Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2009 January 1; 7258: . doi:10.1117/12.811706.

Quantification of Three Dimensional Tongue Motion During Speech Using zHARP

Xiaofeng Liu^a, Jiachen Zhuo^b, Harsh Agarwal^a, Khaled Z. Abd-Elmoniem^a, Emi Murano^c,
Maureen Stone^c, Rao Gullapalli^b, and Jerry L. Prince^a

^a Johns Hopkins University, Baltimore, MD USA

^b University of Maryland School of Medicine, Baltimore, MD USA

^c University of Maryland Dental School, Baltimore, MD USA

Abstract

To understand the role of the tongue in speech production, it is desirable to directly image the motion and strain of the muscles within the tongue. Magnetic resonance tagging—which was originally developed for cardiac imaging—has previously been applied to image both two-dimensional and three-dimensional tongue motion during speech. However, to quantify three-dimensional motion and strain, multiple images yielding two-dimensional motion must be acquired at different orientations and then interpolated—a time-consuming task both in image acquisition and processing. Recently, a new MR imaging and image processing method called zHARP was developed to encode and track 3D motion from a single slice without increasing acquisition time. zHARP was originally developed and applied to cardiac imaging. The application of zHARP to the tongue is not straightforward because the tongue in repetitive speech does not move as consistently as the heart in its beating cycle. Therefore tongue images are more susceptible to motion artifacts. Moreover, these artifacts are greatly exaggerated as compared to conventional tagging because of the nature of zHARP acquisition. In this work, we re-implemented the zHARP imaging sequence and optimized it for the tongue motion analysis. We also optimized image acquisition by designing and developing a specialized MRI scanner triggering method and vocal repetition to better synchronize speech repetitions. Our method was validated using a moving phantom. Results of 3D motion tracking and strain analysis on the tongue experiments demonstrate the effectiveness of this method.

Keywords

Motion quantification; tongue; zHARP

1. INTRODUCTION

The tongue is crucial to many important functions such as speaking, swallowing, and breathing. Therefore, it is important to understand how the tongue muscles contribute to the tongue's function. In speech research, cine magnetic resonance imaging (MRI) has been

used to observe tongue motion during speech.^{1–5} Cine MRI can measure the shape changes of the tongue surface,⁶ but it cannot measure the interior motion because the image intensity inside the tongue is quite homogeneous.

MR tagging,^{7, 8} which places non-invasive and temporary markers (tags) inside the tongue in a pre-specified pattern, can be used to image the motion inside the tongue. Two-dimensional (2D) motion can be measured from tagged images by tracking tag lines⁹ or using the (more automated) harmonic phase (HARP) method.^{10–12} In order to quantify three-dimensional (3D) tongue motion, previous methodology has required the acquisition of multiple images in orthogonal imaging planes.^{13, 14} This is a time-consuming imaging task and is prone to misregistration errors due to patient motion. This data must then be interpolated within the field of view using (for example) spline models^{14–16} on the whole tongue, which is another time-consuming operation. It is highly desirable to image motion and strain in the tongue more directly so that the artifacts caused by patient motion and imperfect speech repetitions can be reduced and that off-line processing is minimized.

Recently, Abd-Elmoniem et al.^{17, 18} developed a new MR imaging and image processing method called zHARP, which can encode and track 3D motion from a single slice without increasing acquisition time. ZHARP also enables convenient and fast 3D strain tensor computation.^{19, 20} Originally developed and applied to cardiac imaging, the application of zHARP to the tongue is not straightforward because the tongue (in repetitive speech) does not move as consistently as the heart (in its very consistent cycle). Therefore, tongue images are more susceptible to motion artifacts than cardiac images, and these artifacts are also exaggerated in zHARP as compared to conventional tagging (explained below). In this work, we re-implemented the zHARP imaging sequence and optimized it for the tongue motion analysis. We also developed specialized triggering method to reduce motion artifacts. Experimental results on tongue images in speech demonstrate the capability of our method on 3D tongue motion quantification and strain analysis.

2. BACKGROUND

2.1 zHARP Imaging and 3D Motion Tracking

zHARP imaging is based on slice-following (SF) CSPAMM.²¹ SF-CSPAMM uses a slice-selective tagging RF pulse and CSPAMM subtraction to compensate for the through-plane motion. In SF-CSPAMM, at the initial time frame, tags are applied only on a thin slice, but a thick slab is imaged later. The slab is thick enough so that it encompasses the moving tagged slice at all times. The imaged signal in the slab but outside the tagged slice is zero immediately after tagging, and it gains over time due to T1 recovery. As in conventional CSPAMM, SF-CSPAMM acquires two image sequences $A(\mathbf{x}, t)$ and $B(\mathbf{x}, t)$, one with a $[+90^\circ, +90^\circ]$ RF pulse pair, and the other with a $[+90^\circ, -90^\circ]$ RF pulse pair. The two tagging RF pulse pairs create a tagging phase shift of π between $A(\mathbf{x}, t)$ and $B(\mathbf{x}, t)$, while the imaged signal in the imaged slab but outside of the tagged slice remains the same. By subtracting the two (complex) images, the signal outside of the tagged slice is canceled, and the resulting SF-CSPAMM image contains only the signal from the tagged slice. The SF-CSPAMM image can be represented as:

$$I_{SF} = A(\mathbf{x}, t) - B(\mathbf{x}, t) = 2M(\mathbf{x}, t) \cos(k\mathbf{n} \cdot \mathbf{p}(\mathbf{x}, t)), \quad (1)$$

where $M(\mathbf{x}, t)$ represents the effective magnetization, $\mathbf{p}(\mathbf{x}, t)$ is the 3D position of the imaged material point $\mathbf{x}$ at the initial time frame, k is the tagging frequency and $\mathbf{n}$ is the normal vector representing the tagging orientation. To compute the 2D in-plane motion, two SF-CSPAMM images with orthogonal tag orientations—e.g., in the x and y directions—must be acquired.

ZHARP imaging modifies SF-CSPAMM by adding a small z -encoding gradient immediately before the imaging readout. This gradient adds to all material points in the tagged slice a phase ϕ_z , which is linearly related to the points' locations in the through-plane (z) direction at the imaging time. In addition, the z -encoding gradients applied in the orthogonally tagged images have the same magnitude but opposite polarity. If the slice is thin enough, the two orthogonally tagged images can be expressed as

$$I_z^x = 2M(\mathbf{x}, t) \cos(k_x \mathbf{n}_x \cdot \mathbf{p}(\mathbf{x}, t)) e^{+j\phi_z(\mathbf{x}, t)} e^{j\phi_e(\mathbf{x})}, \quad (2)$$

$$I_z^y = 2M(\mathbf{x}, t) \cos(k_y \mathbf{n}_y \cdot \mathbf{p}(\mathbf{x}, t)) e^{-j\phi_z(\mathbf{x}, t)} e^{j\phi_e(\mathbf{x})}, \quad (3)$$

where ϕ_z is the phase error caused by susceptibility and magnetic field inhomogeneities, $\phi_z(\mathbf{x}, t) = k_z z(\mathbf{x}, t)$ with k_z be the z -encoding frequency and $z(\mathbf{x}, t)$ be the location of $\mathbf{x}$ in the through-plane direction at time t , and ϕ_e is the phase error caused by susceptibility and magnetic field inhomogeneities. We denote $\phi_x = k_x \mathbf{n}_x \cdot \mathbf{p}(\mathbf{x}, t)$ and $\phi_y = k_y \mathbf{n}_y \cdot \mathbf{p}(\mathbf{x}, t)$ as the tagging phases of the material point $\mathbf{x}$ on the two orthogonal tagging orientations.

Given zHARP images, the displacement-encoded phase images ϕ_x , ϕ_y , and ϕ_z can be separated using the 2-D harmonic phase (HARP) concept.^{10, 11} Instead of applying HARP filtering on only one harmonic peak, however, the bandpass filter is applied to both the positive and negative harmonic peaks of each of the two orthogonally tagged images. The displacement-encoded phase images and ϕ_e can be isolated from the phases of these four resulting complex images via linear combination. The zHARP processing¹⁸ is illustrated in Fig. 1.

With the phase images ϕ_x , ϕ_y , and ϕ_z , the 3D motion of the imaged material points can be computed. The in-plane motion is first determined as in conventional 2D HARP tracking. Given the position of a material point $\mathbf{x}_i$ at some time frame t_i , the HARP method searches its neighborhood at the next time frame t_{i+1} in both tag orientations to find the closest point $\mathbf{x}_{i+1}$ such that

$$\begin{bmatrix} \phi_x(\mathbf{x}_{i+1}, t_{i+1}) \\ \phi_y(\mathbf{x}_{i+1}, t_{i+1}) \end{bmatrix} = \begin{bmatrix} \phi_x(\mathbf{x}_i, t_i) \\ \phi_y(\mathbf{x}_i, t_i) \end{bmatrix}. \quad (4)$$

The through-plane position is computed based on the fact that ϕ_z is a linear function of the material point's location in z -axis—i.e., $k_z(z_{i+1} - z_i) = \phi_z(\mathbf{x}_{i+1}, t_{i+1}) - \phi_z(\mathbf{x}_i, t_i)$.

2.2 3D Strain Computation

Using two parallel zHARP slices, we can quantify the full 3D strain tensor without first calculating the 3D motion.²⁰ Given a material point $\mathbf{x}$ and its location in the reference time frame $\mathbf{p}(\mathbf{x}, t)$, the displacement vector can be expressed as:

$$\mathbf{u}(\mathbf{x}, t) = [u_x, u_y, u_z]^T = \mathbf{x} - \mathbf{p}(\mathbf{x}, t). \quad (5)$$

The *Eulerian strain tensor* is defined as

$$E = \frac{1}{2}[\mathbf{I} - (\mathbf{F}^{-1})^T \mathbf{F}^{-1}] = \frac{1}{2}[\mathbf{I} - (\mathbf{I} - \nabla \mathbf{u})^T (\mathbf{I} - \nabla \mathbf{u})] \quad (6)$$

where $\mathbf{F}^{-1} = (\nabla \mathbf{p}(\mathbf{x}, t))$ is the deformation gradient. Since

$$\Phi(\mathbf{x}, t) = \begin{bmatrix} \varphi_x \\ \varphi_y \\ \varphi_z \end{bmatrix} = \begin{bmatrix} k_x \mathbf{n}_x & k_y \mathbf{n}_y & k_z \mathbf{n}_z \end{bmatrix}^T \mathbf{p}(\mathbf{x}, t) = \mathbf{K} \cdot \mathbf{p}(\mathbf{x}, t), \quad (7)$$

the deformation gradient can be re-written as

$$\mathbf{F}^{-1} = \nabla \mathbf{p}(\mathbf{x}, t) = \mathbf{K}^{-1} \nabla \Phi(\mathbf{x}, t). \quad (8)$$

The computation of 3D strain tensor requires two parallel zHARP slices. The 3D phase vector images $\Phi^{(1)}(\mathbf{x}, t)$ and $\Phi^{(2)}(\mathbf{x}, t)$ are first calculated on a regular grid $\mathbf{x} = (i, j)$, $1 \leq i, j$

N . Then the phase gradient $\nabla \Phi(\mathbf{x}, t)$ is calculated using finite differences. The Eulerian strain tensor is computed by substituting $\nabla \Phi(\mathbf{x}, t)$ into Eqs. (8) and (6).

3. MATERIALS AND METHODS

3.1 Implementation

We implemented zHARP using a gradient echo sequence on a 3T Tim-Trio MRI scanner (Siemens Medical Solutions, Malvern, PA) equipped with twelve receiver channels. The pulse sequence is shown in Fig. 2. Immediately after the trigger signal is detected, the slice-selective tagging is applied on a thin slice, followed by a standard gradient echo pulse train but modified by adding the z -encoding gradient. For one image set, the same acquisition is repeated 4 times with combinations of $\pm 90^\circ$ second tagging RF pulse, and phase encoding direction in x or y -axis—i.e., the readout is in y or x -axis, respectively—with different z -encoding gradient polarity. We also implemented a ramped flip angle technique (cf. ²¹) to improve the tag persistence and achieve a constant tag contrast during acquisition.

3.2 System Setup with a Specialized MRI Triggering System

Motion artifacts are much more severe in tongue imaging than in heart imaging because the tongue does not move as consistently in repeated speech as the heart does in its repeating cycles. To reduce motion inconsistency we designed a triggering system similar to Masaki et al.¹ to help synchronize the tongue motion. The system setup is shown in Fig. 3. In the system, a continuous periodic stereo sound wave is generated and played on a stereo sound system during image acquisition. One channel of the sound simulates the electrocardiogram (ECG) trigger signal, and is connected to an ECG signal receiver, which in turn triggers the scanner to run the zHARP sequence. The other channel is connected to the headphone that the subject wears in the scanner. During the scan, the subject paces his or her utterance based on the noise bursts from the headphone. In a typical setting, the trigger period is 2 sec. The subject talks in the first second, and breathes in and out during the second second. In this way the speech is coordinated to start at the same time relative to the trigger, and the duration of the speech is more consistent. Before the experiment the subjects are trained for 15–30 minutes to follow the sound rhythm.

3.3 Artifact Reduction

In the *in vivo* tongue motion experiments, we found that slice-following tagging images are much more sensitive to motion artifacts than conventional tagging. Fig. 4 shows examples in which the subject is not speaking, yet artifacts due to swallowing are visible around the throat. These images were all acquired in 14 repetitions at 625 ms after tagging. As well, when the imaged slab is thicker the artifact becomes much more severe. This is because slice-following tagging images are computed by subtracting two images that are acquired on a much thicker slab than the slice of interest. Through subtraction, the motion artifacts on the thick slab are carried over to the remaining tagged thin slice and are greatly amplified, because the signal strength of the tagged slice is much weaker than the imaged slab. Besides, motion artifacts are more severe in later time frames than in earlier time frames because in slice-following tagging, the untagged tissue in the thick slab has no MR signal immediately after tagging, but gains signal over time, while the tagging signal decays in the meantime.

To suppress motion artifacts, the imaged slab should be as thin as possible while still encompassing the moving tagged slice in all time frames. In practice since the tongue moves most in the anterior-posterior direction during normal speech, zHARP images of the tongue should not be acquired in the coronal plane. As well, motion artifacts can be lessened by reducing the number of repetitions required to image one slice, although that necessitates a sacrifice of temporal resolution, which may be undesirable.

4. EXPERIMENTS AND RESULTS

Our implementation of zHARP sequence was validated on phantom experiments. The phantom moved back and forth along the main magnetic field direction with a period of 1.02 second, and the images were acquired in the first 460 ms in 20 time frames. The imaging parameters were: FOV = 150 mm × 150 mm, image size = 128 × 128, TR = 23 ms, tagged slice thickness = 8 mm, imaged slab thickness = 24 mm, z-encoding period = 30 mm, tag separation = 16 mm. The zHARP image slice was tilted about 70 degrees from the motion

direction. Therefore both in-plane and through-plane (z) motion exhibited. Fig. 5 shows the orthogonally tagged image pairs and the phase images computed from zHARP processing at two time frames.

For validation, we acquired standard CSPAMM images parallel to the motion direction. Phantom motion was estimated from CSPAMM images using standard HARP processing and was taken as the ground truth. The true motion was then projected onto both the in-plane and through-plane directions of the zHARP image slices, and compared with the motion computed from zHARP 3D tracking. The comparison results are shown in Fig. 6. Over all time frames, the mean error in the in-plane direction was 0.346 mm, and the average error in the through-plane direction was 0.133 mm.

In the *in vivo* experiments, the speech material studied was /iu/. The subject was in a supine position in the scanner with both head and neck coils positioned. The imaging parameters were: tagged slice thickness = 6 mm, imaged slab thickness = 18 mm, slice separation = 7.2 mm, z-encoding period = 30 mm, FOV = 240 mm, tag separation = 16 mm, temporal resolution = 52 ms, time frames = 12. The acquired k -space matrix size was 64×22 and 11 k -space lines were acquired during each repetition. Each zHARP slice was acquired in 4 repetitions. In the first experiment, three sagittal slices were acquired. Tongue regions were manually segmented in all slices at the first time frame then tracked in 3D to all later time frames using zHARP tracking. Examples of 3D displacement maps were shown in Fig. 7.

In another experiment, five axial slices of zHARP images were acquired using the same parameters. These axial slices are perpendicular to the sagittal image shown in Fig. 8(a), and their intersections are shown as red lines. Figs. 8(b–d) show the 3D path lines of points inside the tongue on slices 2, 3, and 4. From the figures we observe that the upper part of the tongue moves more than the lower part, and it moves more in the anterior-posterior and inferior-superior directions than in the medio-lateral direction. We also acquired conventional CSPAMM images in the sagittal orientation for visual assessment. Tongue regions from two of the axial images were manually selected and are shown in Fig. 9(a). Figs. 9(b)–(d) show the locations of those regions at later times, displayed together with the mid-sagittal CSPAMM image. It can be seen that the 3D locations of the axial slices match the tags of CSPAMM image pretty well. Next, we computed the 3 by 3 strain tensors between pairs of axial images; these results are shown in Fig. 10.

5. CONCLUSION

We developed a system for fast imaging and quantification of 3D tongue motion and strain during speech using zHARP. In this system, an optimized MR imaging sequence was implemented. We also used a specialized vocal repetition and MR scanner triggering system to account for the poor repeatability of tongue motion. Experiments on tongue during speech demonstrated the effectiveness of this pulse sequence and system.

References

1. Masaki S, Tiede MK, Honda K, Shimada Y, Fujimoto I, Nakamura Y, Ninomiya N. MRI-based speech production study using a synchronized sampling method. *J Acoustic Science Japan*. 1999; 20:375–379.
2. Demolin, D.; Metens, T.; Soquet, A. Real time MRI and articulatory coordinations in vowels. *Proc. of International Speech Production Seminar*; 2000.
3. Mady, K.; Sader, R.; Zimmermann, A.; Hoole, P.; Beer, A.; Zeilhofer, H-F.; Channig, C. Assessment of consonant articulation in glossectomee speech by dynamic MRI. *Proc. ICSLP*; 2002; 2002.
4. Engwall, O. A revisit to the application of MRI to the analysis of speech production - testing our assumptions. *6th Int. Sem. Spee. Pro. (ISPS)*; 2003.
5. Narayanan S, Nayak K, Byrd D, Lee S. An approach to real-time magnetic resonance imaging for speech production. *J Acoust Soc Amer*. 2004; 115(4):1771–1776. [PubMed: 15101655]
6. Li, M.; Kambhamettu, C.; Stone, M. Snake for band edge extraction and its applications. *Sixth IASTED Int. Conf. Comp. Graph. Imag*; 2003. p. 261-266.
7. 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]
8. Axel L, Dougherty L. MR imaging of motion with spatial modulation of magnetization. *Radiology*. 1989; 171:841–845. [PubMed: 2717762]
9. Guttman MA, Prince JL, McVeigh ER. Tag and contour detection in tagged MR images of the left ventricle. *IEEE Tran on Med Imag* (13). 1994; (1):74–88.
10. Osman NF, Kerwin WS, McVeigh ER, Prince JL. Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. *Mag Res Med*. 1999; 42:1048–1060.
11. Osman NF, Prince JL. Visualizing myocardial function using HARP MRI. *Phys Med Biol*. 2000; 45:1665–1682. [PubMed: 10870717]
12. 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]
13. McVeigh ER, Atalar E. Cardiac tagging with breath-hold cine MRI. *Magn Res Med*. 1992; 28:318–327.
14. Dick, D.; Ozturk, C.; Douglas, A.; McVeigh, ER.; Stone, M. Three-dimensional tracking of tongue motion using tagged-MRI. *8th Int. Soc. Mag. Res. Med*; 2000.
15. Stone, M.; Dick, D.; Douglas, A.; Davis, E.; Ozturk, C. Modelling the internal tongue using principal strains. *Proc. Speech Prod Seminar*; 2000. p. 133-136.
16. Liu, X.; Stone, M.; Prince, JL. Tracking tongue motion in three dimensions using tagged MR images. *IEEE Int. Sym. Biomed. Imag*; 2006. p. 1372-1375.
17. Abd-Elmoniem, KZ.; Stuber, M.; Osman, N.; Prince, JL. ZHARP: three-dimensional motion tracking from a single image plane. *Info. Proc. Med. Imag*; 2005. p. 639-651.
18. Abd-Elmoniem KZ, Osman NF, Prince JL, Stuber M. Three-dimensional magnetic resonance myocardial motion tracking from a single image plane. *Mag Res Med*. 2007; 58:92–102.
19. Abd-Elmoniem, KZ.; Stuber, M.; Prince, JL. Multi-slice three-dimensional myocardial strain tensor quantification using zHARP. *Inf. Proc. In Med. Imag*; 2007. p. 62-73.
20. Abd-Elmoniem KZ, Stuber M, Prince JL. Direct three-dimensional myocardial strain tensor quantification and tracking using zHARP. *Med Imag Anal*. 2008; 12:778–786.
21. Fischer SE, McKinnon GC, Scheidegger MB, Prins W, Meier D, Boesiger P. True myocardial motion tracking. *Mag Res Med*. 1994; 31:401–413.

Figure 1. The zHARP processing. This example shows sagittal slices of the lower part of the head. The anterior-posterior direction is right-left in the images.

Figure 2.
The zHARP pulse sequence.

Figure 3.
The system setup diagram.

Figure 4.

Examples of the motion artifacts on SF-CSPAMM images. These images were acquired using same parameters except the thickness of the imaged slab is (a) 2, (b) 3 and (c) 4 times of the tagged slice.

Figure 5.

The zHARP processing of the phantom data. at time frame Top row: time frame 1; bottom time frame 10. (a–b) are the two horizontally tagged image pairs. (c–d) are the in-plane harmonic phase images, and (e) are the z-phase. The phase images are masked so that only phases inside the phantom are shown.

(a) In-plane motion.

(b) Through-plane (z) motion.**Figure 6.**

The comparison of average in-plane and through-plane motion computed from zHARP with standard CSPAMM tagging.

Figure 7.

The 3D displacement maps of two sagittal slices close to the middle of the tongue at different time during speech. The mesh grid shows the in-plane 2D displacement, and the color of the deformed mesh shows the through-plane displacement.

Figure 8.

(a) The locations of the 5 axial slices in the tongue. (b)–(d) the path lines of points inside the tongue in axial slice 2, 3, and 4, with lip toward the right.

Figure 9.

Illustration of the 3D motion of the tongue on two axial images using zHARP. The tongue regions in the two axial images are shown in meshes. The background image is a sagittal CSPAMM image, with lip to the right. The meshes are color coded with the distance to the sagittal image plane.

Figure 10.

The strain tensor components on one axial slice at different time frames. The tongue tip is pointing upward. x goes from top to bottom, y goes from left to right, and z is in the through plane direction.