

HHS Public Access

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2017 October 27.

Published in final edited form as: *Proc SPIE Int Soc Opt Eng.* 2017 ; 10133: . doi:10.1117/12.2254363.

Author manuscript

A Four-dimensional Motion Field Atlas of the Tongue from Tagged and Cine Magnetic Resonance Imaging

Fangxu Xing^a, Jerry L. Prince^b, Maureen Stone^c, Van J. Wedeen^a, Georges El Fakhri^a, and Jonghye Woo^{*,a}

^aDept. Radiology, Massachusetts General Hospital, Boston, MA, US 02114

^bDept. Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, US 21218

^cDept. Neural and Pain Sciences, University of Maryland School of Dentistry, Baltimore, MD, US 21201

Abstract

Representation of human tongue motion using three-dimensional vector fields over time can be used to better understand tongue function during speech, swallowing, and other lingual behaviors. To characterize the inter-subject variability of the tongue's shape and motion of a population carrying out one of these functions it is desirable to build a statistical model of the fourdimensional (4D) tongue. In this paper, we propose a method to construct a spatio-temporal atlas of tongue motion using magnetic resonance (MR) images acquired from fourteen healthy human subjects. First, cine MR images revealing the anatomical features of the tongue are used to construct a 4D intensity image atlas. Second, tagged MR images acquired to capture internal motion are used to compute a dense motion field at each time frame using a phase-based motion tracking method. Third, motion fields from each subject are pulled back to the cine atlas space using the deformation fields computed during the cine atlas construction. Finally, a spatiotemporal motion field atlas is created to show a sequence of mean motion fields and their intersubject variation. The quality of the atlas was evaluated by deforming cine images in the atlas space. Comparison between deformed and original cine images showed high correspondence. The proposed method provides a quantitative representation to observe the commonality and variability of the tongue motion field for the first time, and shows potential in evaluation of common properties such as strains and other tensors based on motion fields.

Keywords

Speech; tongue; atlas; motion; dynamic MRI; tagged; spatio-temporal; statistical

1. INTRODUCTION

The study of internal motion of the human tongue has been an important topic to oral surgeons, lingual scientists, and speech pathologists. Understanding the tongue motion of

^{*} jwoo@mgh.harvard.edu; Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital, Boston, MA, USA 02114.

Xing et al.

healthy human subjects is an essential first step toward helping post-glossectomy, speech disorder, and sleep-apnea patients in ongoing research. It is a challenging task, however, to acquire the motion data in the form of motion fields because the tongue is capable of making highly deformable and complex motions in short periods of time. Further, when studying multiple subjects, a statistical atlas of the three-dimensional (3D) tongue over time is essential to characterize the inter-subject commonality and variability of a given population.

Magnetic resonance (MR) imaging has been used in previous studies to capture both the tongue's anatomical structure and internal motion^[1]. Cine MR images are taken at multiple slice locations covering the tongue and at specified time instants during repeated pronunciations of a designed utterance. At each time frame, the slices can be combined using super-resolution techniques to form a dense 3D volume^[2], yielding a four-dimensional (4D) time sequence reflecting the change of anatomy. Meanwhile, a tagged MR sequence is collected at the same slice locations and same time frames during additional repetitions of the utterance, and these images can be used to calculate the internal motion of the tongue from the time-varying tag patterns deforming together with the tongue tissue^[3].

Most previous works have focused on the calculation and analysis of individual subject's 4D tongue motion from tagged and cine MR images. Principal component analysis (PCA) has been performed on a collection of subjects^[4]. However, the reported motion was averaged within local tissue regions due to the limitations of PCA. As a result, statistical properties of the entire motion field was not evaluated. On the other hand, there have been methods proposed to construct a 4D intensity atlas (scalar atlas) of the tongue from cine MR images^[5], from which a collected anatomical change can be statistically evaluated. However, an atlas of 4D motion fields (vector atlas) has never been proposed due to the complexity of motion estimation combined with high-dimensional atlas construction.

In this paper, we propose a method to build a 4D atlas of the tongue's motion field using tagged and cine MR images collected from fourteen healthy human subjects. The method is established on the basis of an existing cine atlas and a set of accurate 4D tongue motion estimates from each subject. Individual subject's motion fields are deformed into the cine atlas space using the deformation field obtained during the construction of the cine atlas, and the motion field atlas is estimated based on these deformations. We evaluate the quality of the motion field atlas by using it to propagate the cine atlas through time and comparing the propagation result with the original cine atlas. The motion field atlas provides a visualization of a consensus tongue motion from a number of subjects for the first time, and is a quantitative representation of the commonality and variability of the tongue's motion field.

2. METHODS

In data acquisition, each of the fourteen participants was asked to pronounce a controlled utterance "a geese" in repeated speech cycles while cine (Figure 1(a)) and tagged (Figures 1(b) and 1(c)) MR images were acquired. The tongue motion was captured in one second into 26 time frames, where the tongue is expected to start from a speech-ready position at/ə/, perform an upward motion to /g/, back into the /i:/, and end with a forward motion of /s/.

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2017 October 27.

Xing et al.

The deformation of the tongue in the speech task was recorded in the form of deformed tags. As shown in Figures 1(b) and 1(c), in the sagittal view, horizontal tags were used to record motion components in the superior-inferior direction, while vertical tags were used to record motion components in the anterior-posterior direction. Similarly, on the acquired axial slices, vertical tags were applied to record the remaining left-right motion component. At each time frame, tagged images from all three direction must be processed and combined simultaneously to obtain a desired 3D motion estimate.

In a traditional two-dimensional (2D) scenario, it has been a common practice that tagged images are processed by the harmonic phase (HARP) algorithm^[6], since the use of phase tracking was justified to be more accurate than intensity-based tracking methods in the condition of tag fading^[7]. By processing the filtered phase images using HARP, 2D in-plane motion fields can be computed at every time frame. Similarly, in this application, since the desired motion estimate is a dense 3D motion field, we apply the phase vector incompressible registration algorithm (PVIRA)^[8] to process tagged images. Similar to HARP but established in an image registration framework, PVIRA is also a phase tracking method that computes a 3D motion field by using harmonic phase volumes extracted from interpolated tagged slices. The result is a sequence of dense 3D motion fields $u_{s,t}(x)$ for each subject s at each time t (Figure 1(d)). One important property of PVIRA is that the method respects the tongue's physical incompressibility by incorporating a divergence-free velocity field constraint, yielding a motion estimate that is guaranteed to be incompressible. Since each motion field can also be considered as a deformation between time t and its first time frame, we use symbol ϕ to denote the deformation as $\phi_s(x) = x + u_s(x)$, where each individual subject is labeled by $s(1 \ s \ N, 1 \ t \ 26)$ and N = 14 is the total number of subjects.

Meanwhile, for each subject, cine slices from axial, sagittal, and coronal directions are combined into a dense 3D super-resolution volume at each time frame^[2]. The results from all fourteen subjects are used to construct a 4D intensity atlas^[5]. In this process, the deformations $\Psi_{s,t}$ from each individual subject *s* at time frame *t* to the atlas space at time frame *t* are obtained.

In order to build a motion field atlas, all subjects' motion fields need to be deformed into the atlas space. Since cine and tagged images are aligned both spatially and temporally, we directly use the cine deformations $\Psi_{s,t}$ to find $\phi_{s,t}$ in the atlas space^[9], i.e.,

$$\phi'_{s,t} = \psi_{s,t} \circ \phi_{s,t} \circ \psi_{s,t}^{-1}.$$
 (1)

Note that $\phi'_{s,t}(x) = x + u'_{s,t}(x)$, which is also computed in the form of motion fields $u'_{s,t}(x)$. As a result, all motion fields $u'_{s,t}(x)$ that have been deformed into the atlas space are averaged to construct the mean of the motion field atlas at time frame *t* by

$$\tilde{\boldsymbol{u}}_{t}(\boldsymbol{x}) = \frac{1}{N} \sum_{s} \boldsymbol{u}_{s,t}^{'}(\boldsymbol{x}).$$
(2)

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2017 October 27.

The inter-subject variation can be obtained by performing a PCA on all motion fields $u'_{s,t}(x)$.

3. RESULTS

The motion field atlas constructed from all fourteen subjects has 26 time frames, each visualized with a 3D motion field with respect to the first time frame. In Figure 2, four critical time frames are shown where the positions of /ə/, /g/, /i:/, and /s/ are reached. Since the motion fields are 3D, the figures are sagittal views from the left side of the tongue with multiple layers of vectors. We used color-coded cones to represent the direction of motion, where anterior-posterior is colored green, inferior-superior is colored blue, and left-right is colored red. Few red motions are observed in the whole tongue because a healthy subject's tongue typically moves symmetrically during speech and all fourteen participants are healthy volunteers. Visually, in the pre-speech neutral position $\frac{1}{2}$, the whole tongue show lesser amount of motion and its general trend goes down from the first time frame to move into a speech-ready position, preparing for the following deformations. At /g/, the tongue shows mostly upward motion in blue and its shape becomes vertically elongated to touch the palate. Then the tongue moves to an intermediate position with /i:/, and ends with a forward motion to pronounce /s/ shown mostly in green in the tongue tip. As qualitatively assessed, the motion field atlas shows a statistical global motion expected from the designed utterance "a geese".

In order to quantitatively evaluate the motion field atlas, we utilized the fact that cine and tagged MR images are aligned both spatially and temporally. This in turn indicates a theoretical spatial and temporal alignment between the cine intensity atlas and the motion field atlas. Using this assumption, from computed vector fields $\tilde{a}_{h}(x)$, we automatically obtained the deformation from the first time frame of the cine intensity atlas to its remaining 25 time frames by $\tilde{\phi}_{t}(x) = x + \tilde{u}_{t}(x)$. Therefore, we performed deformations $\tilde{\phi}_{t}(x)$ from the first cine intensity atlas to all of the 25 following time frames, and compared the deformation results (bottom row of Figure 3) with the original cine atlas at corresponding time frames (top row of Figure 3). At each time frame, we subtracted the two cine volumes and box-plotted their intensity difference on the tongue region. As expected, time frame 1 has zero difference. All the rest time frames have a median difference is 0.1017±0.1036 (intensity normalized to 1), indicating a good correspondence between the original cine atlas and the deformed cine volumes.

Finally, we performed a PCA on the motion field atlas to reveal its inter-subject variations. The standard procedure of PCA is to use an orthogonal transformation to convert all motion field estimates from all subjects into a set of linearly uncorrelated principal components (PCs). The degree of freedom is determined by the original dataset. In our case, we treated each time frame independently for an individual PCA, where all fourteen subjects yielded thirteen PCs. Since the first PC accounts for the largest variability in the dataset and each succeeding PC accounts for the remaining variability in a decreasing order, we studied the first two PCs in detail and visualized them in Figure 4. Three critical time frames /g/, /i:/, and /s/ are visualized, where the mean motions are shown in the middle column. Centered at

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2017 October 27.

the mean motion, variations along the positive and negative directions of the first PC and the second PC are shown in the other columns, respectively. All these time frames showed a similar trend of variation. For the first PC that represents the most extensive change among all subjects, the difference is at the tongue tip, where different subjects vary at the amount of moving forward and down (fourth column) against moving back and up (second column). This observation agrees with the muscle orientation and the anatomy of the tongue, where the genioglossus muscle forms a fan shape that roots at the mandible and expands back and up. Therefore, activities of genioglossus deforms the tongue along its fiber direction (from back up to front down). For the second PC, the difference in the whole tongue body (especially the back part of the tongue) is also accounted for besides the difference at the tongue tip. The motion variation shows a major trend in forward-backward at the tongue tip and up-down at the tongue back. Since the weight of the second PC dropped by 67% comparing to the first PC, these variations are not considered major. Further, the weight of the third PC dropped by 97% from the first PC and its contribution along with the remaining PCs was considered trivial.

4. Discussion

Dynamic atlas construction from MR images relies heavily on the acquisition and use of cine images. Intensity atlases are relatively easy to achieve, and have been used in the diagnosis and specification of anatomy-related problems. However, the construction of a motion field atlas is always a complex problem, because it typically involves a combination of multiple post-processing steps such as image registration, spatial and temporal alignment, and efficient multi-subject motion estimation. In this application, we have constructed the first motion field atlas of the tongue during speech. The proposed technique requires a spatio-temporal alignment of the cine images and tagged images during acquisition, and motion information acquired from all three orthogonal directions. Besides, a cine intensity atlas is also needed beforehand to provide deformations between each subject and the atlas space. All of these prerequisite steps are essential for the final result.

Another important problem to be addressed in the future is to achieve time alignment between subjects. Since all subjects speak at a different rates, their critical time frames do not happen at the same instants during the 26 time frames. In this practice, since the critical time instants were already aligned during the construction of the cine atlas, and we assumed cine and tagged images were aligned during acquisition, this alignment information was directly used in the construction of the motion field atlas. However, an independent time alignment directly from tagged imaging itself would be preferred to increase the result's data-adaptiveness. Besides, more subjects are needed in the future to increase the atlas' statistical significance. However, as preliminary result, the atlas has demonstrated the effectiveness of the workflow, and can be adapted to cardiac imaging as well. Also, internal strain tensors as a collective information can be computed based on this statistical motion field to further provide useful information on the study of tissue deformations.

5. CONCLUSION

We have described a method for building a spatio-temporal atlas from the motion field of the tongue during speech. The method combined both sources of information acquired from tagged and cine MR images, and used existing information from cine atlas construction methods and motion estimation methods. A principal component analysis was used to reveal inter-subject variation. Preliminary validations using both visual assessment and deformation of the cine intensity atlas was carried out, demonstrating a good quality of the motion field atlas both qualitatively and quantitatively.

Acknowledgments

This work was supported by NIH/NIDCD grants R00DC012575 and R01DC014717.

References

- Parthasarathy V, Prince JL, Stone M, Murano EZ, NessAiver M. Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. The Journal of the Acoustical Society of America. 2007; 121(1):491–504. [PubMed: 17297803]
- Woo J, Murano EZ, Stone M, Prince JL. Reconstruction of high-resolution tongue volumes from MRI. Biomedical Engineering, IEEE Transactions on. 2012; 59(12):3511–3524.
- Osman NF, Kerwin WS, McVeigh ER, Prince JL. Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. Magnetic resonance in medicine. 1999; 42(6):1048. [PubMed: 10571926]
- Xing F, Woo J, Lee J, Murano EZ, Stone M, Prince JL. Analysis of 3-D tongue motion from tagged and cine magnetic resonance images. Journal of Speech, Language, and Hearing Research. 2016; 59(3):468–479.
- Woo J, Xing F, Lee J, Stone M, Prince JL. A spatio-temporal atlas and statistical model of the tongue during speech from cine-MRI. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. 2016:1–12.
- Osman NF, McVeigh ER, Prince JL. Imaging heart motion using harmonic phase MRI. Medical Imaging, IEEE Transactions on. 2000; 19(3):186–202.
- Prince JL, McVeigh ER. Motion estimation from tagged MR image sequences. IEEE Transactions on Medical Imaging. 1992; 11(2):238–249. [PubMed: 18218378]
- Xing F, Woo J, Gomez AD, Pham DL, Bayly PV, Stone M, Prince JL. Incompressible Phase Registration for Motion Estimation from Tagged Magnetic Resonance Images. Proc International Workshop on Reconstruction and Analysis of Moving Body Organs, MICCAI, LNCS. 2016; 10129
- Ehrhardt J, Werner R, Schmidt-Richberg A, Handels H. Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. IEEE transactions on medical imaging. 2011; 30(2): 251–265. [PubMed: 20876013]

Author Manuscript



Figure 1.

(a) A mid-sagittal cine MR image slice. (b)(c) Mid-sagittal slices of horizontally and vertically tagged MR images. (d) Sagittal view of the 3D motion field estimated from tagged images at one time frame. Note: motion fields are visualized by cones using the diffusion imaging color coding scheme (green for anterior-posterior, blue for inferior-superior, and red for left-right).

Xing et al.



Figure 2.

The motion field atlas constructed from fourteen subjects pronouncing the utterance "a geese". Four critical time frames from 26 total time frames are shown, representing (a) a prespeech neutral position, (b) an upward motion, (c) an intermediate motion between /g/ and /s/, and (d) a forward motion. Note that each image is a sagittal view from the left side of the tongue.



Figure 3.

Deformation of the first cine atlas frame to all following frames using the motion field atlas. (a)(b)(c)(d) Original cine atlas at critical time frames. (f)(g)(h) First cine atlas time frame deformed to critical time frames. (e) Intensity difference in the tongue region between original and deformed images at all 26 time frames.



Figure 4.

Principal component analysis of the motion field of all subjects at three critical time frames. The variation of the first and second principal components (PCs) around the mean motion are shown in columns.