue, a radial-forearm free flap, to replace the resected mass. The speech material was the

Figure 1. Flowchart of the tongue motion analysis pipeline (TMAP). MRI = magnetic resonance imaging; HARP = harmonic phase algorithm; IDEA = incompressible deformation estimation algorithm; PC = principal components; PCA = principal-components analysis.



phrase "a souk" (IPA: /ə'suk/), which was designed to elicit specific tongue motions. The phrase started with /ə/, a centralized tongue position; moved into /s/, where forward tongue motion was expected to be prominent; and ended with /k/, where upward tongue motion was expected to be prominent.

Because subjects needed to repeat the speech materials to a metronome in the MRI scanner, they were trained, prior to entering the scanner, to speak to the same metronome beat for about 15 min. The four-beat metronome sequence represented the two syllables of the speech task plus an inhalation and exhalation. Thus, every motion of the oral cavity was timed as precisely as possible.

Data Acquisition

MRI scanning was performed on a Siemens 3.0T Tim Trio system (Siemens Medical Solutions, Malvern, PA) with a 12-channel head coil and a four-channel neck coil using a segmented gradient-echo sequence. The field of view was 240 × 240 mm with an in-plane resolution of 1.87 × 1.87 mm and a slice thickness of 6 mm. Each data set contained a sagittal, coronal, and axial stack of images encompassing the tongue and surrounding structures. The image sequence was obtained at the rate of 26 time frames per second.

The MRI recording session lasted for about 1 hr. The MRI machine collected a very weak signal, namely the number of hydrogen protons in each unit of tissue. Therefore, multiple repetitions of a speech utterance needed to be collected and summed to generate a single time series showing tongue motion. To sum the cine MRI data, five repetitions were needed per slice. For the tagged MRI data, three repetitions were needed per slice. In addition, data for each tag direction (superior–inferior, anterior–posterior, left–right) were collected twice, once for a sinusoidal and once for a cosinusoidal tag pattern (see NessAiver & Prince, 2003). The number of slices depended on the size of the subject's tongue and ranged as follows: sagittal—five to nine slices; coronal—10 to 14 slices; axial—10 to 14 slices. Pauses were allowed after each set of slices so that consecutive acquisitions contained 15 to 42 repetitions for tagged acquisitions and 25 to 70 repetitions for cine acquisitions (Parthasarathy et al., 2007).

HARP Algorithm

Complex-valued tagged MRI (see Figure 2A) was combined using either the CSPAMM or MICSR (NessAiver & Prince, 2003) method, yielding images with two major harmonic peaks in the Fourier domain (see Figure 2B). The HARP algorithm (Osman et al., 2000) filtered one of the second-order harmonic peaks with a bandpass filter, took the phase part of the resulting complex image (see Figure 2C), and tracked each pixel's phase value over time by assuming that the phase of a fixed tissue point stayed constant. In practice, both horizontally and vertically tagged images were processed to obtain motion components in two in-plane directions so that the HARP algorithm provided a dense in-plane 2-D motion field (see Figure 2D). Due to phase wrapping (Liu & Prince, 2010; Osman et al., 2000), the HARP algorithm could fail by tracking a "jumped" tag when a tissue point made larger movements than commonly seen in the tongue. Thus the shortest-path HARP refinement (Liu & Prince, 2010) method was used to reduce this type of error. In the end, the HARP algorithm yielded a collection of 2-D vector-valued images, each representing the 2-D projection of the 3-D motion occurring from the current time frame to the initial time frame when the tags were applied.

Superresolution Tongue-Volume Reconstruction

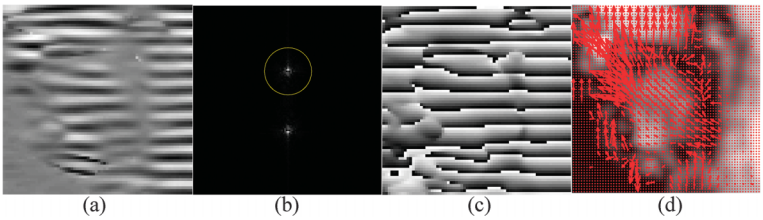
Cine MRI was used to provide anatomical information for tongue segmentation (see Figure 3). Because of the need to acquire data rapidly during speech and to maintain a high overall signal-to-noise ratio, cine images were acquired in the same positions with the same relatively large slice thickness as the tagged MRI data. Any single stack of cine MRI could not be used for high-resolution segmentation because of the poor through-plane resolution. Therefore, the entire collection (axial, sagittal, and coronal) was combined using superresolution methods into a single image on a 3-D grid whose voxel resolution was the same as the original 2-D in-plane resolution. To be specific, the SUPERV algorithm (Woo et al., 2012) was used to obtain one supervolume at each time frame.

Random-Walker Segmentation

In order to constrain the analysis to the tongue region only, a segmentation of the tongue was carried out using the supervolume images. Despite the improved quality of the supervolume over the cine images, it remained helpful to introduce manual guidance at this stage. The random-walker algorithm (Grady, 2006) was applied; it is a graph-based algorithm to find a global optimal probabilistic solution for multilabel image segmentation. In practice, a user specified (by drawing) a small number of pixels as seeds within predefined structures (labels), such as the tongue and the background. Each unlabeled pixel was then assigned to the label with the greatest probability in such a way that a random walker starting at this pixel would reach one of the seeds with this label.

In TMAP, a human user was required to input seeds on a few slices (six to nine) of the cine images at one time frame. The user-given seeds were then propagated by deformable registration (Vercauteren, Pennec, Perchant, & Ayache, 2009) to additional user-determined time frames at the same slice location (in this case, four time frames uniformly distributed over 26 time frames). For the remaining time frames, seeds were automatically generated by (a) segmenting a 3-D temporal stack using the random walker and (b) using the skeleton (Ronse, Najman, & Decencière, 2005) of the temporal segmentation as seeds (see Figure 4A). After seeds were found for all time frames at this

Figure 2. (A) Tagged sagittal tongue image. (B) Fourier domain of the tagged image. (C) Harmonic phase image from the filtered peak. (D) 2-D motion field from the harmonic phase algorithm displayed on a 3-times-sparser grid.

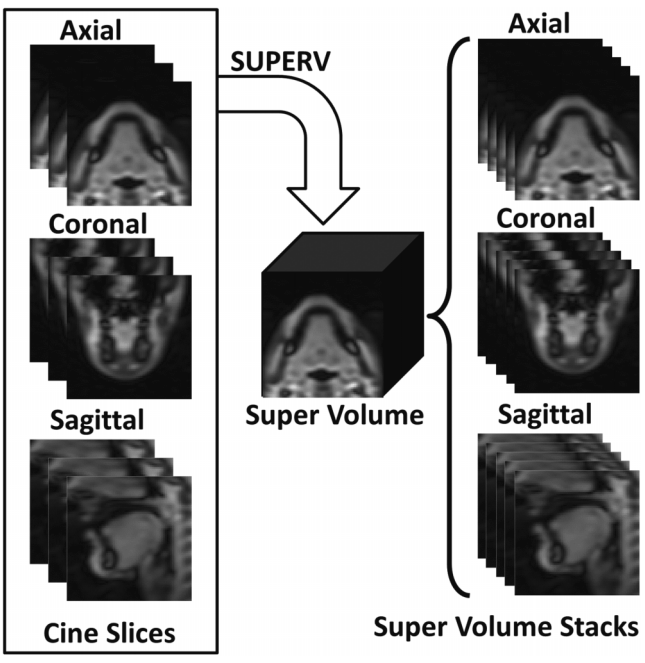


slice location, they were exported to the 3-D supervolume space and the random walker was computed, yielding the final segmentation (see Figure 4B). During this process, the user was allowed to validate and correct the propagated and automatically generated seeds. For more details of the segmentation process, we refer readers to Lee et al. (2014). As a final step, the 3-D tongue masks were used to cut the 2-D in-plane motion at the positions where they intersected the slice plane, leaving the 2-D motion only on the tongue region as input for the incompressible deformation estimation algorithm in the next step.

IDEA

The segmented 2-D motion slices from tagged data were viewed as multiple observations about the underlying 3-D motion (see Figures 5A and 5B). However, each observation was an in-plane 2-D projection of the true 3-D motion and was spatially sparse due to the low through-plane resolution. To reconstruct a dense 3-D motion estimate, interpolation was required. Because the tongue is an incompressible muscular hydrostat (Kier & Smith, 1985), the desired dense 3-D motion field should be volume preserving, and this provided a key constraint. The IDEA incorporated these sparse and incomplete projections as well as the incompressibility constraint by using a divergence-free vector spline (Liu et al., 2012). To be specific, it reconstructed a sequence of divergence-free velocity fields over small time steps so that the integrated velocities yielded the observed HARP displacements to good approximation. Because the IDEA was computationally demanding, the segmented 3-D tongue mask alleviated the problem by letting it compute only on the tongue region, yielding a desired 3-D motion field. Figures 5C and 5D show the estimated 3-D motion from Figures 5A and 5B, respectively. The color diagram at the left shows that anterior–posterior motion is green, superior–inferior motion is blue, left–right motion is red, and intermediate motion directions are RGB-valued combinations of these colors.

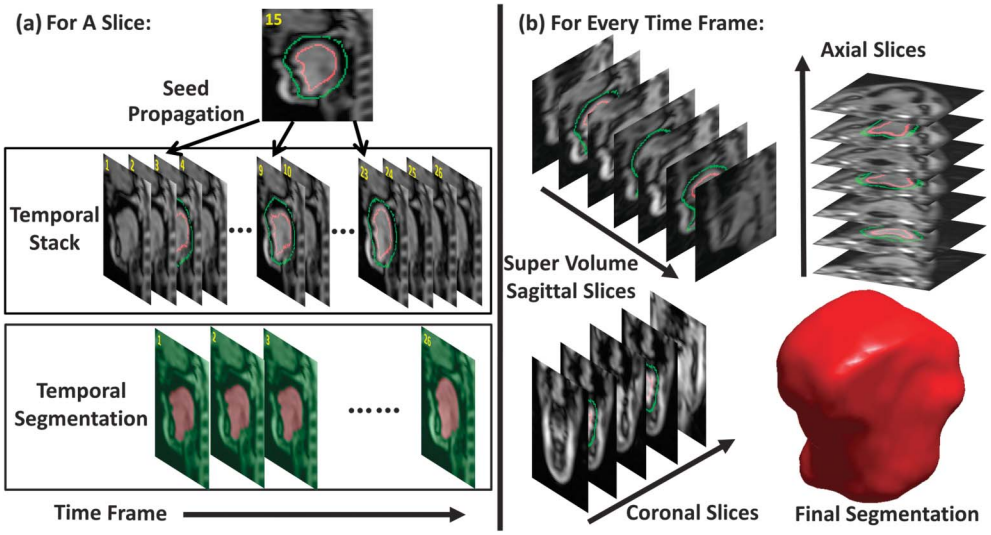
Figure 3. Cine supervolume reconstruction by the SUPERV algorithm.



Multisubject Data Normalization

Due to variable speaking rates among different subjects, displacement fields had to be computed relative to a common position before their motions from the results of IDEA could be compared. Because the first time frame of "a souk" was generally an unpredictable position of the tongue as it moved into /ə/, and because the deformation following /ə/ was a forward motion into /s/ and then an upward motion into /k/, the midcentral schwa /ə/ was used as the common reference frame to compare motion across subjects (Kent & Read, 2002). Therefore, the reference frame was switched from time frame 1 to the maximum /ə/ position. Two speech scientists examined the raw data independently to determine the position of the schwa, which was defined as the time frame prior to the beginning of the forward motion. Afterward, they consulted on the result and a consensus was established. Figure 6 (discussed later) shows the realignment of time frames, on the basis of the schwa, from the original speech for each subject (see Figure 6B) to the aligned time frames (see Figure 6C).

Figure 4. (A) Seed propagation in time and segmentation of temporal stack. (B) Segmentation of the supervolume.



The mathematical details for switching reference time frames are discussed in the following. For each subject, the sequence of 3-D vector fields obtained from the output of the IDEA is denoted $\{\boldsymbol{D}_{1,1}(\boldsymbol{X}_1), \boldsymbol{D}_{1,2}(\boldsymbol{X}_1), \ldots, \boldsymbol{D}_{1,26}(\boldsymbol{X}_1)\}$. Each vector field $\boldsymbol{D}_{1,t}(\boldsymbol{X}_1)$, as visualized in Figure 5, shows the displacement from time frame 1 (default reference frame) to the current time frame t. The symbol $\boldsymbol{X}_1$ represents the 3-D grid located at time frame 1 (a Lagrangian representation is used). If the vector field $\boldsymbol{D}_{1,t}(\boldsymbol{X}_1)$ is considered as arrows, they grow from the grid locations $\boldsymbol{X}_1$ at time frame 1 and end up pointing at the nongrid locations in the current time frame t.

Suppose the centralized prespeech position /ə/ happens at time frame r. At an arbitrary time frame t, related motion fields are $\boldsymbol{D}_{1,r}(\boldsymbol{X}_1)$ and $\boldsymbol{D}_{1,t}(\boldsymbol{X}_1)$. If the inverse field of $\boldsymbol{D}_{1,r}(\boldsymbol{X}_1)$—namely $\boldsymbol{D}_{r,1}(\boldsymbol{X}_r)$—can be found, their intermediate field can be composed by

$$\boldsymbol{D}_{r,t}(\boldsymbol{X}_r) = \boldsymbol{D}_{r,1}(\boldsymbol{X}_r) + \boldsymbol{D}_{1,t}\Big(\boldsymbol{X}_r + \boldsymbol{D}_{r,1}(\boldsymbol{X}_r)\Big) \qquad (1)$$

where $\boldsymbol{X}_r$ is now the grid on the new reference r.

Because the field $\boldsymbol{D}_{1,r}(\boldsymbol{X}_1)$ is discrete, then from the definition of an inverse field, at time frame r it is true that $\boldsymbol{D}_{r,1}(\boldsymbol{X}_1 + \boldsymbol{D}_{1,r}(\boldsymbol{X}_1)) = -\boldsymbol{D}_{1,t}(\boldsymbol{X}_1)$. To find the value of $\boldsymbol{D}_{r,1}$ at $\boldsymbol{X}_r$, a fixed-point method (Chen, Lu, Chen, Ruchala, & Olivera, 2008) was applied by iteratively solving the equation

$$\boldsymbol{D}_{r,1}{}^{(n)}(\boldsymbol{X}_r) = -\boldsymbol{D}_{1,r}\Big(\boldsymbol{X}_r + \boldsymbol{D}_{r,1}{}^{(n-1)}(\boldsymbol{X}_r)\Big) \qquad (2)$$

where n is the iteration. Thus, through substitution of the converged result of Equation 2 into Equation 1 and repetition for every time frame, a new sequence of displacement fields $\{\boldsymbol{D}_{r,1}(\boldsymbol{X}_r), \boldsymbol{D}_{r,2}(\boldsymbol{X}_r), \ldots, \boldsymbol{D}_{r,26}(\boldsymbol{X}_r)\}$ can be found for every subject starting at time frame /ə/.

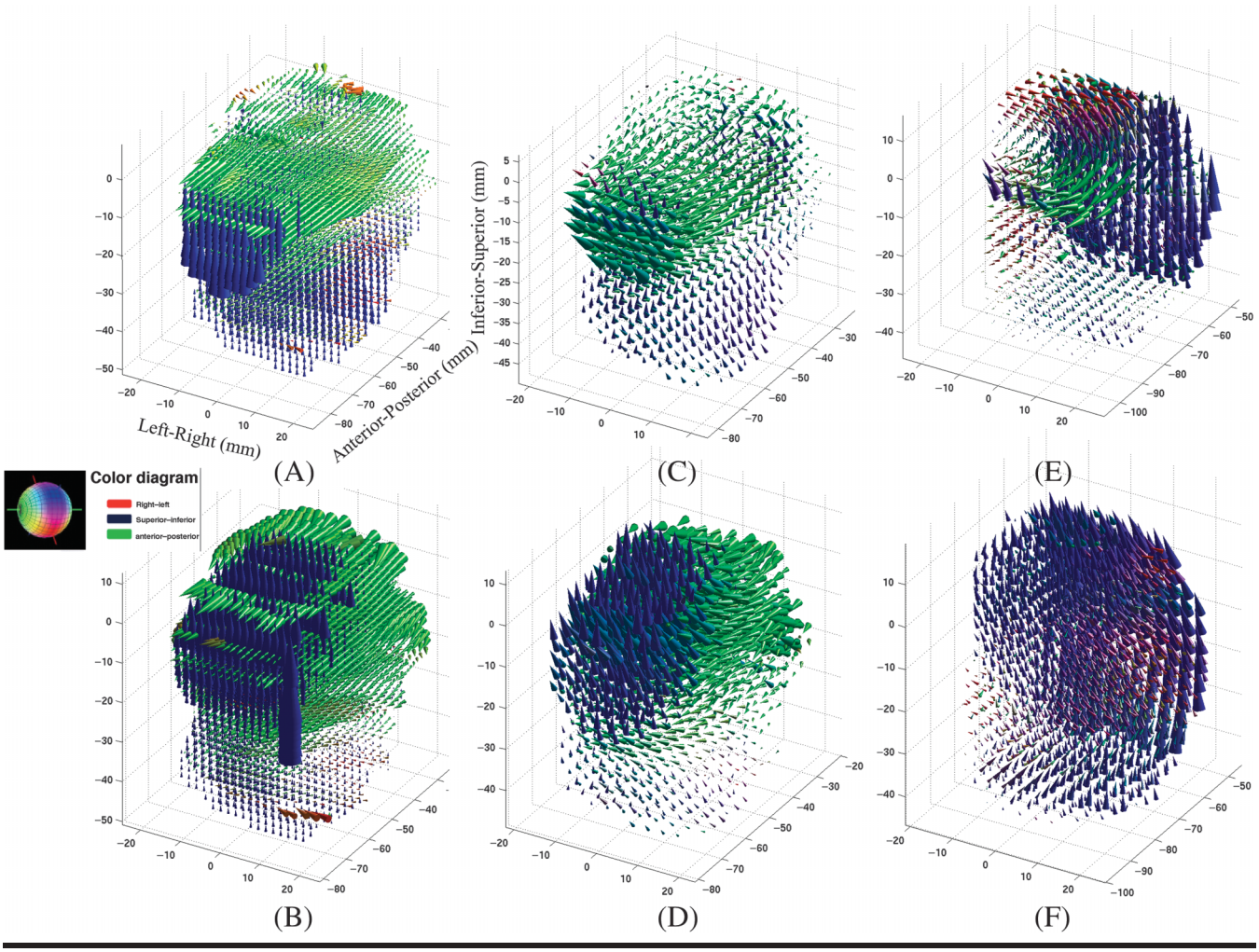
Two-Step Principal-Components Analysis

A two-step principal-components analysis (PCA) was used to differentiate the subjects in the control and glossectomy groups in this study. First, a PCA (PCA-1) was done for the control group only and a control motion PC (principal-components) space was obtained. All of the motions of the subjects in the glossectomy group were then projected onto these PCs to identify and extract the motion patterns identical to those of the subjects in the control group. A second PCA (PCA-2) was performed on the remaining variance of the subjects in the glossectomy group to account for the motion patterns that were unique to them. The mathematical details of the PCA procedure are presented in the following.

First, the PCA required a certain tongue-motion quantity to be in the same frame of reference. Although all displacement fields had been regularized to be with respect to /ə/, different subjects' tongue shapes varied widely. Therefore, the tongue region was divided into eight volumes of interests (VOIs) by separating from the volume's center planes (see Figure 6A). The motion field inside each VOI was averaged to produce one vector to represent its general motion (see Figure 6B), denoted $\{\boldsymbol{d}_{r,1}, \boldsymbol{d}_{r,2}, \ldots, \boldsymbol{d}_{r,26}\}_v$, where v is the VOI number (1 through 8). These VOIs were treated independently in the following processing.

Because only the periods from /ə/ to /s/ to /k/ were of interest, a common time interval was created by taking the average motion between these two periods and then using cubic spline (Fan & Yao, 2003; denoted cspline in Equation 3) to interpolate them into 17 time frames for all subjects, where /ə/ was at time frame 1, /s/ was at time frame 7, and /k/ was at time frame 17. Denoting the time-frame

Figure 5. 3-D motion estimation by the incompressible deformation estimation algorithm. (A) Harmonic phase algorithm input at /s/ for the control group. (B) Harmonic phase algorithm input at /k/ for the control group. (C) 3-D motion at /s/ for the control group. (D) 3-D motion at /k/ for the control group. (E) 3-D motion at /s/ for the glossectomy group. (F) 3-D motion at /k/ for the glossectomy group. Cones are color-coded by motion directions as shown in the color diagram (red for left–right, blue for superior–inferior, green for anterior–posterior). Cone size is motion magnitude.



subscripts of maximum /ə/, /s/, and /k/ positions as *a, s,* and *k,* we have

$$\left\{\hat{\boldsymbol{d}}_{1,1},\ldots,\hat{\boldsymbol{d}}_{1,7},\ldots,\hat{\boldsymbol{d}}_{1,17}\right\}_v = \text{cspline}\left\{\boldsymbol{d}_{a,a},\ldots,\boldsymbol{d}_{a,s},\ldots,\boldsymbol{d}_{a,k}\right\}_v \tag{3}$$

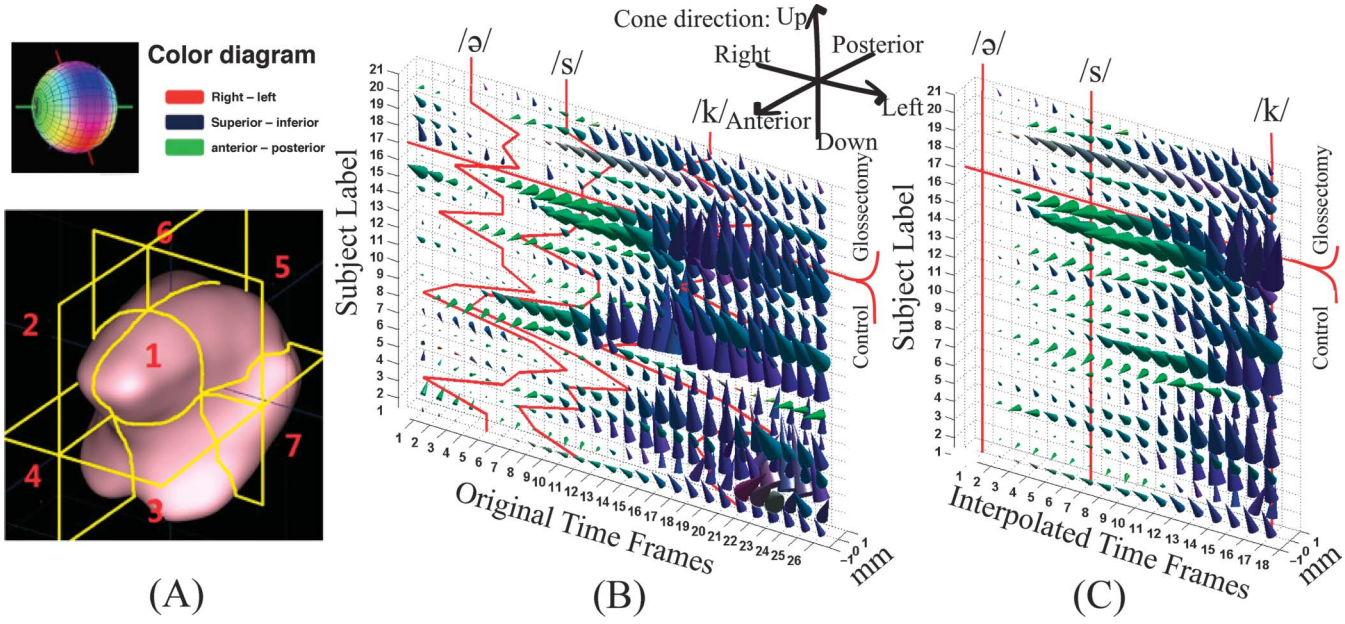
For any VOI, $\hat{\boldsymbol{d}}_{1,t}$ was the interpolated mean motion that put all subjects' motions in the same framework and ready for PCA (see Figure 6C). Denoting the subject number with *i,* the mean motion of all 17 frames was stacked into one vector,

$$\hat{\boldsymbol{d}}^i = \left[\hat{\boldsymbol{d}}^i_{1,1};\ldots;\hat{\boldsymbol{d}}^i_{1,7};\ldots;\hat{\boldsymbol{d}}^i_{1,17}\right] \tag{4}$$

which existed in a $3 \times 17 = 51$-dimensional space. The physical meaning of $\hat{\boldsymbol{d}}^i$ was all motion both in space and in time of subject *i* while performing the complete speech task of "a souk." Note that in this way, the method avoided treating each time frame independently. Instead, the entire speech task was considered as an evaluation of the subject's speech function.

Denoting the control-group subject number with *C,* PCA-1 on these subjects required the following steps: (a) subtracting the mean of controls $\hat{\boldsymbol{s}}^i = \hat{\boldsymbol{d}}^i - \text{mean}\{\hat{\boldsymbol{d}}^1,\ldots,\hat{\boldsymbol{d}}^i,\ldots,\hat{\boldsymbol{d}}^C\}$; (b) computing the covariance matrix of the subtracted motion $COV = [\hat{\boldsymbol{s}}^1,\ldots,\hat{\boldsymbol{s}}^i,\ldots,\hat{\boldsymbol{s}}^C][\hat{\boldsymbol{s}}^1,\ldots\hat{\boldsymbol{s}}^i,\ldots,\hat{\boldsymbol{s}}^C]^T$; and (c) finding the eigen-decomposition of COV to get $C-1$ principal-component directions $\{\boldsymbol{e}^1,\ldots,\boldsymbol{e}^{C-1}\}$ and principal values $\{P^1,\ldots,P^{C-1}\}$. To evaluate this PC space to see if it was able to distinguish motion between subjects in the control and glossectomy groups, motion from a test group of subjects in the control group and all the subjects in the glossectomy

Figure 6. (A) Division of eight volumes of interest in the tongue. (B) An example using VOI-1: Average motion of 26 frames and all subjects. Horizontal line divides control group (bottom) and glossectomy group (top). Vertical curves are at time frames /ə/, /s/, and /k/. (C) Interpolated motion between time frames /ə/ and /k/ from (B). Horizontal line divides control group (bottom) and glossectomy group (top). Vertical lines are at time frames /ə/, /s/, and /k/.



group was projected onto these PCs. This result was unsatisfying, however, because it showed the similarities between subjects in the two groups instead of the unique features of those in the glossectomy group. Therefore, PCA-2 was introduced to solve this problem. Because the first PC space had a rank of $C - 1$ and the entire space had 51 dimensions, the remaining $51 - (C - 1)$ principal directions could be vectors generated by any feasible orthogonalization method (e.g., the Gram–Schmidt process). This remaining space with $51 - (C - 1)$ dimensions contained only the motion of the subjects in the glossectomy group because ideally those in the control group should project a 0 PC score in this space. Denoting the glossectomy-group subject number with P, the motions of those subjects, indexed by j ($1 \le j \le P$) after being subtracted with all controls' mean motion, were used to compute those subjects' part that was identical to the control group by projecting onto the PCA-1 space; that is,

$$\hat{\boldsymbol{s}}^j = \hat{\boldsymbol{d}}^j - \text{mean}\left\{\hat{\boldsymbol{d}}^1, \ldots, \hat{\boldsymbol{d}}^i, \ldots, \hat{\boldsymbol{d}}^C\right\} \quad (5)$$

$$\hat{\boldsymbol{s}}^j_{control} = \left(\left(\hat{s}^j\right)^T \cdot \boldsymbol{e}^1\right)\boldsymbol{e}^1 + \ldots + \left(\left(\hat{s}^j\right)^T \cdot \boldsymbol{e}^{C-1}\right)\boldsymbol{e}^{C-1} \quad (6)$$

The remaining motion was considered unique to the subjects in the glossectomy group and was given by

$$\hat{\boldsymbol{s}}^j_{patient} = \hat{\boldsymbol{s}}^j - \hat{\boldsymbol{s}}^j_{control} \quad (7)$$

Then the covariance matrix of $\hat{\boldsymbol{s}}^j_{patient}$ was computed and its eigen-decomposition was used to get P more vectors as the PC directions for glossectomy-group subjects' motion $\{\boldsymbol{u}^1, \ldots, \boldsymbol{u}^P\}$. Taken together, $\{\boldsymbol{e}^1, \ldots, \boldsymbol{e}^{C-1}, \boldsymbol{u}^1, \ldots, \boldsymbol{u}^P\}$ were generated from a two-step PCA to represent the control-group and glossectomy-group motion parts and are referred to, respectively, as primary and secondary PC spaces in the following. Note that the reason the rank of the secondary PC space was P instead of $P - 1$ is that the mean motion of the control group was subtracted in Equation 5 so that the glossectomy group's motions were not zero centered. The entire purpose of building the secondary PC space was to use it to contain the glossectomy group's unique motion pattern and to separate it from its part that was like the motion of the control group. After the two-step PCA, any new subject's motion was projected onto $\{\boldsymbol{e}^1, \ldots, \boldsymbol{e}^{C-1}, \boldsymbol{u}^1, \ldots, \boldsymbol{u}^P\}$ to compute its primary and secondary PC scores for the purpose of evaluating its control-group-like and glossectomy-group-like motion patterns.

Results

Evaluation of Azimuth Motion Angle

The first purpose of applying TMAP to the 21 subjects was to determine whether those in the glossectomy group had more left–right motion than those in the control group. As an example, the motion of a subject from the control group and one from the glossectomy group at critical time frames /s/ and /k/ are shown in Figures 5C–5F. For the

subject in the control group, both forward and upward motions were symmetrical (only blue or green), whereas the motion of the subject in the glossectomy group contained more left–right motion, shown by red and purple cones. After multisubject data regularization, each subject's average motion was interpolated using Equation 3. For the subjects in the glossectomy group, if their tumors were on the right side at VOI-2, their tongues were symmetrically flipped along the midsagittal plane so that the tumor side was mirrored to the left and became VOI-1. If their tumors were originally on the left side at VOI-1, the data were left unflipped. The azimuth angle ϕ between the motion vector and the midsagittal plane (angle tilted to the left or right of the anterior direction) was computed at every time frame; its magnitude reflected motion asymmetry (see Figure 7). We studied ϕ from time frames 4 through 17 because the motion magnitude in the first three time frames was nearly 0, so that noise dominated over angle computation. Figure 7 shows an example of the $|\phi|$ values at the tongue tip (VOI-1 on the left tip and VOI-2 on the right tip). Note that the $|\phi|$ value stayed small for the control group in general, whereas it became large or inconsistent for most of the glossectomy group. For all eight VOIs, the ϕ angle is shown in degrees in Table 1 as the mean, standard deviation, and median across all time frames. In VOIs-1 through 4 (anterior tongue), the subjects in the glossectomy group had a larger angle and standard deviation than those in the control group. A paired Student's t test was performed with independent variables of the control and glossectomy groups and a dependent variable of ϕ value for all four anterior VOIs and all time frames: $t(55) = -9.57$, $p < .01$, effect size = .79. The test proved that in the anterior tongue, subjects in the control group used less left–right motion compared with the subjects in the glossectomy group, who had strong evidence of left–right asymmetry. However, for VOIs-5 through 8 (posterior tongue), although the standard deviation was mostly higher in the glossectomy group, the mean and median showed less difference, and the results were not significant ($p = .48$).

Test of Two-Step PCA on the Mirrored Hemi-Tongue

The two-step PCA strategy on this data set of 16 subjects in a control group and five in a glossectomy group was first applied on the whole tongue, yielding 15 primary PC directions and five secondary PC directions. In Figures 8A and 8B, the PC directions and PC weights of VOI-1 are shown as an example of the PC space's appearance. Other VOIs were processed independently in the same fashion. Figures 8A and 8B demonstrate that visual assessment of the PC space was difficult, although motions in the glossectomy group looked different from those in the control group and seemed to contain more inconsistent motion patterns. Because this was an ideal setup of the two-step PCA, where all subjects in the control group were used to build the primary PC space, Figures 8C and 8D show a perfect outcome: the control group loaded in only the primary space and the glossectomy group loaded in both spaces. In constructing an efficient test of the method, a result similar to this ideal situation should be expected: Subjects in the control group should load smaller and closer to 0 than those in the glossectomy group in the secondary PC space.

Thus, a mirrored-hemi-tongue experiment was performed to test the efficacy of the two-step PCA strategy. Because the tongue motions of the 16 subjects in the control group were generally symmetric, the right sides of all their tongues (VOIs-2, 4, 6, and 8) were mirrored to the left (VOIs-1, 3, 5, and 7), overlaid on top of the left-side motion, and averaged with the left side for each time frame. This yielded 16 subjects in the control group with only the left four VOIs. For the five subjects in the glossectomy group, because their glossectomy was performed on only one side (either left or right) of the tongue, their tongues were no longer symmetric and their data should not be averaged. Therefore, after their right hemi-tongues were mirrored to the left, the result was five resected hemi-tongues and five native hemi-tongues, both appearing to be the left four VOIs. The resected group was named Patient

Figure 7. Azimuth motion angle $|\varphi|$ of 16 control-group subjects and five glossectomy-group subjects on the resected and native tongue parts. Each curve is a subject. (A) Controls at VOI-1. (B) Controls at VOI-2. (C) Patients at VOI-1. (D) Patients at VOI-2. VOI = volume of interest.

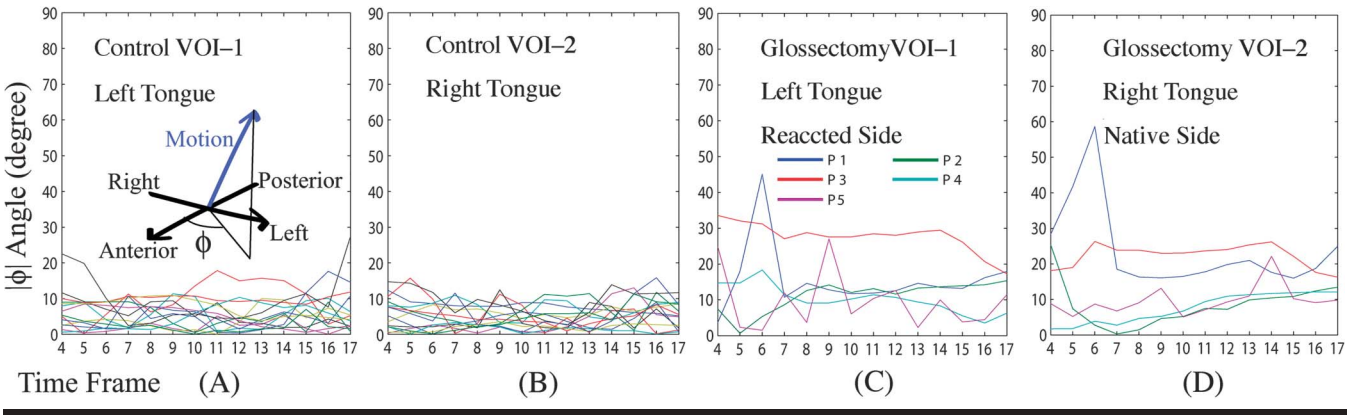


Table 1. Mean azimuth angle of motion (in degrees; *M* ± *SD* [median]) for the control and glossectomy groups.

Group	VOI-1	VOI-2	VOI-3	VOI-4	VOI-5	VOI-6	VOI-7	VOI-8
Control	7.2 ± 5.9 (5.8)	6.4 ± 5.1 (5.6)	9.9 ± 8.3 (7.3)	10.1 ± 7.9 (7.6)	8.5 ± 6.5 (7.2)	8.3 ± 8.8 (4.9)	10.3 ± 9.6 (8.8)	11.9 ± 9.7 (8.0)
Glossectomy	14.8 ± 9.2 (12.7)	14.4 ± 9.8 (12.0)	17.3 ± 19.3 (11.6)	14.2 ± 16.3 (8.8)	11.4 ± 9.6 (9.6)	8.6 ± 9.2 (4.8)	9.9 ± 10.9 (6.2)	11.4 ± 9.7 (8.1)

Note. VOI = volume of interest.

Glossectomy Side (PGS) and the native group was named Patient Native Side (PNS).

With these modified data on hemi-tongues, the two-step PCA was performed four times, once for each VOI, using 11 of the subjects in the control group as training data to build the primary PC space and leaving five as test data (control-group tests). The five PGS subjects were then used to create the secondary PC space. Following the presented procedure, 10 primary PC directions and five secondary PC directions were obtained for each VOI. Then the three subject groups (five control-group tests, five PGSs, and five PNSs) were projected onto the secondary PCs. The first two secondary PC weights of all three subject groups for all four VOIs are shown in Figure 9. The solid blue dot represents zero weight, for no glossectomy-group motion. The green dots, which are control-group tests, show smaller loading values closer to the center 0. The glossectomy group had higher loadings in both the PNS (crosses) and PGS (circles) subgroups because both sides of the tongue showed patterns different from those of the control group (see Figure 7, for example).

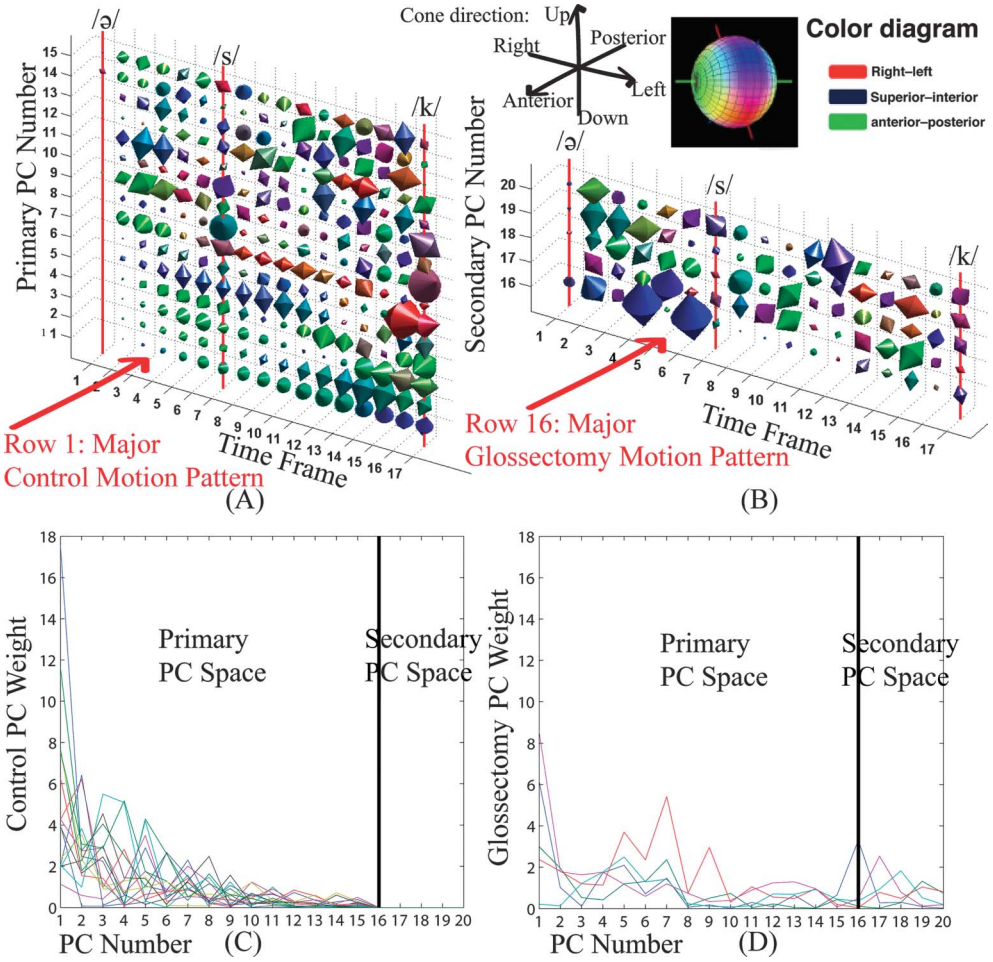
A final PCA was done using all combinations of control-group subjects in the testing and training groups. The choice of training control-group subjects from the whole group (16 choose 5 is 4,368 possibilities) was randomized, thus randomizing the construction of the primary and secondary PC spaces. The hemi-tongue experiment was repeated in 4,368 permutation tests for all possible forms of the PC spaces, keeping five control-group subjects in the test group to compare with the PGS and PNS groups. For each PC space, the three groups' motions were projected onto the secondary PC space to compute their PC weights for the part like the glossectomy group. The average PC weights for the part like the glossectomy group of all combinations and all VOIs are box plotted in Figure 10. The mean of the control-group tests' secondary weights was lower than that of both the PGS and PNS groups for all 4,368 cases, Student's $t = -473.23$, $p < .01$, effect size = .98. Despite the small amount of training data, this analysis was capable of distinguishing motions in the control group from those of both the native and resected hemi-tongues in the glossectomy group.

Discussion

The azimuth motion angle experiment was intended to test the assumption that the subjects in the glossectomy group had more left–right asymmetric motions than those in the control group. The control-group subjects' production of "a souk" used predominantly symmetrical motion as it moved forward to /s/ and then upward to /k/. Figure 8A shows no lateral (red) motion until PCs 6 and 7, after almost 99% of the variance has been accounted for. Thus, the left–right asymmetric motion seen in Table 1, in the floor of the mouth (VOIs-3, 4, 7, and 8), represented a small amount of motion. Figure 10 shows greater asymmetry in the PC loadings for P2 and P5, who had quite dissimilar motion in the PGS versus PNS tongue root (VOI-7). Thus, some but not all of these people with small tumors moved more asymmetrically than the control-group subjects did.

Table 1 and the corresponding statistical test show that for the presented small number of subjects in the glossectomy group, their tongue motion was noisier and less predictable than that of the control group, producing a larger amount of left–right asymmetry. From the temporally averaged motion of all VOIs, the azimuth angle of the glossectomy group's tongues always had greater value and standard deviation than that of the control group at VOIs-1 through 4, the anterior tongue. Because our study used only subjects with a unilateral tumor behind the tongue tip, the motor control of the tongue tip was reduced. Therefore, the anterior VOIs-1 through 4 were expected to be affected more strongly by surgery than the posterior part of the tongue, and this was supported by the azimuth-angle data. For the glossectomy group, the odd-numbered VOIs were the resected side of the tongue and the even numbers were the native side. Although the glossectomy group had more left-to-right rotation than the control group, the difference in rotation between the resected and native sides was small—no more than 3°—and reasonable, because the two sides of the tongue were contiguous and therefore moved with each other. Figure 7 shows that the resected side of the tongue behaved in an equally or more unusual manner than the native side. For example, P1 had a peak angular motion at time frame 6, which was during the motion into /s/, and the peak was larger for the native side. P3 had a large left–right angle on both sides. P5, who had a free flap, showed a rhythmic left–right alternation on the resected-flap side, which was echoed to a lesser extent on the native side. Greater motion on the resected side was also found by Stone, Langguth, Woo, Chen, and Prince (2014) and Bressmann et al. (2006). Bressmann and colleagues have also found poorer motility correlated with reduced intelligibility (Bressmann, Sader, Whitehill,

Figure 8. Example of the two-step principal-component (PC) directions and PC weights. (A) Primary PC directions from 16 control-group subjects. (B) Secondary PC directions from five glossectomy-group subjects. (C) Weights on the two PC spaces for control-group subjects. Each curve is a subject. (D) Weights on the two PC spaces for glossectomy-group subjects.

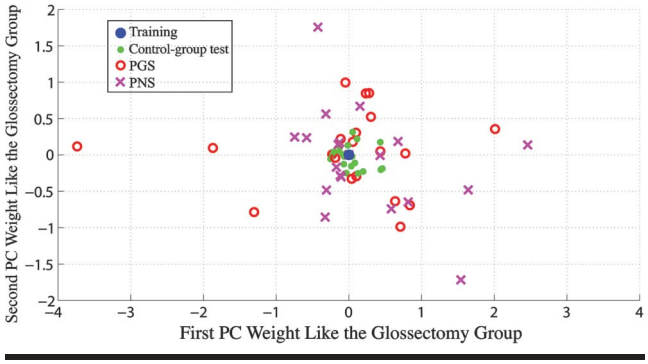


& Samman, 2004) and greater asymmetry in people who have received a glossectomy for tongue shape (Bressmann, Ackloo, Heng, & Irish, 2007) and motion (Bressmann et al., 2006).

The two-step PCA experiment further revealed the unique motion pattern in the glossectomy group. Although the PC space was visually difficult to accurately interpret, a general pattern on the first few primary PCs starting from 1 and the first few secondary PCs starting from 16 were recognizable. The primary directions showed little lateral motion until PC 7, indicating a consistent control-group motion featuring mostly anterior–posterior and inferior–superior (see Figure 8A). The secondary directions contained all glossectomy-group motion patterns, showing many red glyphs that correspond to left–right motion (see Figure 8B). However, the quantitative evaluation of this result was achieved using the mirrored-hemi-tongue experiment.

The purpose of the mirrored-hemi-tongue data was to reveal that both sides of the glossectomy-group subjects' tongues, native and resected, loaded more highly than the control-group subjects' tongues on the secondary PC space. From 4,368 permutation tests, the efficacy of TMAP was confirmed by proving that the control group's motions were more consistent and stable than the glossectomy group's. The mean of the control-group subjects' energy like the glossectomy group's weighted lower than both the PGS and PNS groups in most VOIs, and two-step PCA was capable of distinguishing the subtle motion-variation pattern of the glossectomy group from the control group. However, PGS and PNS motions were not well distinguished by the current approach, suggesting that compensation may be occurring on both sides of the tongue. In a previous study including three people who had received a glossectomy (Stone et al., 2014), they moved their resected side to a greater extent than did the people in the control group, whereas the native side moved similarly to that of the people in the control group. In the current work, however, the subjects in the glossectomy group loaded on the secondary PCs in both the PGS and PNS subgroups, suggesting that both sides of the tongue

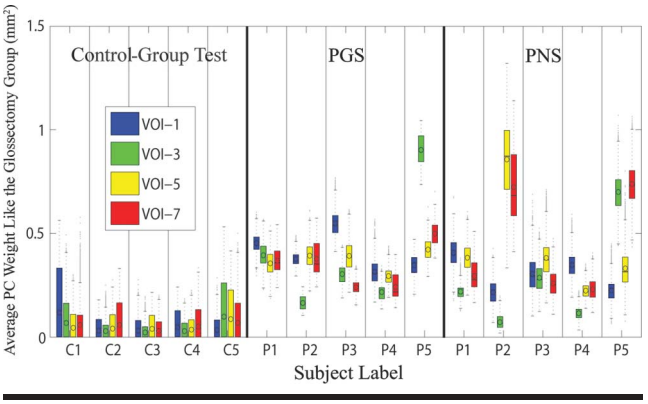
Figure 9. Weights on the two major principal components (PCs) like the glossectomy group for all subjects on all volumes of interest. Blue dot: the origin and all training control-group subjects used to build the PC space. Green dots: control-group subjects used to test the PC analysis. Circles: glossectomy-group subjects on the glossectomy side (PGS). Crosses: glossectomy-group subjects on the native side (PNS).



used different strategies from those of the control group. And the higher PNS loading in some VOIs suggests that these subjects used the native side in a more unusual manner than the resected side for compensation.

Although the two halves of the tongue may move asymmetrically, it is likely that the asymmetry is part of a global control strategy for the tongue, which takes into account acoustical goals and natural morphological asymmetries. We do not believe one hemi-tongue is controlled separately from the other, even for the glossectomy group, although we do believe that differing and even oppositional commands can be sent to the two hemi-tongues as part of a single gesture. This would be similar to walking, skipping, or hopping, where different commands and timing are sent to the legs but they are part of a coordinated gesture. There could also be asymmetries that arise for biomechanical reasons, such as hard-structure asymmetries or task demands, where stiffening and rotation can be used to speed up a motion. Evidence for this comes from two facts besides direct observations. First, chewing requires the tongue to throw food onto the teeth prior to each chew. To do this, the tongue elevates one side, rotates laterally, and pushes the bolus onto the teeth. This rotation is consistent with different levels of activation types to different sides of the tongue. Second, coronal ultrasound movies of the tongue show not only differences in motion on the left and right tongue but also left–right rotation during speech (Slud, Stone, Smith, & Goldstein, 2002). Left–right rotation can be produced as a unified, single motor strategy but requires agonist muscles on either side of the tongue to activate alternately rather than simultaneously.

Figure 10. Box plot of the average weights on the principal components (PCs) like the glossectomy group. Three groups on all volumes of interest are shown: control-group subjects used to test the PC analysis, glossectomy-group subjects on the glossectomy side (PGS), and glossectomy-group subjects on the native side (PNS). In each box, the center bar shows the median and the circle shows the mean.



A closer look at Figure 10 with reference to clinical recordings provided us more information on the potential compensation strategies used by the subjects in the glossectomy group. As mentioned under Data Acquisition, the fifth subject in the glossectomy group (PGS 5 and PNS 5) had different anatomy from the other four subjects in the group due to his T2 tumor and flap closure, which affected his motion pattern. He used the mouth-floor muscles (VOIs-3 and 7) differently, on both sides of the tongue, yielding a higher loading of secondary PCs. The floor muscles helped move the upper tongue appropriately to shape the vocal tract and produce good-quality speech, as seen by the low loadings on the secondary PCs of VOIs-1 and 5 on the upper tongue. This example implies that the secondary PCs allowed an explanation of the degree and nature of a glossectomy-group subject's unique movement in a way that was not possible before.

Conclusion

In this article, a workflow of algorithms—TMAP—was proposed for processing speech MRI data to obtain a reliable estimate of the motion field. It is semiautomatic and easy to operate. Methods to achieve effective segmentation, multisubject data regularization, and a two-step PCA to reveal subtle unique motion patterns were described in detail. TMAP was tested using a varied collection of subjects with permutation tests, showing its promising value in current speech studies and medical applications.

It has been important for the progress of biological research into the human body to have a pipeline method suitable for any user that can process tagged and cine MRI data from the raw images to the final data analysis and interpretation. Collaboration between medical professionals and engineers is crucial to providing systems such as this one. In the past, 3-D motion was difficult to accurately compute, and multiple subjects' 3-D motion was mostly interpreted by visual assessment. The introduction of TMAP enables the computation and two-step PCA-enabled quantitative analysis, which was the major contribution of this work.

The conversion of tissue-point motion from multiple 2-D orientations into a 3-D data set is challenging, and

more improvements to the detailed methods of TMAP are being studied and made. This system involves registration, segmentation, motion analysis, interpretation, and other problems of great interest in medical imaging. An upgrade in any step could lead to an improved system and better results, which is to be proposed in future studies.

Acknowledgments

This project was supported by National Cancer Institute Grant 5R01CA133015 (awarded to Maureen Stone), National Institute on Deafness and Other Communication Disorders Grant K99/R00 DC009279 (awarded to Emi Z. Murano), and National Institute on Deafness and Other Communication Disorders Grant K99 DC012575 (awarded to Jonghye Woo).

References

Abd-El-Malek, S. (1955). The part played by the tongue in mastication and deglutition. *Journal of Anatomy, 89,* 250–254.1.

American Cancer Society. (2016). *What are the key statistics about oral cavity and oropharyngeal cancers?* Retrieved from http://www.cancer.org/cancer/oralcavityandoropharyngealcancer/detailedguide/oral-cavity-and-oropharyngeal-cancer-key-statistic

Bressmann, T., Ackloo, E., Heng, C.-L., & Irish, J. C. (2007). Quantitative three-dimensional ultrasound imaging of partially resected tongues. *Otolaryngology—Head & Neck Surgery, 136,* 799–805.

Bressmann, T., Flowers, H., Ackloo, E., Heng, C.-L., Wong, W., & Irish, J. C. (2006). Static and dynamic 3D ultrasound imaging of the tongue following partial glossectomy surgery: Assessment of grooving and symmetry [Abstract]. *Stem-, Spraak- en Taalpathologie, 14*(Suppl.), 52.

Bressmann, T., Sader, R., Whitehill, T. L., & Samman, N. (2004). Consonant intelligibility and tongue motility in patients with partial glossectomy. *Journal of Oral and Maxillofacial Surgery, 62,* 298–303.

Chen, M., Lu, W., Chen, Q., Ruchala, K. J., & Olivera, G. H. (2008). A simple fixed-point approach to invert a deformation field. *Medical Physics, 35,* 81–88.

Fan, J., & Yao, Q. (2003). *Nonlinear time series: Nonparametric and parametric methods.* New York, NY: Springer.

Grady, L. (2006). Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28,* 1768–1783.

Han, X., Xu, C., & Prince, J. L. (2003). A topology preserving level set method for geometric deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25,* 755–768.

Heller, K. S., Levy, J., & Sciubba, J. J. (1991). Speech patterns following partial glossectomy for small tumors of the tongue. *Head & Neck, 13,* 340–343.

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech.* San Diego, CA: Singular.

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

Lee, J., Woo, J., Xing, F., Murano, E. Z., Stone, M., & Prince, J. L. (2014). Semi-automatic segmentation for 3D motion analysis of the tongue with dynamic MRI. *Computerized Medical Imaging and Graphics, 38,* 714–724.

Liu, X., Abd-Elmoniem, K. Z., Stone, M., Murano, E. Z., Zhuo, J., Gullapalli, R. P., & Prince, J. L. (2012). Incompressible deformation estimation algorithm (IDEA) from tagged MR images. *IEEE Transactions on Medical Imaging, 31,* 326–340.

Liu, X., & Prince, J. L. (2010). Shortest path refinement for motion estimation from tagged MR images. *IEEE Transactions on Medical Imaging, 29,* 1560–1572.

Maton, A. (1997). *Human biology and health.* Englewood Cliffs, NJ: Prentice Hall.

NessAiver, M., & Prince, J. L. (2003). Magnitude image CSPAMM reconstruction (MICSR). *Magnetic Resonance in Medicine, 50,* 331–342.

Nicoletti, G., Soutar, D. S., Jackson, M. S., Wrench, A. A., Robertson, G., & Robertson, C. (2004). Objective assessment of speech after surgical treatment for oral cancer: Experience from 196 selected cases. *Plastic and Reconstructive Surgery, 113,* 114–125.

Osman, N. F., McVeigh, E. R., & Prince, J. L. (2000). Imaging heart motion using harmonic phase MRI. *IEEE Transactions on Medical Imaging, 19,* 186–202.

Parthasarathy, V., Prince, J. L., Stone, M., Murano, E. Z., & NessAiver, M. (2007). Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. *The Journal of the Acoustical Society of America, 121,* 491–504.

Pauloski, B. R., Logemann, J. A., Colangelo, L. A., Rademaker, A. W., McConnel, F. M. S., Heiser, M. A., . . . Esclamado, R. (1998). Surgical variables affecting speech in treated patients with oral and oropharyngeal cancer. *The Laryngoscope, 108,* 908–916.

Ronse, C., Najman, L., & Decencière, E. (Eds.). (2005). *Mathematical morphology: 40 years on—Proceedings of the 7th International Symposium on Mathematical Morphology, April 18–20, 2005.* Dordrecht, the Netherlands: Springer.

Slud, E., Stone, M., Smith, P. J., & Goldstein, M., Jr. (2002). Principal components representation of the two-dimensional coronal tongue surface. *Phonetica, 59,* 108–133.

Stone, M., Langguth, J. M., Woo, J., Chen, H., & Prince, J. L. (2014). Tongue motion patterns in post-glossectomy and typical speakers: A principal components analysis. *Journal of Speech, Language, and Hearing Research, 57,* 707–717.

Stone, M., Liu, X., Chen, H., & Prince, J. L. (2010). A preliminary application of principal components and cluster analysis to internal tongue deformation patterns. *Computer Methods in Biomechanics and Biomedical Engineering, 13,* 493–503.

Takemoto, H. (2001). Morphological analyses of the human tongue musculature for three-dimensional modeling. *Journal of Speech, Language, and Hearing Research, 44,* 95–107.

Vercauteren, T., Pennec, X., Perchant, A., & Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage, 45*(1, Suppl. 1), S61–S72.

Woo, J., Murano, E. Z., Stone, M., & Prince, J. L. (2012). Reconstruction of high-resolution tongue volumes from MRI. *IEEE Transactions on Biomedical Engineering, 59,* 3511–3524.

Zerhouni, E. A., Parish, D. M., Rogers, W. J., Yang, A., & Shapiro, E. P. (1988). Human heart: Tagging with MR imaging—A method for noninvasive assessment of myocardial motion. *Radiology, 169,* 59–63.