3D Tongue Motion from Tagged and Cine MR Images

Fangxu Xing[1], **Jonghye Woo**[1,2], **Emi Z. Murano**[3], **Junghoon Lee**[1,4], **Maureen Stone**[2], and **Jerry L. Prince**[1]

[1] Department of Electrical and Computer Engineering, Johns Hopkins University

[2] Department of Neural and Pain Sciences, University of Maryland Dental School

[3] Department of Otolaryngology, Johns Hopkins School of Medicine

[4] Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

Abstract

Understanding the deformation of the tongue during human speech is important for head and neck surgeons and speech and language scientists. Tagged magnetic resonance (MR) imaging can be used to image 2D motion, and data from multiple image planes can be combined via post-processing to yield estimates of 3D motion. However, lacking boundary information, this approach su ers from inaccurate estimates near the tongue surface. This paper describes a method that combines two sources of information to yield improved estimation of 3D tongue motion. The method uses the harmonic phase (HARP) algorithm to extract motion from tags and diffeomorphic demons to provide surface deformation. It then uses an incompressible deformation estimation algorithm to incorporate both sources of displacement information to form an estimate of the 3D whole tongue motion. Experimental results show that use of combined information improves motion estimation near the tongue surface, a problem that has previously been reported as problematic in HARP analysis, while preserving accurate internal motion estimates. Results on both normal and abnormal tongue motions are shown.

Keywords

Tongue; motion; HARP; registration; 3D; surface

1 Introduction

The human tongue moves rapidly in complex and incompressible motions during speech [1]. In post-glossectomy patients, i.e., people who have had surgical resection of part of the tongue muscle for cancer or sleep apnea treatment, tongue moving ability and its speech functionality may be adversely affected. Therefore, understanding the tongue motion during speech in both normal and post-glossectomy subjects is of great interest to speech scientists, head and neck surgeons, and their patients.

fxing1@jhu.edu.

To capture the tongue's motion during speech, tagged magnetic resonance (MR) images can be acquired over a series of time frames spanning a speech utterance [2, 3]. The two-dimensional (2D) motion information carried in these images can be extracted using the harmonic phase (HARP) algorithm [4]. With a collection of 2D motions from image slices covering the tongue, a high-resolution three-dimensional (3D) motion estimate can be achieved by interpolation with previously reported incompressible deformation estimation algorithm (IDEA) [5].

However, since HARP uses a bandpass filter to extract the harmonic images, object boundaries are blurred and motion estimates near the anatomical surfaces are inaccurate [6, 7]. To make matters worse, HARP measurements near the boundaries are sparse because of the sparseness of image plane acquisition. These two problems severely affect 3D motion estimation near anatomical surfaces, as shown in Fig. 1. Zooming in on the back of the tongue (see Fig. 1(a)), 1(b) shows the sparse 2D motion components from HARP and 1(c) is the IDEA reconstruction of 3D motion that shows inaccurate large motion.

This paper presents a novel approach that combines data from tagged images with surface deformation information derived from cine MR images to dramatically improve 3D tongue motion estimation. At every time frame, the tongue is segmented to achieve a 3D mask, and the deformation between the reference mask at the resting position and the deformed mask is computed using deformable registration. The normal components of surface deformation are then used to augment the HARP measurements within the IDEA estimation framework. Fig. 1(d) shows the additional input and Fig. 1(e) shows the result of proposed method. Comparing with Fig. 1(c), this result is more sensible from a qualitative point of view. Quantitative evaluations provided below also show that this method achieves a more accurate estimate of the whole tongue motion.

2 Methods

2.1 Data Acquisition and HARP Tracking

In this study, subjects repeatedly speak an utterance "a souk" during which tagged and cine MR image sequences are acquired at multiple parallel axial slice locations covering the tongue. The resolution scheme is 1.88 mm in-plane (dense) and 6.00 mm through-plane (sparse). For tagged images, both horizontal and vertical tags are applied on each slice, providing motion components in two in-plane directions (x and y components). To acquire motion components in the through-plane direction (z component), another set of parallel coronal slices orthogonal to axial is also acquired. HARP is then used on every tagged image at every time frame, resulting in a corresponding 2D motion field representing the projection of the 3D motion of every tissue point on the current slice plane. Fig. 1(b) shows such HARP slices for the utterance "a souk" at the moment when /s/ is sounded (current time frame), where the tongue is expected to have moved forward from the /a/ moment (time frame 1) when the tags are applied. Meanwhile, cine images revealing better anatomical structures are going to be used for segmentation and registration to be described in section 2.3.

2.2 IDEA Algorithm

Figs. 2(a) and 2(b) illustrate how HARP data are processed in IDEA [5]. The undeformed tissue at time frame 1 has undeformed reference tag planes. At current time frame, the tag planes have deformed along with the tissue. To each point (pixel location) x_a on an axial image such as Fig. 2(a), HARP produces two vectors representing components of displacement:

$$\begin{cases} \mathbf{q}_x = q_x \mathbf{e}_x, \\ \mathbf{q}_y = q_y \mathbf{e}_y, \end{cases} \quad (1)$$

where $\mathbf{e}_x$ and $\mathbf{e}_y$ are unit vectors in the x and y directions and $\mathbf{q}_x$ and $\mathbf{q}_y$ are the projections of the 3D motion $\mathbf{u}(\mathbf{x}_a)$ on the current axial plane. Similarly, for each point $\mathbf{x}_c$ on a coronal image such as Fig. 2(b), HARP yields the displacement component vector

$$\mathbf{q}_z = q_z \mathbf{e}_z, \quad (2)$$

where $\mathbf{e}_z$ is the unit vector in the z direction.

IDEA takes such data on all pixels $\{\mathbf{x}_a, \mathbf{q}_x(\mathbf{x}_a), \mathbf{x}_a, \mathbf{q}_y(\mathbf{x}_a), \mathbf{x}_c, \mathbf{q}_z(\mathbf{x}_c)\}$ as in put, and estimates an incompressible deformation field $\mathbf{u}(\mathbf{x})$ on a high-resolution grid within the tongue mask. The details are omitted here for lack of space, but are given in [5]. We only note two important aspects. First, IDEA is carried out as a series of smoothing splines, each of which seeks a divergence-free velocity field yielding the deformation field only when integrated. Thus the final field $\mathbf{u}(\mathbf{x})$ is nearly incompressible and its reprojected components at all input points nearly agree with the input measurements. Second, the inputs are observed components of displacements that can arise at any physical position and in any sub-direction of motion. This is the key to utilization of surface deformation measurements within the IDEA framework. In particular, as shown in Fig. 2(c), the tongue surface may deform between time frames, and a point $\mathbf{x}_s$ on the surface at current time frame can be associated with a point on the reference tongue surface. However, like the traditional aperture problem in optical flow, we should not assume to know any tangential information about the surface displacement. This leads to a perfect analogy with HARP data: observations about surface normal deformation, if available, can be used in 3D reconstruction.

2.3 Measuring Tongue Surface Deformation

IDEA requires segmentation of the tongue volume in order to limit the tissue region that is assumed to be incompressible [8]. Cine MR images are used to construct a super-resolution volume [9] at each time frame, which is then manually segmented for the tongue surface mask. We notice that these 3D masks can also be used for deformable registration in order to provide surface deformation information.

The diffeomorphic demons method [10] is applied to the pair of masks between the two time frames where motion is to be computed. Denoting the reference mask at time frame 1 as $I_1 : \Omega_1 \subset \mathbb{R}^3 \to \{0, 1\}$ and the current deformed mask as $I_t : \Omega_t \subset \mathbb{R}^3 \to \{0, 1\}$ defined on open and bounded domains Ω_1 and Ω_t, the deformation field is found and denoted by the

mapping $\mathbf{d}{:}\Omega_t \mapsto \Omega_1$. The estimated displacement field at a point $\mathbf{x}_s$ on the surface of the tongue in current time frame can be denoted as

$$\mathbf{u}\left(\mathbf{x}_s\right) = -\,\mathbf{d}\left(\mathbf{x}_s\right). \quad (3)$$

Although diffeomorphic demons generates a whole 3D displacement volume, we take only tongue surface normal components for the reason stated in the previous section. We represent the 3D tongue mask at current time frame by a levelset function $\varphi(\mathbf{x})$ that is zero on the surface, positive outside the tongue, and negative inside the tongue. The normal directions of the surface are given by

$$\mathbf{n}\left(\mathbf{x}_s\right) = \frac{\nabla\phi\left(\mathbf{x}_s\right)}{\left|\nabla\phi\left(\mathbf{x}_s\right)\right|}. \quad (4)$$

The normal components of motion—serving as additional input to IDEA—are

$$\mathbf{q}_n\left(\mathbf{x}_s\right) = \left(\mathbf{u}\left(\mathbf{x}_s\right) \cdot \mathbf{n}\left(\mathbf{x}_s\right)\right) \mathbf{n}\left(\mathbf{x}_s\right). \quad (5)$$

An example of such a field is shown in Fig. 1(d).

2.4 Enhanced IDEA

With the enhanced input $\{\mathbf{x}_a,\ \mathbf{q}_x(\mathbf{x}_a),\ \mathbf{x}_a,\ \mathbf{q}_y(\mathbf{x}_a),\ \mathbf{x}_c,\ \mathbf{q}_z(\mathbf{x}_c),\ \mathbf{x}_s,\ \mathbf{q}_n(\mathbf{x}_s)\}$, our proposed method computes the 3D motion over the super-resolution grid points $\{\mathbf{x}_i\}$ and all the surface points $\{\mathbf{x}_s\}$. The algorithm is summarized below.

3 Results

We evaluated E-IDEA on 50 tongue volumes (25 from a normal control and 25 from a patient) during the utterance "a souk". Conventional IDEA was also computed for comparison. We computed motion fields relative to time frame 1 which was the /a/ sound, because the resting tongue serves as a good reference configuration, is the natural reference frame for the MR tags, and also fits into continuum mechanics framework for deforming bodies.

Firstly, we visually assessed the motion fields. The results of both subjects are shown in Figs. 1(c), 1(e) and Fig. 3 on two critical time frames: at the /s/, when forward motion is prominent, and at the /k/, when upward motion is prominent (Fig. 1 is for control at time frame /s/). Knowing that the internal muscular structure of tongue prevents its back from performing either too large or zero motion [1], at tongue's back, we see E-IDEA has reduced the erroneous large motions for the control, and has captured those small motions where IDEA mistakenly interpolates as zero for the patient. We also see E-IDEA can straighten up the motion at the top of the tongue to better estimate the displacement when the tongue hits the palate vertically (Figs. 3(a), 3(d)). In general, the boundary estimation agrees more with tongue physical mechanics [1].

Secondly, to obtain a numerical comparison, we manually tracked the motions of 15 surface points distributed 5 each on the front, top, and back parts of the tongue (labeled in Fig. 4(a)).

We then computed their trajectories with IDEA and E-IDEA motion fields. The tracks of three methods are shown in Fig. 4(a) and errors from manual tracking at each point are shown in Figs. 4(b) and 4(c), boxplotted across all time frames. The error magnitude has been reduced by E-IDEA, especially on the back part of the tongue. Also, the mean error (circles in boxes) is reduced by E-IDEA at all 15 points. The improvement is significant ($p = 0.00003$).

Lastly, we took the estimated 3D motions at input sample locations and reprojected them onto input directions using Eqns. (1) and (5). We then computed a *reprojection error* that gives the error in distance in the input directions between the estimated sample components and the input sample components. This measure assumes input motion components (HARP and surface normal motions) are the truth. We compare four types of reprojection errors in histograms of Fig. 5: on IDEA internal points, on E-IDEA internal points, on E-IDEA boundary points, and on IDEA boundary points as indicated in the legend. For the control, on a total of 105455 internal points and 108853 boundary points, the mean of the four errors are: 0.32 mm, 0.35 mm, 0.65 mm, and 1.33 mm, respectively. The boundary error has been reduced by 0.68 mm and the internal error has been raised by 0.03 mm. For the patient, on 133302 internal points and 100523 boundary points, the mean of the four errors are: 0.22 mm, 0.24 mm, 0.96 mm and 3.11 mm. The boundary error has been reduced by 2.15 mm and the internal error has been raised by 0.02 mm.

4 Conclusion and Discussion

We have proposed a novel algorithm for estimating the tongue's motion field in 3D. The major innovation is in the incorporation of surface motion as additional information, which compensates for the well-known deficiencies of HARP in estimating boundary motions. Both qualitative and quantitative improvements are evident using two independent metrics. Especially, from reprojection error, we see that boundary error is substantially reduced while internal error is only minimally increased.

This method is still being improved. Aspects that will be addressed in the future include optimizing the segmentation and registration methods, studying intra-subject volume dependency, and adding data reliability terms to balance HARP and registration information. Also, choice of different reference frames can be explored. And fitting the "internal plus surface motion" idea into other motion estimation frameworks can be an interesting topic.

Acknowledgments

References

1. Kier WM, Smith KK. Tongues, Tentacles and Trunks: the Biomechanics of Movement in Muscular-hydrostats. Zool. J. Linnean Soc. 1985; 83:307–324.

2. Zerhouni EA, Parish DM, Rogers WJ, Yang A, Shapiro EP. Human Heart: Tagging with MR Imaging — a Method for Noninvasive Assessment of Myocardial Motion. Radiology. 1988; 169:59–63. [PubMed: 3420283]

3. Parthasarathy V, Prince JL, Stone M, Murano E, Nessaiver M. Measuring Tongue Motion from Tagged Cine-MRI Using Harmonic Phase (HARP) Processing. J. Acoust. Soc. Am. 2007; 121(1): 491–504. [PubMed: 17297803]

4. Osman NF, McVeigh ER, Prince JL. Imaging Heart Motion Using Harmonic Phase MRI. IEEE Trans. Med. Imaging. 2000; 19(3):186–202. [PubMed: 10875703]

5. Liu X, Abd-Elmoniem K, Stone M, Murano E, Zhuo J, Gullapalli R, Prince JL. Incompressible Deformation Estimation Algorithm (IDEA) from Tagged MR Images. IEEE Trans. Med. Imaging. 2012; 31(2):326–340. [PubMed: 21937342]

6. Tecelao SR, Zwanenburg JJ, Kuijer JP, Marcus JT. Extended Harmonic Phase Tracking of Myocardial Motion: Improved Coverage of Myocardium and Its Effect on Strain Results. J. Magn. Reson. Imaging. 2006; 23(5):682–690. [PubMed: 16570245]

7. Liu X, Prince JL. Shortest Path Refinement for Motion Estimation From Tagged MR Images. IEEE Trans. Med. Imaging. 2010; 29(8):1560–1572. [PubMed: 20304720]

8. Xing, F.; Lee, J.; Murano, EZ.; Woo, J.; Stone, M.; Prince, JL. Estimating 3D Tongue Motion with MR Images.. Asilomar Conference on Signals, Systems, and Computers; Pacific Grove, California. 2012.

9. Woo J, Murano EZ, Stone M, Prince JL. Reconstruction of High Resolution Tongue Volumes from MRI. IEEE Trans. Biomedical Engineering. 2012; 59(12):3511–3524.

10. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic Demons: Efficient Non-parametric Image Registration. NeuroImage. 2008; 45(1):61–72.
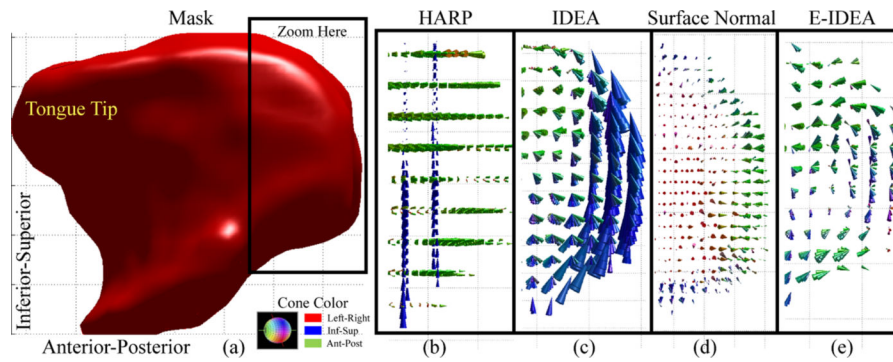
Fig. 1.
(a) Tongue mask of a normal control subject (sagittal view). (b) HARP field on axial and coronal slices as input for IDEA, zoomed in at the tongue back. (c) IDEA result at the tongue back. (d) Surface normal deformation component at tongue back surface. (e) Proposed method result. Note: In this paper *cones* are used to visualize motion fields, where cone size indicates motion magnitude and cone color follows conventional DTI scheme (see cone color diagram).
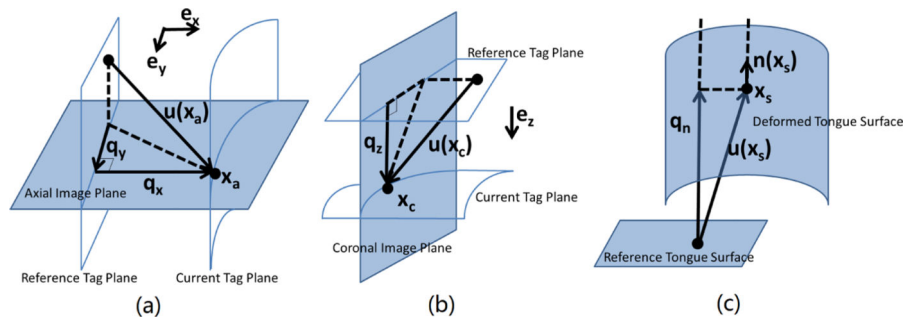
Fig. 2.
Relationship between 2D motion components and 3D motion on (a) an axial slice, (b) a coronal slice and (c) the tongue surface.
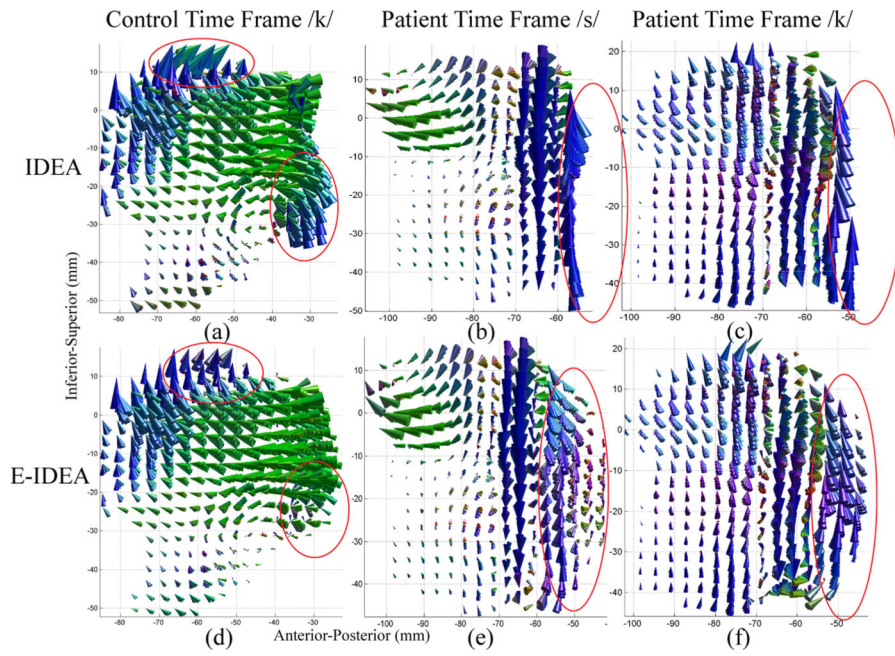
Fig. 3.
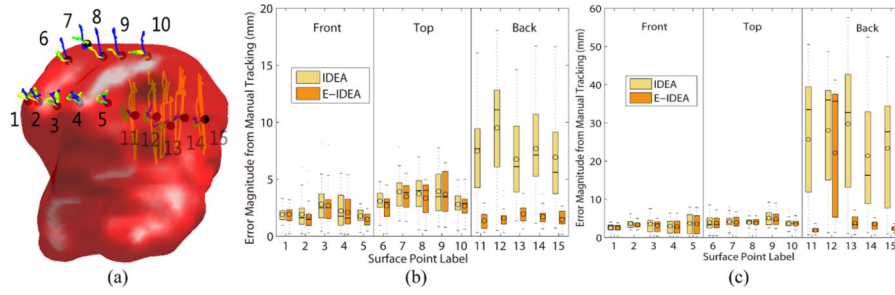Visual comparison of conventional IDEA result and E-IDEA result.

Fig. 4.

Comparison of IDEA and E-IDEA with manually tracked surface points. (a) Tracks of the control surface points by manual (*blue*), IDEA (*yellow*), and E-IDEA (*green*). (b) Error magnitude for the control (*bar* is median and *circle* is mean). (c) Error magnitude for the patient.
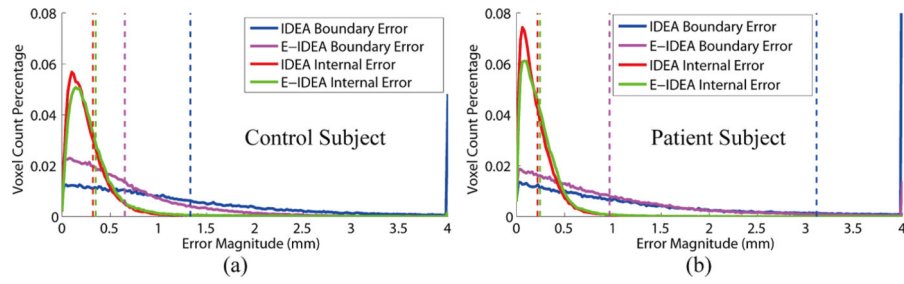
Fig. 5.
Regularized histogram of IDEA and E-IDEA's reprojection error on internal and surface points. *Dotted lines* show the mean of four types of reprojection error.

Algorithm

Enhanced Incompressible Deformation Estimation Algorithm

---

1. Set $u(x_i) = 0$ and $u(x_s) = 0$.

2. Set M time steps, **for** $m = 1$ to M **do**

3. Project currently computed displacement onto input directions by $p_x(x_a) = u(x_a) \cdot e_x$, $p_y(x_a) = u(x_a) \cdot e_y$, $p_z(x_c) = u(x_c) \cdot e_z$, $p_n(x_s) = u(x_s) \cdot n(x_s)$.

4. Compute remaining motion projection by $r_x(x_a) = q_x(x_a) - p_x(x_a)$, $r_y(x_a) = q_y(x_a) - p_y(x_a)$, $r_z(x_c) = q_z(x_c) - p_z(x_c)$, $r_n(x_s) = q_n(x_s) - p_n(x_s)$.

5. Use part of the remaining motion to approximate velocity: $v_x(x_a) = r_x(x_a)/(M - m + 1)$, $v_y(x_a) = r_y(x_a)/(M - m + 1)$, $v_z(x_c) = r_z(x_c)/(M - m + 1)$, $v_n(x_s) = r_n(x_s)/(M - m + 1)$.

6. Update estimation: $u(x_i) = u(x_i) + \text{DFVS}\{v_x(x_a), v_y(x_a), v_z(x_c), v_n(x_s)\}$, $u(x_s) = u(x_s) + \text{DFVS}\{v_x(x_a), v_y(x_a), v_z(x_c), v_n(x_s)\}$.

7. **end for**

Here DFVS stands for divergence-free vector spline, which is also the key algorithm "workhorse" of IDEA [5]. M is typically set to 20 which provides a proper trade-off between accuracy and computation time. Enhanced IDEA, which we refer to as E-IDEA below, typically takes about 5 hours on 26 time frames.