Three-Dimensional Muscle Motion Reconstruction Using Tagged MR Images

by

Xiaofeng Liu

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

October, 2009

Abstract

For clinical and scientific studies, it is important to understand the internal muscle motion of the tongue during speech and the heart in its beating cycle, which are made entirely of soft tissue, mostly muscle. Magnetic resonance (MR) tagging places non-invasive and temporary markers (tags) inside the soft tissues in a pre-specified pattern, yielding images that carry information about motion in the tagging direction. These images can be processing using harmonic phase (HARP) method to compute the in-plane motion. The dissertation studies the three-dimensional (3D) muscle motion using MR tagging with a focus on tongue imaging, and addresses the technical challenges in both 2D and 3D motion estimation.

In the dissertation, we developed HARP tracking refinement methods to reliably and automatically track the whole tissue from tagged MRI even when traditional HARP tracking fails. We measured 3D tongue motion during speech by re-implementing and optimizing the zHARP imaging sequence, and using a specialized MR triggering and vocal repetition method. We developed a thin plate spline based 3D tongue motion tracking method using tagged MR images by extending the 3D-

HARP method for cardiac motion tracking. We developed a method to reconstruct a 3D, dense, incompressible deformation field from tagged MR images based on the divergence-free vector spline with incomplete data samples, and applied it to both tongue and cardiac motion reconstruction. Finally, we performed preliminary studies of the internal tongue motion pattern and muscle mechanisms of glossectomy patient and compared them with normal speakers.

Primary Reader: Dr. Jerry L. Prince

Secondary Readers: Dr. Gregory D. Hager and Dr. Maureen Stone

Acknowledgements

This dissertation would not be possible without the great support and help from many people. I would like to thank my research advisor, Dr. Jerry L. Prince, for his guidance and support. Over the past five years, he directed me to this field, guided me with his insights and passion, and taught me how to be a good researcher in all aspects. It is my great fortune to have his supervision during my graduate study.

I would like to thank Dr. Maureen Stone for her financial support and her guidance on my research. I would also like to thank Dr. Maureen Stone and Dr. Gregory Hager for patiently reading my thesis and serving in my thesis committee. I thank Dr. Emi Murano for her friendship and helpful discussions. I would like to thank Dr. Rao Gullapalli and Ms. Jiachen Zhuo for their collaborative work and valuable help in my research.

I am grateful to all the past and present members of the Image Analysis and Commnunications Lab. Dr. Vijay Parthasarathy gave me many useful advices during the early stage of my PhD study. Dr. Xiaodong Tao was my first mentor in the lab. I thank the HARP group members: Dr. Khaled Abd-Elmoniem, Harsh Agarwal, and

Sahar Soleimanifird, for many useful and helpful discussions and also being my good friends. I want to thank Aaron Carass and Laura Libertini for their supports on technical and other issues.

I want to thank my parents Shuxin Liu and Lianmei Wang for everything they have done for me. I thank my wife Dr. Kun Lu for her endless patience, support, and confidence on me, without which this thesis could not be possible. I dedicate this thesis to her. Finally, I thank my children, Rachel and Luke, who bring so much joyful and wonderful time to me, and also provide me great momentum to finish this work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Muscle motion directly contributes to the movement of human body, and also causes the movement of certain internal organs. This is especially true for the organs that are almost completely made of muscles, i.e., the heart and the tongue, whose functions are determined by the underlying muscle motion. Therefore a good knowledge of the muscle motion can not only benefit scientific studies of the organ functions, but also directly contribute to early diagnosis of certain diseases and better surgical planning.

The tongue is crucial for speaking, swallowing, and breathing. It contributes to speech by shaping the vocal tract through tongue shape changes. Since there is no bone in the tongue, the speech and the tongue shape change are produced

through tongue muscle activation. A better understanding of the tongue function requires knowledges of its internal deformation patterns and their relationship to the underlying muscle activity. In addition, oral cancers have the 7^{th} highest incidence in the United States. Among them tongue cancer incidence has shown a recent five to six-fold increase in young adults aging 20 to 44 years and a two-fold increase in older adults [1]. Although tongue cancer is not considered fatal [2], it may cause significant problems in speech, mastication, and swallowing, and affect the patient's life quality. Glossectomy is often performed to treat tongue cancer by removing the cancerous tissue in the tongue, followed by primary surgical closure or reconstruction with flaps or skin grafts. However, it is not very well understood how the patients adapt to the surgery or which types of surgical reconstruction are better for the patients to retain tongue functions, especially speech. To help answer these questions, we need to better understand the motion patterns of glossectomy patients and compare them with those of normal speakers.

The human heart is responsible for pumping blood through the circulatory system. Its correct functioning is essential to the human body. Cardiovascular disease is the number one cause of death and disability in the United State and most European countries. It is estimated that 80,000,000 Americans adults (approximately one in three) have at least one cardiovascular diseases, and every 37 seconds an American dies of cardiovascular disease [3]. About 5.3 million Americans are affected by heart failure. The treatment of these diseases costs more than $448.5 billion a year in

America in 2008. The left ventricular function of the heart is known to be a sensitive indicator of cardiovascular diseases, for example ischemia and infarction [4–7]. Therefore regional ventricular function analysis can help characterize and detect certain heart diseases [8], and it is often performed by measuring the motion of the heart.

Among medical imaging techniques, magnetic resonance imaging (MRI) has shown advantages for imaging muscle motion because it has no radiation, superior temporal and contrast resolution in soft tissues, and adequate SNR. MRI has been capable of capturing information about motion within the interior of muscles since the development of tagging. MR tagging was originally developed on cardiac imaging and has been applied to tongue motion imaging because of the similar tissue properties between the heart and the tongue. Tagged MRI contains detailed motion information in the image plane, and can be post processed to achieve useful measurements of muscle motion for both scientific and clinic purposes.

This thesis focuses on the study of three-dimensional (3D) muscle motion of the tongue using MR tagging. We address the technical challenges of both 2D and 3D tongue motion estimation, including reliable 2D tracking, fast imaging of 3D motion, 3D tracking, and dense 3D motion field reconstruction. We perform preliminary studies on the motion patterns of the tongue during speech for both normal speakers and glossectomy patients. We also study cardiac imaging because the tagged MR images of the heart and the tongue share many common properties; as well, historically most

MRI motion measurement techniques were developed in the heart and subsequently adapted for the tongue.

1.2 Anatomy of the Tongue and the Heart

The tongue has a very complex muscular architecture, and is incompressible (volume-preserving). It consists of eight muscle pairs, four extrinsic muscles (hyoglossus, styloglossus, palatoglossus, and genioglossus) and four intrinsic muscles (verticalis, transversalis, and superior and inferior longitudinalis) with a unique and complex muscle fiber organization. The intrinsic muscles lie completely inside the tongue, and function to change the tongue shape by lengthening and shortening, flattening and rounding its surface, and so on. The extrinsic muscles insert into the tongue and attach the tongue to outside structures, and function to reposition the tongue and make it protrude, retract, depress, and elevate. The intrinsic and extrinsic muscles work together to create the complex deformations for speech and swallowing. Fig. 1.1 shows the shape of the tongue and its muscle structure.

The human heart is made of cardiac muscles that are self-exciting and capable of continuous beating. It beats about 100,000 times per day. The human heart consists four chambers: left atrium, left ventricle (LV), right atrium, and right ventricle (RV) (see Fig. 1.2). The ventricles are separated from the atria by one-way valves which keep blood flowing through the heart in the correct direction. Each cardiac cycle can

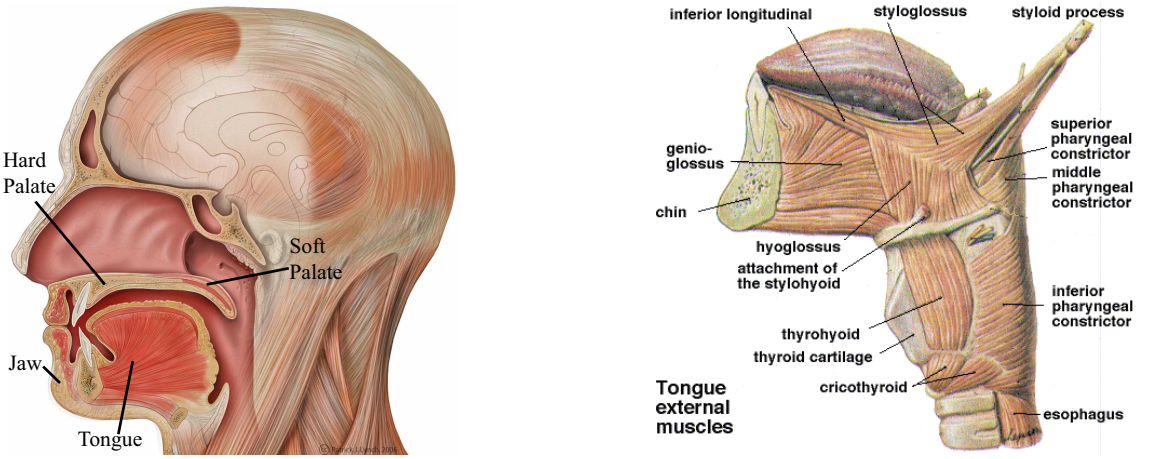


Figure 1.1: Left: The mid sagittal diagam of the tongue (from Patrick J. Lynch, Radiopaedia.org). Right: The muscle structure of the tongue (from http://www.yorku.ca/earmstro/).

be divided into two stages: systole and diastole. During systole, the muscles of both the left and right ventricles contract. The blood in the left ventricle is pumped out into the aorta and circulates through the entire body, while the blood in the right ventricle is sent to the lungs through the pulmonary artery. During diastole, the ventricles relax and are refilled with circulating blood. The blood pressure increases during systole, and decreases during diastole.

1.3 Imaging the Tongue and the Heart

MRI is considered the gold standard for quantitative assessment of cardiac structure and function [9, 10]. Regional changes in myocardial performance (contractile function) can be assessed using tagged MRI [11–13], phase contrast MRI [14], and strain rate MRI [15]. Although the heart and the tongue have different structures

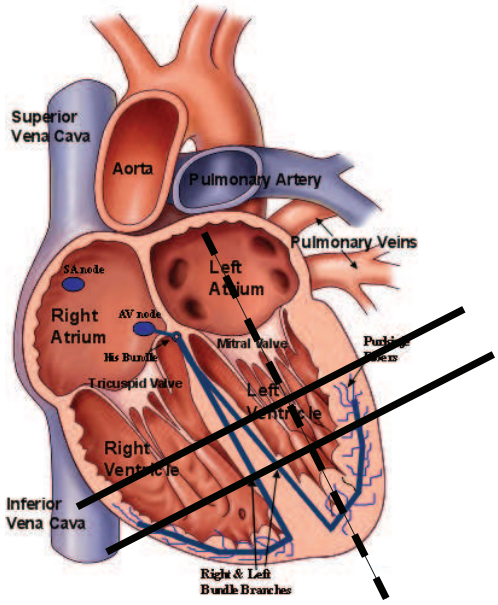
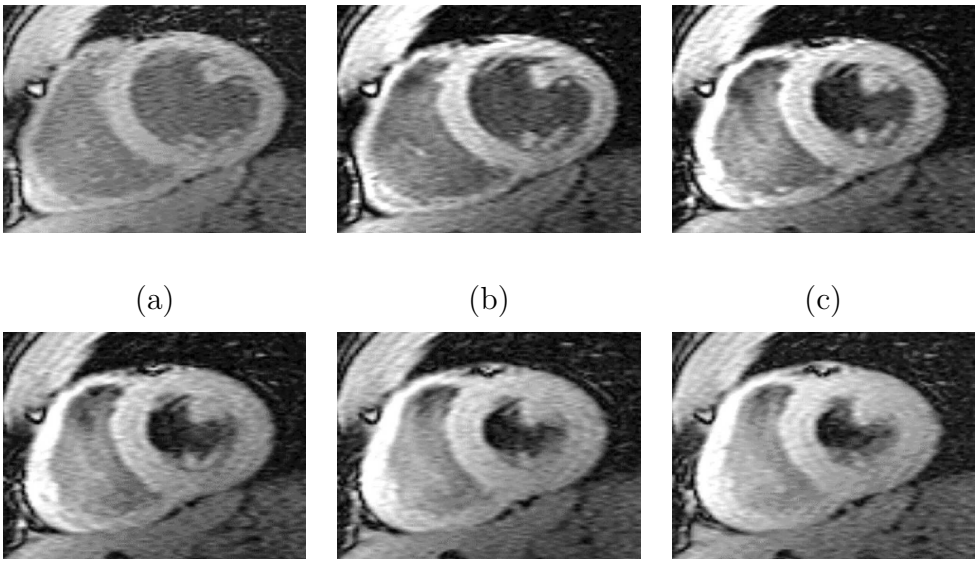


Figure 1.2: The anatomy of the heart. The long axis of the left ventricle is represented by the dashed line. The solid lines represent short axis image planes (from http://www.washingtonhra.com/2.html).

and different functions, they are both made of muscles and have similar contrast in MR images. Therefore many MR imaging methods that were first developed for one application can be carried over to the other.

In cardiac imaging, the image planes are defined relative to the orientation of the heart, and are oblique to the conventional sagittal, axial, and coronal planes. The long axis (LA) image planes go through the long axis, which is defined as the line that passes through the apical point of the left ventricle and the mitral valve orifice (see Fig. 1.2). LA images are usually acquired in radial orientations. The short axis (SA) image planes are perpendicular to the LA. Examples of cardiac images are shown in Fig. 1.3.



(a) (b) (c)

(d) (e) (f)

Figure 1.3: A sequence of cine cardiac MR images on a short axis image plane during systole in the order from (a) to (f).

To capture cardiac motion during a heartbeat, a temporal series of images are acquired and this is called cine MRI. Cine cardiac MRI is acquired in multiple heart beats using ECG gating. From cine MRI, one can measure the progression of heart motion by tracking the myocardial boundary, from which other useful quantities — myocardial wall thickness, ejection fraction, and so on — can be measured. However, since the myocardium appears homogeneous it does not provide information to quantify the internal muscle deformation in the myocardium, i.e., it cannot differentiate the motion variation between the epicardium and endocardium. This drawback is overcome by tagged MR imaging [11–13]. Tagged MRI places temporary planar markers (tags) within myocardium. The tags are usually placed at end-diastole, and

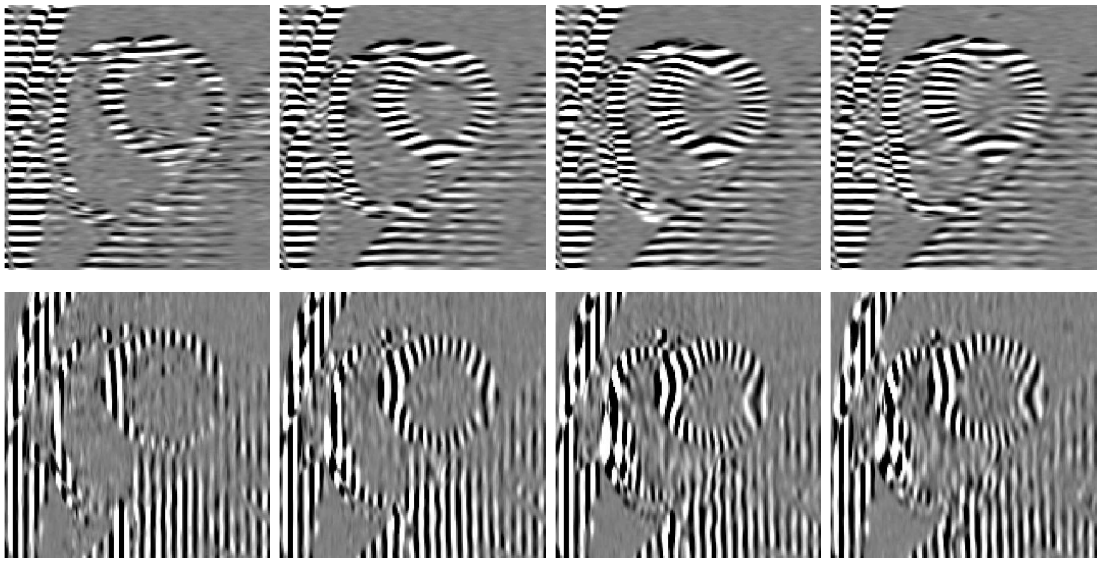


Figure 1.4: Tagged MR image sequence of a human heart on a short axis plane. The tags deform as the heart contracts during systole (shown from left to right). Top row: horizontally tagged images. Bottom row: vertically tagged images.

images are captured throughout the cardiac cycle. As the tags move with tissue, the internal deformation of the myocardium can be observed by tracking the tags. Fig. 1.4 shows examples of tagged MR images of a human heart on an SA plane. The detailed information about myocardial deformation makes it possible to compute many other regional function measurements, e.g., displacement, velocity, rotation, translation, elongation, strain, and twist.

In tongue imaging, multiple repetitions of a speech utterance are used (instead of multiple heartbeats in cardiac imaging) to generate a time-series of images showing its motion. Measuring the tongue dynamics is challenging because because (1) the tongue moves deep within the vocal tract, (2) tongue motion has a high number of degrees of freedom, and (3) tongue motion is rapid during speech and swallowing. In

speech research, MRI was first introduced to capture static images of steady state vowels [16] in anatomical studies [17, 18], disease examinations [19, 20], and building acoustic tube models by extracting the vocal tract [21–27]. Cine MRI has also been applied to capture the tongue motion during speech. The images are usually acquired over multiple repetitions of a speech utterance to achieve adequate spatial and temporal resolution [28–30], or using one repetition per slice with relatively low temporal and spatial resolution [31–34]. The images are usually collected in standard image orientations, i.e., sagittal, coronal or axial. As in cardiac imaging, cine MRI is valuable in the study of tongue surface motion, but not in measuring the muscle deformation of the tongue body.

The ability of tagged MRI to image internal tongue motion was first demonstrated by Niitsu et al. [35]. Later Napadow et al. [36, 37] applied tagged MRI to compute principal strains of tongue muscles during non-speech motions. These approaches applied the tags at the rest position, and an image was captured with deformed tags during or after the movement. Later this was improved by combining tagged MRI with cine MRI to create cine tagged MRI [38, 39]. These images can be processed similar to the cardiac tagged MR images to compute useful measurements of tongue motion.

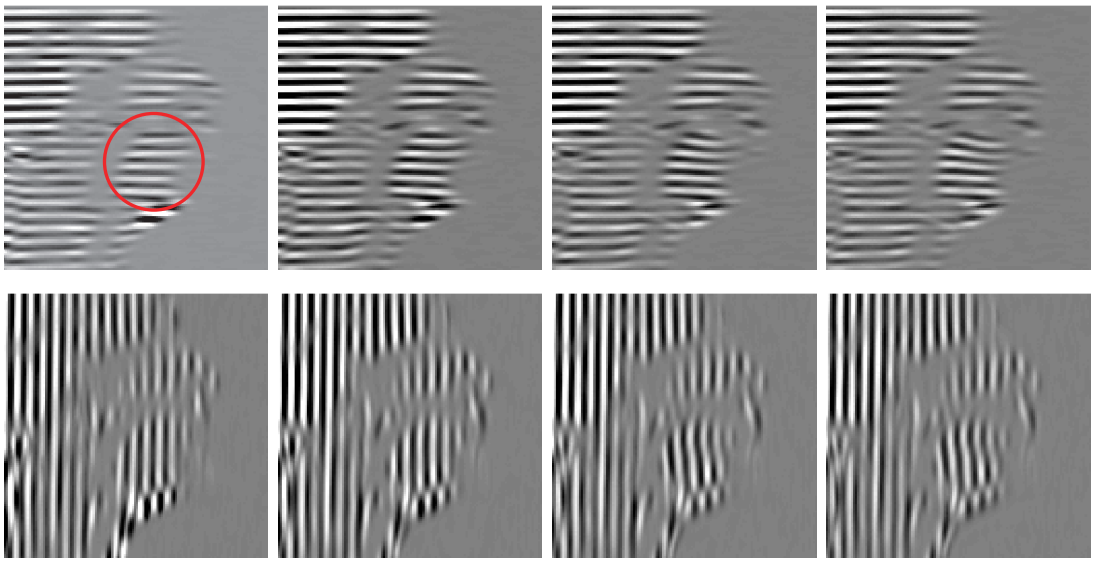


Figure 1.5: Tagged MR image sequence of a tongue on a sagittal plane. The images were acquired when the subject said "eeoo" repetitively, and are shown from left to right in order. The tongue region is circled in the top left image. Top row: horizontal tag. Bottom row: vertical tag.

1.4 3D Tagged MR Image Acquisition

Tagged MR images contain only motion information in the normal direction of the tag planes as illustrated in Fig. 1.6. Point $\mathbf{x}$ on the image plane moves from $\mathbf{p}$ at the reference time frame. From the planar image, it can only be determined that $\mathbf{x}$ originates from somewhere on the tag plane, but its location on the plane is unknown. Thus the full displacement vector $\mathbf{u}(\mathbf{x})$ cannot be determined. Instead only the component of $\mathbf{u}$ in the normal direction ($\mathbf{e}$) of the tag plane can be measured (shown as w in Fig. 1.6). In fact the components of in-plane motion can be computed by acquiring two images with different tag directions at the image plane, but the through-plane motion cannot be computed from this image plane alone.

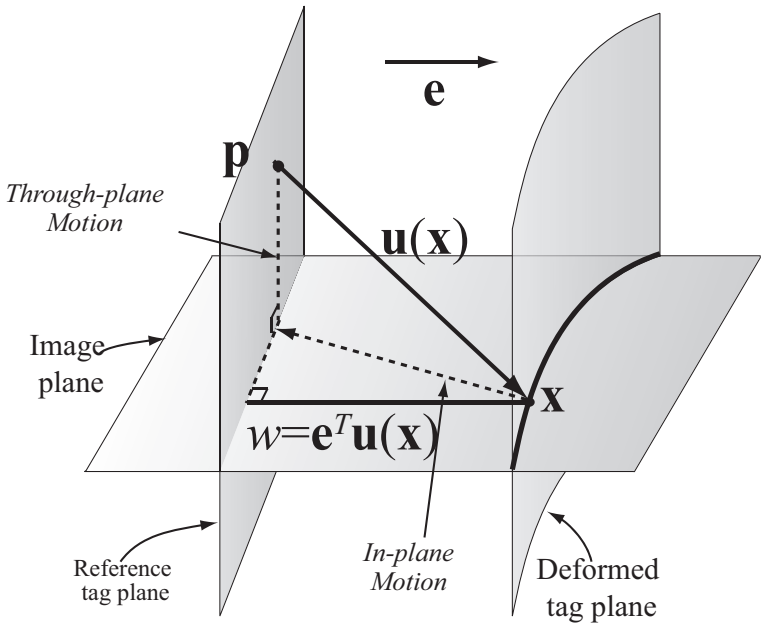


Figure 1.6: Displacement measurements from tagged images.

To measure the motion components in three dimensions, multiple images must be acquired in different image orientations with tags applied in different directions. In cardiac applications, tagged MR images are acquired in both short axis (SA) and long axis (LA) image planes. The relative locations of LA and SA images are illustrated in Fig. 1.7. On each SA plane, typically two images are acquired; one is with horizontal tags and contains motion information in the x-axis, and the other is acquired with vertical tags and contains motion information in the y-axis. The LA images are tagged in the horizontal direction only, and contain motion information in the z-axis (Fig. 1.7).

The tongue images are usually acquired in two orthogonal orientations, i.e., axial

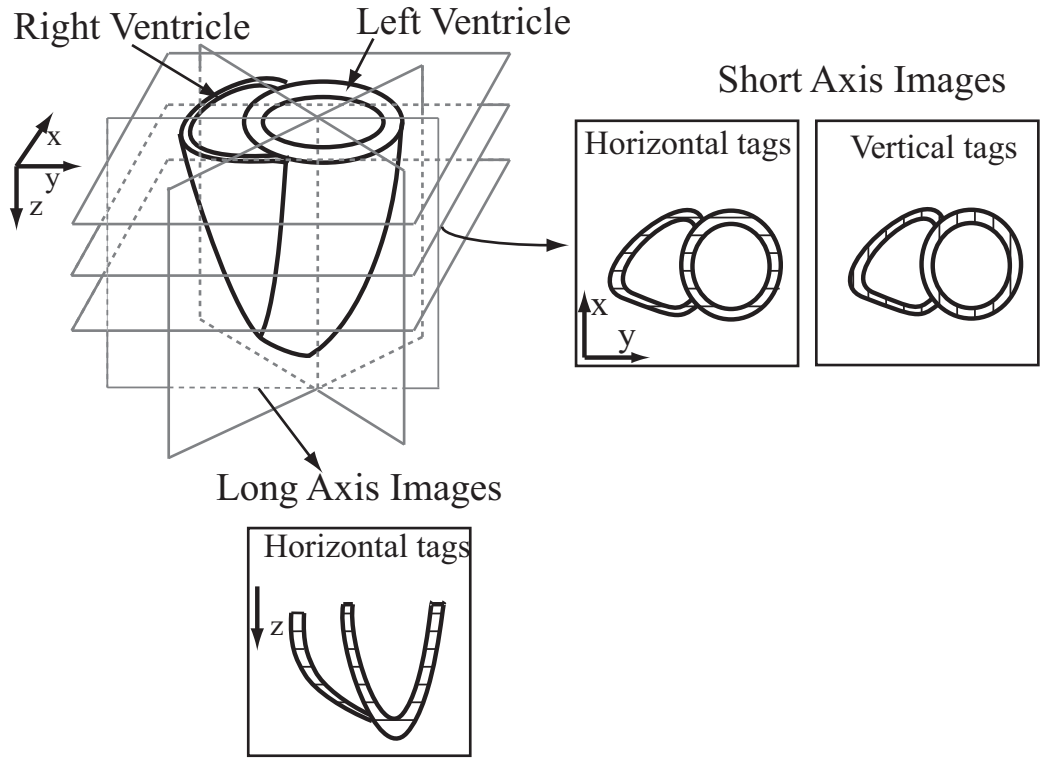


Figure 1.7: Cardiac tagged MR image acquisition for 3D motion.

and sagittal, axial and coronal, or sagittal and coronal orientations. In one orientation, two sets of images are acquired on multiple parallel planes with horizontal and vertical tags respectively. In the other orientation images are acquired with tags in the through-plane direction of the first image orientation. Fig. 1.8 shows an example configuration in which both horizontally and vertically tagged images are acquired on axial planes, and horizontally tagged images are acquired on sagittal planes.

The motion that can be measured directly from either of the cardiac or tongue acquisition protocols is incomplete and sparse. It is incomplete because only partial motion information is available for points on the images. We take the configuration

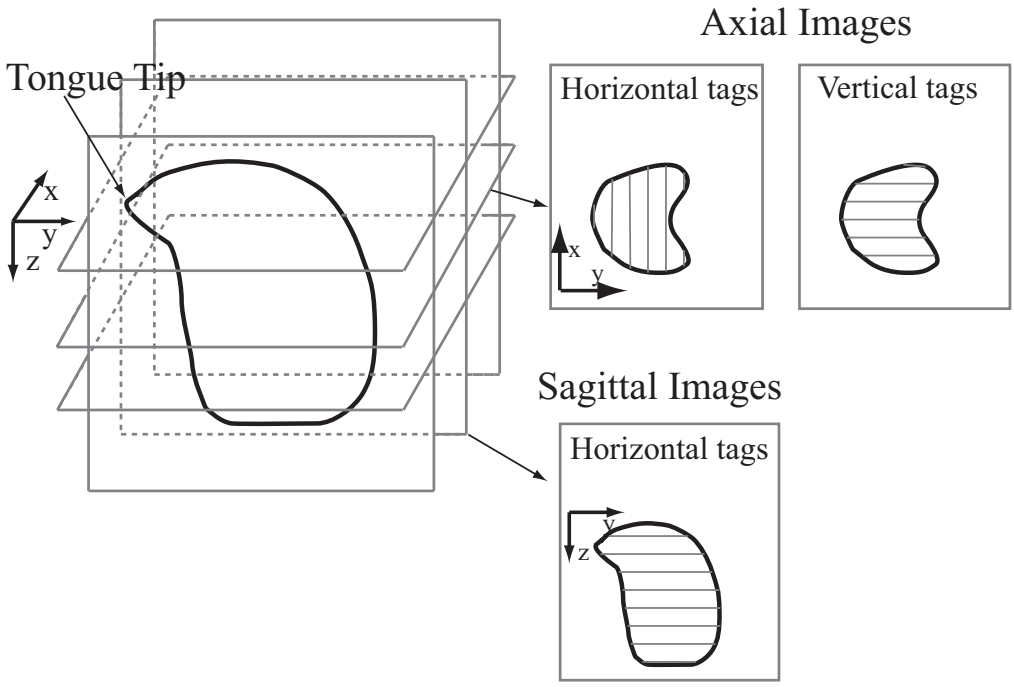


Figure 1.8: Tongue tagged MR image acquisition for 3D motion.

of tongue image acquisition in Fig. 1.8 as an example and show the collection of acquired images within a volume in Fig. 1.9. For points on the axial image planes we can only know the 2D motion projections in the x and y axes from the orthogonally tagged image pairs, and for points on the sagittal image planes we can only know the 1D motion projections in the z axis. The complete 3D motion is measured only for points on the intersection lines of the image planes (the red lines in Fig. 1.9). In addition, the imaged points (whether completely or partially measured) are sparsely distributed in space. Because of the limitation of imaging time, the image planes in each orientation are usually acquired far apart such that the slice separation is much larger than the in-plane image resolution. No motion measurements are directly

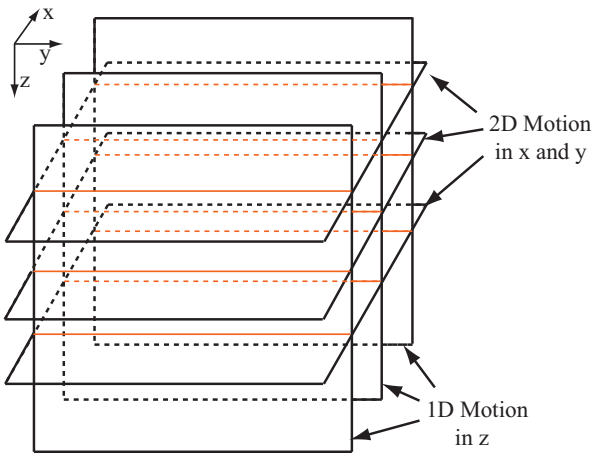


Figure 1.9: The motion measurements that can be directly computed from 3D tagged MR image acquisition. The 3D displacement vector can be directly computed only on the intersection lines of the images (marked in red in the figure). Only 1D or 2D projections of the actual 3D displacements can be computed from other imaged points.

acquired in the spaces between the slices.

For complete regional function analysis of the tongue and the heart, these sparse and incomplete motion measurements must be combined together to generate dense and complete 3D muscle motion.

1.5 Challenges in 3D Motion Analysis using Tagged MRI

The analysis of muscle motion using tagged MRI is challenging in several aspects. First, it is required to *reliably* compute in-plane motion from the planar images for 2D motion analysis, and also for 3D motion analysis since the 3D motion recon-

struction requires the in-plane motion be correctly computed from all the imagess. The harmonic phase (HARP) method [40–42] is advantageous over other tagged MRI processing methods because it can accurately track every pixel in the image and is not limited to tracking points on tag lines. However, HARP tracking suffers from mistracking because it implicitly assumes that tissue points do not move much from one time frame to the next. When this assumption is violated, HARP tracking will fail. Manual correction of mistracked points is an unpractical task especially in the research of tongue motion, because parts of the tongue move very fast in some utterances relative to the temporal resolution of the scan and are likely to be mistracked. Therefore an automatic and fast method for robust 2D tracking is required to improve subsequent 3D motion analysis.

The second challenge is to reconstruct dense 3D muscle motion from the sparse and incomplete motion measurements while preserving important mechanical properties of the muscles. Incompressibility is an important physical property of biological tissues including the tongue and cardiac muscles. However this property has been largely ignored by previous 3D motion reconstruction and interpolation methods. Existing methods that do take incompressibility into consideration assume that the deformation is small, which is not correct for large motion and may result in large error. The reconstruction of 3D, incompressible, dense motion fields from the tagged MR images remains a challenging problem.

Third, because of the importance of the heart, regional functional analysis of the

LV has been a hot topic for years and lots of innovative research and new imaging techniques have been developed, such as 3D-HARP, a 3D tracking method of points in the LV of the heart using HARP, and zHARP, a new imaging and image processing technique that can image 3D cardiac motion on one single image slice. Many of these techniques developed on the heart can be adapted to the motion analysis of the tongue. However, the adaption is often not straightforward because of the different tissue structures and motion patterns between the tongue and the heart.

The fourth challenge is to explore the tongue motion patterns and mechanism and relate them to certain tongue functions especially for glossectomy patients. Though it has been of interest for many years, the internal tongue motion pattern during speech is not fully understood. In order to better interpret clinical observations and to provide data that can help predict optimal surgical outcomes, we need a better understanding of the tongue motion patterns and the underlying mechanisms of tongue muscles in both glossectomy patients and normal speakers.

1.6 Thesis Contributions

The contributions of the thesis are summarized as follows.

- We developed the shortest path HARP refinement method to address the HARP mistracking problem by formulating a single source shortest path problem, and solving it using Dijkstra's algorithm. In this method, the image is represented as

a graph and the points are tracked by following the resolved optimal refinement paths. We also use synthetic phase images at the reference time and a two-step tracking procedure to further prevent the mistracking caused by through-plane motion. The method can reliably track every point inside the interested tissue, and are very fast. Compared to the original HARP refinement idea [40], our method is automatic, does not require a circular geometry or organized mesh of points defined on the region.

As a preliminary work we also developed the region growing HARP refinement method, in which the tissue points are tracked in a region growing process in an order based on local HARP phase smoothness. The method works well for tongue motion tracking, but it sometimes fails in cardiac motion tracking because the tracking order is not optimal.

- We measured the 3D tongue motion on a single slice using a fast imaging method called zHARP. We re-implemented zHARP with a gradient echo sequence and Cartesian sampling on Siemens scanners, and optimized the imaging parameters specifically for tongue imaging. To reduce motion artifacts, we also optimized the image acquisition by designing and developing a specialized MRI scanner triggering and vocal repetition method to better synchronize speech repetitions.

- We developed a method to reconstruct a dense representation of the 3D, incompressible displacement fields based on the sparse and incomplete motion

observations computed from tagged MR images. At each time frame, the images are first processed using HARP and HARP refinement to measure the partial motion information at each pixel from each time frame back to the reference time. The approach then applies a smoothing, divergence-free vector spline to interpolate velocity fields at intermediate discrete times such that the collection of velocity fields integrate over time to match the observed displacement components. Through this process, the method yields a dense estimation of the displacement field that matches the incomplete and sparse observations and also corresponds to an incompressible motion.

We also developed a method to track 3D tongue motion on a sparse rectangular mesh using a thin-plate spline(TPS). The method extends the 3D-HARP method [43] from the heart to the tongue. It iteratively tracks in-plane motion on the intersection points of the mesh and the image planes using 2D-HARP followed by TPS interpolation to extend the 2D motion to the whole mesh.

- We performed preliminary studies of the tongue motion patterns of glossectomy patients and compare them with normal speakers. The first study applied principal component analysis to examine the statistical motion pattern of the midsagittal section of the tongue during elevation of the tongue body. By comparing patient speakers to a set of normal speakers, we were able to quantitatively characterize the motion differences between the normal and patient speakers. The second study analyzed the mechanical behavior of the inferior

longitudinal (IL) muscle during speech. We looked at the muscle deformation for a normal speaker, a patient with a partial glossectomy and a radial forearm free flap (RFFF), and examined whether their different tongue motion patterns could be explained by the changes in muscle mechanics.

1.7 Thesis Organization

The thesis is organized as follows.

Chapter 2 includes background knowledge on MR tagging, the HARP method, and strain calculation. In Chapter 3, we describe the region growing HARP refinement method to address the mistracking problem of 2D HARP, and show the results on tongue motion tracking. In Chapter 4, we describe the shortest path HARP refinement method, a method that further improves the HARP refinement results by finding the optimal paths that connect tissue points with the seed. In Chapter 5, we present a method for fast imaging and measurement of 3D tongue motion using zHARP. In Chapter 6, we describe a method for 3D tongue motion tracking using an iterative method based on thin plate spline interpolation. In Chapter 7, we present an approach to reconstruct a 3D, dense, incompressible displacement field using the incomplete and sparse sample data computed from tagged MR images, and apply the approach to both the heart and the tongue. In Chapter 8, we describe preliminary studies of the internal tongue motion patterns and muscle mechanics and compare

between glossectomy patients and normal speakers. Finally, Chapter 9 summarizes the conclusions and future work.

Chapter 2

Background

In this chapter we provide a brief overview of MR tagging and tagged MR image analysis. In Section 2.1, we first introduce the basics of MR tagging, including a mathematical description of image formation. Section 2.2 describes the HARP processing and tracking methods. In Section 2.3 we review existing 3D muscle motion reconstruction methods. Section 2.4 overviews the computation of strain, a common measurement for tissue functional analysis. Finally, Section 2.5 summarizes this chapter.

2.1 MR Tagging

Magnetic resonance (MR) imaging is capable of directly imaging motion of soft tissues. Traditional MR images carry information about motion only at the bound-

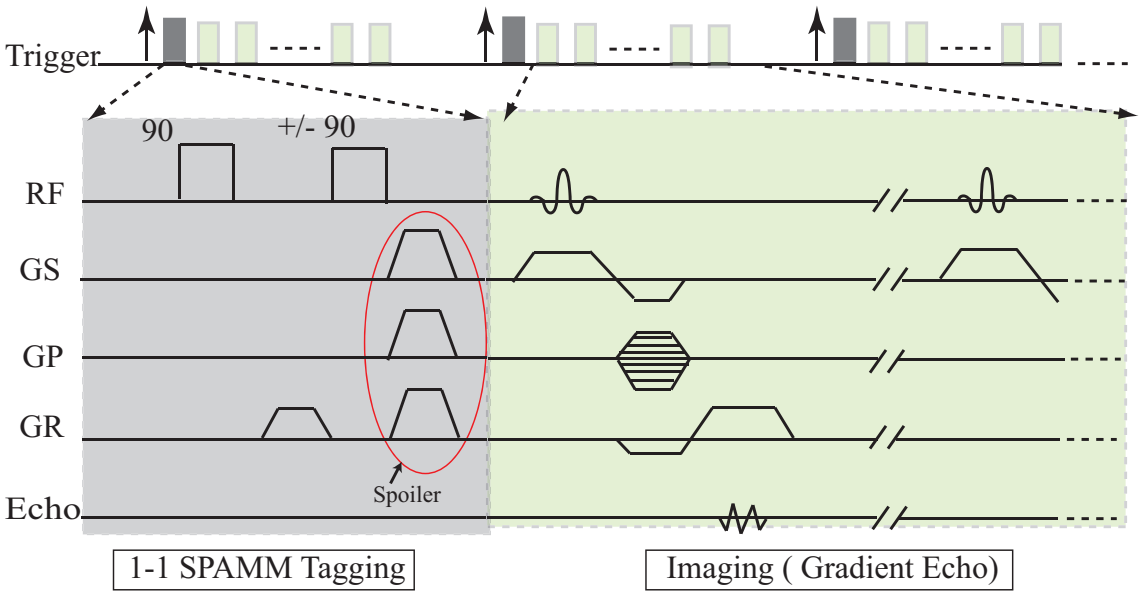


Figure 2.1: 1-1 SPAMM tagging pulse sequence.

aries of tissues because the image intensity of the interior of the tissues is largely homogeneous. MR tagging [12,13] places non-invasive and temporary markers (tags) inside the soft tissues in a pre-specified pattern, yielding images that carry information about motion within homogeneous tissues. This complements traditional anatomical images, and enables the detailed imaging of the motion of tissues such as the heart and the tongue throughout the time. Displacement, velocity, rotation, elongation, strain, and twist are just some of the quantities that can be computed from it.

MR tagging and spatial modulation of magnetization (SPAMM) have been important imaging protocols for visualization and quantification of motion and strain since their developments by Zerhouni et al. [11] in 1988 and Axel and Dougherty [12, 13] in 1989, especially in myocardial imaging.

As the most basic form of SPAMM, the 1-1 SPAMM approach generates a smoothly varying sinusoidal tag pattern in the tissue. A typical pulse sequence with SPAMM tagging is shown in Fig. 2.1. At each repetition, the tagging pulse is applied immediately after a trigger signal is detected, followed by the acquisition of a sequence of images. The trigger signal in cardiac imaging is the R-wave from the ECG signal, which happens at the end of diastole. The tagging pulse consists of two equal RF pulses with a modulating gradient pulse in between. To achieve the best tag contrast and tag persistence, the flip angle of the RF pulses is set to 90^o. After that, the transverse magnetization is spoiled using large gradients.

The acquired image can be understood as the multiplication of the underlying tissue anatomy with a two-dimensional sinusoidal function. The direction and the frequency of the sinusoid is determined by the modulating gradient. Higher order SPAMM tag patterns (for example 1-4-6-4-1) can be produced using a linear combination of several 1-1 SPAMM sequence, and can produce a thin and sharp tag pattern on the tissue that is good for visualization and tag line detection. Herein we consider only 1-1 SPAMM because this is optimal for use in HARP analysis. With a $[+90^o, +90^o]$ tagging pulse, the acquired image at time t after the tag is applied can be expressed as

$$A(\mathbf{x}, t) = M_0(\mathbf{x}, t)e^{-t/T_1} \cos(\phi(\mathbf{x}, t)) + M_0(\mathbf{x}, t)(1 - e^{-t/T_1}) \,, \tag{2.1}$$

where e^{-t/T_1} represents the T_1 decay of the magnetization, and $\phi(\mathbf{x}, t)$ is the phase of point $\mathbf{x}$ at time t. When applying a $[+90^o, -90^o]$ tagging pulse, the image can be

expressed as

$$B(\mathbf{x},t) = -M_0(\mathbf{x},t)e^{-t/T_1}\cos(\phi(\mathbf{x},t)) + M_0(\mathbf{x},t)(1-e^{-t/T_1})\,. \tag{2.2}$$

The tagging patterns in these two images differ by a phase shift of π radians. Examples of SPAMM image pairs are shown in Figs. 2.2(a) and (b) and Figs. 2.2(d) and (e). To improve the tag contrast, the image pair $A(\mathbf{x},t)$ and $B(\mathbf{x},t)$ are acquired separately at the same time of a repeated motion, and combined to yield a so-called complementary SPAMM (CSPAMM) [44] image

$$I_{\text{CSPAMM}} = A(\mathbf{x},t) - B(\mathbf{x},t) = 2M_0(\mathbf{x},t)e^{-t/T1}\cos(\phi(\mathbf{x},t))\,. \tag{2.3}$$

The tag contrast of the CSPAMM images is twice that of the SPAMM images. When we are only able to acquire the magnitude of the MR images, another method called MICSR (magnitude image CSPAMM reconstruction) [45] can be used to improve the tag contrast over plain SPAMM:

$$I_{\text{MICSR}} = |A(\mathbf{x},t)|^2 - |B(\mathbf{x},t)|^2 = 4M_0^2(1-e^{-t/T_1})e^{-t/T_1}\cos(\phi(\mathbf{x},t))\,. \tag{2.4}$$

Besides enhancing tag contrast, the CSPAMM and MICSR methods also remove the constant bias through subtraction and produce images with pure sinusoidal tags. Examples of MICSR images on the tongue are shown in Figs. 2.2(c) and (f).

The tagging phase $\phi(\mathbf{x},t)$ is a material property, and it does not change when a point moves in space. Immediately after the tag is applied and the object has not deformed, i.e., $t = 0$, the tag is a linear function of the spatial location. Let us denote

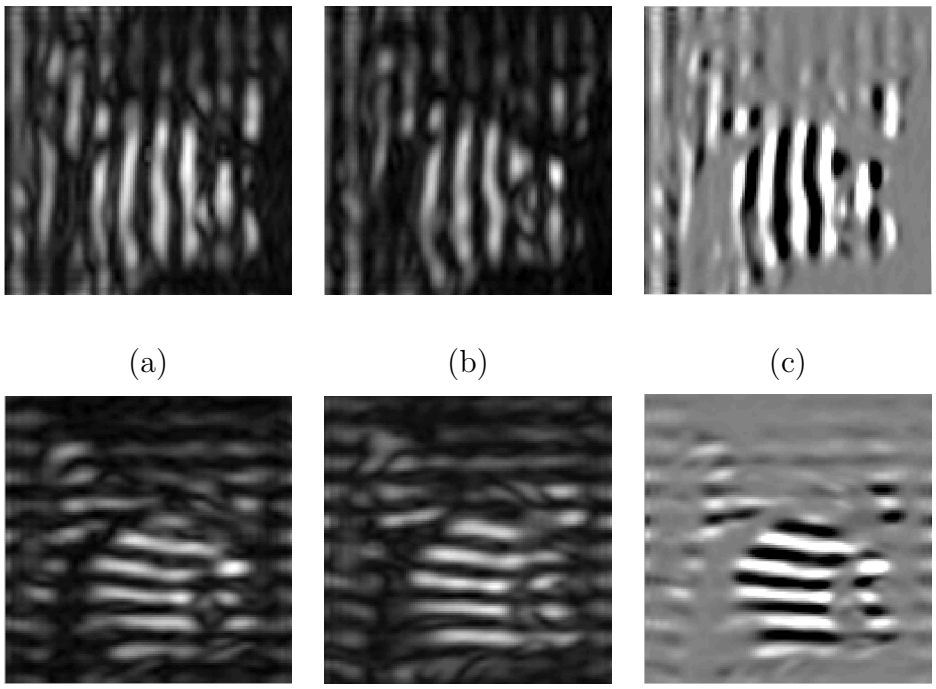


Figure 2.2: Tongue images (a) and (b) form a pair of vertically tagged SPAMM images with phase shifted tags, and (c) is the MICSR combination of images (a) and (b). Images (d) and (e) form a pair of horizontally tagged SPAMM images and (f) is the MICSR combination of images (d) and (e). (c) and (f) have been thresholded for better visualization of the tags.

a material point at time 0 as $\mathbf{X}$, and its location at time t as $\mathbf{x}(t) = \mathbf{X} + \mathbf{u}(\mathbf{x}, t)$.

Then we have

$$\phi(\mathbf{X}, 0) = k\mathbf{e} \cdot \mathbf{X}\,, \tag{2.5}$$

where $\mathbf{e}$ is a unit vector normal to the tag planes, and k is the tagging frequency. At later time after tissue deformation, the tagging phase retains, i.e.,

$$\phi(\mathbf{x}, t) = \phi(\mathbf{X}, 0) = k\mathbf{e} \cdot (\mathbf{x} - \mathbf{u}(\mathbf{x}, t))\,. \tag{2.6}$$

2.2 Harmonic Phase (HARP) Method

Existing methods for tagged MR image processing and motion tracking can be divided into two categories. The first category identifies and tracks either the intersections of the tag lines [46–48] or the whole tag lines [49–52] from the 2D images. The motion on other points within the image can be estimated from these sparse measurements using model-based or model-free interpolation [47–50, 53–60]. The methods in the second category include the harmonic phase (HARP) method [40–42] and the Gabor filter bank-based method [61]. These methods compute the wrapped tagging phases from the images, and track points based on the fact that the tagging phases of material points are constant. Since the tagging phases are available on every image pixel inside the tissue, a dense 2D displacement field is directly achieved without interpolation.

The HARP method has been successfully applied in both the heart [40] and the tongue [62], and has proven to be useful for both scientific and clinical applications. It has shown to be advantageous to other methods because it is fast and accurate, and therefore is used throughout this thesis for planar tagged MR image processing. The HARP method includes two components: HARP processing and HARP tracking. HARP processing computes the harmonic phase — the wrapped value of the tagging phase — images by applying bandpass filters in the Fourier domain, and HARP tracking computes the 2D motion based on the property that the harmonic phase value is retained when a point moves.

2.2.1 HARP Processing

The Fourier transforms of both CSPAMM and MICSR images have two harmonic peaks which result from the sinusoidal function (see Eqns. (2.3) and (2.4)). The Fourier transforms of SPAMM images have an additional spectral peak at the center that results from the second term in Eqns. (2.1) and (2.2). Figs. 2.3(a) and (b) show an example SPAMM image and its Fourier transform, respectively. To estimate the phase value $\phi(\mathbf{x}, t)$ from the tagged images, the HARP method applies a bandpass filter to extract just one of the harmonic peaks (see the circle in Fig. 2.3(b)). The resulting filtered complex image can be expressed as

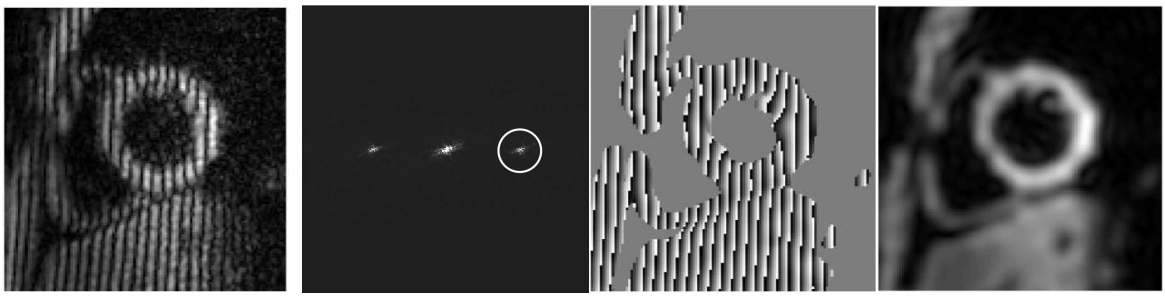
$$I(\mathbf{x}, t) = D(\mathbf{x}, t)e^{j\phi(\mathbf{x},t)}\,, \tag{2.7}$$

where $D(\mathbf{x}, t)$ is called the harmonic magnitude image, and $\phi_(\mathbf{x}, t)$ is the harmonic phase (HARP) image. The magnitude image reflects the tissue anatomy, and the HARP image contains the tissue motion information. Because the HARP phase must be computed using an arctangent operation, its value $\theta(\mathbf{x}, t)$ is the principal value of the true phase. It is therefore restricted to take on values in the interval $[-\pi, +\pi)$, and is related to the true phase by

$$\theta(\mathbf{x}, t) = W(\phi(\mathbf{x}, t))\,, \tag{2.8}$$

where $W(\cdot)$ is the wrapping function defined as:

$$W(\phi) = \mathrm{mod}(\phi + \pi, 2\pi) - \pi\,. \tag{2.9}$$



(a) (b) (c) (d)

Figure 2.3: (a) Part of a SPAMM tagged MR image of the heart, (b) the magnitude of its Fourier transform, (c) the harmonic phase image and (d) the harmonic magnitude image after applying the bandpass filter as in (b). The harmonic phase image is masked for visualization purpose so that only phases of the tissue points are shown.

The true phase and harmonic phase are both material properties of tissue points. Thus, the HARP values of a material point do not change as the point moves around in space. This property is called *phase invariance* and is the basis of HARP motion tracking.

2.2.2 HARP Tracking

The HARP image $\theta(\mathbf{x}, t)$ contains information about tissue motion in the normal direction of the tag plane $\mathbf{e}$. To track the 2D apparent motion of a material point, one needs two HARP images which are generally acquired with orthogonal tag directions. Let $\Phi = [\phi_1, \phi_2]^T$, and ϕ_1 and ϕ_2 be the tagging phases of the two orthogonally tagged images. For a material point located at $\mathbf{x}(t_i)$ at time frame t_i, we look for a point

$\mathbf{x}(t_{i+1})$ at time t_{i+1} such that

$$\Phi(\mathbf{x}(t_{i+1}), t_{i+1}) = \Phi(\mathbf{x}(t_i), t_i)\,. \tag{2.10}$$

Solving for $\mathbf{x}(t_{i+1})$ is a multidimensional root finding problem, which can be solved iteratively using the Newton-Raphson technique, as follows

$$\mathbf{x}^{(n+1)}(t_{i+1}) = \mathbf{x}^{(n)}(t_{i+1}) - \nabla\Phi(\mathbf{x}^{(n)}(t_{i+1}), t_{i+1})^{-1}(\Phi(\mathbf{x}^{(n)}(t_{i+1}), t_{i+1}) - \Phi(\mathbf{x}(t_i), t_i))\,, \tag{2.11}$$

with $\mathbf{x}^{(0)}_{t_{i+1}} = \mathbf{x}_{t_i}$.

Since the corresponding HARP values $\Theta = [\theta_1, \theta_2]^T$ are just the principal values of the true phase Φ, Eq. (2.11) cannot be used directly. If one assumes, however, that any material point moves less than half of the tag separation from one time frame to the next in both tag orientations—i.e., $|\phi_k(\mathbf{x}(t_i), t_{i+1}) - \phi_k(\mathbf{x}(t_i), t_i))| < \pi, k = 1, 2$—then it can be shown that

$$\Phi(\mathbf{x}^{(n)}(t_{i+1}), t_{i+1}) - \Phi(\mathbf{x}(t_i), t_i) = W(\Theta(\mathbf{x}^{(n)}(t_{i+1}), t_{i+1}) - \Theta(\mathbf{x}(t_i), t_i))\,. \tag{2.12}$$

Moreover, the gradient of Φ can be written as

$$\nabla\Phi = \nabla^*\Theta = \nabla^*[\theta_1, \theta_2]^T\,, \tag{2.13}$$

where

$$\nabla^*\theta_k = \begin{cases} \nabla\theta_k, & ||\nabla\theta_k|| \leq ||\nabla W(\theta_k + \pi)|| \\ \nabla W(\theta_k + \pi)||, & \text{otherwise} \end{cases} \tag{2.14}$$

for $k = 1, 2$. Eq. (2.11) can then be written as

$$\mathbf{x}^{(n+1)}(t_{i+1}) = \mathbf{x}^{(n)}(t_{i+1}) - \nabla^* \Theta(\mathbf{x}^{(n)}(t_{i+1}), t_{i+1})^{-1} W(\Theta(\mathbf{x}^{(n)}(t_{i+1}), t_{i+1}) - \Theta(\mathbf{x}(t_i), t_i)) \,, \tag{2.15}$$

which is computable from the underlying images.

Traditional HARP tracking is therefore just the iteration of Eq. (2.15) until $||\mathbf{x}^{(n+1)}(t_{i+1}) - \mathbf{x}^{(n)}(t_{i+1})||$ is below a pre-specified small number, or until a specified number of iterations is reached.

2.3 3D Motion Reconstruction

In a 3D acquisition as described in Section 1.4, tagged MR images provide sparse and incomplete measurements of the 3D muscle motion, from which the dense and smooth 3D motion must be estimated. In cardiac imaging, many 3D motion reconstruction approaches have been developed using various structural models or interpolation models, especially for the left ventricle. Here we provide a brief review.

B-splines have been used together with spatial smoothness constraints to model the cardiac deformation [55,56,58,59,63,64]. B-spline models are defined on Cartesian coordinates, cylindrical coordinates [64], or planispheric coordinates [56]. Huang et al. [59], Declerck et al. [56], and Ozturk et al. [58] used 4D deformable B-spline models to reconstruct the motion both spatially and temporally. Instead of directly estimating the deformation, several other approaches used finite element models (FEM) to

represent the LV shape. Young et al. [48,57] modeled the tag surfaces using FEM and reconstructed the 3D cardiac deformation without prior identification of ventricular boundaries or tag line locations. Park et al. [60] and Haber et al. [49] computed the motion of both left ventricle and right ventricle using a 3D biventricular deformable FEM. Methods have also been proposed using other models or direct interpolation. O'Dell et al. [53] used a displacement field fitting method. Denny and Prince [50] reconstructed the 3D motion of the LV using an estimation theoretic approach. Kerwin and Prince [47] used a deformation model to track a sparse set of material points in 3D. Suter and Chen [65] interpolated the 3D left ventricle motion using elastic vector splines. Pan et al. [43] placed a sparse cup-shaped mesh model inside the left ventricle and iteratively tracked it in 3D based on HARP tracking. All of the mentioned methods were based on the sparse and incomplete motion measurements from the 2D images and therefore can be viewed as some kinds of interpolation.

Methods have also been developed to directly image 3D myocardial motion without the need of interpolation. Ryf et al. [66] extended the 2D-CSPAMM tagging sequence to 3D to acquire 3D tagged image volume. The 3D motion was then computed by extracting spectral peaks in 3D Fourier space in a way similar to 2D HARP processing. This method requires a very long acquisition time ($\sim$ 16 minutes) and is not feasible in regular clinical scan. Perman et al. [67] and Kuijer et al. [68] combined in-plane tagging and phase-contrast imaging together to quantify the in-plane motion and through-plane velocity simultaneously. Abd-Elmoniem [69, 70] recently

developed the zHARP method that can encode and track 3D motion from a single slice without increasing acquisition time.

The above methods were all originally developed for cardiac motion, and some of them have been extended to the tongue [71,72]. These 3D motion reconstruction and interpolation methods have either ignored the fact that tongue and cardiac muscles are incompressible, or assumed that the deformation is small and may result in large error for large motion.

2.4 Strain Calculation

Strain is a commonly used measure for deformation characterization and functional analysis of tissues. It defines the amount of stretch or compression along particular directions, and measures how much a given displacement differs locally from a rigid-body displacement.

Consider a material point $\mathbf{X}$ that deforms to $\mathbf{x}$ under a displacement $\mathbf{u}(\mathbf{X})$, i.e., $\mathbf{x} = \mathbf{X} + \mathbf{u}(\mathbf{X})$. The *deformation gradient tensor* F is defined by

$$F = \frac{d\mathbf{x}}{d\mathbf{X}} = I + \frac{d\mathbf{u}}{d\mathbf{X}}\,, \tag{2.16}$$

with I being the identity matrix. F is a 3×3 matrix in 3D, and a 2×2 matrix in 2D. $d\mathbf{u}/d\mathbf{X}$ is called the *material displacement gradient tensor*. We also have

$$F^{-1} = \frac{d\mathbf{X}}{d\mathbf{x}} = I - \frac{d\mathbf{u}}{d\mathbf{x}}\,, \tag{2.17}$$

where $d\mathbf{u}/d\mathbf{x}$ is the *spatial displacement gradient tensor*.

- Lagrangian Strain

  The *Lagrangian strain tensor* is defined in terms of the material coordinate $\mathbf{X}$, and is given by

$$E_l = \frac{1}{2}(F^T F - I) = \frac{1}{2}\left(\frac{d\mathbf{u}}{d\mathbf{X}} + \frac{d\mathbf{u}}{d\mathbf{X}}^T + \frac{d\mathbf{u}}{d\mathbf{X}}^T \frac{d\mathbf{u}}{d\mathbf{X}}\right). \tag{2.18}$$

- Eulerian Strain

  The *Eulerian strain tensor* is defined on the spatial coordinate $\mathbf{x}$, and is given by

$$E_e = \frac{1}{2}(I - F^{-T}F^{-1} = \frac{1}{2}\left(\frac{d\mathbf{u}}{d\mathbf{x}} + \frac{d\mathbf{u}}{d\mathbf{x}}^T - \frac{d\mathbf{u}}{d\mathbf{x}}^T \frac{d\mathbf{u}}{d\mathbf{x}}\right). \tag{2.19}$$

- Stretch Ratio and Normal Strain

  The *stretch ratio* measures the strain of a differential line element. Consider a vector $\mathbf{v} = ||\mathbf{v}||\mathbf{n}$ at material coordinate with $\mathbf{n}$ being a unit vector, and suppose it deforms into vector $\mathbf{w}$, the stretch ratio in the material coordinate is defined as

$$\Lambda(\mathbf{n}) = \frac{||\mathbf{w}||}{||\mathbf{v}||} = \frac{||\mathbf{F}\mathbf{v}||}{||\mathbf{v}||}. \tag{2.20}$$

  We can also have

$$\Lambda^2(\mathbf{n}) = \frac{\mathbf{v}^T\mathbf{F}^T\mathbf{F}\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \mathbf{n}^T\mathbf{F}^T\mathbf{F}\mathbf{n}. \tag{2.21}$$

  Similarly the stretch ratio for the spatial coordinate is defined as

$$\frac{1}{\Lambda} = \frac{||\mathbf{v}||}{||\mathbf{w}||}. \tag{2.22}$$

  The *normal strain* is defined as

$$e(\mathbf{n}) = \Lambda(\mathbf{n}) - 1. \tag{2.23}$$

In cardiac imaging, the circumferential and radial directions are the two common directions for cardiac function analysis. The circumferential strain represents the myocardial shortening, and the radial strain represents the myocardial thickening in short axis plane. Let $\mathbf{n}_c$ and $\mathbf{n}_r$ be the circumferential direction and radial direction, respectively. The circumferential and radial strains can be expressed as

$$e_c = \mathbf{n}_c^T E \mathbf{n}_c \,, \text{ and}$$

$$e_r = \mathbf{n}_r^T E \mathbf{n}_r \,.$$

2.5 Summary

In this chapter we briefly described the background knowledge of MR tagging and the HARP method which computes in-plane motion from tagged MRI. We also reviewed the 3D motion reconstruction methods the 2D measurements, and described the computation of strain.

Chapter 3

HARP Mistracking and Region Growing HARP Refinement

3.1 Introduction

Two-dimensional (2D) in-plane motion tracking is an important part of the HARP method because other quantities are often computed using these tracking results. HARP tracking implicitly assumes that tissue points do not move very far from one time frame to the next. If the tissue moves too fast, the temporal resolution is too low, or the MR tag parameters are selected incorrectly, this assumption is violated, and HARP tracking will fail. HARP tracking may also fail at points close to tissue boundary, and at points moving in or out from the image plane due to through-plane motion. Although such failures are relatively rare in typical well-

designed applications, careful scientific studies and robust clinical applications require that the user manually identify and correct mistracked points. This can be very time-consuming, to the point where large research studies take too much time and clinical throughput is too low. In research on tongue motion, there are some utterances in which parts of the tongue move quite fast relative to the temporal resolution of the scan, causing inevitable HARP tracking errors. Efforts to track a very large number of points thereby become extremely time consuming, as manual correction is routinely required.

There have been some previous efforts to identify and automatically correct mistracked points. Khalifa et al. [73] used an active contour model to correct HARP tracking for cardiac motion. The approach is limited to the circular geometry, however, and is therefore not easily generalized for non-cardiac applications such as imaging the tongue in speech. Tecelao et al. [74] proposed an extended HARP tracking method to correct the mistracking caused by through-plane motion and boundary effects, but it did not completely address the mistracking problem. Use of spatial continuity of motion, generically called *refinement*, was described in Osman et al. [40] as a process that could be employed to alleviate the mistracking described above. At the time, refinement was thought to be overly time consuming to employ on a routine basis. It was also developed on circular geometries, and is not straightforward to extend to arbitrary tissue points.

To address the mistracking problem, we develop two new refinement methods

for HARP tracking. These methods automatically track every point (pixel) inside a given tissue (or even the whole image) and do not require a circular geometry or organized mesh of points defined on the region. The first method is called *region growing HARP refinement* (RG-HR) method. The method is based on seeded region growing [75] with a tracking order determined by the local HARP phase smoothness. The RG-HR method works well in tongue motion tracking, but may fail occasionally in cardiac motion tracking. This is because the region growing process implicitly connects every point to the seed through a path that is determined by the tracking order. The local HARP smoothness is not enough to promise an optimal path, and thus RG-HR may yield incorrect tracking. This is especially a problem in cardiac motion tracking because the heart has a ring structure in the SA images, and in some cases points in the blood pool can be tracked before all points in the myocardium are tracked.

To address the problem of erratic tracking path, we develop another method called *shortest path HARP refinement* (SP-HR) method. It further improves the RG-HR method by representing the image as a graph and finding the optimal refinement paths for all tissue points by solving a single source shortest path problem. In this way each tissue point will be accessed via an optimal path from the seed point (which is assumed to be correctly tracked). Cost functions defined on both the edges and vertices of the (image-based) graph encourage the shortest path to stay within the same tissue region as the seed point so that incorrect estimation from long paths in

adjacent regions is much less likely.

In this chapter, we describe the RG-HR method in details. The SP-HR method is described in Chapter 4. This chapter is organized as follows. Section 3.2 summarizes the three cases of mistracking in 2D HARP. Section 3.3 briefly describes the original HARP refinement idea. Section 3.4 describes the region growing HARP refinement method in details. Section 3.5 shows experiment results on tongue motion tracking. Section 3.6 provides a discussion on the results, and Section 3.7 summarizes the chapter.

3.2 Mistracking in 2D HARP

Though HARP tracking works well in most scenarios, points can be mistracked when the underlying assumptions are not satisfied. In general, mistracking can be classified into three categories. First, mistracking can occur when a tissue point has a large motion between two successive time frames. In this case, HARP tracking (see Eq. (2.15)) will converge, but it finds a point that is one or more tag periods away from the truth—this is called a "tag jumping" error. This kind of mistracking can be alleviated by either improving the temporal resolution or decreasing the tag spatial frequency, but this comes at the cost of decreased image resolution and poorer HARP phase estimation. It may be impossible to take either of these steps in some scenarios such as imaging the rapid movement of the tongue in speech.

Second, the tissue point can be mistracked because of through-plane motion. In the Lagrangian framework of HARP tracking, a material point is specified in the first time frame and tracked through all later time frames successively. Because of through-plane motion, the point may disappear in some time frames, and (possibly) re-appear in some later time frames. Because the standard tracking approach moves successively from one time frame to the next, when this occurs HARP tracking will converge to an incorrect point during the lost frames and will not generally find the correct point when it reappears.

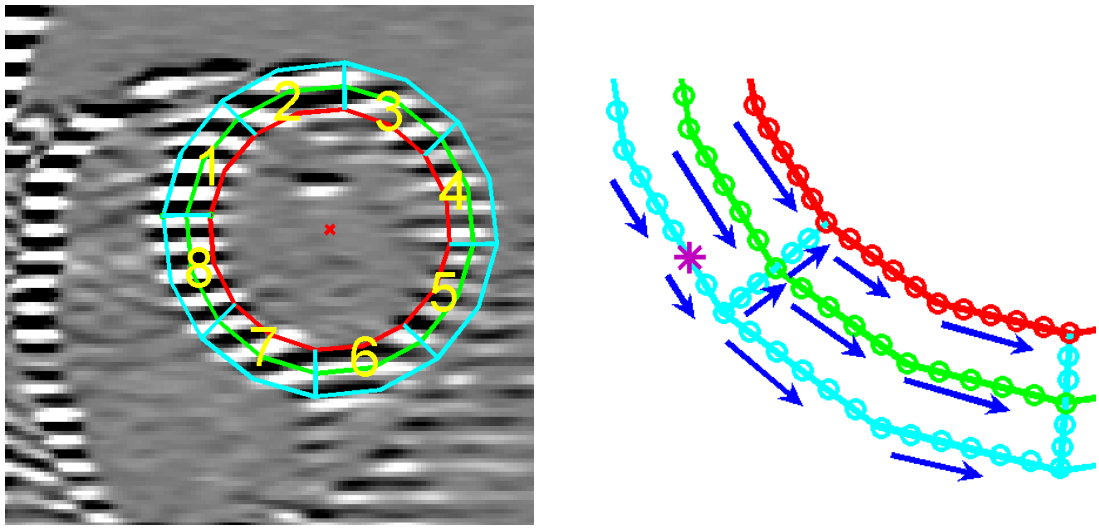
Third, mistracking can happen at points close to the tissue boundary. The problem is that the tracking equation (2.15) must start with an initial "guess" as to where the point goes in the second frame in order to initialize the iterative process. If that initial guess happens to be outside the tissue in the second frame, then phase noise, caused by the lack of sufficient signal, will yield highly erratic, usually erroneous results.

In scientific and clinical applications, mistracked points must be manually identified and corrected by the user. This can be very time-consuming especially when tracking a large number of points in rapidly moving tissue. For routine clinical applications and large-scale scientific studies, it is therefore imperative to find a method that can correctly track all tissue points including these three classes of points that are mistracked in standard HARP tracking.

3.3 HARP Refinement

Standard HARP tracking relies on the temporal continuity of a point trajectory—i.e., that a point should not move much from one time frame to the next. The idea of using spatial continuity of motion to improve HARP tracking was first proposed by Osman et al. [40] and was called *HARP refinement*. In the original HARP refinement, points are tracked on concentric "circles" that are placed manually within the left ventricular (LV) myocardium, as shown in Fig. 3.1(a). One point on this geometric construct is manually identified as an "anchor point"—the asterisk in Fig. 3.1(b)—which we will refer to as a *seed point*. This point generally has a very small displacement over the entire cardiac cycle and can be certified by the user to be correctly tracked by standard HARP. Starting from the seed point, an adjacent point (about a pixel away) on the concentric circle is tracked next, wherein it is explicitly assumed that its initial displacement, which defines the initialization of the tracking algorithm (2.15), is equal to that of the seed point. The entire circle is tracked this way by assuming the initial displacement of a given point is equal to the estimated displacement of the previous point. This process, which is illustrated in Fig. 3.1(b), can also be applied on radial paths in order to facilitate the tracking of all concentric circles.

This refinement method has several limitations. First, its success is overly tied to the locations of the circles. In particular, if a single point on a circle is mistracked, it is very likely that all the remaining points on the circle will also be mistracked. This means that circles placed close to the edges of the myocardium may produce a large



(a) (b)

Figure 3.1: (a) Concentric circles placed on the LV. (b) Processing order for a conventional HARP refinement procedure. The asterisk shows the location of the anchor point.

fraction of mistracked points despite refinement. Second, correctly tracked points are limited to the circles and radial lines, which means that strain computations must be Lagrangian in nature and are not densely computed on the object of interest. Finally, although the approximately circular shape is useful for the LV, alternative shapes would have to be developed for other objects of interest such as the RV and the tongue.

3.4 Region Growing HARP Refinement

The *region growing HARP refinement* (RG-HR) method is developed to address limitations of HARP refinement. It is observed that the motion field within the tissue

is smooth — e.g., the displacements of neighboring tissue points are similar. For two neighboring 2D points $\mathbf{x}(t_1), \mathbf{y}(t_1)$ at time frame t_1, and at time t_2 they move to $\mathbf{x}(t_2)$and $\mathbf{y}(t_2)$, then the difference between their displacements

$$\Delta(\mathbf{x}, \mathbf{y}, t_1, t_2)) = |(\mathbf{x}(t_2) - \mathbf{x}(t_1)) - (\mathbf{y}(t_2) - \mathbf{y}(t_1))| \tag{3.1}$$

is small. So if $\mathbf{y}(t_2)$ can be tracked correctly from $\mathbf{y}(t_1)$, a good estimation of $\mathbf{x}(t_2)$ is

$$\mathbf{x}'(t_2) = \mathbf{x}(t_1) + (\mathbf{y}(t_2) - \mathbf{y}(t_1)) \,. \tag{3.2}$$

Therefore $\mathbf{x}'(t_2)$ can be used as the starting point when tracking $\mathbf{x}(t_1)$ to prevent HARP tracking from failing.

There is usually some part of the tissue that has relatively small motion and can be tracked correctly using HARP over all time frames. For example, the bottom part of the tongue is relatively stationary in speech. Hence, it is easy to manually identify a few seed points where the traditional HARP tracking succeeds.

The RG-HR algorithm is based on the above facts, and is described in details below.

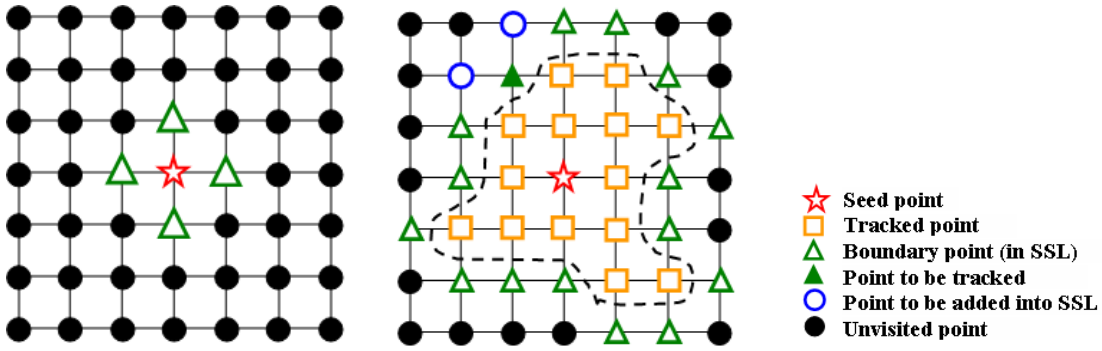
3.4.1 RG-HR Algorithm

The RG-HR algorithm is based on seeded region growing [75]. It starts with one manually identified seed that can be correctly tracked between two time frames using traditional HARP tracking. Since the tracking between different time frames is done

independently, we consider only the tracking from one time t_1 to another time t_2 in the following.

A data structure called sequentially sorted list (SSL) [75] is used in this algorithm. The points in the SSL are sorted based on a cost function (described below) that reflects the likelihood that a point lies inside the tissue and is good to be tracked next. The SSL is maintained throughout the algorithm, and it stores the points that have at least one tracked neighbor point. When the algorithm starts, the neighbor points of the tracked seed are identified and put into the SSL. At each iteration, the first point in the SSL, which has the lowest cost value, is taken off the list and tracked, and its neighbors are inserted into the SSL based on the cost function value. The process is repeated until the list is empty.

The points in the SSL are called *boundary points*. Each entry on the list contains the following information: the point's 2D coordinate $\mathbf{x}(t_1)$ at time t_1, an initial estimate of its location in the next time frame $\mathbf{x}'(t_2)$, and a cost value $C(\mathbf{x}(t_1))$ (defined below). $\mathbf{x}'(t_2)$ is computed using (3.2), assuming that $\mathbf{x}(t_1)$ has the same displacement as its tracked neighbor when it is inserted. The list is sorted in the ascending order of the cost value. At each iteration, the first entry is fetched and removed from the SSL. The point in this entry, $\mathbf{x}(t_1)$, is tracked using traditional HARP tracking but starting from $\mathbf{x}'(t_2)$ instead of $\mathbf{x}(t_1)$. Because the actual location $\mathbf{x}(t_2)$ is close to $\mathbf{x}'(t_2)$ (as discussed above), HARP tracking is highly likely to converge to the correct position. After that, those neighbors of $\mathbf{x}(t_1)$ that have not been tracked and are



(a) (b)

Figure 3.2: Illustration of region growing process on the image grid. At the beginning of the algorhtim ((a)), a seed point is given. Its four neighbors are then put into the SSL. (b) shows the status at an intermediate step. First the first point in the SSL (solid triangle) is removed from the SSL and tracked. Then its neighbors that are not tracked and not in the SSL are inserted into the SSL (hollow circle) based on their cost values.

not already labeled as boundary points are inserted into the list based on their cost values. The estimated locations of newly inserted neighboring points at time t_2 are computed using (3.2). An example is displayed in Fig. 3.2. In this example 4-neighbor connectivity is assumed, though 8-neighbor connectivity can be used instead without much difference.

It is important to track every point inside the tissue first before growing the region into non-tissue regions. This is because the HARP value is quite noisy in the air, blood, or bone, where no tag pattern can be found in the image. To encourage the region growing process to visit tissue points first, the cost function that we use to sort the SSL is defined as the phase similarity function:

$$C(\mathbf{x}(t_1)) = \sum_{i=1}^{2} |W(\phi_i(\mathbf{x}(t_1), t_1) - \phi_i(\mathbf{x}'(t_2), t_2))| \,. \tag{3.3}$$

Since the phase values inside the tissue are smooth, for a point inside the tissue the wrapped difference between its actual phase ($\phi_i(\mathbf{x}(t_1), t_1)$) and the phase at its estimated location is small, and the cost function has small value. Therefore the point is placed at the front of the SSL and is visited early.

3.4.2 Implementation of RG-HR Algorithm

This algorithm can be implemented as in Algorithm 3.1.

3.5 Results on Tongue Motion Tracking

Our method was applied on the tagged MR images of the tongue. The images were collected on a 1.5T Marconi scanner when the subject uttered "eeoo" repeatedly. The images were acquired in 12 time frames with a temporal resolution of 66 msec. The interpolated spatial resolution was 1.09 mm $\times$ 1.09 mm and the slice thickness was 7 mm. Four sets of SPAMM images were collected: horizontal tagging with [$+90^0$ $+90^0$] and [$+90^0$ -90^0] , and vertical tagging with [$+90^0$ $+90^0$] and [$+90^0$ -90^0] tagging pulses. As preprocessing, the MICSR [45] images were reconstructed from these 4 sets of data. Our method was implemented in C, and compiled in Matlab 7 (Mathworks, Natick MA). On a computer with Intel Core Duo 1.83 GHz processor and 1.0 G ram, our implementation took about 0.2 second to track an 128 by 128 image for one time frame.

Algorithm 3.1 Region growing HARP refinement (RG-HR)

1: Track manually selected seed point(s) over all time frames.
2: **for all** time frame t_i **do**
3: Create the SSL, and insert the seed.
4: **while** the SSL is not empty **do**
5: Remove the first node $\mathbf{x}(t_i)$ from the SSL.
6: Find $\mathbf{x}(t_{i+1})$ using HARP tracking.
7: Label $\mathbf{x}(t_i)$ as *tracked point*.
8: **for all** neighbor $\mathbf{y}_k(t_i)$ of $\mathbf{x}(t_i)$ **do**
9: **if** not *tracked point* and not *boundary point* **then**
10: Calculate $\mathbf{y}'_k(t_{i+1})$, and the cost $C(\mathbf{y}(t_i))$.
11: Insert $\mathbf{y}_k(t_i)$ in the SSL based on $C(\mathbf{y}(t_i))$.
12: Label $\mathbf{y}_k(t_i)$ as *boundary point*.
13: **end if**
14: **end for**
15: **end while**
16: **end for**

Fig. 3.3 shows the intermediate results of the region growing process on the mid-sagittal slice. Starting from the seed point, the points inside the tongue were tracked first. The outside points were tracked only after all points inside the tongue were tracked.

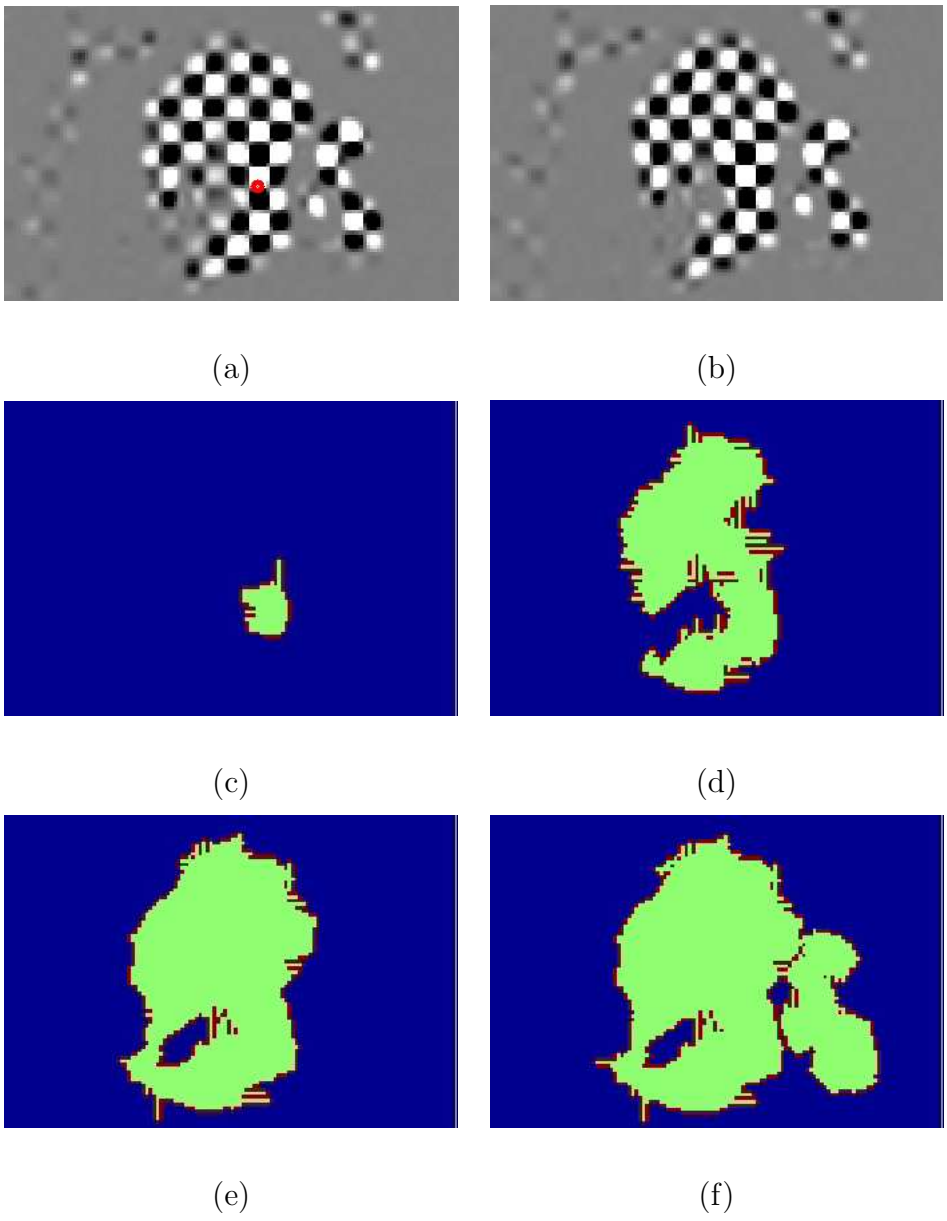
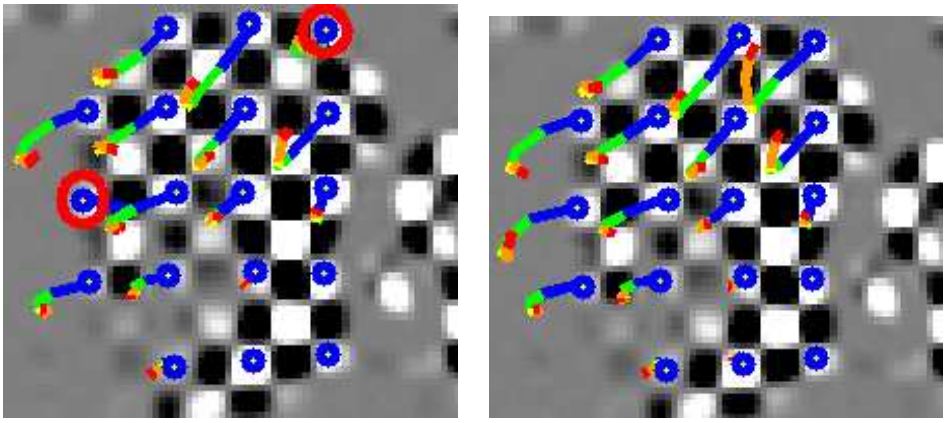


Figure 3.3: Illustration of the region growing process in RG-HR. (a) is the checkboard image at the first time frame by overlaying the two tagged images with different orientations, (b) is the checkboard image at the second time frame. The red dot in (a) is the manually selected seed point. (c-f) shows how the region grows. The green color means tracked points, brown means boundary points, and blue means points that are not tracked and not boundary points.



(a) (b)

Figure 3.4: The trajectories of selected points on tagged MR image over 12 time frames. (a) is the results of traditional HARP tracking. The points in the circle were mistracked. (b) is the RG-HR results.

In Fig. 3.4 a grid of points was placed inside the tongue and tracked through the sequence, and their trajectories are shown. Some points were mistracked (Fig. 3.4(a)) when using the traditional HARP tracking, but correctly tracked using RG-HR (Fig. 3.4(b)).

Fig. 3.5 illustrates how our refinement method improves the Lagrangian strain calculation. The Lagrangian strain is computed as the length change of line segments with respect to the length at first time frame. It has a positive value when stretching and a negative value when contracting. Five line segments were manually placed in the tongue and they represented the fan-shaped genioglossus muscle. In Fig. 3.5(b), the number 1 and number 4 line segments were mistracked in traditional HARP tracking, which made the Lagrangian strain calculation wrong (Fig. 3.5(d)). However, they were correctly tracked using RG-HR (Fig. 3.5(c) and (e)).

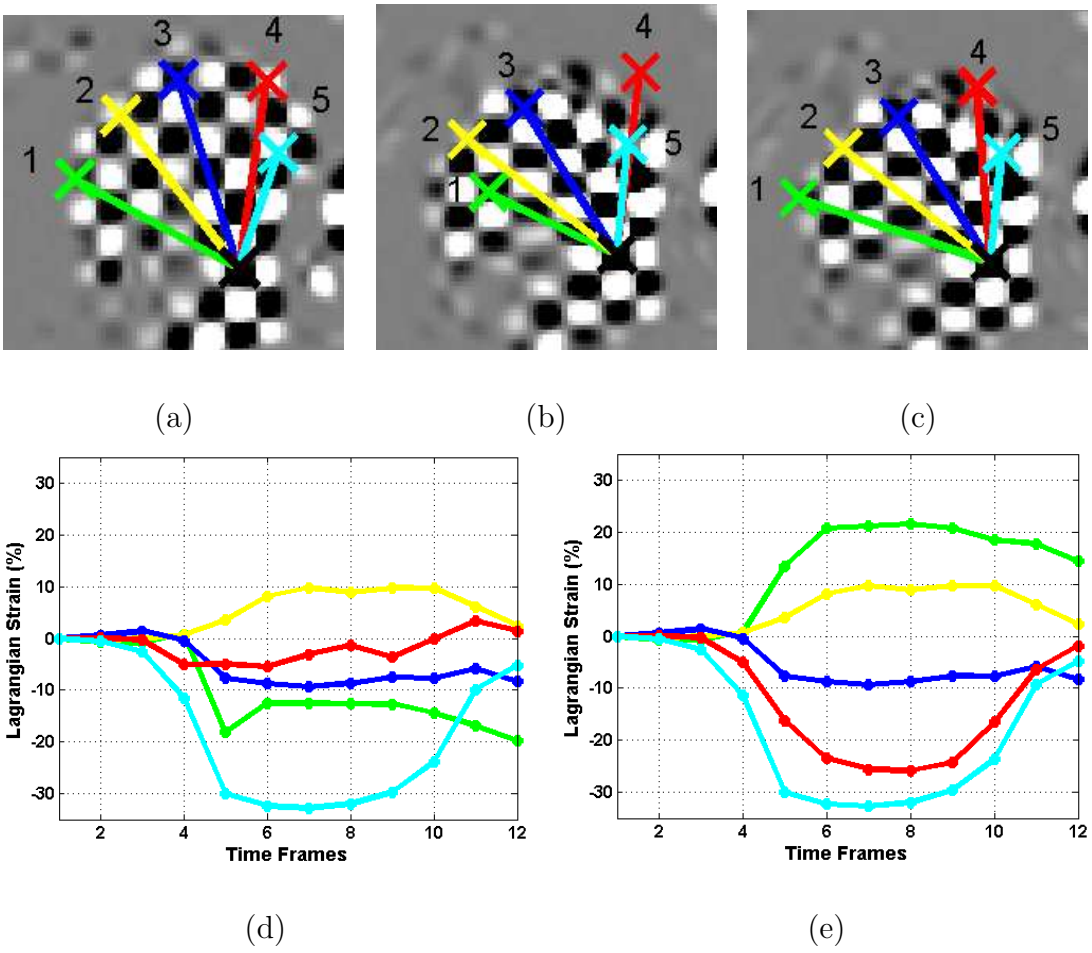


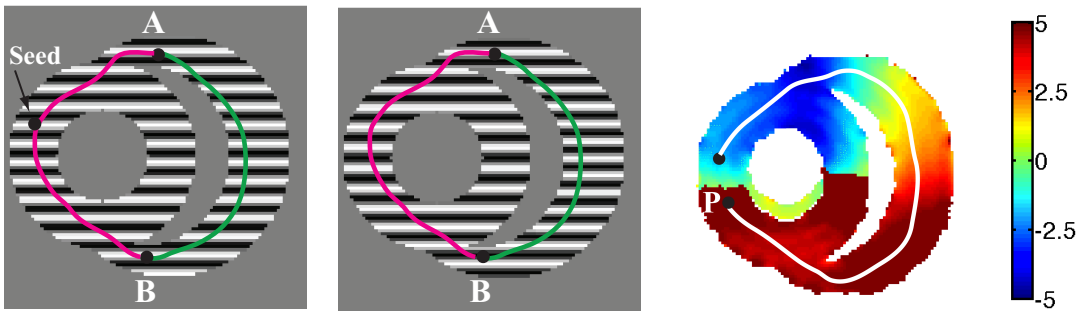
Figure 3.5: The action of genioglossus (GG). (a) shows five segments of GG at the first time frame. They are then tracked using both traditional HARP method and RG-HR throughout all the time frames, The tracked position at the last time frame is shown in (b) and (c) respectively. (d) and (e) show the Lagrangian strains of the 5 line segments change with time. (d) is the result of traditional HARP method, and (e) is the result of RG-RG.

3.6 Discussion

Although RG-HR is applied to directly compute motion, the region growing refinement process can also be thought of as an application-specific harmonic phase unwrapping process. With this interpretation, it is clear that the flood-fill algorithm

used to unwrap DENSE phase images (cf. [76]), reported in Spottiswoode et al. [77], is another example of a refinement algorithm for improved motion estimation. In both the RG-HR refinement method and the flood-fill algorithm for DENSE phase unwrapping, tissue points are tracked or phase unwrapped in an order that is based solely on the local smoothness of the phase images. Because of image noise and the existence of other tissues near to that of primary interest (e.g., the heart or tongue), the spatial paths from the seed to any given points of interest can be quite erratic, and incorrect tracking may result.

RG-HR generally works very well on tongue images, but less reiably in cardiac motion tracking experiments. This is mainly because of the different structures of the tongue and the heart. It is possible, for example, for a point within the free wall of the LV to be assigned a displacement based on a seed in the septum and a path that travels through the liver or the blood pool for some distance rather than entirely through the myocardium. This is illustrated in Fig. 3.6 using a simulated example. The example shows a ring-shaped object of interest on the left and a second object that touches the object of interest. It is assumed the object only moves in the top-bottom direction. The points on the object are tracked from Frame 1 (Fig 3.6(a)) to Frame 2 (Fig. 3.6). The touching parts of the two objects change from Frame 1 to Frame 2 so that the second object contains different numbers of tags at the two frames because the two objects do not move at the same pace. I.e., the two paths connecting points **A** and **B** in Figs. 3.6(a) and (b) pass through different numbers of



(a) Frame 1 (b) Frame 2 (c) Displacement field

Figure 3.6: An simulated example that demonstrates erratic tracking path in RG-HR. In (a) and (b), the two curves are two different paths connecting points A and B. The unit of displacement field is pixel. (c) shows the displacement field computed using RG-HR.

tags at Frame 2, and same numbers of tags at Frame 1. We then apply the RG-HR method to track the motion from Frame 1 to Frame 2 with a seed whose location is shown in Fig. 3.6, and the resulting displacement field is shown in Fig. 3.6(c). The computed displacement field in the object of interest is not continuous and "breaks" in the middle. This is caused by the incorrect tracking order of the region growing process. For example, point **P** in Fig. 3.6(c) is located in the object of interest and is close to the seed point. Therefore it should be tracked early. However in the RG-HR method it is not tracked until all points on the long path passing through the second object are tracked, which is not correct.

Since the phase unwrapping strategy using the DENSE imaging framework employed a very similar strategy for motion and strain estimation as RG-HR, it will likely suffer from the same problem of erratic tracking path.

The performances of RG-HR can be improved by enforcing a tissue mask in the

algorithm. The mask can be manually determined and contains only the regions of interested tissue. By limiting the region growing process inside the mask, both methods will produce better results because the chance that the region growing "leaks" into the non-tissue regions is greatly reduced.

3.7 Summary

In this chapter we presented a HARP tracking refinement method based on seeded region growing. HARP tracking suffers from mistracking when there is large motion, low temporal resolution, through-plane motion, or the points are close to tissue boundary. The proposed method prevents mistracking through a region growing process so that the tissue points with reliable harmonic phase values are tracked first. Experimental results showed that this method can reliably track every point in- side the tissue even in the case of large motion when the traditional HARP tracking fails. This method is also computationally fast and makes it feasible to compute Lagrangian strain between arbitrary points in real time.

Chapter 4

Shortest Path HARP Refinement

4.1 Introduction

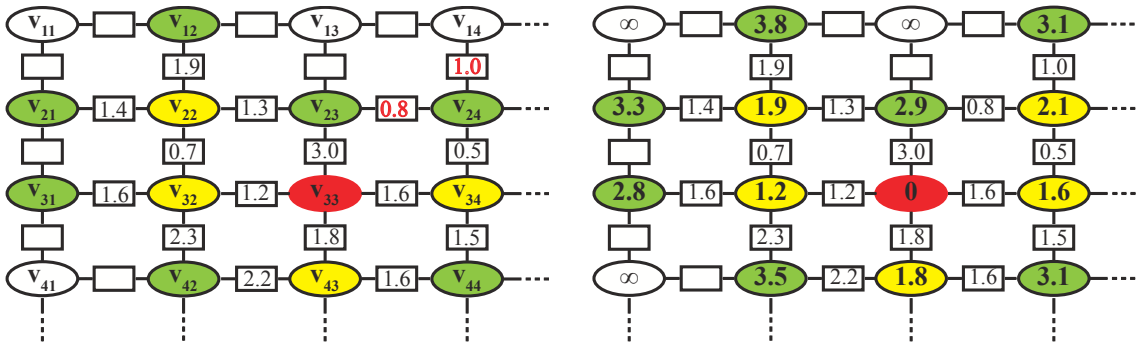
The region growing HARP refinement described in Chapter 3 addresses the mis-tracking problem of HARP tracking using a region growing process, with a tracking order determined by the local smoothness of the HARP phase images. It generally works very well in tongue motion tracking but not in cardiac motion tracking. This is because the motion estimate computed at any given point may depend on the specific path over which the motion or phase values are estimated from the seed to the given point, and region growing according to local HARP phase smoothness is not enough to promise an optimal path especially in the cardiac tracking.

We posit that the overall region growing strategy is quite good—it covers the entire field of view and is computationally fast. Better results should be produced if an

optimal path from the seed to each point can be found. Paths that are unnecessarily long or go through multiple tissues should be avoided. So, rather than making local decisions on how the region should be grown relative to the region's current boundary, the entire path from the seed to each point should be determined optimally. In this way, the path's entire length goes through points that are correctly tracked with high reliability. Based on this concept, we develop another HARP refinement method by formulating the problem as a single source shortest path problem; we call this the *shortest path HARP refinement* (SP-HR) method. Experiments on cardiac motion tracking have showed that this method is more robust in preventing HARP mistracking, and is computationally as fast as the RG-HR method.

By defining synthetic phase images at the reference time and applying refinement methods between each time frame and the reference time, it also becomes possible to automatically track points that may appear and disappear due to through plane motion, providing an extra level of relief from manual intervention.

This chapter is organized as follows. Section 4.2 describes the shortest path HARP refinement method in details. Section 4.3 explains the two-step tracking procedure using reference time frame that helps track points between any two time frames conveniently and reduce the occurrence of mistracking caused by through-plane motion. Section 4.4 introduces a semi-automatic way of seed selection. Section 4.5 describes experiment results on both numerical simulations and cardiac motion tracking. Section 4.6 provides a discussion on the results, and Section 4.7 summarizes the chapter.



(a) (b)

Figure 4.1: A graph representation of shortest path HARP refinement procedure (a) before and (b) after some particular iteration. Red oval: seed vertex. Yellow ovals: tracked vertices. Green ovals: boundary vertices. White ovals: unvisited vertices.

4.2 The SP-HR Method

In SP-HR, the image is represented as an undirected graph $G = (V, E)$, where V is the set of vertices (pixels), and E is the set of edges (that connect pixels), as shown in Fig. 4.1(a). Consider tracking points within an image at time t_1 to another time frame t_2. Each point (pixel) $\mathbf{x}(t_1)$ in the image at time t_1 is represented as a vertex v in the graph and each edge $e_{ij} = \langle v_i, v_j \rangle$ in E corresponds to a neighboring vertex pair v_i and v_j. Each edge has an edge cost $C_E(e_{ij})$, which is non-negative and measures the dissimilarity between the two end vertices. As well, each vertex is associated with a vertex cost $C_V(v)$. Both types of costs are defined below. With this framework, the HARP tracking refinement problem can be formulated as a single source shortest path problem, which lends itself to an optimal solution that can be computed very efficiently.

4.2.1 Cost Functions

We define the edge cost using both the harmonic magnitudes and the implied motions at the two end vertices. Let $\mathbf{x}_i(t_1)$ and $\mathbf{x}_j(t_1)$ be the adjacent points associated with the two end vertices at time t_1, and at time t_2 they move to $\mathbf{x}_i(t_2)$ and $\mathbf{x}_j(t_2)$, respectively. The first part of the edge cost is from the phase similarity function (cf. Eqn. (3.3))

$$\Delta(\mathbf{x}_i, \mathbf{x}_j) = C(\mathbf{x}(t_1)) = \sum_{k=1}^{2} |W(\phi_k(\mathbf{x}_j(t_1), t_1) - \phi_k(\mathbf{x}'_j(t_2), t_2))| \,. \tag{4.1}$$

It depends on the smoothness of harmonic phases, and should take small value when the edge lies in regions with smooth HARP values.

Pixels that fall outside of tissue—e.g., air or bone—have very weak MR signals and therefore possess unreliable harmonic phases. We use the harmonic magnitude images to determine whether adjacent points are likely to be within tissue or not. Let $D_1(\mathbf{x}, t)$ and $D_2(\mathbf{x}, t)$ be the two harmonic magnitude images computed from the two tagged images with orthogonal tags, respectively. The weighting functions are defined as

$$w_1(\mathbf{x}_i, \mathbf{x}_j) = (\bar{D}(\mathbf{x}_i(t_1), t_1) + \bar{D}(\mathbf{x}_j(t_1), t_1))/2 \,, \quad \text{and} \tag{4.2}$$

$$w_2(\mathbf{x}_i, \mathbf{x}_j) = (\bar{D}(\mathbf{x}_i(t_2), t_2) + \bar{D}(\mathbf{x}'_j(t_2), t_2))/2 \,, \tag{4.3}$$

where $\bar{D}(\mathbf{x}, t) = D_1(\mathbf{x}, t) + D_2(\mathbf{x}, t)$ is normalized to the interval $[0, 1]$.

When $\Delta(\mathbf{x}_i, \mathbf{x}_j)$ is small and both $w_1(\mathbf{x}_i, \mathbf{x}_j)$ and $w_2(\mathbf{x}_i, \mathbf{x}_j)$ are large, then the edge cost should be small. Accordingly, we define the *edge cost function* associated

with the two neighboring points $\mathbf{x}_i$ and $\mathbf{x}_j$ as

$$C_E(e_{ij}) = \frac{\Delta(\mathbf{x}_i, \mathbf{x}_j)}{w_1(\mathbf{x}_i, \mathbf{x}_j) w_2(\mathbf{x}_i, \mathbf{x}_j)} \,. \tag{4.4}$$

The edge cost function penalizes both a lack of harmonic phase smoothness and edges that cross between tissue and background.

Given a seed vertex v_0, we define the *vertex cost function* $C_V(v_i)$ at any other vertex v_i as the accumulated edge cost along the shortest path from v_0 to v_i. For any path $p = \langle v_0, v_1, v_2, ..., v_i \rangle$ in G from v_0 to v_i, its accumulated edge cost is $C(p) = \sum_{k=1}^{i} C_E(\langle v_{k-1}, v_k \rangle)$. Therefore the vertex cost function at v_i is

$$C_V(v_i) = \min\{C(p) : p \text{ is any path from } v_0 \text{ to } v_i\} \,, \tag{4.5}$$

where $C_V(v_0)$ is set to zero. The vertex cost function serves to establish the best path starting from the seed by which to define the initial estimated displacement at any pixel within the image. The actual estimated displacement at any pixel is found by iterating (2.15) as in standard HARP tracking but starting from the initial estimate determined by the best path.

This can be explained using Fig. 4.1(a) as an example. In this example v_{33} is the seed. Two of the paths that connect vertex v_{22} to v_{33} are: $p_1 : v_{33} \rightarrow v_{32} \rightarrow v_{22}$ and $p_2 : v_{33} \rightarrow v_{23} \rightarrow v_{22}$. The costs of the two paths are: $C(p_1) = 1.2 + 0.7 = 1.9$ and $C(p_2) = 3.0 + 1.3 = 4.3$, respectively. p_1 is shorter than p_2 because $C(p_1) < C(p_2)$. In fact $C(p_1)$ is less than the cost along any other path that connects v_{33} to v_{22}. Therefore p_1 is the shortest path and thus $C_C(v_{22}) = 1.9$.

4.2.2 Motion Tracking via Shortest Path Following

In SP-HR, the shortest path from the single manually specified seed point to every other point is found using Dijkstra's algorithm [78]. The overall algorithm is still a region growing algorithm in the sense that boundary pixels are successively added to the growing list of points comprising a region. But in addition to keeping track of the region itself, for points on the region boundary the vertex costs and the nearest neighbors (toward the seed)—we call them the *predecessors*—along the shortest paths are also computed and stored. When a point is tracked, the traditional HARP tracking is initialized using the displacement of the point's predecessor on the shortest path.

To carry out Dijkstra's algorithm, the vertices are classified into three disjoint sets: the boundary vertex set V_b, the tracked vertex set V_t, and the unvisited vertex set V_u. The boundary vertices are maintained in a linked list structure that is sequentially sorted based on their vertex costs—i.e., the first vertex in the list has the smallest vertex cost. We denote $N(v)$ to be the predecessor of v on its shortest path, and $\mathbf{u}(v) = \mathbf{x}(t_2) - \mathbf{x}(t_1)$ to be the displacement of the point $\mathbf{x}$ associated with v. Given these notations, the SP-HR tracking refinement algorithm is summarized in Algorithm 4.1.

A numerical example is given in Fig. 4.1(a) and 4.1(b). Fig. 4.1(a) shows that at some iteration, four vertices (v_{22}, v_{32}, v_{43}, and v_{34}) besides the seed have been tracked and are marked in yellow. Among the boundary vertices (marked in green),

Algorithm 4.1 Shortest Path HARP Refinement (SP-HR)

1: Pick a seed vertex v_0. Set $N(v_0) = v_0$, and $\mathbf{u}(v_0) = 0$.

2: Initialize $V_t = \emptyset$, $V_b = \{v_0\}$, and $V_u = V \backslash \{v_0\}$.

3: Set $C_V(v_0) = 0$, and $C_V(v) = \infty$, for $\forall v \in V_u$.

4: **repeat**

5: Remove the first vertex v_k in V_b. Set $V_t = V_t \cup \{v_k\}$.

6: Track the point $\mathbf{x}_k$ associated with v_k using the traditional HARP method, but starting from the initialization of $\mathbf{x}'_k = \mathbf{x}_k(t_1) + \mathbf{u}(N(v_k))$ instead of $\mathbf{x}_k(t_1)$.

7: **for all** v_i such that $\langle v_k, v_i \rangle \in E$ and $v_i \notin V_t$ **do**

8: Compute the new cost $C'_V(v_i) = C_V(v_k) + C_E(e_{ki})$.

9: **if** $C'_V(v_i) < C_V(v_i)$ **then**

10: Set $C_V(v_i) = C'_V(v_i)$, and $N(v_i) = v_k$.

11: **if** $v_i \in V_b$ **then**

12: Set $V_b = V_b \backslash \{v_i\}$, and re-insert v_i into the sorted list V_b based on $C_V(v_i)$.

13: **else if** $v_i \in V_u$ **then**

14: Set $V_u = V_u \backslash \{v_i\}$, and insert v_i into V_b based on $C_V(v_i)$.

15: **end if**

16: **end if**

17: **end for**

18: **until** $V_b = \emptyset$

v_{24} has the lowest cost ($1.6+0.5=2.1$) via path $\langle v_{33}, v_{34}, v_{24}\rangle$. Therefore it is tracked, and the edge costs on $\langle v_{24}, v_{23}\rangle$ and $\langle v_{24}, v_{14}\rangle$ are computed. The vertex costs of v_{14} and v_{23} are updated accordingly, and v_{23} becomes the predecessor of both vertices. In addition, v_{14} is changed to a boundary vertex. The edge and vertex costs of the graph at the end of this iteration is shown in Fig. 4.1(b).

4.2.3 Advantages of SP-HR over RG-HR

The RG-HR method uses an edge criterion alone to decide what point to track next. The inherent problem with this approach is that determination of what point to track next depends on what points are currently on the boundary, and these points are (potentially) far from the seed and have no tight relationships with the seed. This permits potentially unnatural *effective paths* to determine how points in the image are tracked. In contrast, the SP-HR algorithm ties the tracking of every point throughout the image directly to the seed point through its own optimal path determined by motion smoothness and reluctance to cross boundaries. Since the seed's displacement is certified by the user to be correctly tracked, it is far less likely that gross tracking errors will occur within the region of interest defined by the seed—e.g., the myocardium or the tongue. This approach also permits us to correctly track points that are very near to a tissue boundary, because these points will almost always be tied back to the seed through the region of interest defined by the seed.

4.3 Two-Step Tracking Using Reference Time Frame

There is an additional benefit to the SP-HR approach (as well as the RG-HR approach). Suppose that the seed is certified by the user to have undergone a very large displacement. In this case, all neighbors of the seed will be initialized with this displacement in order to find displacements. This overall large displacement can then propagate to all corners of the image, permitting HARP to track very large displacements. Because of this capability, we can then track directly between any pair of images in the image sequence (provided that there is a seed that is correctly tracked throughout the entire sequence). This frees us from the previous modes of operation which were limited to sequentially tracking either forward or backward in time.

Through-plane motion can cause tissues to appear and disappear in an image sequence (cf. [74]). Because of this, it is sometimes problematic to try to track a given point all the way through the sequence. When tracking fails due to disappearing tissue, it is difficult to find a correct correspondence when it appears again later. The authors of [74] addressed this problem by defining active and inactive points corresponding to those that appear and disappear. Here we solve this problem in a much simpler way by using synthetic harmonic images at a reference time and the large-displacement, two-frame tracking procedure mentioned above. We describe this

overall approach now.

4.3.1 Reference Time Frame

The reference time t_0 is defined as the time immediately after tissue tagging is applied and the tissue has not deformed. Because it takes some time in order to acquire the first image, we can never actually obtain an image of the anatomy at the reference frame. However, we know from the tagging parameters what the gradients of harmonic phase should be at the reference time—and we know this throughout the entire image, not just within the tissue. In particular, given the two normal vectors of the tag planes $\mathbf{e}_1$ and $\mathbf{e}_2$, synthetic HARP images can be defined as follows

$$\begin{aligned} \theta_1(\mathbf{x}, t_0) &= W(k_1 \mathbf{x} \cdot \mathbf{e}_1 + \theta_0^1)\,, \\ \theta_2(\mathbf{x}, t_0) &= W(k_2 \mathbf{x} \cdot \mathbf{e}_2 + \theta_0^2)\,, \end{aligned} \tag{4.6}$$

where k_1 and k_2 are the (known) tagging frequencies in the $\mathbf{n}_1$ and $\mathbf{n}_2$ directions, respectively, and θ_0^1 and θ_0^2 are unknown phase offsets. The authors of [74] describe an approach to accurately estimate the unknown phases θ_0^1 and θ_0^2. In our approach this is not necessary since we simply use this image as a reference frame—not a representation of the true configuration—so we set $\theta_0^1 = \theta_0^2 = 0$. With this image in hand, tracking from the seed in the first time frame to this reference time frame is carried out, so that the seed now has a position in the reference frame.

4.3.2 Two-Step Tracking

Tracking between any two time frames t_j and t_i is now performed in two steps. First HARP refinement is applied to track *directly* from time frame t_j to the reference time t_0. Since the phase value is available everywhere in the reference time, all tissue points at t_j can be tracked back to the reference time. In the second step, HARP refinement is applied to track *directly* from t_0 to time frame t_i. In this step, tissue points can be tracked correctly to t_i as long as they have not moved out of the image plane due to through-plane motion. We note that direct tracking quite deliberately means that no intervening time frames are used to establish a smaller motion estimates between each frame. Instead, we rely on knowledge of the motion of the seed throughout all time frames and the spatial continuity and shortest path algorithm to track potentially very large motions between these time frames.

We see that this two-step procedure does not depend on any particular tracking result except between the reference frame and either frame at times t_i or t_j. Therefore, if a tissue point disappears and then reappears due to through plane motion, it can still be successfully tracked between t_i and t_j as long as it appears in both of these images. We also see that the two-step procedure (with HARP refinement at its core) makes it possible to directly track between any two time frames, which is generally not possible in the conventional HARP tracking. This changes how one goes about tracking all the points in an image sequence, as we see next.

4.3.3 Tracking Through an Image Sequence

So far we have described a process that tracks between pairs of images. When the goal is to track all points in an entire image sequence, three approaches seem reasonable given the general framework we have proposed. First, it might seem that the most general approach would be to stack up all the images as a three-dimensional image and apply the fundamental region-growing strategy to the entire stack. It turns out that this approach is problematic because it is difficult to properly scale the changes in phase and magnitude that one might expect between pixels in the spatial dimensions and pixels in the temporal dimension. We experimented with this technique but soon abandoned the approach entirely.

The second approach to tracking an entire image sequence is to simply track each image pair sequentially through time after having established the validity of the seed's trajectory through all time. This approach has the disadvantage that tissues might appear and disappear throughout the image sequence causing very erroneous tracks to appear. But it has the advantage that every image has a pixel-specific displacement field associated with it. Tracks for any point that might be picked anywhere on the image, including between voxel centers, can be computed using interpolation of this time-varying spatial vector field.

The third approach is to use the reference frame tracking approach described above. In this case, we track the image at every time frame directly to the reference time frame, and also track the reference image directly to every time frame. With

these computed motion fields, the motion between any two time frames can be readily achieved using the two-step procedure. For example, by computing the motion from the first time frame to all the later time frames, this approach provides a direct Lagrangian track for each point (pixel) in the first time frame—i.e., this result tells us precisely where these first pixels are at every time frame. It will generate erroneous tracks when the tissue disappears but will pick up the correct positions when the tissue reappears. This approach is ideal for determination of Lagrangian strain over time. With the motion field from every time frame to the reference time frame in hand, this approach also makes it straightforward to compute the Eulerian strain at any time frame.

4.4 Semi-Automatic Seed Selection

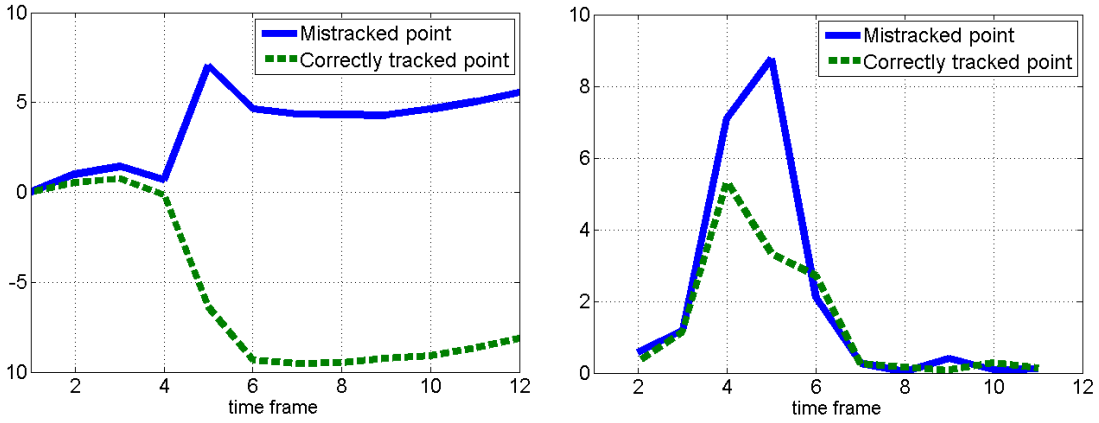
The seed point should lie within the tissue of interest and must be correctly tracked by conventional HARP tracking through all time frames. For most applications—e.g., scientific and clinical—it is essential that the seed's trajectory be manually checked before applying HARP refinement. Trial-and-error is a straightforward approach to finding a suitable seed. In this strategy, the user manually clicks on a point in one image, the point is tracked forward and backward in time using traditional HARP tracking, and then the user verifies the appropriateness of the trajectory by observing the path through all time frames. A slightly more efficient approach to finding a seed

is to have the computer suggest a putative seed and then the user needs only verify its appropriateness. We describe an approach in this section that has proven to be 100% reliable in our tests to date.

The approach begins with the user outlining a region of interest, a step that is nearly always carried out in motion analysis anyway. In the heart, the region of interest is usually the LV myocardium; in the tongue, it is the body of the tongue muscle. A list of candidate seeds $\mathbf{p}_i(t_1), i = 1, 2, ...N$ within the region of interest is then automatically produced. In the heart, the candidate seeds are equally spaced pixels on the mid-ventricular contour (which is found using morphological thinning of the ventricular wall). In the tongue, the candidate seeds comprise pixels on a coarse rectangular grid.

All candidate seeds are tracked forward to all later times $t_k, k = 2, 3, ...n$ using traditional HARP tracking. Let the location of the i^{th} candidate seed at time t_k be $\mathbf{p}_i(t_k)$. The points $\mathbf{p}_i(t_n)$ are again tracked backward to the first time frame, yielding the points $\mathbf{p}'(t_k), k = n-1, n-2, ..., 1$. Points that can be correctly tracked with traditional HARP method must satisfy the forward-backward tracking identity $\mathbf{p}_i(t_1) = \mathbf{p}'_i(t_1)$. Therefore, all tracked seeds that violate this condition are removed from the candidate seed list.

In order to choose the best point from the remaining seeds, we look at the magnitudes of their second derivatives over time. We note that mistracked points typically involve a sudden displacement at some point in the trajectory, which corresponds



(a) (b)

Figure 4.2: Comparison of correctly tracked and mistracked points in traditional HARP method. (a) The y components of the 2D displacements of a mistracked and a correctly tracked points among 12 time frames, and (b) the L1-norm of their second derivatives.

to a sudden large acceleration—i.e., second derivative. Therefore, mistracked points tend to have large second derivatives at some point in their trajectory while correctly tracked points tend to have smaller second derivatives throughout, as illustrated in Fig. 4.2. Following this observation, we pick the seed from the remaining candidate points as the one that has the minimum maximum second derivative over all time, as follows

$$\mathbf{p}_{\text{seed}} = \arg\min_{\mathbf{p}_i} \left(\max_t \left(\left\| \frac{d^2 \mathbf{p}_i(t)}{dt^2} \right\| \right) \right) . \tag{4.7}$$

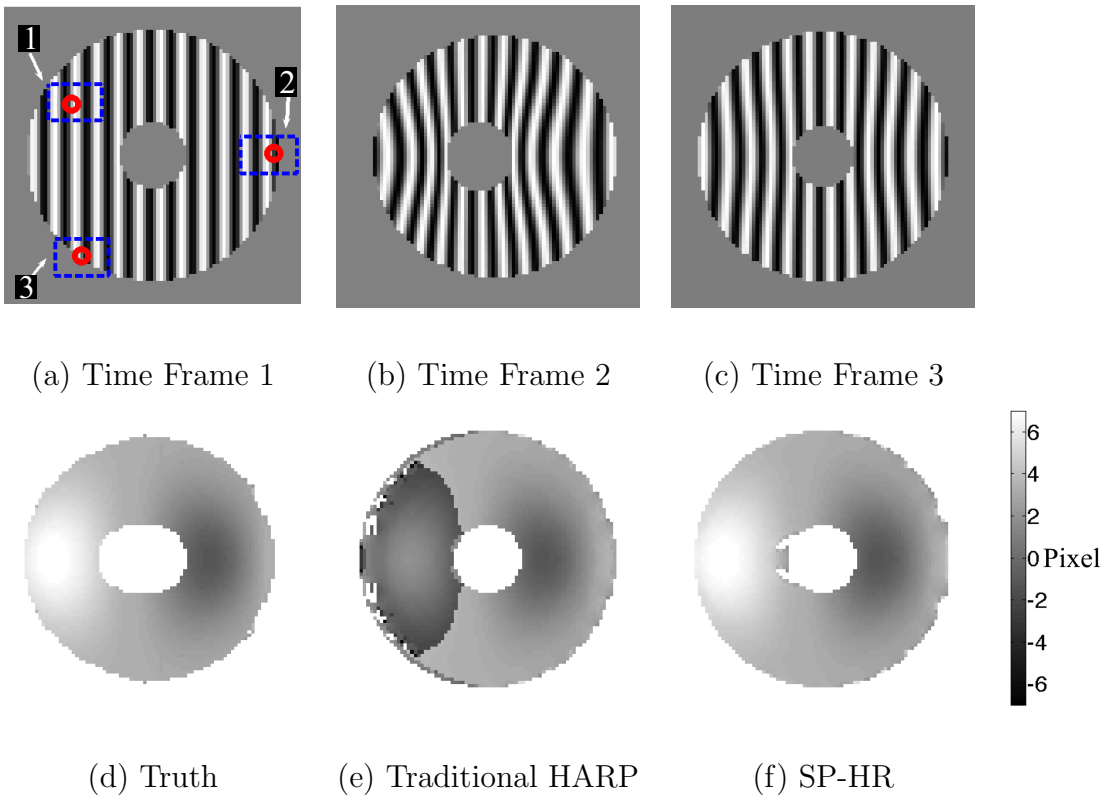
Numerical finite differences are used to approximate this derivative.

4.5 Experiment Results

4.5.1 Numerical Simulation

We first demonstrate the effectiveness of the SP-HR method using a simulated image sequence. In this simulation, the tissue is moving only in the left-right direction, so only one tag orientation is used. Three time frames were simulated, as shown in Figs. 4.3(a)–(c). Both SP-HR and traditional HARP tracking were applied and compared. In SP-HR, the two-step procedure was applied, while in the traditional HARP tracking, the points were tracked sequentially in time. The computed displacement fields from the first to the second time frames are shown in Figs. 4.3(d)–(f). We observe very large tracking errors in the traditional HARP result [Fig. 4.3(e)] on the left side of the "tissue," which are not present in the SP-HR result [Fig. 4.3(f)].

This example also serves to illustrate the three classes of mistracking that occur in traditional HARP. The three rectangular regions 1, 2, and 3 depicted in Fig. 4.3(a) are expanded and shown in Figs. 4.4(a), (b), and (c) on the left. The circles shown in Time Frame 1 were tracked to the second and third frames using the two methods and the results are shown in the second and third columns. In Region 1 the point was mistracked in traditional HARP because of a large motion between the frames, but it was correctly tracked using SP-HR. In Region 2 the point moves out of the plane in the second time frame and re-appears in the last frame due to through-plane motion. It was mistracked by both methods in the second time frame because the



(a) Time Frame 1 (b) Time Frame 2 (c) Time Frame 3

(d) Truth (e) Traditional HARP (f) SP-HR

Figure 4.3: Tracking results for simulated data, horizontal motion only. Three simulated images at time frames (a) $t = 1$, (b) $t = 2$, and (c) $t = 3$. Displacement fields (horizontal only, in units of pixels) from the first to the second time frames: (d) the true field, (e) that computed using traditional HARP tracking, and (f) that computed using SP-HR.

corresponding point does not exist. It remained mistracked by traditional HARP in Time Frame 3, but was correctly tracked by SP-HR because of its use of the reference frame. In Region 3, the tracked point is very close to the boundary. In this case, traditional HARP failed in the second time frame but recovered in the third, while SP-HR worked in all cases.

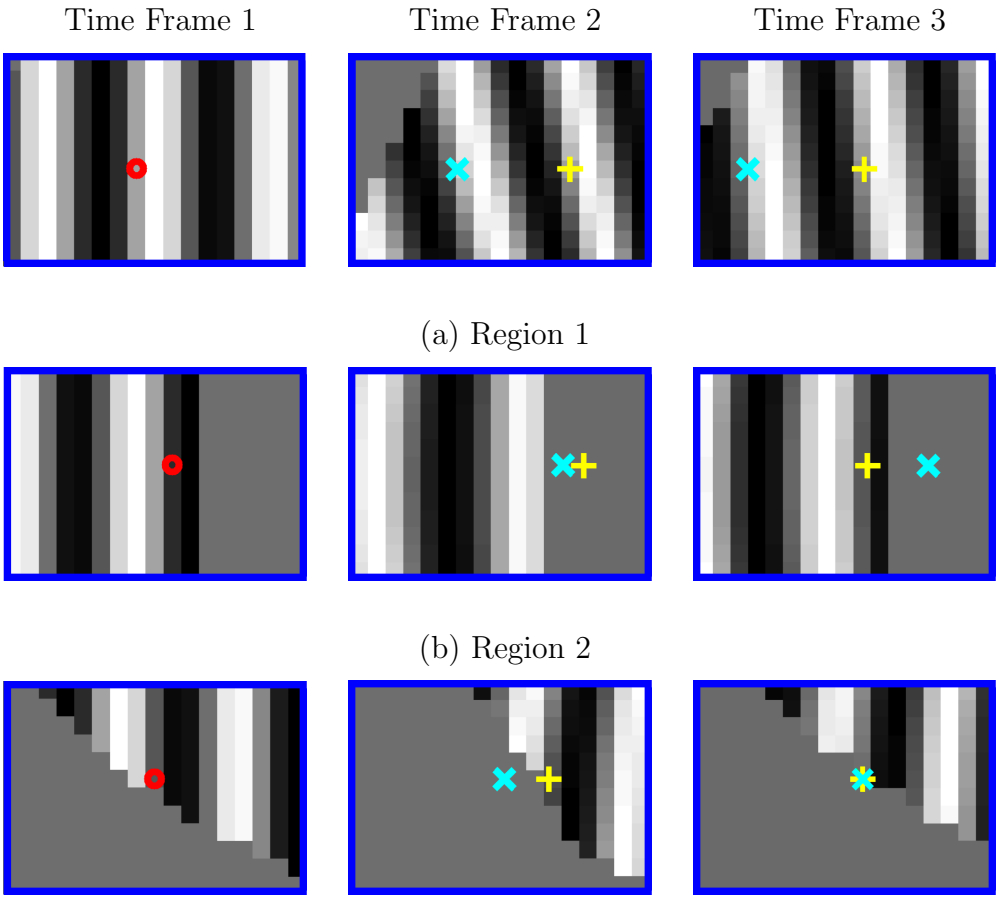


Figure 4.4: Examples of the three kinds of mistracking in traditional HARP, depicted in (a) Region 1, (b) Region 2, and (c) Region 3. The circles in Time Frame 1 are tracked into Time Frames 2 and 3 using traditional HARP ("x" symbols) and SP-HR ("+" symbols).

4.5.2 Cardiac Motion Tracking

We applied SP-HR to track the motion of the LV of the human heart. Experiments were carried out on 13 short axis cine tagged image sequences acquired over time from four different normal human subjects covering different parts of the heart. All human

subjects data were obtained with informed consent under an approved IRB protocol. In each case the tag period was 12 mm and the slice thickness was 8 mm. The numbers of time frames varied from 15 to 25, and the temporal resolutions varied from 20 ms to 43 ms. The image sizes were all 256×256, with FOVs either 300 mm or 320 mm.

To quantitatively evaluate our refinement method, we asked 20 volunteers to delineate the LVs from all the image sequences. We picked one time frame from each of the 13 image sequences. After some training (on other images), each volunteer manually drew the epicardial and endocardial contours of the LVs together with two insertion points of the RV on each of the 13 selected images. For each delineation, the myocardium region was automatically divided into epi-, mid-, and endo-cardial regions. The septal wall was then automatically divided into two pie-shaped sectors and the free wall was automatically divided into four sectors. An example of the resulting delineation is shown in Fig. 4.5.

4.5.2.1 Comparison of SP-HR and Traditional HARP

We first compared the performance of SP-HR with traditional HARP tracking in the 18 parts of the LV. The displacement fields from the first time frame to the last time frame in all image sequences were computed using both methods. In SP-HR, the seeds were automatically determined as described in Section 4.4. For all 260 delineations, we computed the average ratios of correctly tracked points to the

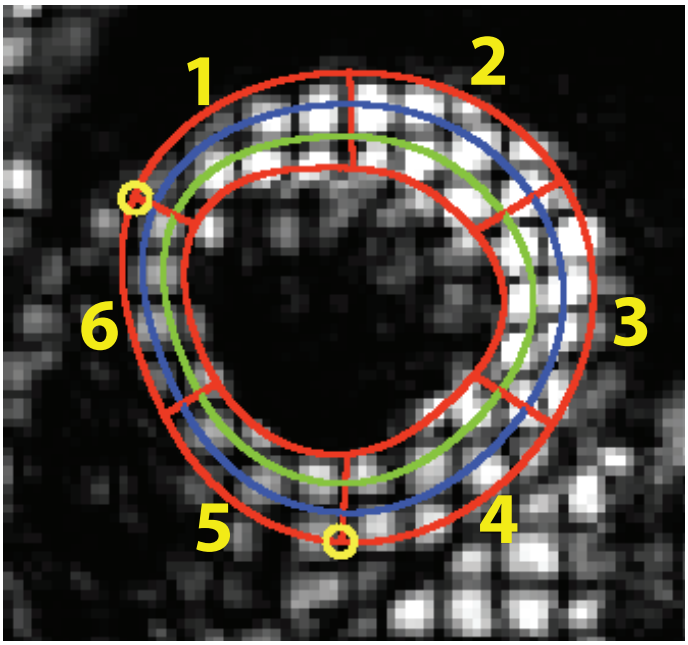
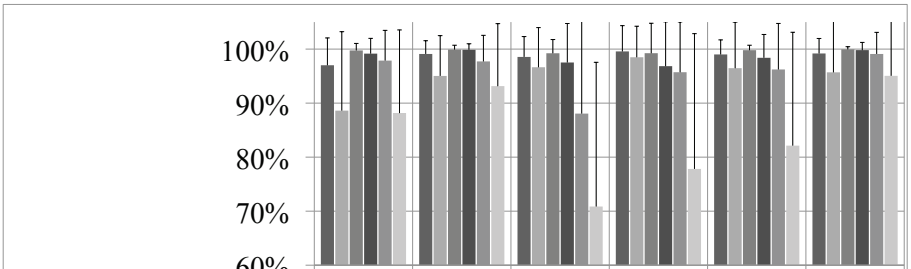


Figure 4.5: Illustration of the 18 sectors of the LV. The two small circles mark the insertion points between the left and right ventricles.

total points in each of the 18 parts. The results are shown in Fig. 4.6. It is observed that refinement worked better in all 18 regions although both methods worked almost perfectly in the midwall regions. Traditional HARP tracking generally performs worse in the endocardium than in the epicardium, while the SP-HR works almost equally well in all regions. Regarding sectors, traditional HARP tracking performs most poorly in sectors 3 and 4 on the free wall. This is because the free wall has larger motion both in the in-plane directions as well as in the through-plane direction.

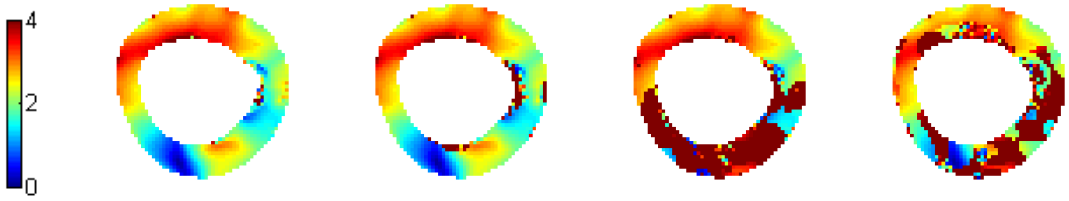
To evaluate the effectiveness of our method when there is large motion and low temporal resolution, we used subsets of the image sequences obtained by dropping intermediate images. In particular, for a time step n, the points were tracked in the



	1	2	3	4	5	6
Epicardium, SP-HR	0.9701	0.991	0.9855	0.9957	0.9901	0.9919
Epicardium, HARP	0.8861	0.9503	0.9664	0.9849	0.9647	0.957
Midcardium, SP-HR	0.9975	0.9991	0.9923	0.9924	0.998	0.9996
Midcardium, HARP	0.9916	0.9987	0.9752	0.9684	0.9839	0.9981
Endocardium, SP-HR	0.9787	0.9772	0.8806	0.9572	0.9623	0.991
Endocardium, HARP	0.8817	0.9315	0.7086	0.7782	0.8213	0.9505

Figure 4.6: The average ratios of correctly tracked points to total points in all epi-, mid-, and endo-cardial regions over the six sectors, for both SP-HR and traditional HARP tracking. Variances are represented by black lines. The numbers 1–6 represent sector number.

image series at time frames 1, $n+1$, $2n+1$, ... only. The images at the last time frames were always included so the final computed displacements could be directly compared (which means that the final time step might be less than n). By using time step other than $n = 1$, we mimic the situation of large motion and/or poor temporal resolution. In our experiments we computed the displacement fields from the first time frame to the last time frame using both SP-HR method and traditional HARP tracking with time steps 1, 2, and 3 on all 13 image sequences. For each image sequence, we determined the LV myocardium region by averaging all 20 rater's epicardial and endocardial contours, and including all pixels within these average boundaries. From the tracking results, we computed the ratio of correctly tracked points inside the LV myocardium. It is important to keep in mind that time steps are irrelevant in SP-HR

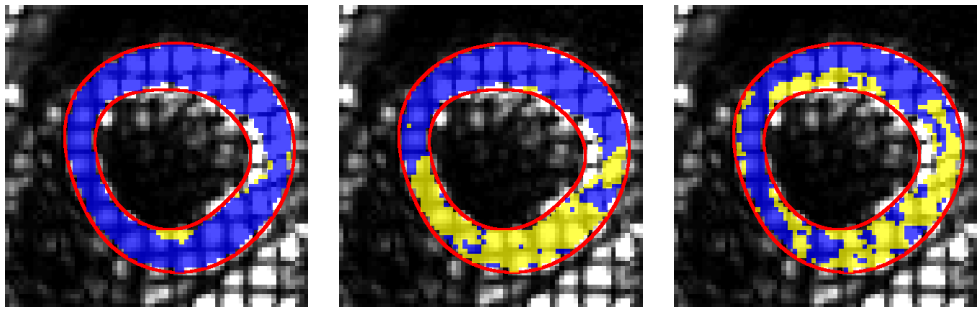


(a) (b) (c) (d)

Figure 4.7: Examples of the magnitude of displacement field computed using (a) SP-HR and traditional HARP tracking with time steps (b) 1, (c) 2, and (d) 3 respectively.

because the two-step approach is used. We note that the seed point tracked correctly using traditional HARP in all cases including the subsampled sequences.

Fig. 4.7 shows the magnitude of the motion field on one of the 13 test image sequences computed using SP-HR and traditional HARP with time steps 1, 2, and 3. Figs. 4.8 (a)–c) compare the correctly tracked regions on the same image sequence. We observe that with a time step greater than 1 traditional HARP mistracks a large portion of the LV region and performs much worse than SP-HR; for a time step equal to 1, traditional HARP performs slightly worse than SP-HR. Fig. 4.8(d) illustrates the ratios of correctly tracked points in the LV myocardium of 13 image sequences when using SP-HR and traditional HARP tracking with time steps 1, 2, and 3. It can be seen from Fig. 4.8(d) that SP-HR worked better than traditional HARP in all image sequences. On average, 98.4% points were correctly tracked using SP-HR, 93.1% in traditional HARP tracking with time step 1, 79.3% with time step 2, and 58.7% with time step 3.



(a) Time step 1 (b) Time step 2 (c) Time step 3

(d)

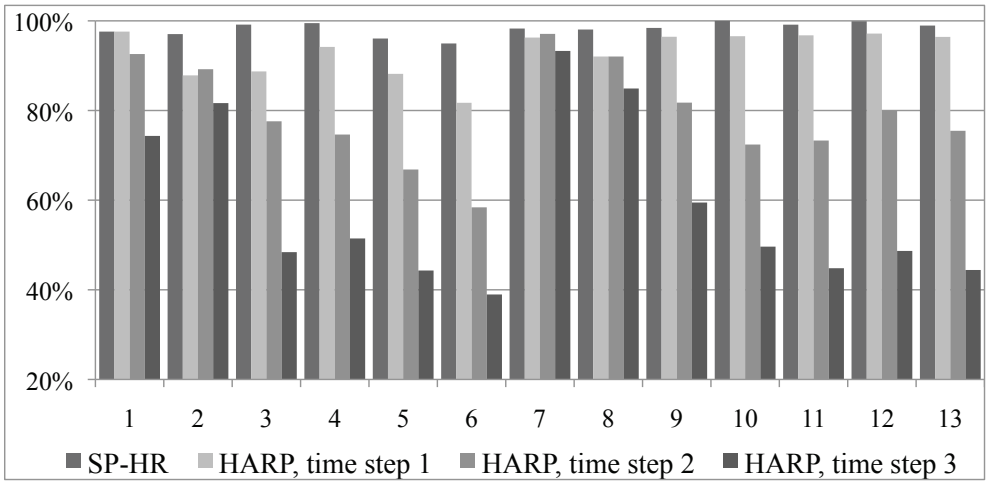


Figure 4.8: Comparison of SP-HR and traditional HARP with different time steps. Correctly tracked regions on one data set using SP-HR and traditional HARP with time steps (a) 1, (b) 2, and (c) 3. Blue: correctly tracked regions using traditional HARP with different time steps; yellow+blue: correctly tracked regions using SP-HR; Red: average LV myocardium boundaries. (d) shows the percentage of correctly tracked points on the 13 test image sequences, which is shown from left to right.

4.5.2.2 Comparison of SP-HR and RG-HR

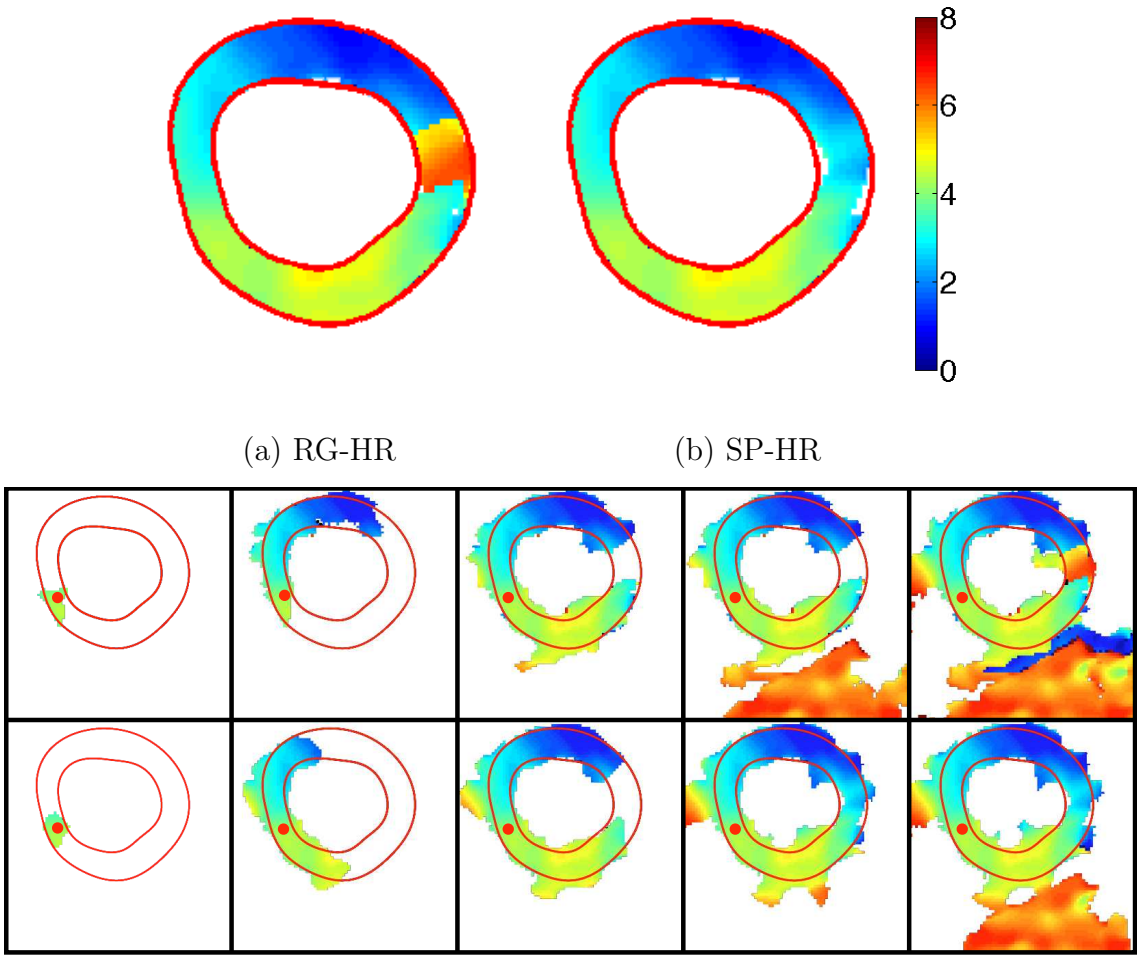
The same seed points were picked in both methods. For each image sequence, the displacement field from the time at which the contours and regions were determined to all other times was computed using both methods. In total each method was applied 208 times. Tracking was considered to be successful if the computed motion field

inside LV has no visible abrupt jump. The success rate of RG-HR was 93.8%, while SP-HR was 99.5%. Fig. 4.9 shows an example in which RG-HR fails while SP-HR succeeds. Figs. 4.9(a) and (b) show the magnitude of the displacement field inside the LV using RG-HR and SP-HR methods, respectively. We observe that RG-HR result contains abrupt jumps of displacement inside the myocardium region, which indicates errors in tracking. In contrast, the SP-HR method result is smooth throughout the whole tissue.

Fig. 4.9(c) shows the intermediate results of the two methods on the same example in Figs. 4.9(a) and (b). The evolution of the tracked regions is depicted as iteration increases from left to right. It is observed that in RG-HR, the liver—at the lower part of the image—is tracked before the LV is completely tracked while in SP-HR, points in the liver are tracked only after nearly all LV points are tracked. Because of the globally defined vertex cost function in SP-HR, the tracked region grows inside the LV with a similar speed on all sides of the seed, which is not true in RG-HR. Thus, points inside the LV are reached via an optimal path in SP-HR and are more likely to be correctly tracked.

4.5.2.3 Tracking with Two-Step Procedure

The effect of introducing the reference time frame and two-step procedure in SP-HR is shown in Fig. 4.10. In this experiment, we computed the displacement fields from the first to the last time frame using SP-HR in two ways: 1) by tracking



(a) RG-HR (b) SP-HR

(c) Tracking Order

Figure 4.9: The magnitude of the resulting displacement field inside the LV using (a) RG-HR and (b) SP-HR methods. (c) The tracking order of the (top row) RG-HR and (bottom row) SP-HR methods shown from left to right. The red dots mark the seed point.

through successive time frames and 2) by also using the two-step procedure with synthetic phase images in reference frame. It can be seen from Figs. 4.10(a) and (b) that both procedures get the same results almost everywhere inside the myocardial region. However, the two-step strategy works much better for points close to the

tissue boundary. The displacement of one point over time is shown in Fig. 4.10(c) as an example. Due to through-plane motion, this point moved out of the imaged plane from the 5th to the 10th time frames and moved back afterwards. When tracked successively, it was mistracked from the 5th time frame until the last time frame (green dashed lines). By using the two-step procedure, this point was correctly tracked after the 11th time frame (blue lines). For comparison, the displacement of one of its neighbors that was correctly tracked in both procedures is also shown (dotted lines).

4.5.3 Computation Time

The SP-HR works as fast as the RG-HR method because the computation required for the shortest path calculation is negligible in comparison to the remaining computations. I.e., it takes less than 0.2 second to track an entire 128 by 128 image for one time frame in our implementation.

4.6 Discussion

The contribution of HARP refinement methods (both SP-HR and RG-HR) is to correct the mistracking in traditional HARP tracking and to extend the regions that can be tracked correctly. For points that are correctly tracked in both HARP refinement and traditional HARP tracking, the computed motion is the same. This

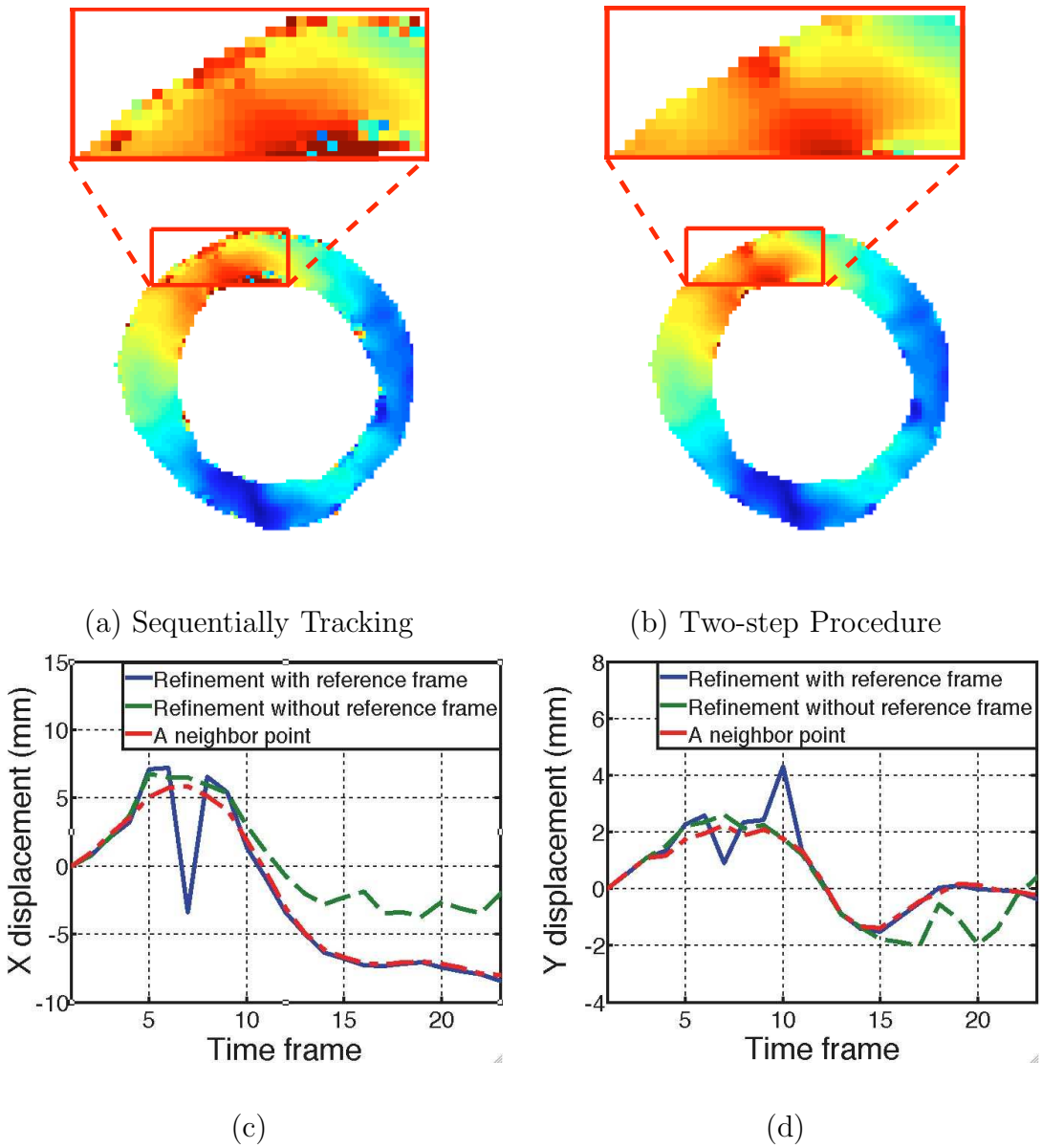


Figure 4.10: Effectiveness of the two-step approach in SP-HR tracking. Magnitude of displacement in the myocardium achieved (a) by SP-HR tracking successively from the first to last time frames and (b) by tracking first from the first time frame to the reference time frame and then from the reference time frame to the last time frame. The colormap is the same as in Fig. 4.9. The computed (c) x and (d) y displacement components between the first and last image frame for a point near the LV boundary and a (more interior) neighboring point.

is because they all rely on the same harmonic phase images and are based on the phase invariance property. This is to say, the tracking accuracy in HARP refinement methods is the same as traditional HARP, i.e., about 0.1–0.3 mm [40]. As well, with the displacement fields computed using refinement methods, the other useful quantities for functional analysis—velocity fields, strain rates, Eulerian strain, and so on—can be computed in the same way as with traditional HARP [40–42].

The two-step procedure provides a way to reduce the error caused by through-plane motion when computing point trajectories. If only the displacements from one particular time frame to all other times are of interest, the introduction of a reference time frame is not necessary. The refinement method can be directly applied between that particular time frame and all other times one by one after the seed trajectory is computed. However, we typically want to compute the displacement of a point between any two time frames, backward or forward. The introduction of reference time frame can both reduce the computation time and reduce gross tracking errors in this task.

The computed dense displacement fields from every time frame to the reference time provides the possibility of directly computing the Eulerian strain from the displacement field. In conventional HARP practice, the Eulerian strain is computed directly from the slope of the phase images [41, 42]. Now, the availability of our fast HARP refinement method provides an alternative way of Eulerian strain calculation, and this procedure may provide more insights into it. For example it has been known

that the strain computation is susceptible to noise and HARP artifacts, especially the radial strain, and it is often necessary to smooth the strain map. Previously, strain smoothing is performed by smoothing the spatial derivative of the harmonic phase images [79–81], because smoothing directly the phase images is difficult due to phase wrapping. With SP-HR, it is now possible to directly smooth the displacement field and a smoothed Eulerian strain map can be subsequently computed.

We applied our refinement methods (both RG-HR and SP-HR) only on the harmonic phase images computed using HARP method. Our methods should work with the phases images computed with other methods too, e.g., Gabor filter banks [61]. In addition, as mentioned in Chapter 3, the refinement process can be thought of as an application-specific harmonic phase unwrapping process. Therefore though not demonstrated here, both refinement methods should be applicable to DENSE motion tracking as well with very little modification (see [76, 77]).

4.7 Summary

In this chapter, we presented a new refinement method for HARP tracking by formulating a single source shortest path problem and solving it using Dijkstra's algorithm. We also developed a two-step procedure by introducing a reference time frame to help conveniently track points between any two time frames even when large displacements exist. Experimental results showed that this method can reliably track

every point inside the tissue even when there is large motion, low temporal resolution, through-plane motion, or the points are close to tissue boundary. This method is also computationally fast and makes it feasible to reliably compute other useful quantities in functional analysis of motion, for example 2D strain.

Chapter 5

Measuring 3D Tongue Motion During Speech Using zHARP

5.1 Introduction

In order to measure 3D tongue motion, existing methods all required the acquisition of multiple images in orthogonal imaging planes, which 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 [71, 72] on the whole tongue. It is 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. [69, 70] developed a new MR imaging and image

processing method for cardiac imaging 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 [70, 82]. 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 tongue motion imaging and analysis. We also used a specialized MR triggering and vocal repetition method to reduce motion artifacts. Experimental results on imaging the tongue in speech demonstrated the capability of our method on 3D tongue motion measurement and strain analysis.

This chapter is organized as follows. Section 5.2 provides background knowledge on slice-following tagging and the zHARP method. Section 5.3 describes the methodology of 3D tongue motion imaging using zHARP, which includes the implementation of zHARP on a Siemens scanner and parameter optimization, and a specialized vocal repetition and MR triggering system for artifact reduction. Section 5.4 shows results on phantom validation and *in vivo* tongue experiments. Section 5.5 provides discussion and future work. Finally, Section 5.6 summarizes this chapter.

5.2 Slice-Following and zHARP Imaging

5.2.1 Slice-Following Tagging

In contrast to standard tagging, slice-following CSPAMM (SF-CSPAMM) technique [83, 84] takes into account through-plane motion when imaging moving tissue. In standard SPAMM/CSPAMM tagging, tags are applied on the whole tissue at the initial time frame, and a thin image slice is acquired at later time frames. The location of the image slice is fixed relative to the scanner. Due to through-plane motion, the tissue that is imaged in earlier time frames may move out of the image plane so that the imaged material points at different time frames may not be the same. Thus the motion computed from SPAMM/CSPAMM images is just the apparent motion.

The principle of SF-CSPAMM is illustrated in Fig. 5.1, and Fig. 5.2 shows the pulse sequence of SF-CSPAMM. In SF-CSPAMM, only a thin slice is tagged at the initial time frame, but a thick slab is imaged later. The slab is thick enough so that the moving tagged slice is always encompassed by the slab at later time frames.

As in conventional CSPAMM, SF-CSPAMM acquires two image sequences, one with a $[+90^o, +90^o]$ tagging pulse pair, and the other with a $[+90^o, -90^o]$ tagging pulse pair (see Fig. 2.1 and Eqns. (2.1) and (2.1)). The two tagging RF pulse pairs create a tagging phase shift of π between the two image sequences, while the imaged signal in the imaged slab but outside of the tagged slice remains the same. At time

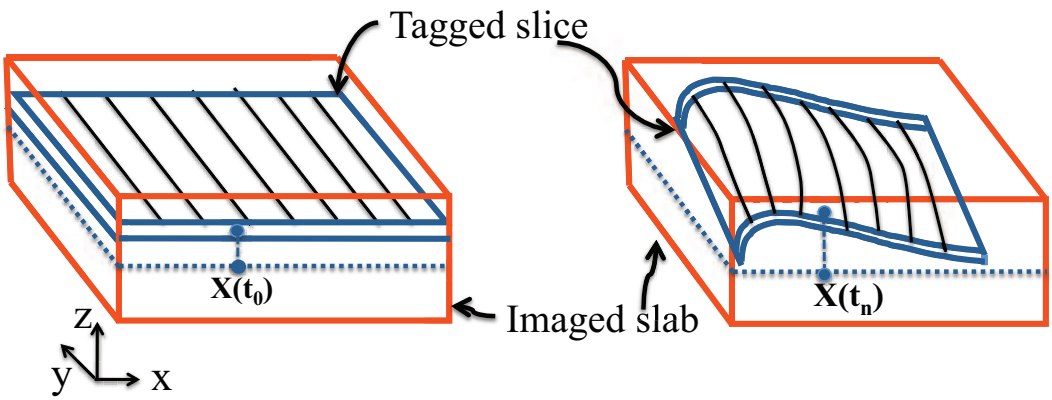


Figure 5.1: Illustration of SF-CSPAMM. The tag is applied in a thin slice at the initial time frame (left), and the image is acquired on a thick slab. At later time frame t_n (right), the tagged slice is deformed in 3D. The location of the material point $\mathbf{x}(t_n)$ at t_n in the image is its projected location on the image plane.

t, the acquired images from the two image sequences can be written as:

$$I_1(\mathbf{x}, t) = A(\mathbf{x}, t) + T(\mathbf{x}, t)\,, \tag{5.1}$$

$$I_2(\mathbf{x}, t) = B(\mathbf{x}, t) + T(\mathbf{x}, t)\,, \tag{5.2}$$

where $A(\mathbf{x}, t)$ and $B(\mathbf{x}, t)$ are the magnetization in the tagged slice as defined in Eqns. (2.1) and (2.2), and $T(\mathbf{x}, t)$ is the magnetization in the imaged slab but out of the tagged slice. By subtracting the two images, the signal that is not in the tagged slice ($T(\mathbf{x}, t)$) is canceled. The SF-CSPAMM image can be represented as:

$$I_{\mathrm{SF}} = A(\mathbf{x}, t) - B(\mathbf{x}, t) = 2M_0(\mathbf{x}, t)e^{-t/T_1}\cos(\boldsymbol{\omega}^T\mathbf{p}(\mathbf{x}, t))\,, \tag{5.3}$$

where $M_0(\mathbf{x}, t)$ represents the transverse magnetization, $\boldsymbol{\omega}$ is the tag frequency, and $\mathbf{p}(\mathbf{x}, t)$ is the 3D position of the imaged material point $\mathbf{x} = (x, y)$ at the initial (reference) time frame. This is similar to the CSPAMM image (Eqn. (2.3)) except

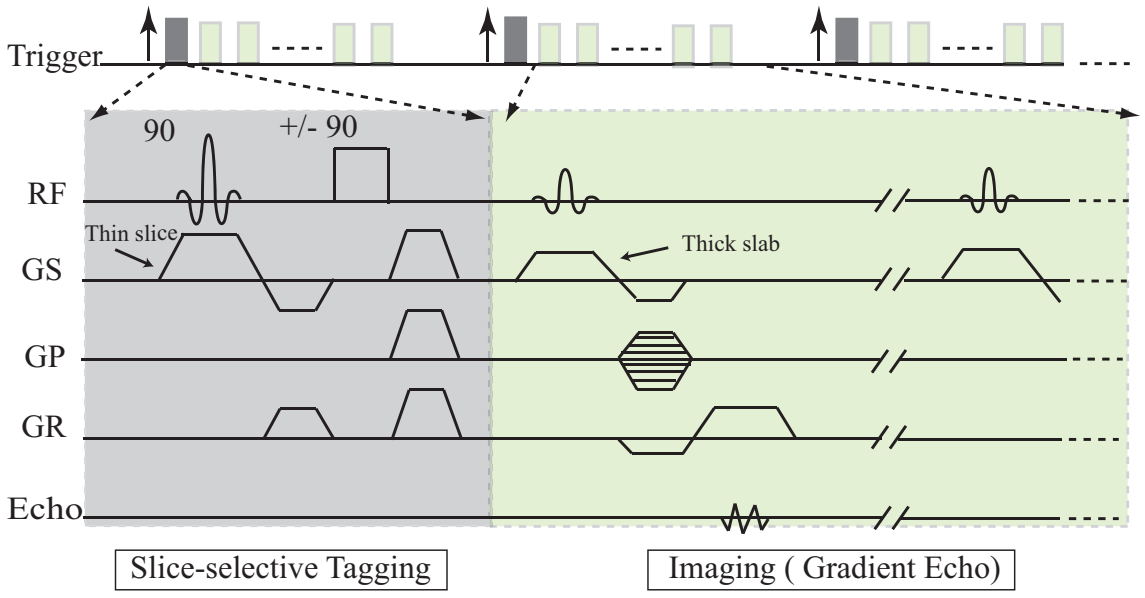


Figure 5.2: The pulse sequence of SF-CSPAMM.

that the imaged point $\mathbf{p}$ is the material point and moves in 3D in space, while in CSPAMM the imaged points are located at fixed spatial locations. In this way the SF-CSPAMM encodes the true in-plane motion of the material points by following the moving slice. 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.

5.2.2 zHARP Imaging

ZHARP [69, 70] is built on SF-CSPAMM technique and it encodes both in-plane (x and y components) and through-plane motion (z component) in one single image slice without affecting the image acquisition speed. The zHARP imaging sequence adds to the SF-CSPAMM tagging sequence a small z-encoding gradient immediately

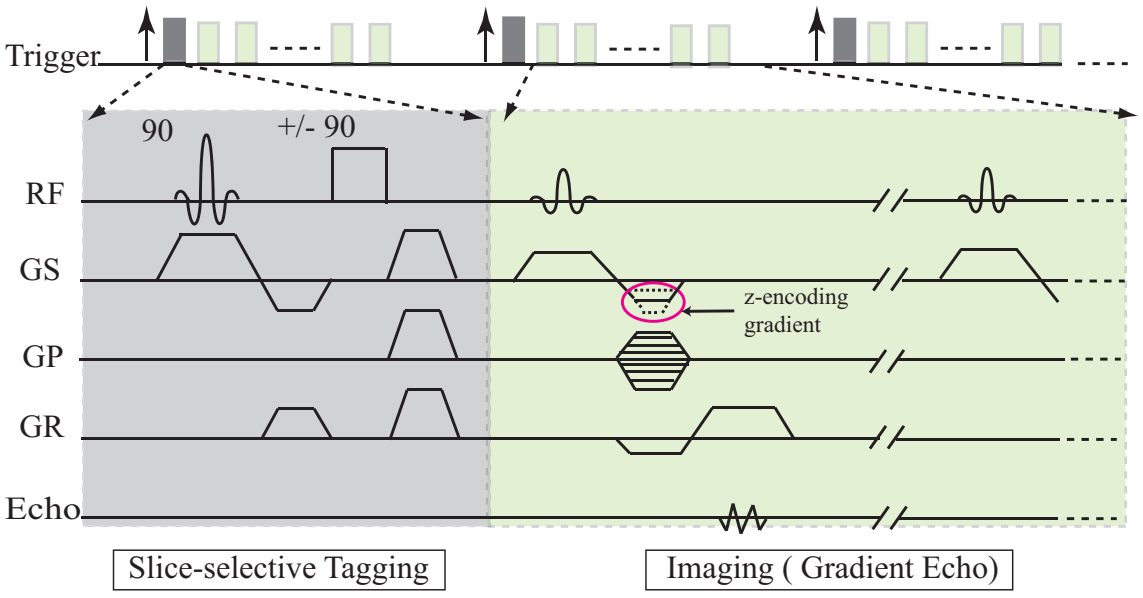


Figure 5.3: The zHARP pulse sequence.

before the readout. This gradient adds to all material points in the tagged slice a z-position dependent phase ϕ_z, which is linearly related to the location of the point in z direction at the time when the image is acquired. The zHARP pulse sequence is illustrated in Fig. 5.3.

As well, the z-encoding gradients applied in the orthogonally tagged images have the same magnitude but opposite polarity. The two acquired images can be expressed as:

$$I^x_{\text{zHARP}} = 2M(\mathbf{x},t)\cos(\boldsymbol{\omega}_x^T\mathbf{p}(\mathbf{x},t))e^{j\phi_z(\mathbf{x},t)}e^{j\phi_e(\mathbf{x})}\,, \tag{5.4}$$

$$I^y_{\text{zHARP}} = 2M(\mathbf{x},t)\cos(\boldsymbol{\omega}_y^T\mathbf{p}(\mathbf{x},t))e^{-j\phi_z(\mathbf{x},t)}e^{j\phi_e(\mathbf{x})}\,, \tag{5.5}$$

where $\boldsymbol{\omega}_x$ and $\boldsymbol{\omega}_y$ are the tag frequencies in the two orthogonal directions and $\phi_e(\mathbf{x})$

is the phase error that is caused by susceptibility and magnetic field inhomogeneities. Letting k_z be the z-encoding frequency, and $z(\mathbf{x})$ be the z location of the material point $\mathbf{x}$ at time t, then

$$\phi_z(\mathbf{x}, t) = k_z z(\mathbf{x}, t)\,. \tag{5.6}$$

5.2.3 zHARP Image Processing

The phase images that encode x, y, and z components of 3D motion can be separated from the zHARP images using the 2D HARP concept [69, 70]. Example zHARP images are shown in Figs. 5.4(a) and (b). Instead of applying the bandpass filter on only one of the harmonic peaks, the bandpass filter is applied to both positive and negative harmonic peaks of each of the two orthogonally tagged images I^x_{zHARP} and I^y_{zHARP} as shown in Figs. 5.4(c) and (d). We then get four harmonic images from filtering of the four harmonic peaks (A-D in Figs. 5.4(c) and (d)):

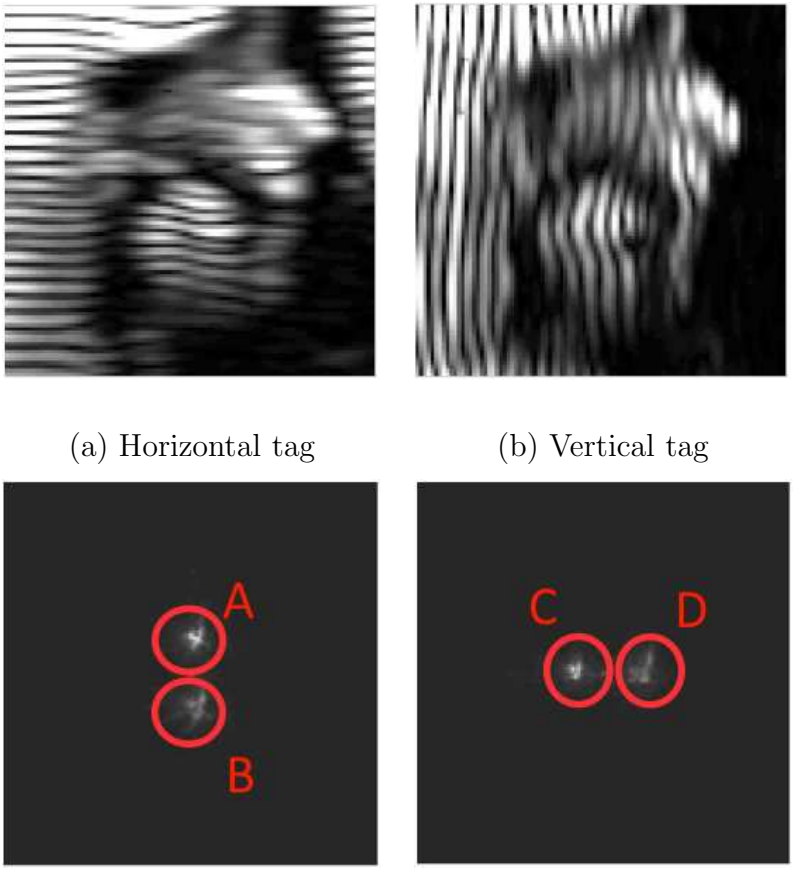
$$I_x^-(\mathbf{x}, t) = D(\mathbf{x}, t)e^{j(-\phi_x(\mathbf{x},t)+\phi_z(\mathbf{x},t)+\phi_e(\mathbf{x},t))}\,, \tag{5.7}$$

$$I_x^+(\mathbf{x}, t) = D(\mathbf{x}, t)e^{j(\phi_x(\mathbf{x},t)+\phi_z(\mathbf{x},t)+\phi_e(\mathbf{x},t))}\,, \tag{5.8}$$

$$I_y^-(\mathbf{x}, t) = D(\mathbf{x}, t)e^{j(-\phi_y(\mathbf{y},t)-\phi_z(\mathbf{x},t)+\phi_e(\mathbf{x},t))}\,, \tag{5.9}$$

$$I_y^+(\mathbf{x}, t) = D(\mathbf{x}, t)e^{j(\phi_y(\mathbf{y},t)-\phi_z(\mathbf{x},t)+\phi_e(\mathbf{x},t))}\,, \tag{5.10}$$

where ϕ_x and ϕ_y are the harmonic phases from the horizontal and vertical tagging, and ϕ_z is the phase arising from the through-plane motion. The phase images can



(a) Horizontal tag (b) Vertical tag

(c) Fourier transform of (a) (d) Fourier transform of (b)

Figure 5.4: The example tongue (a)-(b) zHARP images and (c)-(d) their Fourier transform. The four harmonic peaks marked as A, B, C, and D are extracted by bandpass filtering and used to compute the horizontal, vertical and z phase images.

then be computed from these four complex images:

$$\phi_x = \frac{1}{2}\angle(I_x^+(\mathbf{x},t)\cdot\overline{I_x^-(\mathbf{x},t)})\,, \tag{5.11}$$

$$\phi_y = \frac{1}{2}\angle(I_y^+(\mathbf{x},t)\cdot\overline{I_y^-(\mathbf{x},t)})\,, \tag{5.12}$$

$$\phi_z = \frac{1}{4}\angle(I_x^+(\mathbf{x},t)\cdot I_x^-(\mathbf{x},t)\cdot\overline{I_y^-(\mathbf{x},t))}\cdot\overline{I_y^-(\mathbf{x},t)})\,, \tag{5.13}$$

$$\phi_e = \frac{1}{4}\angle(I_x^+(\mathbf{x},t)\cdot I_x^-(\mathbf{x},t))\cdot I_y^+(\mathbf{x},t)\cdot I_y^-(\mathbf{x},t))\,, \tag{5.14}$$

where $\angle$ is the function that calculates the phase of a complex number, and $\overline{I}$ is the complex conjugate. With the phase images ϕ_x, ϕ_y, and ϕ_z, the 3D motion of the material points can be computed. The random phase error ϕ_e is not used for the motion tracking.

5.2.4 3D Motion Tracking

3D motion tracking from zHARP images is executed in two steps. First, the in-plane motion is computed using conventional 2D HARP tracking and HARP refinement. For a material point $\mathbf{x}(t)$ that is tracked from time t_i to t_{i+1}, denote its locations at the two times as $\mathbf{x}_i = [x_i, y_i, z_i]$ and $\mathbf{x}_{i+1} = [x_{i+1}, y_{i+1}, z_{i+1}]$, and the displacement as $\mathbf{u} = [u_x, u_y, u_z] = \mathbf{x}_{i+1} - \mathbf{x}_i$. The in-plane displacement components u_x and u_y are computed from the phase images $\phi_x(x, y, t_i)$, $\phi_y(x, y, t_i)$, $\phi_x(x, y, t_{i+1})$, and $\phi_y(x, y, t_{i+1})$ based on the phase invariant property using HARP tracking. Hence we can get $x_{i+1} = x_i + u_x$ and $y_{i+1} = y_i + u_y$.

In the second step, the through-plane motion is computed based on the fact that the z-phase ϕ_z is a linear function of the material point's location in the through-plane direction, i.e.,

$$\phi_z(x_{i+1}, y_{i+1}, t_{i+1}) - \phi_z(x_i, y_i, t_i) = k_z(z_{i+1} - z_i)\,. \tag{5.15}$$

Since the phase image is wrapped, and assuming the material point does not move much between neighboring time frames, the through plane motion u_z can be computed

by rewriting Eq. (5.15) as:

$$u_z = z_{i+1} = z_i + \frac{1}{k_z} W\{\phi_z(x_{i+1}, y_{i+1}, t_{i+1}) - \phi_z(x_{i+1}, y_i, , t_i)\}\,. \tag{5.16}$$

5.2.5 3D Strain Computation

The full 3D strain tensor can be measured using two parallel zHARP slices [82,85] without explicitly computing the displacement $\mathbf{u}$. Instead, the gradient of displacement $\nabla\mathbf{u}$ can be readily computed from the HARP images. From Eqs. (5.3) and (5.6), we have

$$\mathbf{\Phi}(\mathbf{x}, t) = \begin{bmatrix} \phi_x \\ \phi_y \\ \phi_z \end{bmatrix} = \begin{bmatrix} \omega & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & k_z \end{bmatrix} \mathbf{p}(\mathbf{x}, t)\,. \tag{5.17}$$

Therefore, the displacement gradient $\mathbf{F}$ can be directly computed from the phase images as follows:

$$\mathbf{F}^{-1} = \nabla\mathbf{p}(\mathbf{x}, t) = \begin{bmatrix} \omega^{-1} & 0 & 0 \\ 0 & \omega^{-1} & 0 \\ 0 & 0 & k_z^{-1} \end{bmatrix} \nabla\mathbf{\Phi}(\mathbf{x}, t)\,. \tag{5.18}$$

Given two parallel zHARP slices, the 3D HARP vector images $\mathbf{\Phi}^{(1)}(\mathbf{x}, t)$ and $\mathbf{\Phi}^{(2)}(\mathbf{x}, t)$ are first calculated on a regular grid $\mathbf{x} = (i, j), 1 \le i, j \le N$. Then the phase gradient is calculated using finite differences as follows:

$$\nabla\phi_k(\mathbf{x}, t) = \frac{1}{2}[\nabla\phi_k^{(1)}(\mathbf{x}, t) + \nabla\phi_k^{(2)}(\mathbf{x}, t)]\,, \tag{5.19}$$

where

$$\nabla\phi_k^{(n)}(\mathbf{x},t) = \begin{bmatrix} \phi_k^{(n)}(i+1,j,t) & - & \phi_k^{(n)}(i,j,t) \\ \phi_k^{(n)}(i,j+1,t) & - & \phi_k^{(n)}(i,j,t) \\ \phi_k^{(2)}(i,j,t) & - & \phi_k^{(1)}(i,j,t) \end{bmatrix}^T . \tag{5.20}$$

The 3D strain tensor can be computed using Eqs. (5.18) and (2.19).

5.3 Measuring 3D Tongue Motion Using zHARP

5.3.1 zHARP Pulse Sequence Implementation

We implemented zHARP using a gradient echo sequence on a 3T Siemens Tim-Trio MRI scanner (Siemens Medical Solutions, Malvem, PA) equipped with twelve receiver channels. The pulse sequence is shown in Fig. 5.3. 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 before readout.

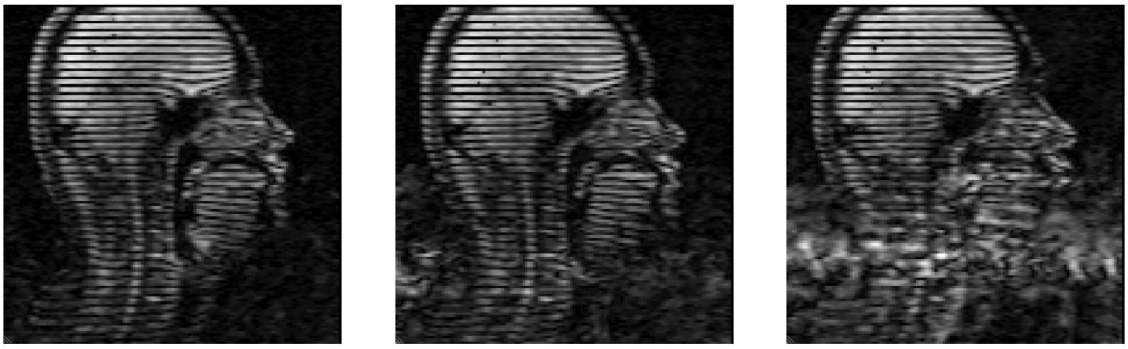
For one image set, the same acquisition is repeated 4 times with combinations of +/- 90^o second tagging RF pulse, and phase encoding direction in x or y-axis—i.e., the readout is in y or x-axis, respectively—with different z-encoding gradient polarity. We also implemented a ramped flip angle technique (see [83]) to improve

the tag persistence and achieve a constant tag contrast during acquisition.

5.3.2 Motion Artifact and Parameter Optimization

Slice-following tagging images have lower SNR comparing to conventional tagging because the signals from unexcited tissue in slice-selective tagging remain out of the imaged slice through time. They are also much more sensitive to motion artifacts than conventional tagging. ZHARP suffers the same problems because it is build upon slice-following tagging.

Motion artifacts are caused by inconsistent repetitions. Fig. 5.5 shows example SF-CSPAMM images 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. From the images it is clear that the artifacts are more severe for a thicker slab, although a thicker slab is often desired because it can capture more through-plane motion. 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 they are greatly amplified because the signal strength of the tagged slice is much weaker than the imaged slab. Thus motion artifacts are more severe when the imaged slab is thicker.



(a) (b) (c)

Figure 5.5: 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.

In addition, motion artifacts are more severe in later time frames than earlier time frames. The tissue in the thick slab but out of the tagged slice has no signal immediately after tagging, since the magnetization is rotated onto the transverse plane by the non-selective 90^o RF pulse in the slice-selective tagging. Because of T1 recovery, the signal from untagged tissue gains over time, while the tagging signal decays in the meantime. Therefore the artifacts are amplified much more in later time frames.

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 least in the left-right direction and most in the anterior-posterior direction during normal speech, it is best to acquire zHARP images in the sagittal plane and acquiring images in the coronal plane should be avoided. As well, motion artifacts can be lessened by reducing the number of repetitions required to image one

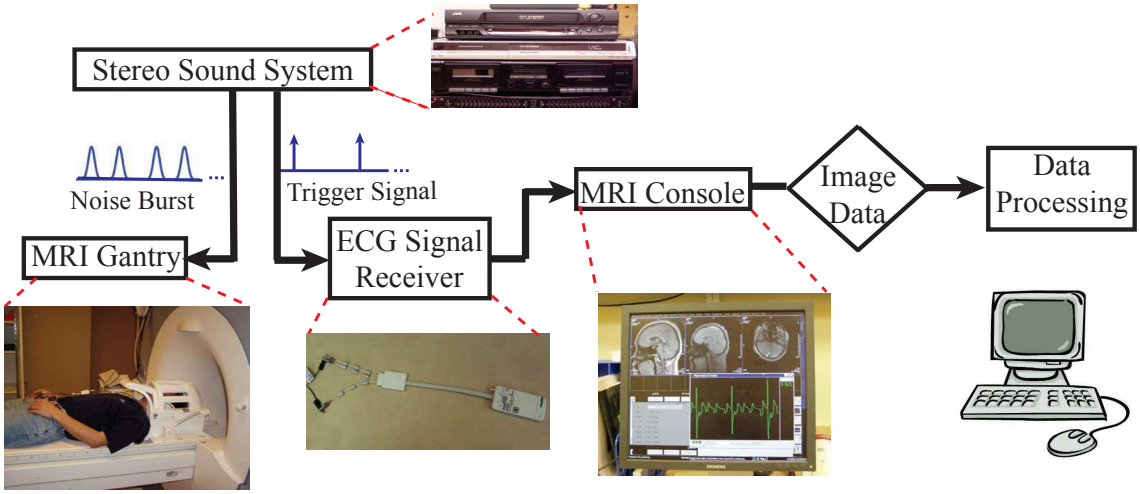


Figure 5.6: The system setup diagram.

slice, although that necessitates a sacrifice of temporal resolution or spatial resolution, which may be undesirable.

5.3.3 MR Trigging and Vocal Repetition Method

To reduce the inconsistency in vocal repetitions, we design a vocal repetition and MR triggering system similar to Masaki et al. [28] to help synchronize the tongue motion in multiple repetitions. This system is also used to trigger the scanner so as to coordinate the image acquisition. The system setup is shown in Fig. 5.6. In the system, a continuous periodic stereo sound wave is generated and played on a stereo sound system during image acquisition.

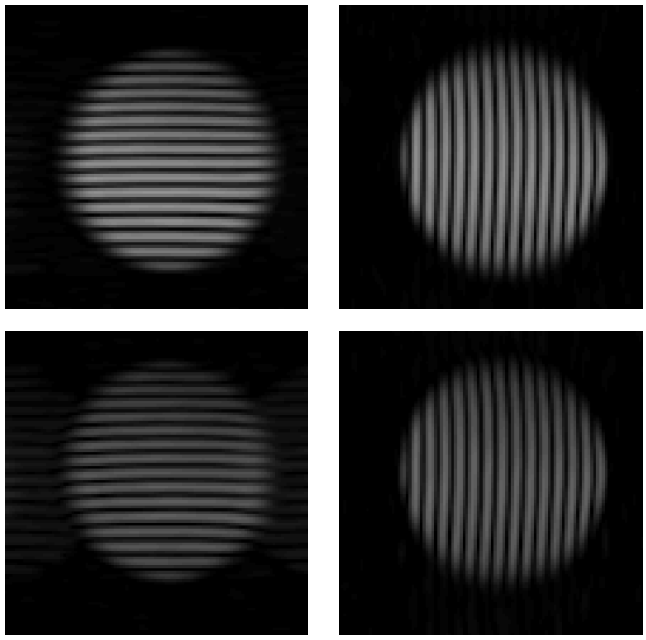
One channel of the sound simulates the electrocardiogram (ECG) trigger signal, and is connected to a ECG signal receiver, which in turn triggers the scanner to run

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 seconds. 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.

5.4 Experiments and Results

5.4.1 Phantom Validation

Our implementation of the zHARP sequence was first validated on phantom experiments. The phantom was connected to a motor, and moved approximately sinusoidally back and forth along the main magnetic field direction with a period of 1.02 second. ZHARP 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 was exhibited. Fig. 5.7 shows the orthogonally tagged image pairs at two time frames, and Fig. 5.8 shows the phase images computed from zHARP processing



(a) Time Frame 1

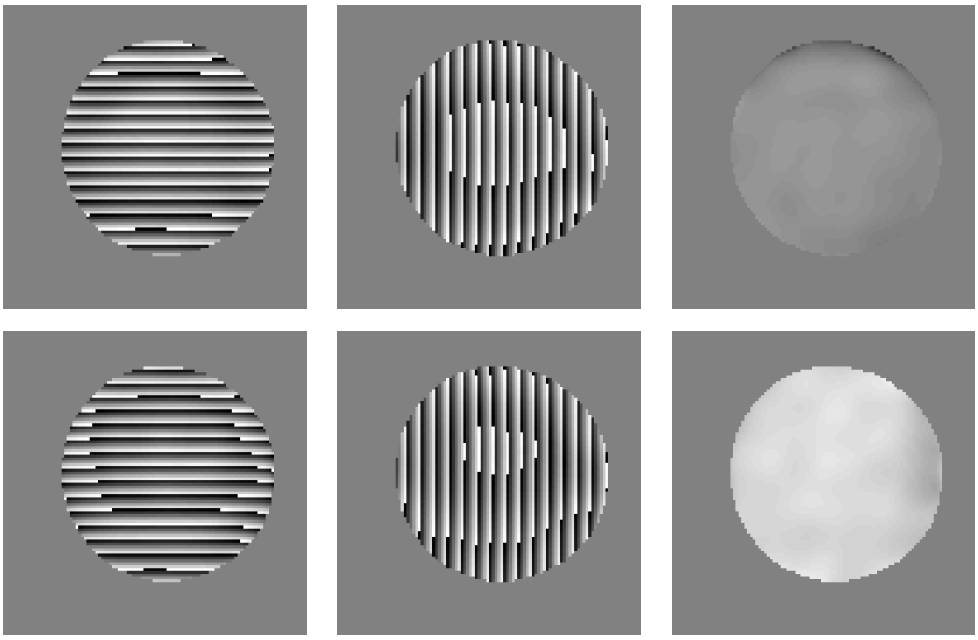
(b) Time Frame 10

Horizontal tag Vertical tag

Figure 5.7: The zHARP images of the phantom at two time frames. The images with horizontal tags are acquired with positive z-encoding, and the images with vertical tags are acquired with negative z-encoding.

at these two time frames.

For validation, we also 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 3D motion computed from zHARP 3D tracking. The comparison results are shown in Fig. 5.9. 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.

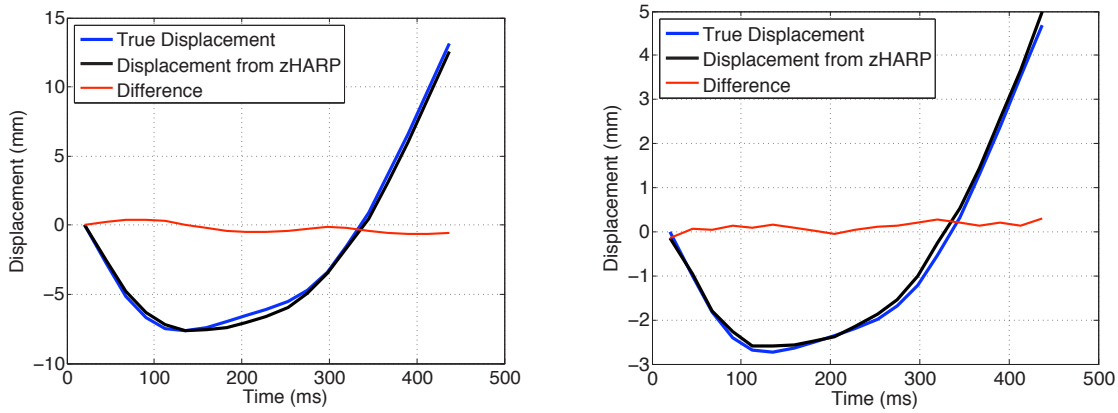


(a)ϕ_x (b) ϕ_y (c) ϕ_z

Figure 5.8: The zHARP processing results of the phantom images at two time frames. Top row: time frame 1; bottom row: time frame 10. (a) and (b) are the in-plane harmonic phase images, and (c) is the z-phase. The phase images are masked so that only phases inside the phantom are shown.

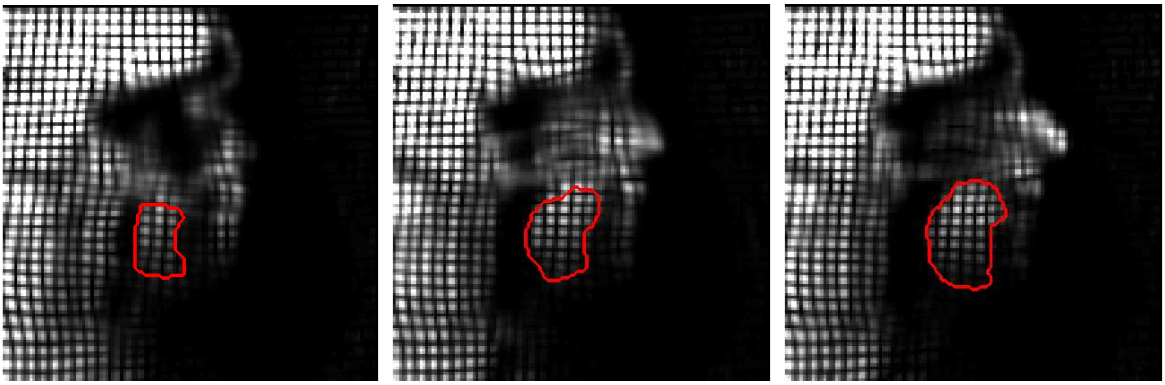
5.4.2 3D Tongue Motion Imaging

In our *in vivo* experiments, the speech material studied was "eeoo". The subjects were 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, number of time frames = 12. The acquired k-space matrix size was 64×22 and 11 k-space lines were acquired during each repetition. The images were then reconstructed into 128×128 matrices



(a) In-plane motion. (b) Through-plane (z) motion.

Figure 5.9: The comparison of average in-plane and through-plane motion computed from zHARP with standard CSPAMM tagging.



(a) (b) (c)

Figure 5.10: The three slices of the acquired zHARP images on a normal subject at the first time frame after the tag is applied. The grid pattern is generated by overlaying the two images tagged in orghogonal orientations. The red contours are the manually selected tongue regions that were analyzed.

by zero padding k-space. Each zHARP slice was acquired in 4 vocal repetitions in each of the two tag directions.

In the first experiment, three sagittal slices were acquired. The three acquired

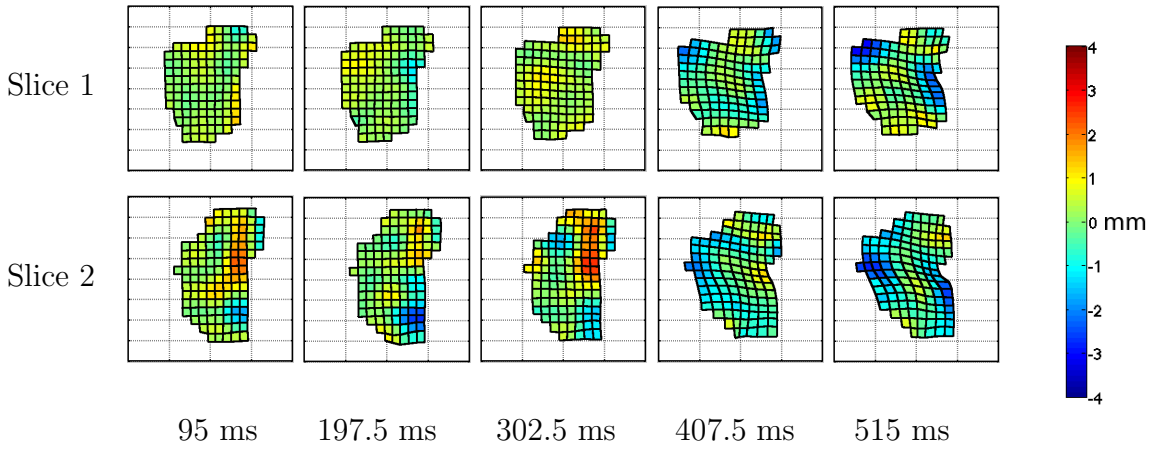


Figure 5.11: The 3D displacement maps of two sagittal slices close to the mid-sagittal plane of the tongue at different time during speech. The mesh grid shows the in-plane 2D displacement, and the color of the deformed mesh shows the through-plane displacement.

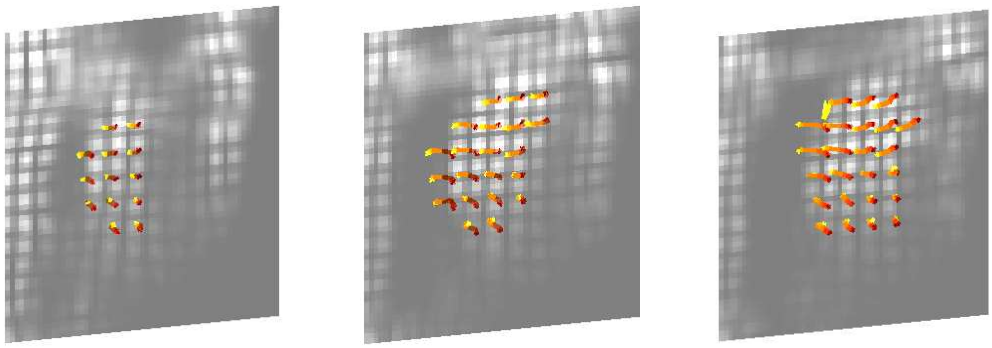


Figure 5.12: The path lines of selected points inside the tongue in the three sagittal slices. The brighter color means later time frame. The lip is toward the right.

sagittal slices at the first time frame are shown in Fig. 5.10. 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 are shown in Fig. 5.11. Since the tongue does not move much in the left-right direction, the through-plane motion in these sagittal slices is small. Fig. 5.12 shows the 3D path

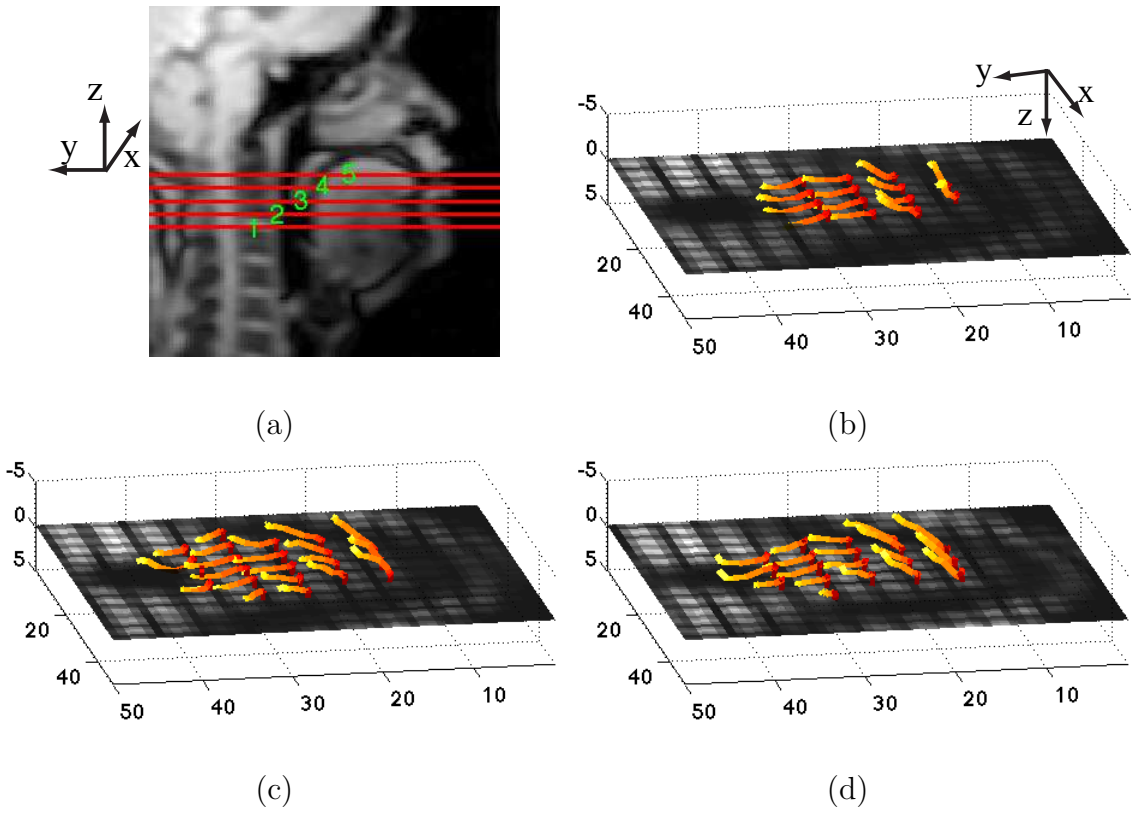
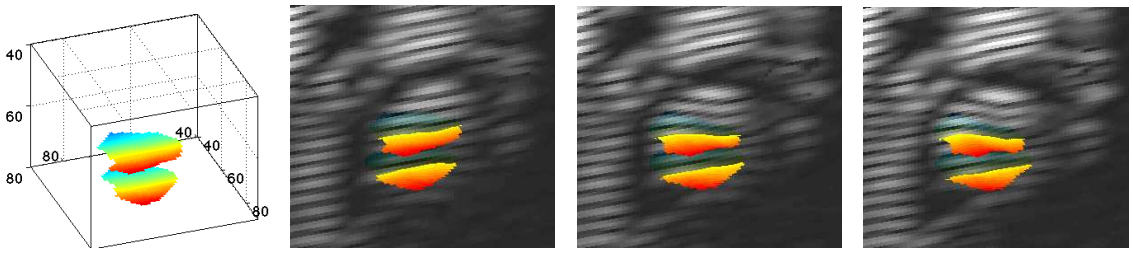


Figure 5.13: (a) The locations of the 5 axial slices in the tongue. (b)-(d) the path lines of selected points inside the tongue in axial slice 2, 3, and 4, with lip toward the right.

lines of selected points inside the tongue.

To better reveal zHARP's ability to measure through-plane motion, we carried out another experiment in which five axial slices were acquired (although this orientation is generally not recommended). The imaging parameters were the same as in the first experiment. These axial slices are perpendicular to the sagittal image shown in Fig. 5.13 (a), and their intersections are shown as red lines. Figs. 5.13(b)-(d) show the 3D path lines of selected 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



(a) 95 ms (b) 305 ms (c) 410 ms (d) 512.5 ms

Figure 5.14: The 3D tongue motion shown on two axial images. 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 shown sagittal image plane.

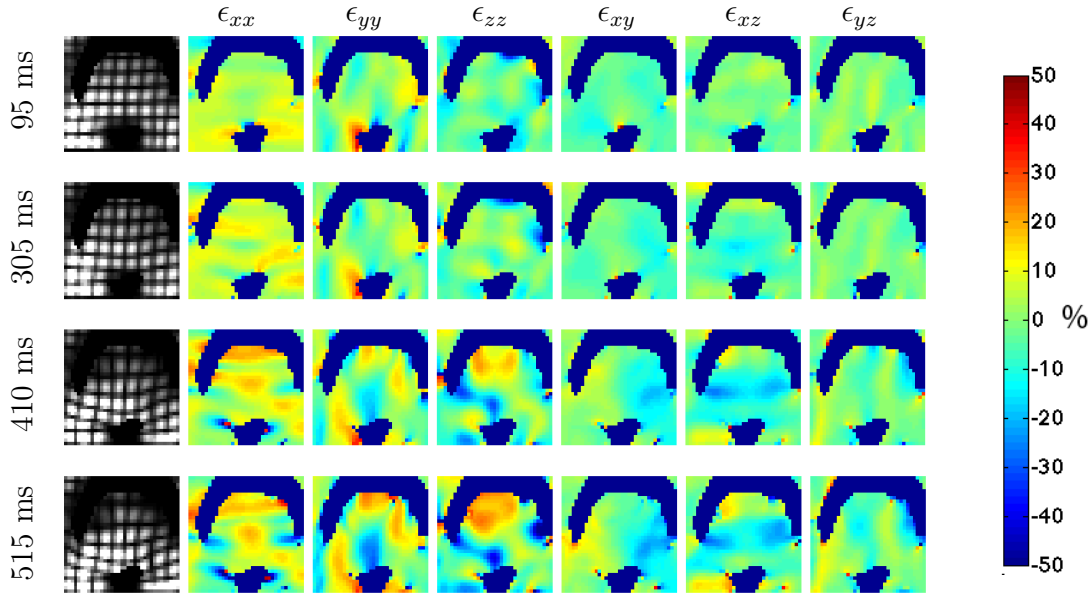


Figure 5.15: The strain tensor components on one axial slice at different time frames. The tongue tip is pointing upward. x goes from top to bottom, y goes from left to right, and z is in the through plane direction.

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. 5.14(a). Figs. 5.14(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. 5.15.

5.5 Discussion

The zHARP imaging and 3D strain computation has been validated by Abd-Elmoniem et al. [69,70,70,82]. Therefore in this work we did not perform an extensive validation of the zHARP method. Instead, a single phantom experiment was executed to just validate that our implementation of the zHARP sequence using gradient echo on a Siemens scanner runs correctly.

As shown, zHARP imaging is very sensitive to motion artifacts because it uses slice-following tagging. The motion consistency of the tongue is improved using the specialized triggering system, but it remains a problem especially for patient studies because the patients often have trouble making consistent repetitions of the utterance. The zHARP image quality may be further improved by reducing the number of repetitions using advanced imaging techniques. However, since the computation of the three phase images in zHARP requires at least four image sets (two in each tag direction), the number of vocal repetitions required for zHARP imaging and motion

tracking cannot be smaller than 4 and the motion artifacts cannot be completely removed.

As demonstrated a full 3D strain tensor can be computed from the 3D motion of two neighboring slices. The 3D strain provides another tool to help better understand the tongue function and tongue muscle activation, which may not be possible or accurate without the 3D information.

5.6 Summary

In this chapter, we presented an approach for fast imaging and measurement of 3D tongue motion and strain during speech using zHARP. In this approach, an zHARP imaging sequence was implemented using gradient echo on a Siemens scanner, and imaging parameters were optimized to reduce motion artifacts. To account for the poor repeatability of tongue motion, a specialized vocal repetition and MR scanner triggering system was designed. The pulse sequence and the triggering system was validated using a phantom study and its effectiveness was demonstrated using experiments on tongue during speech.

Chapter 6

Tracking Tongue Motion in 3D Using HARP

6.1 Introduction

Despite the existence of fast 3D motion imaging method, i.e., zHARP, tagged MRI remains the prevailing method in imaging muscle motion because of high image quality. To measure the 3D motion using tagged MRI, images must be acquired with different tag orientations and in different image slice orientations, as explained in Section 1.4. In scientific research or clinical medicine these images are usually acquired sparsely in space. Although dense methods for directly imaging three-dimensional motion have been developed [66,70,86], these methods require too much time for routine acquisition of dense, three-dimensional motion. Therefore interpolation methods

are likely to be required in practice and may well be the critical element in promoting routine imaging of 3D muscle function in the heart and the tongue.

The dense 3D motion of the tissue can be inferred from tagged MRI [43,47,53,57–59,64,65] through model-based or model-free interpolation of the sparse imaging data (see also [87]). These methods compute the 3D motion of every point in the tissue, and often require extensive computation time. The 3D-HARP method developed by Pan et al. [43] used a sparse 3D mesh model created specifically for the cardiac shape, and tracked the 3D cardiac motion only on the sparse mesh instead of the complete myocardium. The method needs no complicated modeling, and thus it is fast and requires minimal human interaction.

In this work, we develop a method to track 3D tongue motion by extending the 3D-HARP method. In the original 3D-HARP method, the mesh was cup-shaped in order to resemble the shape of the left ventricle of the heart. In contrast, our approach uses small planar rectangular meshes that can be placed anywhere in the tongue. In order to track motion, the approach of Pan et al. formed a deformation field on their mesh by extrapolating sparse and partial displacement information measured at intersections of the mesh with the (tagged) image planes. The extrapolation was formed using a fixed Gaussian kernel whose size was chosen empirically based on the mesh size and separation of the observed images. In our approach, we use a thin plate spline in order to interpolate (and extrapolate) the sparse displacement information. This methodology is more appropriate to the geometry of our meshes

and less arbitrary in its selection of parameters. Results on 3D tongue images acquired during speech show that 3D motion can be tracked smoothly and consistently using our new method.

This chapter is organized as follows. Section 6.2 describes 3D tongue image data set required for 3D tracking. Section 6.3 presents the thin plate spline based 3D tongue motion tracking method in details. Section 6.4 shows the experiment results on 3D tongue motion tracking during speech. Section 6.5 provides a brief discussion and Section 6.6 summarizes this chapter.

6.2 3D Tongue Image Data

The 3D tagged image data acquisition of the tongue is explained in Section 1.4. Multiple parallel image slices are acquired sparsely in each of the two image orientations to cover the tongue. On one image orientation, two sets of images are acquired separately with orthogonal tag orientations, while the slices in the other image orientation are tagged in one direction perpendicular to the image planes of the first image orientation. The image stacks are illustrated in Fig. 6.1.

We assume that the three tag directions (two on one image orientation, and one on the other) are orthogonal. The image coordinate system's axes are defined as follows (Fig. 6.1): The x and y axes are normal directions to the tag planes of the two tag orientations in the first image orientation, and the z axis is normal to the tag planes

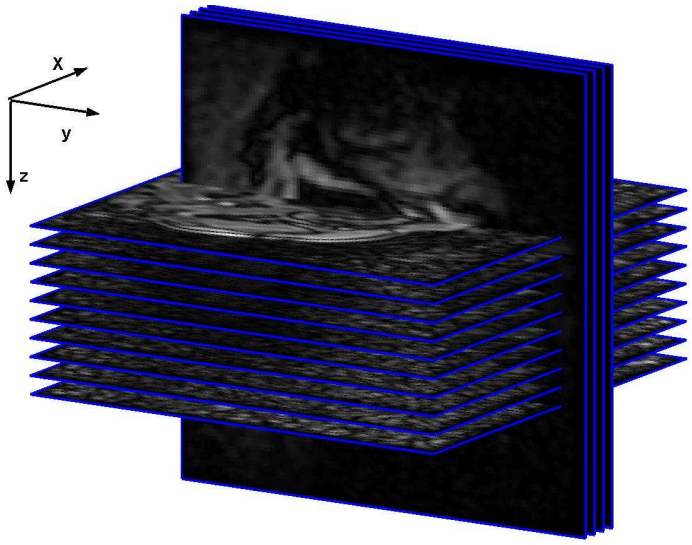


Figure 6.1: 3D tongue MR images.

in the second image orientation. The image planes of the first view are parallel to the xy plane, and the image planes of the second view are parallel to the yz plane.

6.3 Thin Plate Spline Based 3D Tracking

Our thin plate spline based 3D tracking method extends the 3D-HARP method of Pan et al. [43]. Because of the structure of tongue body is different from the heart, a rectangular planar mesh is adopted instead of the cup-shaped mesh in the original method. The mesh is manually placed completely inside the tongue and is tracked in 3D in an iterative manner. The mesh is perpendicular to both image orientations, i.e., it is parallel to the xz plane, so that it intersects with the image planes from

both orientations. The mesh is initialized in the reference time, and the initial HARP phase values of grid points of this mesh are computed based on their positions in space and the knowledge of tag pattern.

Fig. 6.2 illustrates the flow of our thin plate spline based 3D tracking algorithm. HARP methods are first applied on each of the image sequences, thereby giving partial knowledge of the components of 3D motion on all image planes. 2D in-plane motion components are computed for images parallels to the xy plane and 1D motion component in the normal direction of the tag planes is computed for images parallels to the yz plane. At each time frame, the intersection points between the mesh and the image planes are then computed, which give a sparse set of motion observations on the mesh (from the previous HARP computations). Each component of motion is then interpolated using a thin plate spline and based on the partial motion observations on the intersection points so that the motion on all nodes of the mesh are known. The mesh position is updated until no motion is implied anywhere on the mesh (within a small tolerance) and then time is incremented, and these steps are repeated. We now provide details on this overall approach.

6.3.1 Thin-Plate Spline Interpolation

Thin plate spline (TPS) [88] is a widely used interpolation method for scattered data samples. The interpolated function retains values at the data samples while minimizing a *bending energy* function. For 2D space, the bending energy function is

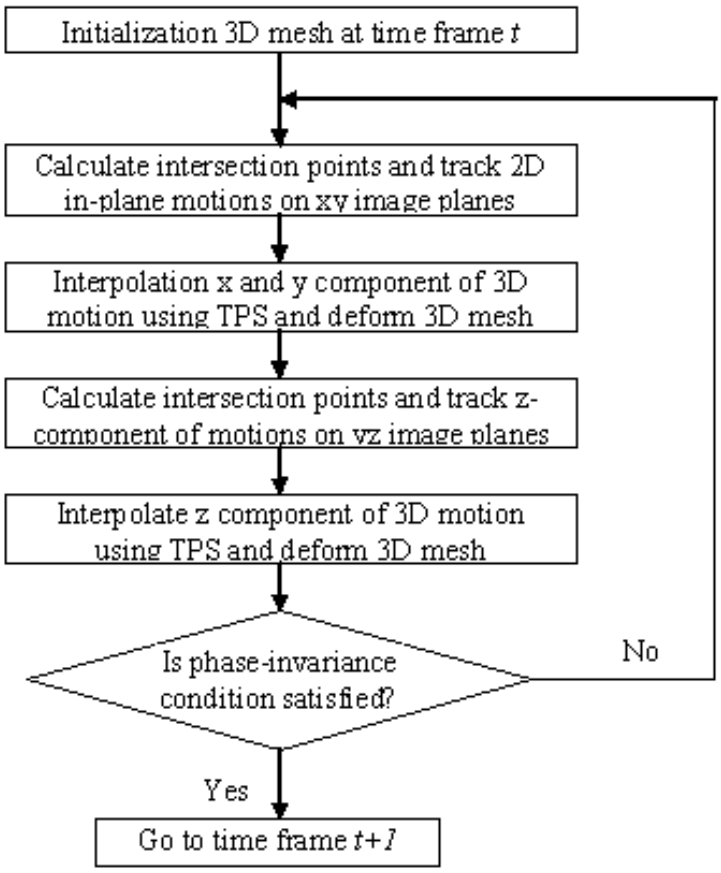


Figure 6.2: The flow chart of the TPS based 3D-HARP algorithm.

defined as

$$J(\mathbf{x}) = \int\int_{\mathbb{R}^2} \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial xy} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dx\ dy\,. \tag{6.1}$$

Let the sample points located at $\mathbf{x}_1, ..., \mathbf{x}_N$. The solution of TPS is in the form:

$$f(\mathbf{x}) = a + \mathbf{b} \cdot \mathbf{x} + \sum_{i=1}^{N} w_i U(||\mathbf{x} - \mathbf{x}_i||)\,, \tag{6.2}$$

with

$$U(r) = r^2 \log(r^2), \text{ where } r^2 = x^2 + y^2 \,, \tag{6.3}$$

and a, $\mathbf{b}$, and w_i are the coefficients that can be solved based on the knowledge of data samples.

In our case of interpolating the motion of the mesh, the data points are located on a planar mesh although the mesh is located in the 3D space, as shown in Fig. 6.3(a). We use 2D thin plate splines instead of 3D to interpolate the three motion components separately. Let the displacement vector be $\mathbf{u} = (u_x, u_y, u_z)^t$. If the motions of a set of material points $\mathbf{x_i} = (x_i, y_i, z_i)^t$ are known as $\mathbf{u_i} = (u_x^i, u_y^i, u_z^i)^t$, the three components of 3D motion on any point $\mathbf{x}$ on the mesh are interpolated using

$$u_x(x, y, z) = a_0^x + a_1^x x + a_2^x y + a_3^x z + \sum_{i=1}^{N} w_i^x r_i^2 \ln r_i^2 \,, \tag{6.4}$$

$$u_y(x, y, z) = a_0^y + a_1^y x + a_2^y y + a_3^y z + \sum_{i=1}^{N} w_i^y r_i^2 \ln r_i^2 \,, \tag{6.5}$$

$$u_z(x, y, z) = a_0^z + a_1^z x + a_2^z y + a_3^z z + \sum_{i=1}^{N} w_i^z r_i^2 \ln r_i^2 \,, \tag{6.6}$$

where $r_i^2 = (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2$. The coefficients can be solved in the similar way as 2D TPS in [88].

In our method, the TPS works better than the Gaussian diffusion used in the original 3D-HARP method [43]. First, TPS does not require parameter tuning because there is no free parameters, while the performance of Gaussian diffusion depends largely on the choice of kernel size. Second, the values on the sample points are preserved in TPS. This is not true in Gaussian diffusion especially when used with large

kernels. Third, Gaussian diffusion may produce "bumpy" results when using small kernels, while TPS does not because the minimized energy function is defined over the whole space.

It should be noted that in our method the TPS is only applied to interpolate the motion on the mesh, and not in the whole 3D space. Because of the limited number of points on the mesh, the TPS interpolation is very fast.

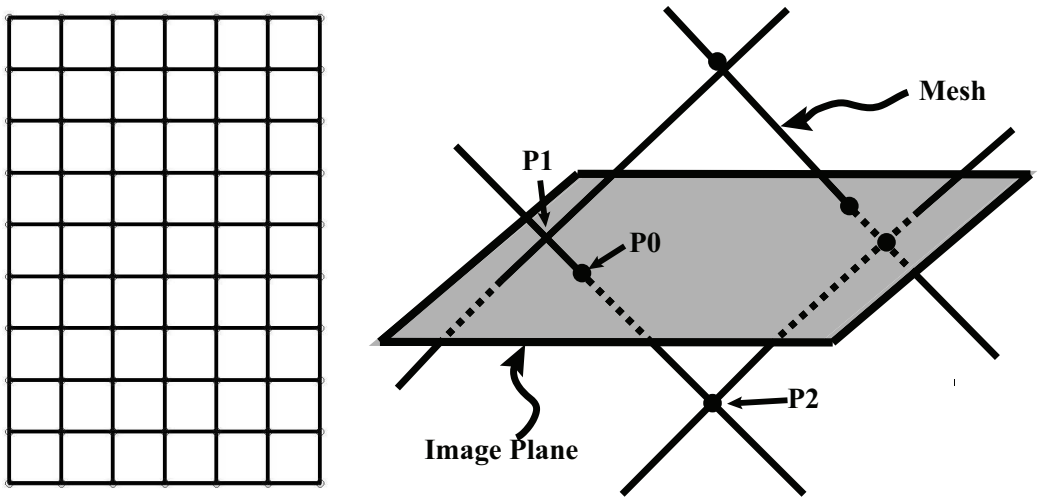
6.3.2 Algorithm

Now we describe the algorithm of TPS-based 3D tongue motion tracking in details.

Step 1. Mesh Initialization

The initial mesh M^0 is a rectangle in the reference time frame, as depicted in Fig. 6.3(a). It is parallel to the xz plane so that it intersects both of the two view images. The mesh is placed completely inside the tongue in order to avoid tracking errors. The mesh is represented by m by n grid points connected by straight lines. We denote the initial 3D phase of grid point $\mathbf{p}$ as $\phi^0(\mathbf{p}) = [\phi_x^0(\mathbf{p})), \phi_y^0(\mathbf{p}), \phi_z^0(\mathbf{p})]^T$. For every grid point, its initial phase is achieved either by knowledge from the imaging process or from the reference time frame. At the reference time the tags are just applied and have not deformed. Therefore the HARP phases are linear functions of $\mathbf{p}$, i.e.,

$$\phi_i(\mathbf{p}) = W(k_i\mathbf{p} \cdot \mathbf{n}_i + \phi_i^{(0)}), i = 1, 2, 3, \tag{6.7}$$



(a) (b)

Figure 6.3: (a) The shape of a 3D tongue mesh. (b) A 3D mesh intersects with an image plane.

where $\mathbf{n}_i$ represents the (three) tag orientation, $\phi_i^{(0)}$ is a constant phase shift, k_i is the tag frequency, and $W(\cdot)$ is a wrapping operator.

For every time frame $t > 0$, the mesh is initialized as the computed mesh from the previous time frame $t - 1$. The following steps are then executed iteratively.

Step 2. Calculation of the intersection points on the xy image planes.

For each image slice parallel to xy plane, the intersection points of the mesh and the image plane are calculated. This is done by finding mesh edges whose vertices are on different sides of the image plane, as illustrated in Fig. 6.3(b). The intersection points are then calculated using linear interpolation. The initial phase values of these points are also calculated as linear interpolation of the initial phase values of the edge vertices.

Step 3. 2D motion tracking in x and y axes and interpolation.

For every intersection point $\mathbf{x}_i = [x_i, y_i, z_i]$ calculated in the previous step, its 2D in-plane motion in x and y axes, $u_x(\mathbf{x}_i)$ and $u_y(\mathbf{x}_i)$, are tracked using the basic 2D HARP tracking and HARP refinement. Then for every grid point $\mathbf{x}_i^t$ of the mesh, the x and y motion, $u_x(\mathbf{x}_i^t)$ and $u_y(\mathbf{x}_i^t)$, is interpolated using Eqns. (6.4) and (6.5).

Assuming the motion in z-direction is zero, the mesh is updated by

$$\mathbf{x}^t = \mathbf{x}^t + \mathbf{u}(\mathbf{x}^t)\,, \tag{6.8}$$

where $\mathbf{x}^t$ is any mesh grid point at time t, and $\mathbf{u}(\mathbf{x}^t) = [u_x(\mathbf{x}^t), u_y(\mathbf{x}^t), 0]^T$. By updating the mesh after each coordinate update (unlike [43]), there is less risk of a tracking error, and the algorithm should converge faster.

Step 4. Calculation of the intersection points on yz image planes.

The intersection points are calculated in the same way as in Step 2.

Step 5. 1D motion tracking in z axis and interpolation.

Since there are only horizontal tags in the yz image slices, only the 1D motion component in z axis can be computed from these images. The z-motion of the intersection points is approximated by assuming the intersection points do not move in the y direction, and the z component of motion on every mesh grid point is calculated using (6.6). The mesh is updated using (6.8) where $\mathbf{x}^t$ is any mesh grid point at time t, and $\mathbf{u}(\mathbf{x}^t) = [0, 0, u_z(\mathbf{x}^t)]^T$.

Step 6. Phase invariance checking.

Because HARP phase is a material property, the correct positioning of the mesh is found when the phases at all intersection points agree with the data. Therefore, the phase invariance property is checked on all intersection points of the updated mesh and all image planes. The initial phase of the k-th intersection point at the current time is denoted as $\phi^{\mathbf{t}}_{\mathbf{k}} = [\phi^t_{x,k}, \phi^t_{y,k}, \phi^t_{z,k}]^T$. If this point is on the first view image planes, only the x and y phase can be calculated. The phase difference of this point is defined as:

$$\delta_k = \max(|\phi^t_{x,k} - \phi^0_{x,k}|, |\phi^t_{y,k} - \phi^0_{y,k}|)\,. \tag{6.9}$$

If this point is on the second view image planes, only the z phase can be calculated. The phase difference is then

$$\delta_k = (|\phi^t_{z,k} - \phi^0_{z,k}|)\,. \tag{6.10}$$

If phase differences of all intersection points are all less than a threshold ϵ, the phase invariance condition is satisfied.

Steps 2–6 are repeated until the phase invariance condition is satisfied. Then time is incremented and the process is repeated. When all time frames have been used, the process is complete.

6.4 Experimental Results

We applied the algorithm on the 3D tongue motion tracking of a normal speaker. The MR image slices were acquired on a Philips Eclipse 1.5 T scanner when the

subject uttered "deGoose." This word was chosen because it involves large front-to-back motion of the tongue. The acquired image data includes horizontally and vertically tagged images in the axial view, and horizontally tagged images in the sagittal view. The imaging parameters were: image size = 128 × 128, pixel spacing = 1.56 mm, slice thickness = 5 mm, tag spacing = 5 mm, flip angle = 10 degrees. Ten axial and four sagittal image slices were acquired at each of 18 time frames. The separation between sagittal images is 7 mm, and between axial images is 8 mm. The interval between the time frames is 49 ms. The 10 axial slices covered the whole tongue, and the 4 sagittal slices went from the middle to the left side of the tongue.

Three parallel meshes were placed in the middle of the tongue and tracked. The meshes were initialized at the reference time frame as 19 mm by 38 mm rectangles parallel to the xz plane, and they were placed 2 mm apart. The meshes were represented as 32 by 32 grids. At the reference time frame, these meshes intersect with 5 axial slices and 3 sagittal slices. The relative locations of these meshes and the image slices are illustrated in Fig. 6.4. The displayed image is on the mid-sagittal plane passing through the center of the tongue. For visualization purpose, the displayed images are untagged images that were acquired with the same imaging parameters and spatial locations of the tagged MR images. After initialization, the meshes were tracked in all 18 time frames using our algorithm.

Fig. 6.5 shows the tracking results of one mesh at four different time frames. . Fig. 6.6 shows the pathlines of several selected points that are tracked in time on the

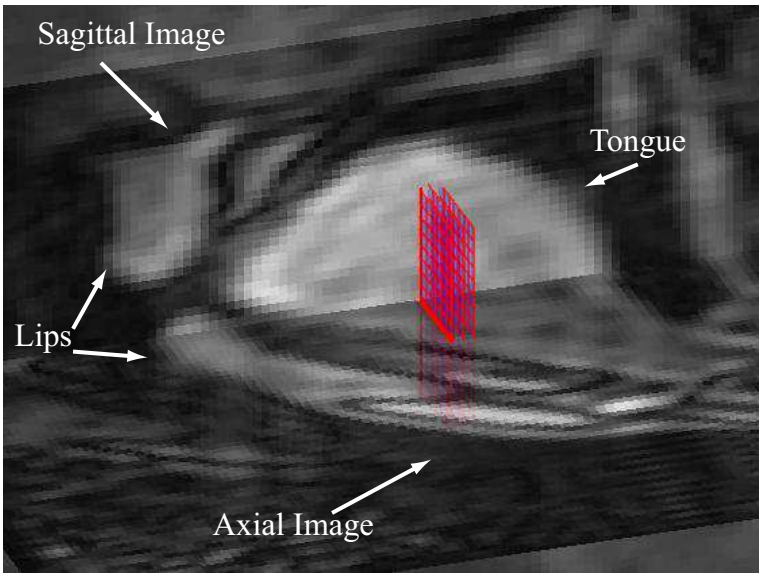
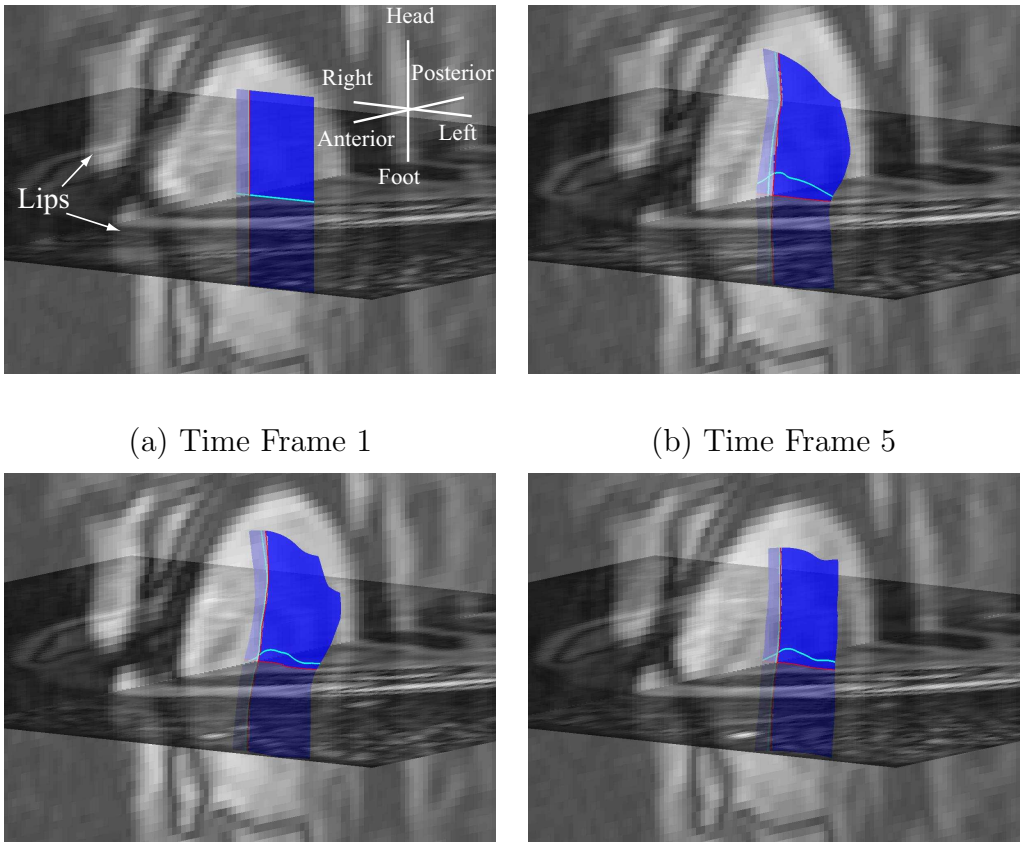


Figure 6.4: The three rectangular meshes displayed together with non-tagged sagittal and axial images.

same mesh. From these figures we can see the tongue moves more in the anterior-posterior and head-foot directions, and less in the left-right direction. The upper part of the tongue moves more than the lower part.

We also computed the 3D Lagrangian strain from the 3D tracking results of the meshes. At each time frame, the strain in x and z axes was computed as the edge lengths between grid points divided by the lengths at the reference time frame, while the strain in y axis was computed as the distance between corresponding grid points at two neighboring meshes divided by the distance at the reference time frame. Fig. 6.7 shows the strain map on the second (middle) mesh in the three axes. The mesh is projected onto xz plane for display purposes.

For visual assessment, we also acquired vertically tagged images in the sagittal



(a) Time Frame 1 (b) Time Frame 5

(c) Time Frame 9 (d) Time Frame 13

Figure 6.5: The mesh motion at four different time frames. The image orientations are shown in (a). The red lines are the intersection of the mesh and the displayed image planes. The intersection line on the first time frame is tracked in the subsequent frames and displayed in cyan.

view. These images were not used in the mesh tracking. It was observed visually that the meshes conformed to the deformation of the tag lines in these images at all time frames. This enables direct comparison of automatic mesh tracking with visual assessment of tag motion. Fig. 6.8 shows the positions of the intersections of one of the tracked meshes with one tagged sagittal image at different time frames.

This algorithm was implemented in Matlab (Mathworks, Natick, MA) on a com-

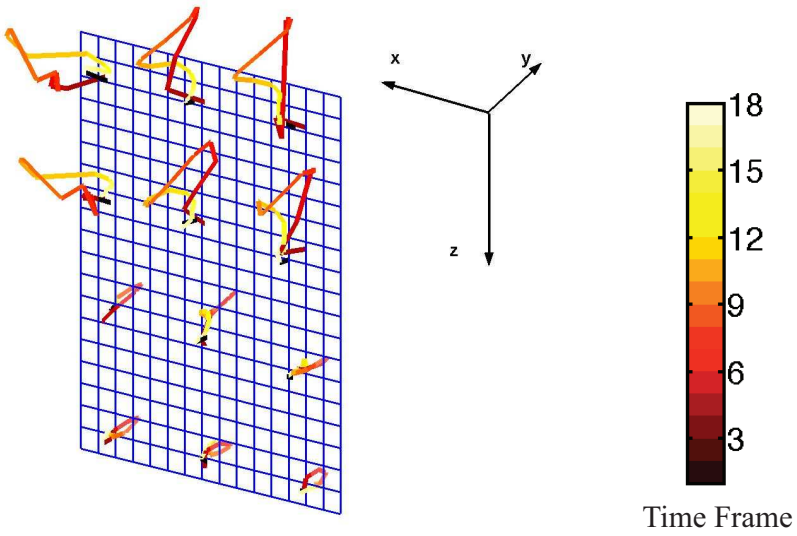
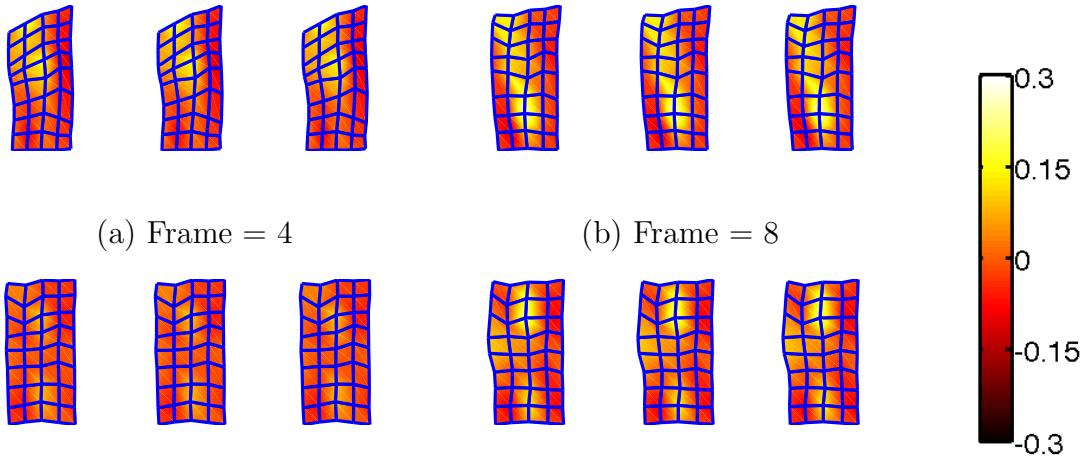


Figure 6.6: The 3D pathlines of selected points on the mesh tracked temporally.

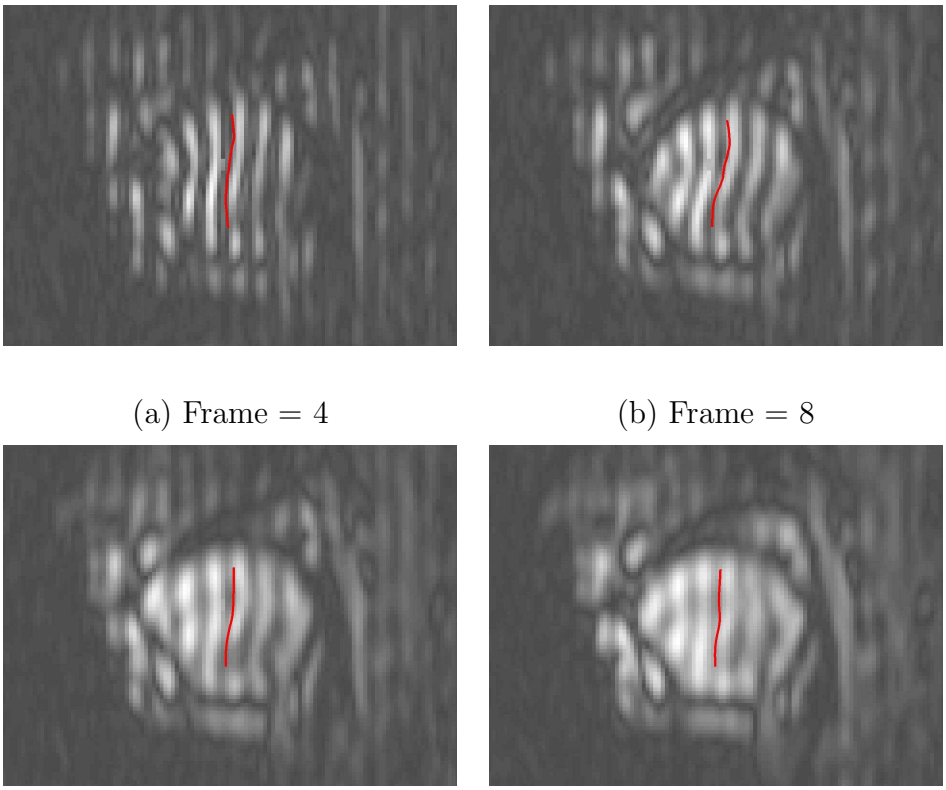


(a) Frame = 4 (b) Frame = 8

(c) Frame = 12 (d) Frame = 16

Figure 6.7: The strain map along three axes directions at four time frames. At each time frame, the strain maps along x, y, and z axes are displayed from left to right.

puter with a 2.8 GHz Intel Pentium 4 processor and 512 MB RAM. In our implementation it took about 10 seconds for each time frame.



(a) Frame = 4 (b) Frame = 8

(c) Frame = 12 (d) Frame = 16

Figure 6.8: The intersection lines of the tracked mesh one vertically tagged sagittal image in different time frame. The image orientation is the same as in Fig. 6.5. The intersection lines are shown in red.

6.5 Discussion

To guarantee that our algorithm converges, the mesh must be placed entirely inside the tongue and not too close to the tongue surface. This is because the convergence relies on the correct 1D/2D tracking of the intersect points between the mesh and the image planes. If the intersect points lie outside of the tongue, HARP tracking will fail and the wrong motion will be transfered to the intermediate deformation field

through TPS interpolation. In this case our algorithm will fail.

Though developed to track 3D meshes, the method can also be used as a way of tracking arbitrary 3D point of interest in the tissue. This is achieved by automatically building a small rectangular mesh centered by the point and oriented in the same way as described in the method. The mesh should be big enough so that it can intersect at least two image planes in each of the two image orientations at all time frames. This ensures the intersection points of the mesh with the image planes surround the point of interest to promise the algorithm can converge. Our method can then be applied to track the mesh, and the 3D trajectory of the point of interest can be easily computed from the tracked mesh.

Moreover, it is possible to extend the method to compute the 3D motion of the whole tongue. The 3D tongue volume can be viewed as a 3D mesh grid with 6-connected voxels. Instead of computing the intersection of the intersection points of the mesh with the image planes, we find the intersection points of the volume with the image planes. The motion on the 3D volume can be iteratively solved in the similar way by extending the sparse motion information from the image planes.

6.6 Summary

In this chapter we introduced a fast method to track tongue motion in three dimensions from tagged MR images using thin plate spline interpolation and 3D-

HARP concept. This method tracks a rectangular planar mesh that is placed inside the tongue and runs in an iterative fashion. In each iteration, the in-plane motion components of intersection points of the mesh and the image planes are computed using 2D HARP, and propagated to the whole mesh through TPS interpolation. By placing two parallel meshes close to each other, one can compute 3D Lagrangian strain from the tracking results. Experiments showed this method can track tongue motion smoothly and accurately.

Chapter 7

3D Incompressible Motion Estimation From Tagged MRI

7.1 Introduction

The soft tissues, including cardiac muscles and tongue muscles, are considered incompressible because they are mainly composed of water, which is incompressible. It is widely accepted that the volume change of myocardium during the cardiac cycle is no more than 4% [89, 90], and the tongue muscle motion is incompressible. Yet, this fact is largely ignored by previous approaches when reconstructing the 3D motion field from tagged MR images [47, 53, 57–59, 64, 65]. It is also not considered in the 3D-HARP method of Pan et al. [43] or our method of 3D tongue motion tracking described in Chapter 6.

Since materials that are incompressible undergo deformations that preserve volumes at all scales and have divergence-free velocity fields, one could improve interpolation results to reflect the physical properties of muscle motion by exploiting this constraint. Song et al. [91] first applied this property in building the 3D velocity of the heart from cine CT images. Denney et al. [92] directly applied the divergence-free constraint to reconstruct the 3D displacement field of the LV in an estimation theoretic approach. In 2007, Bistoquet et al. [93] used an incompressible deformable model to recover the motion of LV from anatomical cine MR images. Recently, Bistoquet et al. [94] constructed nearly incompressible cardiac motion field from non-tagged MR images using a vector spline with a divergence-free matrix-valued radial basis function.

There is a key problem with these previous approaches of reconstructing incompressible motion, however. Because the temporal resolution of the image sequences are relatively large, the deformation between two neighboring time frames may be large. A velocity field that is approximated as the displacement field divided by the time interval is not theoretically predicted to be divergence-free. When this fact is ignored and the underlying field is interpolated in a divergence-free fashion this can lead to considerable errors when reconstructing motion fields in a time sequence since the errors in earlier time frames propagate to later time frames. In [94], this error is reduced by interpolating both forwards and backwards in time and then computing a weighted average of these solutions. However, solutions generated this way are not

guaranteed to yield motions that have divergence-free velocity fields or correspond to incompressible motions.

In this chapter, we propose a framework to reconstruct a 3D, dense, incompressible deformation field from tagged MR images based on divergence-free vector spline with incomplete data samples. In this framework, incomplete observations of the 3D displacement vectors on the imaged tissue points are first computed from the tagged MR images using HARP [40, 41] tracking and HARP refinement methods. From the partial and non-uniform samples we seek a sequence of divergence-free velocity fields from which the final displacement field is computed by integration. From the reconstructed dense displacement field, we can also compute the dense 3D Eulerian strain tensor everywhere inside the tissue, which is not possible in the 3D strain computation in zHARP because it can only compute the strain on the image slices.

This chapter is organized as follows. Section 7.2 introduces background knowledge on the properties of incompressible motion and divergence-free vector spline interpolation. Section 7.3 describes our framework of 3D incompressible motion reconstruction. Section 7.4 demonstrates our method using a simple 2D example, includes the validation of our method with a cardiac motion simulator, and also shows the experiment results on both cardiac motion estimation and tongue motion estimation. Section 7.5 provides a discussion, and finally, Section 7.6 summarizes this chapter.

7.2 Divergence-Free Vector Spline

7.2.1 Properties of Incompressible Motion

The body tissues, including the heart and the tongue, can be approximately considered as homogeneous, isotropic, 3D, incompressible elastic bodies. From continuum mechanics [95], one can learn that the motion of such materials exhibits particular physical properties. We denote $\mathbf{x} = \mathbf{x}(t) = \mathbf{X} + \mathbf{u}(\mathbf{X}, t)$ as the location of a material point $\mathbf{x}$ at time t with $\mathbf{x}(0) = \mathbf{X}$. The physical properties constrain the motion of incompressible tissues to satisfy these following equivalent conditions:

1. The displacement field $\mathbf{u}$ at any time t from reference time 0 is *volume preserving*, i.e., $\det\left[I + \nabla_{\mathbf{X}}\mathbf{u}_{\mathbf{X}}(\mathbf{X}, t)\right] = 1$.

2. The velocity field at any time t is *divergence-free*, i.e., $\mathrm{div}_{\mathbf{x}}\mathbf{v}(\mathbf{x}, t) = 0$, with $\mathbf{v}(\mathbf{x}, t) = d\mathbf{u}/dt$ be the velocity.

3. For small motion, and when subjecting to a known force field $\mathbf{f}$, at equilibrium the displacement field satisfies the *Navier equilibrium partial differential equation (PDE)* (described later).

Among these conditions, the volume-preserving condition is nonlinear, so it is difficult to directly apply it to reconstruct incompressible displacement field. But this condition can be used to verify the incompressibility of a given displacement field. The second condition — divergence-free velocity — was used in our method to reconstruct

incompressible displacement fields through building a sequence of divergence-free velocity fields. About the third condition, it has been shown that under certain forces and small deformations the Navier PDE has analytical solutions that are equivalent to the second condition. These are explained in details in the rest of this section.

7.2.2 Vector Spline

Vector splines (VS) were first proposed by Amodei et al. [96] to interpolate a vector field using known vector-valued data samples. Comparing to scalar-valued spline interpolation, e.g., the thin-plate spline, B-spline and so on, the vector spline couples the components of the vector field together instead of treating them separately. In this way, it can better represent the underlying physical properties of a vector field. Given N points in space $\mathbf{x}_n = [x_n, y_n, z_n]^T$, $n = 1, \ldots, N$, and vector-valued observations $\mathbf{v}_n$, $n = 1, \ldots, N$, at these points, the VS interpolates a smooth vector field over the whole space. Specifically, the VS finds a vector field $\mathbf{v}(\mathbf{x})$ that minimizes

$$J_{\alpha,\beta}(\mathbf{v}) = \int [\alpha ||\nabla^k(\mathrm{div}\mathbf{v}(\mathbf{x}))||^2 + \beta \sum_{i=1}^{3} ||\nabla^k(\mathrm{rot}\mathbf{v}(\mathbf{x}))_i||^2] d\mathbf{x}\,, \qquad (7.1)$$
$$\text{subject to } \mathbf{v}(\mathbf{x}_n) = \mathbf{v}_n, n = 1, 2, \ldots, N\,,$$

where α and β are the weighting coefficients, $\mathrm{div}\mathbf{v} = \nabla \cdot \mathbf{v}$ yields the divergence of a vector field, and $\mathrm{rot}\mathbf{v} = \nabla \times \mathbf{v}$ yields the curl. By weighting the two terms in Eqn. (7.1) differently, a vector spline can control the divergence and vorticity of the

vector field separately. It has been shown that (7.1) has the closed form solution [96]

$$\mathbf{v}(\mathbf{x}) = \sum_{n=1}^{N} \mathbf{K}(\mathbf{x} - \mathbf{x}_n) \cdot \mathbf{c}_n + \mathbf{p}(\mathbf{x}) \,, \tag{7.2}$$

where $\mathbf{c}_n$ are the unknown coefficients and $\mathbf{K}(\mathbf{x})$ is the matrix-valued kernel function given by

$$\mathbf{K}_{\mathrm{VS}}(\mathbf{x}) = [\frac{1}{\beta}\triangle\mathbf{I} + (\frac{1}{\alpha} - \frac{1}{\beta})\nabla\nabla^T]h(\mathbf{x}) \,, \tag{7.3}$$

where $\mathbf{I}$ is the identity matrix, ∇ is the gradient operator, $\triangle$ is the Laplacian operator, and $h(\mathbf{x}) = ||\mathbf{x}||^{2k+1}$ is the solution to $\triangle^{k+1}h(\mathbf{x}) = \delta(\mathbf{x})$. Also, $\mathbf{p}(\mathbf{x})$ is the polynomial function of order k given (for $k = 1$) by

$$\mathbf{p}_{\mathrm{VS}}(\mathbf{x}) = \begin{bmatrix} 1 & x & y & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x & y & z & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & x & y & z \end{bmatrix} \cdot \mathbf{d} = \mathbf{A}(\mathbf{x}) \cdot \mathbf{d} \,, \tag{7.4}$$

where $\mathbf{d}$ is a 12 by 1 vector of unknown coefficients.

The coefficients $\mathbf{c}_n$ and $\mathbf{d}$ in VS are solved using the known vector values on the sample points by formulating

$$\begin{bmatrix} \mathbb{K} & \mathbb{P} \\ \mathbb{P}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{C} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \mathbf{V} \\ 0 \end{bmatrix} \,, \tag{7.5}$$

where $\mathbb{K}$ is a $3N \times 3N$ matrix with $(\mathbb{K})_{ij} = \mathbf{K}(\mathbf{x}_i - \mathbf{x}_j)$, $\mathbb{P}^T = [\mathbf{A}^T(\mathbf{x}_1), ...\mathbf{A}^T(\mathbf{x}_N)]^T$, $\mathbf{V} = [\mathbf{v}_1^T, ..., \mathbf{v}_N^T]^T$, and $\mathbf{C}$ is a $3N \times 1$ vector with $\mathbf{C} = [\mathbf{c}_1^T, ..., \mathbf{c}_N^T]^T$. After the unknown coefficients $\mathbf{C}$ and $\mathbf{d}$ are solved, the vector value at any point $\mathbf{x}$ can be computed using Eqn. (7.2).

The sample data are often corrupted by noise and therefore the spline should not precisely match the data but should use it instead to guide the fitting of a smoothing spline. Assuming Gaussian noise, the smoothing VS solves

$$\arg\min_{\mathbf{v}} C(\mathbf{v}) = \rho J(\mathbf{v}(\mathbf{x})) + \frac{1}{N}\sum_{n=1}^{N} ||\mathbf{v}(\mathbf{x}_n) - \mathbf{v}_n||^2 , \tag{7.6}$$

where ρ is a smoothing parameter, and J is the regularization term as defined in Eqn. (7.1). The smoothing VS solution has the same form as Eqn. (7.2), and its coefficients are solved using

$$\begin{bmatrix} \mathbb{K} + \rho\mathbf{I} & \mathbb{P} \\ \mathbb{P}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{C} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \mathbf{V} \\ 0 \end{bmatrix} . \tag{7.7}$$

Note when the weighting coefficients $\alpha = \beta$ and $k = 1$, the function in Eqn. (7.1) becomes the same as the bending energy function in the thin plate spline. In this case the three components of the interpolated vector field (Eqn. (7.2)) are decoupled because the kernel matrix $\mathbb{K}$ is diagonal. So the VS solution is equivalent to interpolating the three components independently using a thin plate spline. For this reason, we can call the VS with kernel as in Eqn. (7.3) the *thin plate vector spline*.

7.2.3 Divergence-Free Vector Spline

As a special case of VS, the *divergence-free vector spline* (DFVS) constrains the VS to be divergence-free by solving

$$\arg\min_{\mathbf{v}} J_{\mathrm{DF}}(\mathbf{v}) = \int \sum_{i=1}^{3} ||\nabla^k (\mathrm{rot}\mathbf{v}(\mathbf{x}))_i||^2 d\mathbf{x}\,, \tag{7.8}$$
$$\text{subject to: } \mathrm{div}\mathbf{v}(\mathbf{x}) = 0, \text{ and } \mathbf{v}(\mathbf{x}_n) = \mathbf{v}_n, n = 1, 2, ..., N.$$

The *divergence-free vector spline* (DFVS) solution is similar to that of VS except that the kernel matrix becomes

$$\mathbf{K}_{\mathrm{DF}}(\mathbf{x}) = [\triangle \mathbf{I} - \nabla\nabla^T] h(\mathbf{x})\,, \tag{7.9}$$

and $\mathbf{p}(\mathbf{x})$ is also constrained to be divergence-free

$$\mathbf{p}_{\mathrm{DF}}(\mathbf{x}) = \begin{bmatrix} 1 & x & y & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x & y & z & 0 & 0 & 0 \\ 0 & -z & 0 & 0 & 0 & 0 & -z & 0 & 1 & x & y \end{bmatrix} \cdot \mathbf{d}\,. \tag{7.10}$$

Similarly, the coefficients in the DFVS and smoothing DFVS can be solved using Eqns. (7.5) and (7.7), respectively.

7.2.4 Navier Equilibrium PDE and Vector Spline

The Navier equilibrium PDE describes the physical property of a homogeneous isotropic elastic body with small motion. When subject to a force field $\mathbf{f}(\mathbf{x})$ and assuming small deformations at equilibrium, the displacement field $\mathbf{u}(\mathbf{x})$ satisfies

$$\mu \triangle \mathbf{u}(\mathbf{x}) + (\mu + \lambda) \nabla \mathrm{div}\mathbf{u}(\mathbf{x}) = \mathbf{f}(\mathbf{x})\,, \tag{7.11}$$

where μ and λ are the Lamé coefficients, and for an incompressible object $\mu = 0$. For certain force field, Eqn. (7.11) has closed-form solution. For example, for a smooth field $\mathbf{f}(\mathbf{x}) = f(\mathbf{x}) \cdot \mathbf{c}$ with constant $\mathbf{c}$, the solution is given by

$$\mathbf{u}(\mathbf{x}) = [\frac{1}{\mu}\triangle\mathbf{I} - \frac{\mu+\lambda}{\mu(2\mu+\lambda)}\triangle\triangle^T]g(\mathbf{x}) \cdot \mathbf{c} = \mathbf{K}_{\text{EBS}}(\mathbf{x})\mathbf{c}\,, \tag{7.12}$$

where $f(\mathbf{x}) = \triangle g(\mathbf{x})$. This gives the *elastic body spline* (EBS) [97]. $\mathbf{K}_{\text{EBS}}$ is the basis kernel matrix, and is determined by the choice of the force field. Davis et al. [97] derived the kernel matrix of EBS for the force fields in the form of $f(\mathbf{x}) = ||\mathbf{x}||^{2k-1}$, and later Kohlrausch et al. [98] extended the EBS using Gaussian forces.

The EBS is closely related to the VS. In fact, when $\alpha = 2\mu+\lambda$, $\beta = \mu$, and $f(\mathbf{x}) = ||\mathbf{x}||^{2k-1}$, then $\mathbf{K}_{\text{EBS}}$ and $\mathbf{K}_{\text{VS}}$ are equal, which means EBS and VS are equivalent. Therefore, when $\mu \to 0$ EBS interpolates the displacement field of an incompressible object under small deformations, and it becomes DFVS (cf. Eqn. (7.9)).

The Navier equilibrium PDE is valid only for small deformations. In this case the displacement field can be considered divergence-free (for incompressible motion) because the velocity can be approximated to be constant during the deformations. For large deformations, neither EBS nor DFVS can be directly applied to interpolate incompressible deformation fields. However, since the velocity field of an incompressible object is divergence-free then the incremental displacement over a small time interval δt is $\mathbf{u}(\mathbf{x}, t) \approx \mathbf{v}(\mathbf{x}, t)\delta t$, and is approximately divergence-free. We will use this principle in the next section.

7.2.5 Smoothing VS from Incomplete Samples

The motion computed from tagged MR images are sparse and incomplete because (1) the imaged tissue points are non-uniformly and sparsely distributed in space, and (2) only selected components of the displacement vectors are known, i.e., the components in the normal directions of tag planes. The incomplete sample data of a vector field can be written as: $\{\mathbf{x}_n, \mathbf{e}_n, w_n\}$ for $n = 1, 2, \ldots, N$, where $\mathbf{e}_n$ is a unit vector representing the projection direction, and $w_n = \mathbf{e}_n \cdot \mathbf{v}(\mathbf{x}_n)$ is the projection of the 3D vector $\mathbf{v}(\mathbf{x}_n)$ on $\mathbf{e}_n$. Given N sample points, the minimization problem of a smoothing VS from incomplete data can be expressed as

$$\arg\min_{\mathbf{v}} C(\mathbf{v}) = \rho J(\mathbf{v}(\mathbf{x})) + \frac{1}{N}\sum_{n=1}^{N}(\mathbf{e}_n^T\mathbf{v}(\mathbf{x}_n) - w_n)^2 \,. \tag{7.13}$$

Arigovindan [99] showed that the solution to this problem is

$$\mathbf{v}(\mathbf{x}) = \sum_{n=1}^{N}\mathbf{K}(\mathbf{x} - \mathbf{x}_n)\mathbf{e}_n c_n + \mathbf{p}(\mathbf{x}) \,, \tag{7.14}$$

where the coefficients c_n are scalars, and $\mathbf{K}(\mathbf{x})$ and $\mathbf{p}(\mathbf{x})$ are the same as in VS. By replacing $\mathbf{K}$ with $\mathbf{K}_{\mathrm{DF}}$, and $\mathbf{p}$ with $\mathbf{p}_{\mathrm{DF}}$, Eqn. (7.14) describes the solution to a smoothing DFVS with incomplete samples. The coefficients c_n are found using Eqn. (7.7) where in this case $\mathbb{K}$ is a $N \times N$ matrix with $(\mathbb{K})_{ij} = \mathbf{e}_i^T\mathbf{K}(\mathbf{x}_i - \mathbf{x}_j)\mathbf{e}_j$, $\mathbb{P}^T = [\mathbf{A}^T(\mathbf{x}_1)\mathbf{e}_1, \ldots, \mathbf{A}^T(\mathbf{x}_N)\mathbf{e}_N]$ and $\mathbf{V} = [w_1, .., w_N]^T$.

7.3 Incompressible Motion Reconstruction

7.3.1 3D Data Set and Preprocessing

The 3D image data set for the heart and the tongue is acquired as described in Section 1.4. All of the images are processed using HARP [40, 41] to yield sequences of harmonic phase images. The HARP phases at the reference time t_0 are estimated from the HARP images at the first time frame. At t_0, the tagging phase ϕ is a linear function of the point's coordinate $\mathbf{x}$, and wrapped to the range $[-\pi, \pi)$, i.e.,

$$\phi(\mathbf{x}, t_0) = W(k\mathbf{x} \cdot \mathbf{e} + \phi_0)\,, \tag{7.15}$$

where k is the known tagging frequency, $\mathbf{e}$ is the normal vector of the tag planes, ϕ_0 is an unknown phase offset, and W is a phase wrapping operator. By assuming the tags do not deform much at the first time frame, ϕ_0 can be estimated from the HARP images at the first time frame in the same way as in Tecelao et al. [74]. Therefore, we can construct synthetic phase images at t_0 using Eqn. 7.15. The tissue points at each time frame are then tracked back to t_0 using standard HARP tracking [40] and HARP refinement. As illustrated in Fig. 1.6, for tagging direction $\mathbf{e}$ and a 3D spatial point $\mathbf{x_j}$ that is imaged at time t, if this point comes from $\mathbf{X}_j = \mathbf{x}_j(t) - \mathbf{u}(\mathbf{x}_j(t), t)$ at reference time t_0, then HARP tracking computes the projection of its displacement $\mathbf{u}(\mathbf{x}, t)$ onto $\mathbf{e}$ as:

$$w_j = \mathbf{e}^T \mathbf{u}(\mathbf{x_j}, t) = \mathbf{e}^T (\mathbf{X}_j - \mathbf{x}_j)\,. \tag{7.16}$$

On the image planes acquired with two tag orientations, for example the SA images in cardiac imaging, two projections of the displacement of each tissue point are computed. On the image planes acquired with one tag orientations, for example the LA images in cardiac imaging, only one projection is computed. Therefore except for points at the intersections of the image planes from the two orientations, only partial knowledge of the displacement is available for any other pixel on the observed images. (Of course, no observations are available at 3D points that do not lie on an observed image plane.) Because of the irregular geometry of the combined positions of image planes—which are sometimes not even acquired with uniform spacing—the locations of the sample points, each having partial information, are typically quite irregular.

7.3.2 3D Incompressible Displacement Field Reconstruction

7.3.2.1 Problem Statement

HARP tracking provides N incomplete and non-uniform data samples $\{\mathbf{x}_n, \mathbf{e}_n, w_n\}$ at any time frame (see Eqn. (7.16)). Our goal is to reconstruct a 3D, dense, incompressible displacement field $\mathbf{u}(\mathbf{x})$ from these data samples such that $\mathbf{e}_n \cdot \mathbf{u}(\mathbf{x}_n) = w_n$. As discussed earlier, the spatial velocity field $\mathbf{v}(\mathbf{x})$ giving rise to such a deformation must be divergence-free, i.e., $\text{div}\mathbf{v}(\mathbf{x}) = 0$. Based on this physical property, we

want to reconstruct the incompressible displacement field $\mathbf{u}(\mathbf{x})$ through integrating divergence-free velocity fields.

We denote the integration variable as s, which takes on values in the interval $[0, 1]$, and the velocity of $\mathbf{x}(s)$ at s be $\mathbf{v}(\mathbf{x}(s), s)$. We define

$$\mathbf{w}(\mathbf{x}, s) = \int_0^s \mathbf{v}(\mathbf{x}(\tau), \tau) d\tau, \quad \text{and} \quad \mathbf{v}(\mathbf{x}(\tau), \tau) = \frac{d\mathbf{w}(\mathbf{x}, \tau)}{d\tau} \ , t \in [0, 1] \,, \tag{7.17}$$

with $\mathbf{x}(s) = \mathbf{x}+\mathbf{w}(\mathbf{x}, s)$. We then have $\mathbf{u}(\mathbf{x}) = \mathbf{w}(\mathbf{x}, 1)$. Thus the dense incompressible deformation reconstruction problem can be formulated as follows: Given N sample points $\mathbf{x}_n$ for $n = 1, 2, ..., N$, and at each sample point the projection of its 3D displacement $\mathbf{u}(\mathbf{x}_n)$ on $\mathbf{e}_n$ is known as $w_n = \mathbf{e}_n^T \mathbf{u}(\mathbf{x}_n)$, find a dense displacement field $\mathbf{u}(\mathbf{x})$ such that

$$\mathbf{u}(\mathbf{x}) = \int_0^1 \mathbf{v}(\mathbf{x}(\tau), \tau) d\tau, \quad \text{with } \mathrm{div}_{\mathbf{x}} \mathbf{v}(\mathbf{x}, \tau) = 0 \ , \forall \tau \in [0, 1] \,. \tag{7.18}$$

7.3.2.2 Discretization

The incompressible displacement field reconstruction problem can be reduced to a finite-dimensional problem by dividing the integration into discrete steps, i.e., $s_m = m\delta$ for $m = 0, 1, \ldots, M$ with $\delta = 1/M$. The discretization is illustrated in Fig. 7.1. Within each interval the velocity is assumed to be constant, so

$$\mathbf{w}(\mathbf{x}(s), s) = \mathbf{w}(\mathbf{x}(s_m), s_m) + \mathbf{v}(\mathbf{x}(s_m), s_m) * (s - s_m) \,, \tag{7.19}$$

for $s \in [s_m, s_{m+1})$, and $\mathbf{u}(\mathbf{x}) = \mathbf{w}(\mathbf{x}, 1) = \delta \sum_{m=0}^{M-1} \mathbf{v}(\mathbf{x}(s_m), s_m)$. $\mathbf{v}(\mathbf{x}(s), s)$ is not the true myocardial velocity, but rather a computational tool for the estimation of the

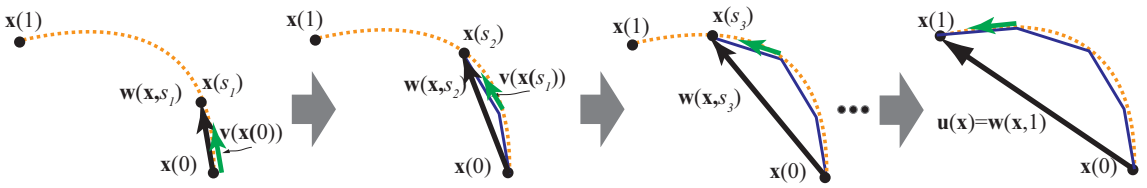


Figure 7.1: Illustration of the process of integrating velocity fields at discrete steps to compute the displacement field.

displacement field.

7.3.2.3 Velocity Computation

We use DFVS from incomplete data samples to interpolate the divergence-free velocity fields separately over each integration interval. The velocity fields are computed sequentially starting from $s_0 = 0$ through $s_M = 1$. Let us denote $r_n(s_m) = \mathbf{e}_n \cdot \mathbf{v}(\mathbf{x}_n(s_m), s_m)$ for any step s_m, and the data samples at s_m are written as $\{\mathbf{x}_n(s_m), \mathbf{e}_n, r_n(s_m)\}$ for $n = 1, \ldots, N$.

At step s_m, the velocity fields at steps between s_0 and s_{m-1} have been computed. So we have $\mathbf{x}(s_i) = \mathbf{x} + \mathbf{w}(\mathbf{x}, s_i)$ for $i = 0, ..., m-1$, and $\mathbf{w}(\mathbf{x}, s_m) = \delta \sum_{i=0}^{M-1} \mathbf{v}(\mathbf{x}(s_i), s_i)$. The velocity at any sample point $\mathbf{x}_n(s_m)$ at s_m is approximated by taking the first order expansion

$$\mathbf{u}(\mathbf{x}_n) - \mathbf{w}(\mathbf{x}_n(s_m), s_m) \approx \hat{\mathbf{v}}(\mathbf{x}_n(s_m), s_m)(1 - \delta m)\,. \tag{7.20}$$

Since the complete knowledge of $\mathbf{u}(\mathbf{x}_n)$ is not available and only its projection on $\mathbf{e}_n$, i.e., w_n, is known, the projection of the velocity on $\mathbf{e}_n$ can be approximated using

$$r_n(s_m) = \mathbf{e}_n \cdot \hat{\mathbf{v}}(\mathbf{x}_n(s_m), s_m) = \frac{w_n - \mathbf{e}_n \cdot \mathbf{w}(\mathbf{x}, s_m)}{1 - \delta m}\,. \tag{7.21}$$

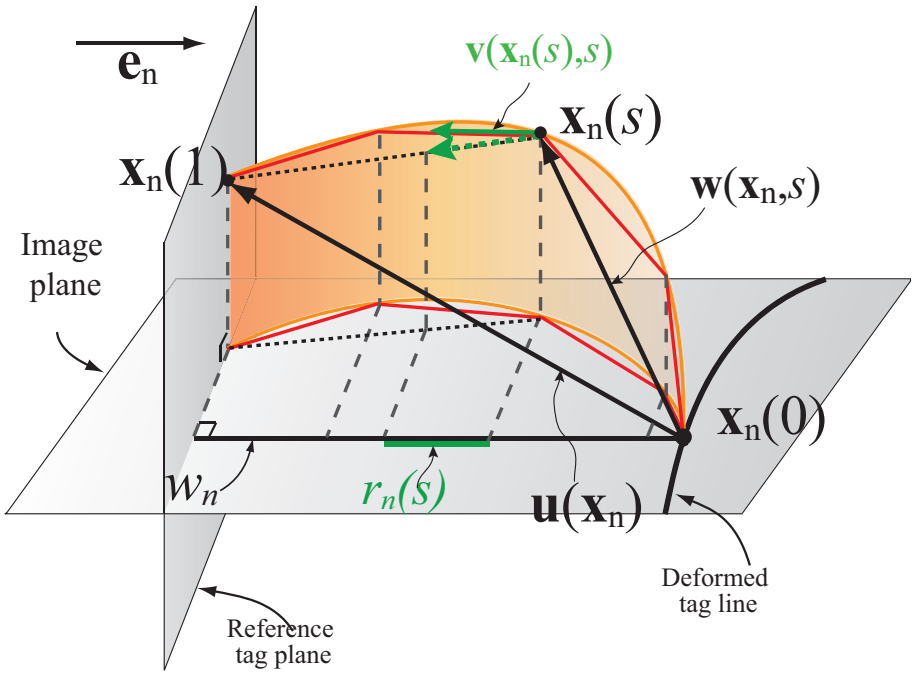


Figure 7.2: The estimation of velocity projection at a sample point $\mathbf{x}_n$. The velocity at s is estimated using approximated approximation, and its projection in $\mathbf{e}_n$ is used to interpolate the 3D velocity vector at any point.

The linear approximation of velocity at a sample point is illustrated in Fig. 7.2.

With the N data samples, the continuous velocity field $\mathbf{v}(\mathbf{x}, s_m)$ is interpolated with smoothing DFVS using Eqns. (7.13) and (7.14) instead of exact DFVS, i.e., $\rho > 0$ in Eqns. (7.13) and (7.7).

From Taylor's expansion, the first order approximation of the velocity is accurate up to the order $(1-\delta m)^2$. Therefore it is less accurate at smaller s and more smoothing is required at earlier steps. So the smoothing parameter ρ should be chosen to be large at small s and grow smaller as s approaches 1. At s_{M-1}, ρ should be set to 0 so that the final displacement $\mathbf{w}(\mathbf{x}_n, 1)$ matches the original data samples exactly—i.e., $\mathbf{e}_n \cdot \mathbf{u}(\mathbf{x}_n) = \mathbf{e}_n \cdot \mathbf{w}(\mathbf{x}_n, 1) = w_n$ for $n = 1, \ldots, N$. In practice, we choose the smoothing

parameter at t_m as $\rho_m = \frac{M-m-1}{M-1}\rho_0$, where ρ_0 is determined empirically as described below.

7.3.2.4 Multi-Resolution

The computation of our algorithm is dominated by solving Eqn. (7.7), because the computational complexity of direct matrix inversion is $O(N^3)$. To reduce computation time, a multi-resolution scheme is adopted. The sample points are subsampled for smaller m, so that only a subset of the samples is used in the interpolation. The subsample rate increases as the algorithm progresses, and the complete set of samples is used only in the last few integration steps. This multi-resolution scheme can greatly reduce the computation while not affecting the accuracy of the displacement field reconstruction.

7.3.2.5 The Algorithm

Given sample points $\mathbf{x}(t_0) = \mathbf{x}_n$, the projection directions $\mathbf{e}_n$, and $w_n = \mathbf{e}_n \cdot \mathbf{u}(\mathbf{x}_n)$ for $n = 1, \ldots, N$, and a dense 3D grid of data points $\mathbf{y}_k$ for $k = 1, \ldots, K$ of which the 3D displacement vectors are to be computed, the dense displacement field is reconstructed as in Algorithm 7.1.

Algorithm 7.1 Dense 3D Incompressible Displacement Field Reconstruction

1: Initialize ρ_0, M. Set $\mathbf{y}_k(0) = \mathbf{y}_k$, $\mathbf{w}(\mathbf{y}_k, 0) = 0$, and $\mathbf{w}(\mathbf{x}_n, 0) = 0$ for all k and n.

2: **for** $m = 0$ to $M - 1$ **do**

3: Set $s_m = m\delta$, $\rho_m = \rho_0(M - m - 1)/(M - 1)$;

4: Downsample the data points $\mathbf{x}(t_m)$ if needed;

5: Compute $r_n(s_m)$ using Eqn. (7.21) for all sample points;

6: Solve Eqn. (7.7) to compute the interpolating coefficients with samples $\{\mathbf{x}_n(s_m), \mathbf{e}_n, r_n(s_m)\}$ and $\rho = \rho_m$;

7: Compute the velocities $\mathbf{v}(\mathbf{y}_k(s_m), s_m)$ and $\mathbf{v}(\mathbf{x}_n(s_m), s_m)$ using Eqn. (7.14);

8: Set $\mathbf{w}(\mathbf{x}_n, s_{m+1}) = \mathbf{w}(\mathbf{x}_n, s_m) + \delta\mathbf{v}(\mathbf{x}_n(s_m), s_m)$, $\mathbf{w}(\mathbf{y}_k, s_{m+1}) = \mathbf{w}(\mathbf{y}_k, s_m) + \delta\mathbf{v}(\mathbf{y}_k(s_m), s_m)$, $\mathbf{x}_n(s_{m+1}) = \mathbf{x}_n(s_m) + \mathbf{w}(\mathbf{x}_n, s_{m+1})$, and $\mathbf{y}_k(s_{m+1}) = \mathbf{y}_k(s_m) + \mathbf{w}(\mathbf{y}_k, s_{m+1})$;

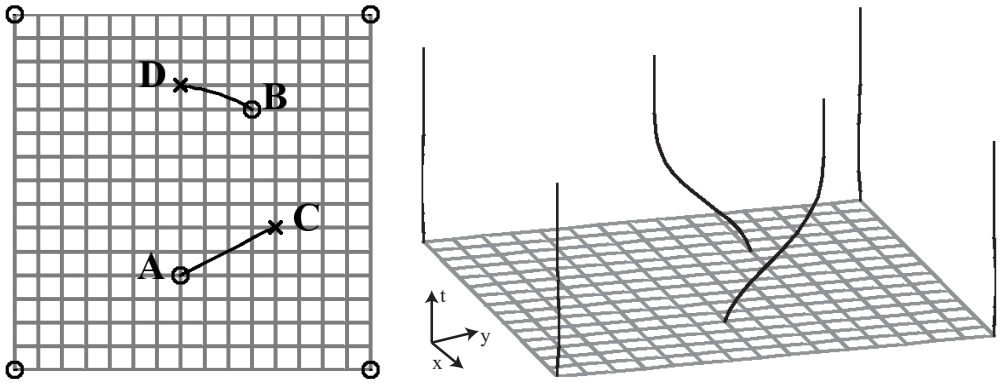
9: **end for**

10: Set $\mathbf{u}(\mathbf{x}_n) = \mathbf{w}(\mathbf{x}_n, s_M)$ and $\mathbf{u}(\mathbf{y}_k) = \mathbf{w}(\mathbf{y}_k, s_M)$.

7.4 Experiment Results

7.4.1 A 2D Example

We demonstrate our approach using a simple example as shown in Fig. 7.3(a). In this example, we want to reconstruct the 2D dense deformation field using six landmarks. Four landmarks are the corner points of the grid, and are fixed during the deformation. The other two landmarks A and B move to C and D respectively. For



(a) (b)

Figure 7.3: A simple 2D example. (a) The 6 matched landmarks marked in circles: four fixed corner points, point A matching to point C, and point B matching to point D. (b) The temporal trajectories of the 6 landmarks computed using our approach.

comparison, we reconstructed the deformation fields using direct TPS interpolation, direct DFVS interpolation, Bistoquet's method, and our approach.

The trajectories of points A and B computed using our method are shown as the lines in Fig. 7.3(a) that connect A to C and B to D. Fig. 7.3(b) shows the temporal trajectories of the 6 landmarks.

The deformation fields computed from all the four methods are shown in Fig. 7.4. It is observed that the areas of the mesh grids are not preserved in TPS and DFVS methods, and they are better reserved in Bistoquet's approach and our approach. We also computed the Jacobian determinant of the deformation computed using the four methods, and show them in Fig. 7.5. The Jacobian determinant measures the compressibility of the deformation field, and a value of 1 indicates incompressible motion. The TPS interpolation results a smooth deformation field, but the incompressibility is

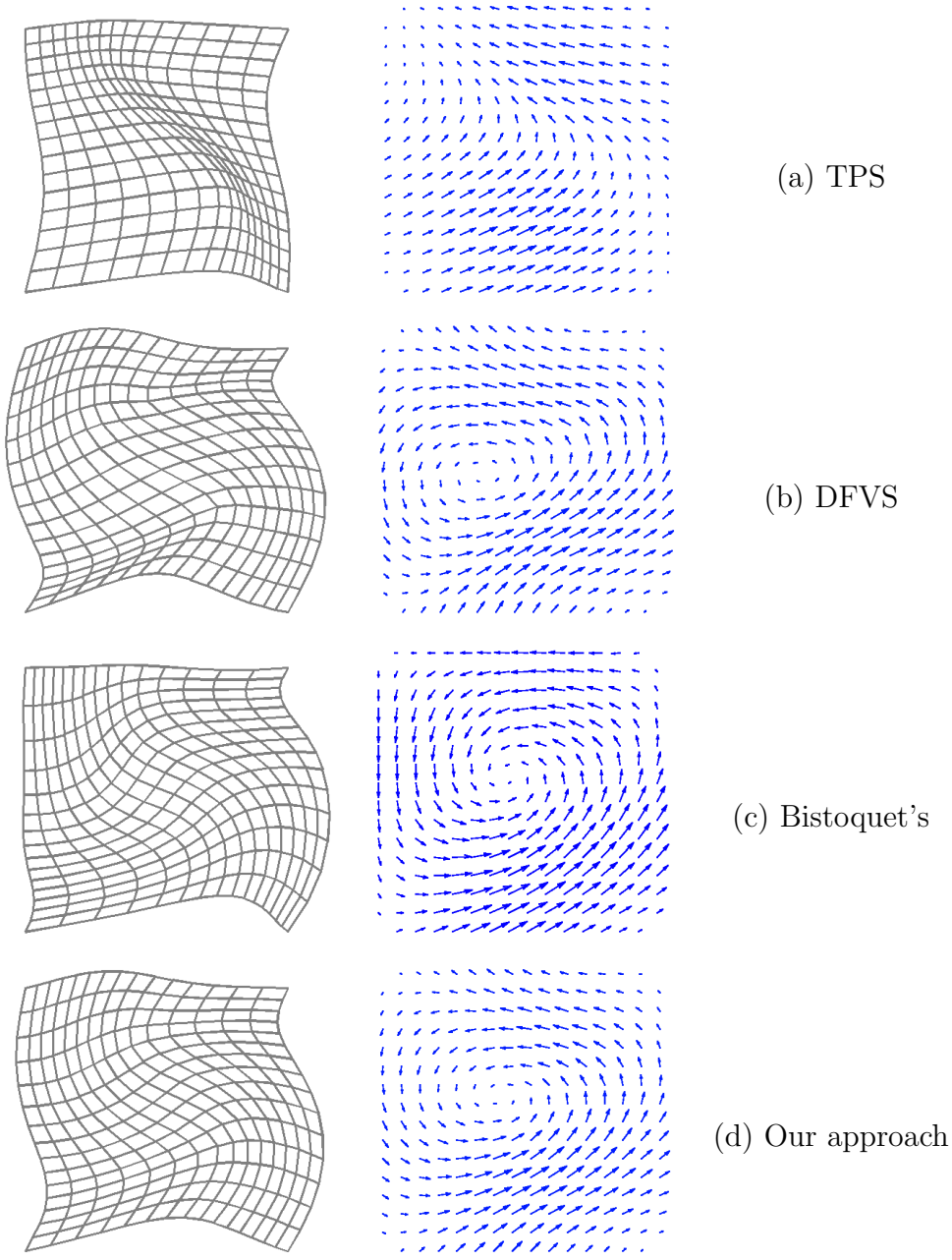
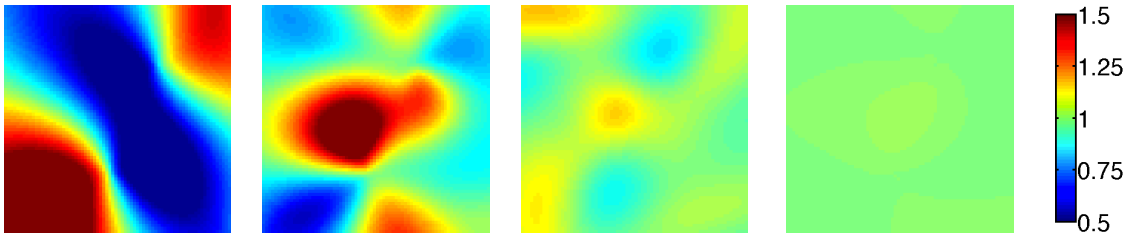


Figure 7.4: The deformation fields computed using (a) thin-plate spline interpolation, (b) direct DFVS interpolation, (d) Bistoquet's approach, and (d) our approach. Left column: the deformed meshes. Right column: the vector fields.



(a) TPS (b) DFVS (c) Bisotquet's (d) Our approach

Figure 7.5: The Jacobian determinant of the deformation computed using (a) TPS, (b) DFVS, (c) Bistoquet's approach, and (d) our approach. (1 means incompressible deformation.)

not preserved because the two components of motion are computed independently. In the DFVS result, the incompressibility is not preserved because for large deformations it is not correct to assume the deformation field to be divergence-free. Bistoquet's approach improved the incompressibility using the backward-forward averaging strategy with a considerable amount of errors. Our method successfully reconstructed the incompressible deformation field with negligible error resulting from discretization.

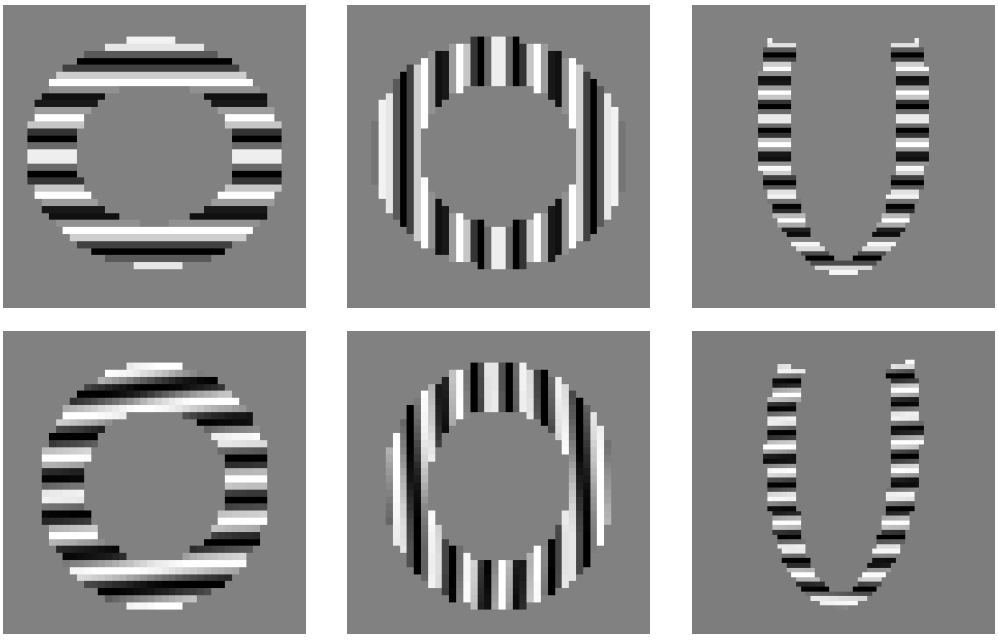
7.4.2 Validation with Cardiac Motion Simulator

To validate the method, we performed a simulation experiment using a cardiac motion simulator for tagged MRI [100]. In the simulator, the LV is defined as a deformed prolate spheroidal shell, with focal radius 35 mm, inner radius 0.35 and outer radius 0.55. The motion is controlled by 13 time-varying parameters that define rotation, translation, shears, ellipticallization, torsion, and radially dependent compression. In our simulation, the parameter that controls compression was set

to zero so that the simulated motion is incompressible. The values of the other parameters came from a bead experiment on a dog heart [101].

We simulated tagged images on six SA image planes with 10 mm separation and six radial LA image planes with a 30 degree separation. On each SA image plane two images were generated with horizontal and vertical tags, respectively. One image was generated on each LA image plane with tag planes parallel to the SA image planes. The tag spacing was 10 mm, and the pixel size of the images was 1.17 mm. The tags were applied at end-diastole and the sequences of images were generated throughout the cardiac cycle. The period of cardiac cycle was 1 second, and 16 time frames were generated with a temporal separation of 66 msec. Example images in both undeformed and deformed states are shown in Fig. 7.6. We then computed the incomplete 1D/2D displacements of all points inside the LV and on these images from every time frame back to the reference time, and used them as the samples in our algorithm.

The six SA images encompassed a slab with 50 mm thickness. For validation, we sampled the slab into 46 slices with 1.11 mm separation. The LV motion inside the slab can be directly computed from the simulator and is taken as the ground truth. Our method was then applied to compute the 3D displacement vectors of all LV points in the 46 slices from the sample points on the simulated six SA and six LA slices with M=20 steps, and the smoothing parameter $\rho_0 = 1.15$, and the results were compared to the ground truth. To compare, we also computed the

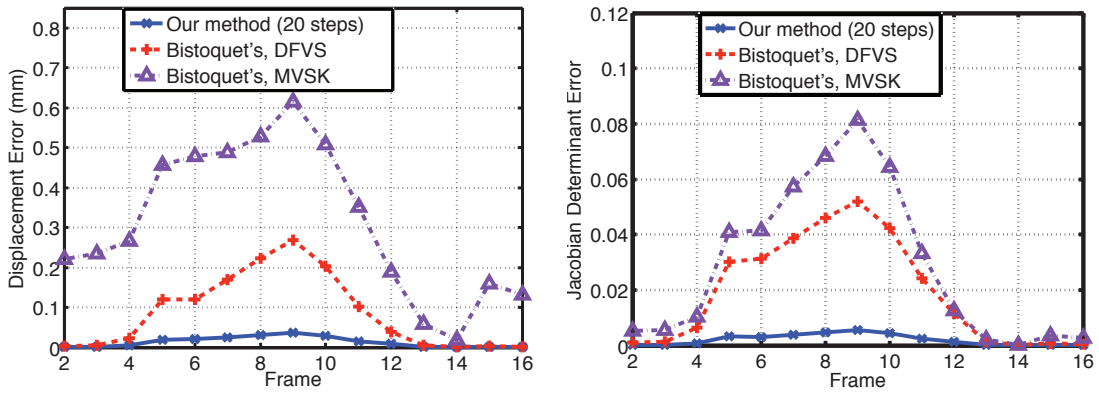


(a) (b) (c)

Figure 7.6: Simulated tagged MR images. SA images with (a) horizontal tags and (b) vertical tags, and (c) LA image with horizontal tags at (top row) undeformed state and (bottom row) deformed state.

displacement field using Bistoquet's method [94], i.e., direct divergence-free spline interpolation with backward-forward averaging. In Bistoquet's method, we tested both the divergence-free matrix-valued smooth kernel (MVSK) in Bistoquet et al. [94] $K(\mathbf{x}) = [\triangle \mathbf{I} - \nabla\nabla^T]e^{-\alpha||\mathbf{x}||^2}$, and the DFVS kernel. The results were then compared to the ground truth.

We first computed the error between the reconstructed displacement field and the ground truth at all LV points on the 46 slices. The RMS error of the x, y, and z components of the three approaches at all time frames are shown in Table 7.1. The mean magnitudes of the error are shown in Fig. 7.7(a). We observed that Bistoquet's



(a) (b)

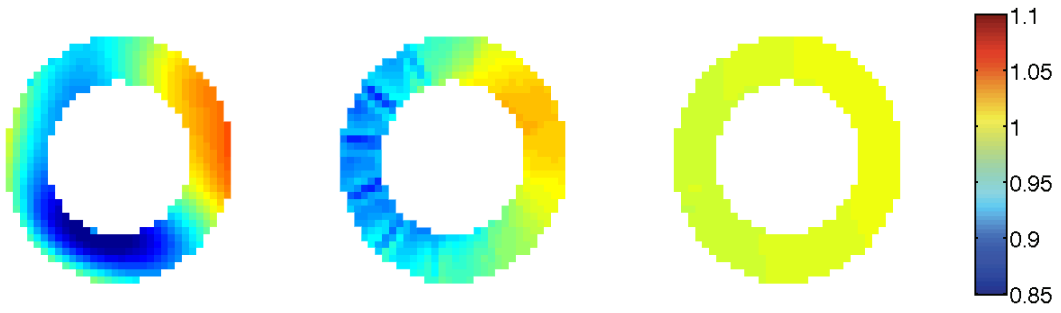
Figure 7.7: The comparison of our method against Bistoquet's method with both DFVS and MVSK. (a) The average displacement error and (b) the average error of the Jacobian determinant over all time frames.

method with MVSK produced much larger error at all time frames than with DFVS. At the beginning and end of the cardiac cycle when the cardiac motion is small, both Bistoquet's method with DFVS and our method give very small error. However, in the middle of the cardiac cycle when the cardiac motion is big, the error of Bistoquet's method with DFVS is about 10 times greater than our method.

To evaluate the incompressibility of the resulting displacement fields, we also computed the determinant of Jacobian using central difference operators. The average absolute difference between the Jacobian determinant and unity at all the time frames were shown in Fig. 7.7(b). It can be seen that incompressibility is much better preserved in our method than Bistoquet's approach at all time frames. At time frame 9 when all methods have the largest error, the average error of the Jacobian determinant of our method was 0.0056, Bistoquet's method with DFVS was 0.052,

Table 7.1: RMS error (mm) on the deformation components.

Time	Our method			Bistoquet's, DFVS			Bistoquet's, MVSK		
	x	y	z	x	y	z	x	y	z
2	0.0005	0.0006	0.0002	00032	0.0033	0.0017	0.1637	0.1707	0.1038
3	0.0014	0.0012	0.0027	0.0038	0.0039	0.018	0.1743	0.1734	0.1185
4	0.0039	0.0062	0.0016	0.0205	0.0208	0.0098	0.2191	0.1969	0.1114
5	0.0109	0.0193	0.0075	0.1008	0.1004	0.0426	0.2410	0.2842	0.3684
6	0.0149	0.0250	0.0073	0.1034	0.0968	0.0502	0.3125	0.3673	0.2799
7	0.0161	0.0173	0.0060	0.1993	0.1888	0.0572	0.3602	0.3934	0.2230
8	0.0227	0.0180	0.0077	0.2947	0.2788	0.0650	0.3902	0.3901	0.2849
9	0.0269	0.0203	0.0098	0.3658	0.3473	0.0716	0.4503	0.4377	0.3588
10	0.0212	0.0158	0.0063	0.2644	0.2500	0.0599	0.3750	0.3872	0.2593
11	0.0114	0.0110	0.0043	0.1075	0.1014	0.0362	0.2511	0.2872	0.1537
12	0.0052	0.0094	0.0033	0.0339	0.0336	0.0169	0.1349	0.1446	0.0851
13	0.0009	0.0013	0.0003	0.0046	0.0048	0.0019	0.0340	0.0449	0.0336
14	0.0003	0.0003	0.0000	0.0004	0.0003	0.0000	0.0079	0.0164	0.0111
15	0.0007	0.0007	0.0002	0.0025	0.0026	0.0011	0.1219	0.1054	0.0890
16	0.0006	0.0005	0.0001	0.0015	0.0015	0.0007	0.0929	0.0744	0.0948
mean	0.0085	0.0092	0.0033	0.0941	0.0896	0.0260	0.2081	0.2171	0.1610



(a) (b) (c)

Figure 7.8: The Jacobian determinant of deformation fields. (a) Bistoquet's method with MVSK, (b) Bistoquet's method with DFVS, and (c) our method.

and Bistoquet's method with MVSK was 0.081. This is consistent with the fact that the divergence-free assumption of the displacement field is only valid for small deformations, and is farther from the truth for larger deformations. Fig. 7.8 shows the Jacobian determinant computed from the three methods on one of the slices at time frame 9.

7.4.2.1 Parameter Optimization

The accuracy of our method depends on the choice of the number of steps M and the smoothing parameter ρ_0. To determine the optimal parameters, we applied our method to time frame 9 with varying M and ρ_0. The time frame 9 was picked because the deformation was the largest among all the time frames.

We first fixed the step number $M = 20$, and reconstructed the displacement fields with different ρ_0. The average displacement error is shown in Fig. 7.9. The error is small for all the selections of ρ_0 (< 0.045 mm). It is minimal for $\rho_0 = 1.15$ and slowly

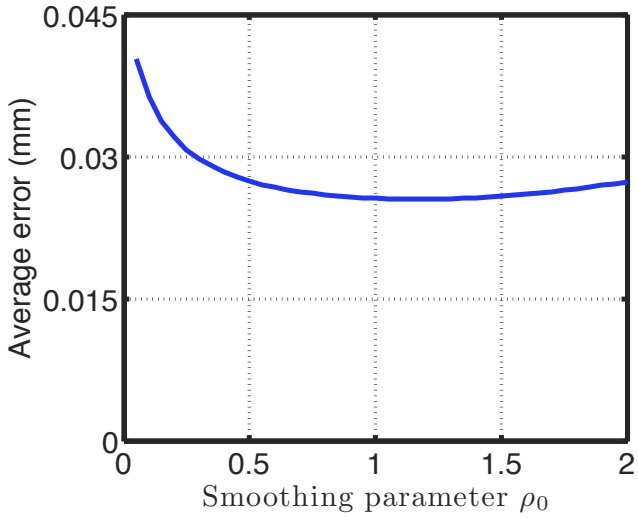


Figure 7.9: The average displacement error of our method with varying smoothing parameter ρ_0.

increases when ρ_0 increases. In addition, it is also shown that our method is pretty robust to the selection of ρ_0 because the error remains small for a reasonably large ρ_0.

Next, we computed the error by varying the number of steps and fixing $\rho_0 = 1.15$, and the results are plotted in Fig. 7.10. The displacement error rapidly decreases when the the number of steps increases. This is expected because when the number of integration steps increases, the step interval decreases. Therefore the linear approximation of velocity at each step is more accurate. Because the computation time directly depends on the number of steps, in practice we choose step number $M = 20$ as a tradeoff.

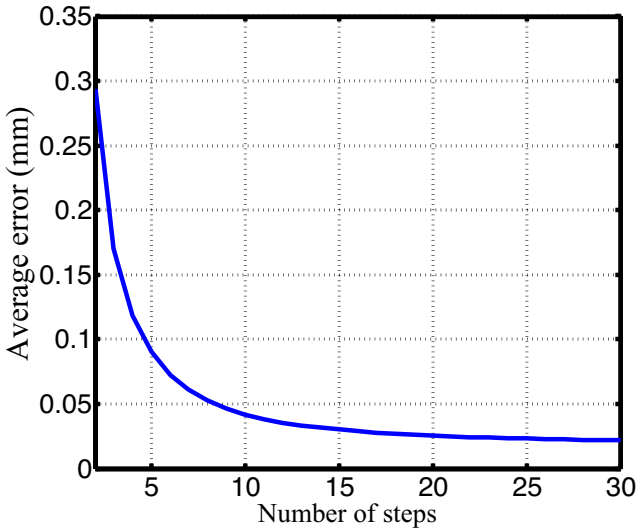
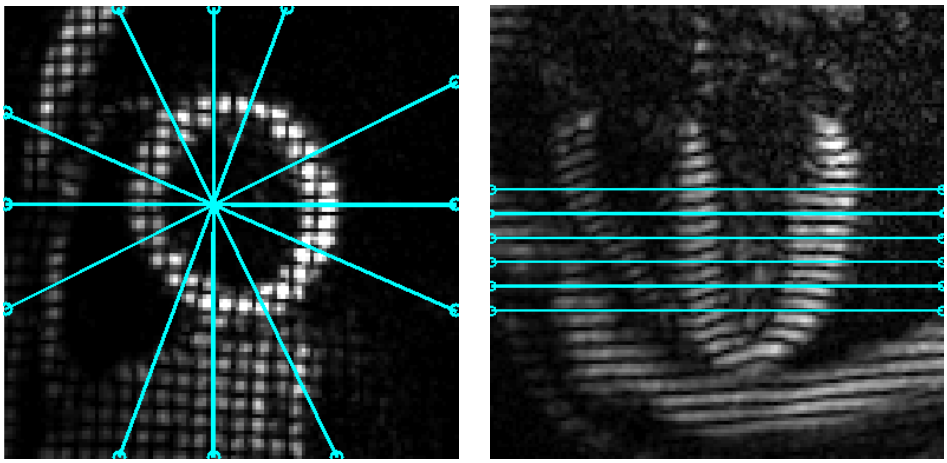


Figure 7.10: The average displacement error of our method with varying numbers of integration steps.

7.4.3 *In Vivo* Cardiac Motion Estimation

CSPAMM tagged MR images of the heart were acquired on a normal subject using a spiral sequence on a Phillips 3T Achieva MRI scanner (Philips Medical Systems, Best, NL). An approved IRB protocol was used and informed consent was obtained. The imaging parameters were: tag spacing = 12 mm, image size = 256×256, FOV = 320 mm, temporal resolution = 30 msec. The number of time frames was 20. We acquired both horizontally and vertically tagged images on twelve parallel SA image planes with a 4 mm slice separation. Eight LA images were acquired with horizontal tags; however, only six of these were used in our experiments because the other two were corrupted with artifacts. We divided the SA slices into two interleaved groups (even and odd slice numbers) so that the slice separation within each group is 8 mm. The first group of six SA slices and all six LA slices were used to reconstruct a 3D,

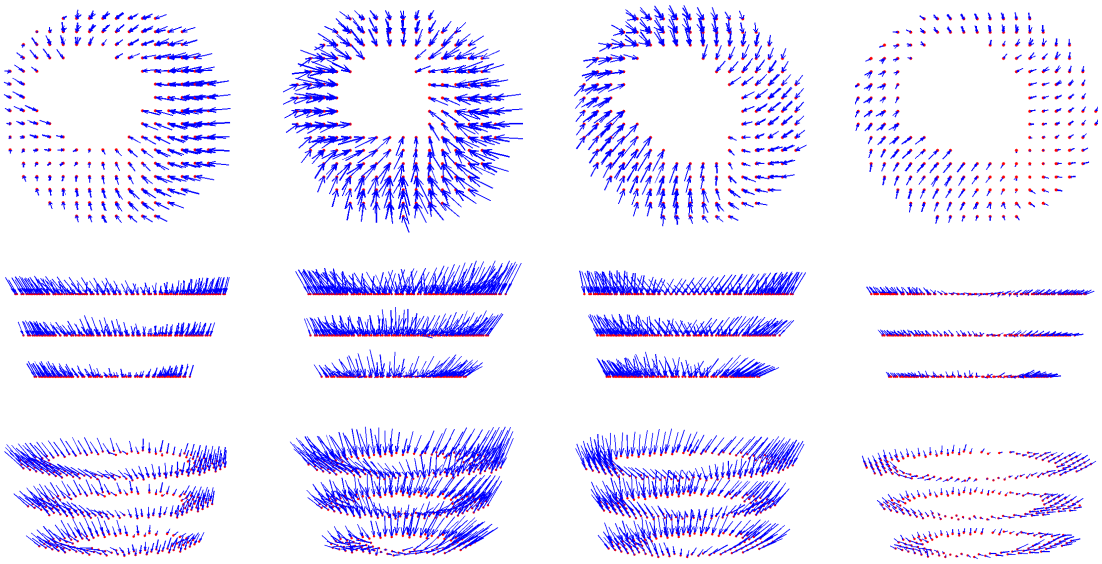


(a) Short axis (b) Long axis

Figure 7.11: A tagged (a) SA and (b) LA image. The SA image shown is the product of the separately acquired horizontal and vertical tagged images only for visualization purpose. The lines overlaying these images depict the geometry of the acquired (a) LA and (b) SA images.

dense, incompressible displacement field of the LV using our approach. The six SA slices in the second group were used for validation. The relative slice locations of the six SA slices in the first group and the six LA slices are illustrated in Fig. 7.11. All these images were then processed with HARP and HARP refinement to get the harmonic phase images and to compute the 2D motion of all the imaged tissue points on the SA slices from each imaged time frame back to the reference time, and the 1D motion of all the imaged tissue points on the LA slices. The LV myocardium regions in all the tagged images were manually delineated.

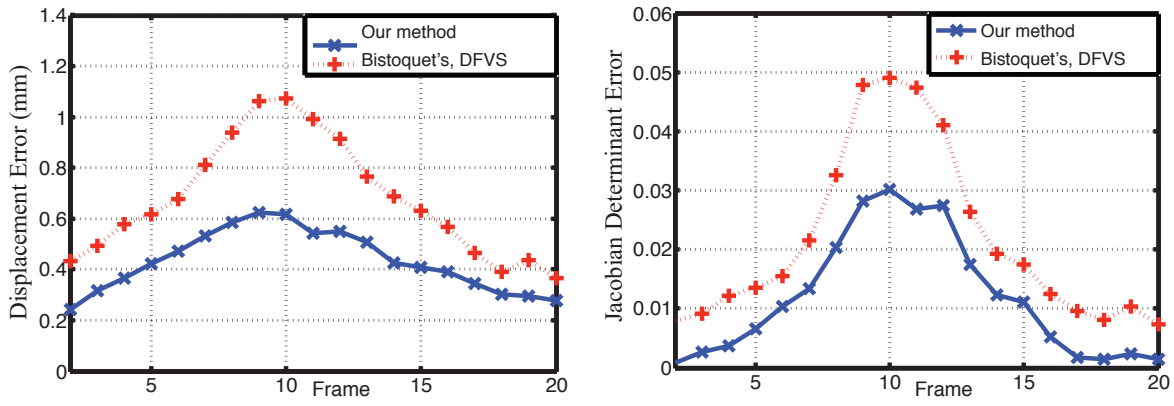
We applied Algorithm 7.1 with $M = 20$ integration steps and a smoothing parameter $\rho_0 = 1.15$. The reconstructed 3D displacement field in the LV regions of three SA slices are shown in three views in Fig. 7.12 at time frames 5, 10, 15, and 20. Since the



(a) Time 5 (b) Time 10 (c) Time 15 (d) Time 20

Figure 7.12: The 3D displacement fields illustrated using three SA slices. From top to bottom: three different views; From left to right: the displacement fields at different time frames.

displacements fields are Eulerian, the displacement vectors shown in Fig. 7.12 end on the spatial slices from which the data is collected, and start from where these material points are located at the reference time. The reconstructed displacements of points on the LV in the validation slices were compared with the 2D displacement projection computed using HARP. For comparison, we also computed the displacement fields at all time frames using Bistoquet's approach [93] with DFVS. The average displacement errors on the validation slices of the two methods are shown in Fig. 7.13(a). In both methods, we observe larger error at the end of systole (time frame 10) because the displacement is larger. At time frame 10, the average displacement error of our



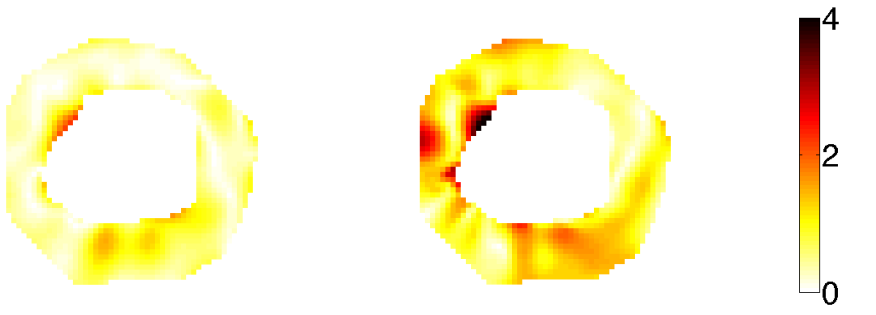
(a) Displacement error (b) Jacobian determinant error

Figure 7.13: The comparison of our method with Bistoquet's method using DFVS. (a) The average displacement error and (b) the average error of Jacobian determinant over all time frames.

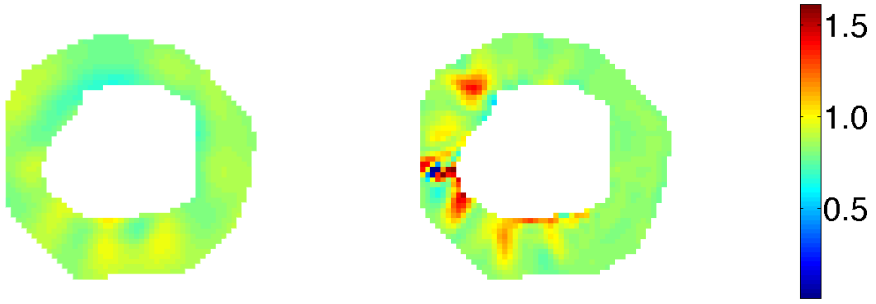
method was 0.61 mm, and Bistoquet's method was 1.08 mm. Figs. 7.14(a) and (b) show the displacement error maps on the 5th validation slice at time frame 10 of our method and Bistoquet's method, respectively.

We also compared the incompressibility of the reconstructed motion fields using using the proposed method and Bistoquet's method with DFVS, and the average absolute error between Jacobian determiant and unity at all time frames is shown in Fig. 7.13(b). At time frame 10 when the heart deforms the most, the average absolute difference between Jacobian determinant and unity of our approach was 0.043. The deviation was mainly caused by both spatial and temporal discretization. The average deviation of Bistoquet's method was 0.081. Figs. 7.14(c) and (d) show the Jacobian determinant resulting from the two methods at the 5th validation slice.

With the dense 3D displacement field, we can readily compute the 3D Eulerian



(a) Our method (b) Bistoquet's method



(c) Our method (d) Bistoquet's method

Figure 7.14: (a-b) The displacement error maps on one slice for (a) our method and (b) Bistoquet's method with DFVS. (c-d) The Jacobian determinant of the deformation field computed on the same slice for (c) our method and (d) Bistoquet's method with DFVS.

strain tensor (cf. [95]). We computed the 2D circumferential and radial strain, and the longitudinal strain in the through-plane dimension on the SA slices. The results on one of the SA slices are shown in Fig. 7.15. One possible benefit of our new approach is to reduce the so-called "zebra artifacts" which are known to yield artifactual circumferential patterns in traditional HARP strains [102].

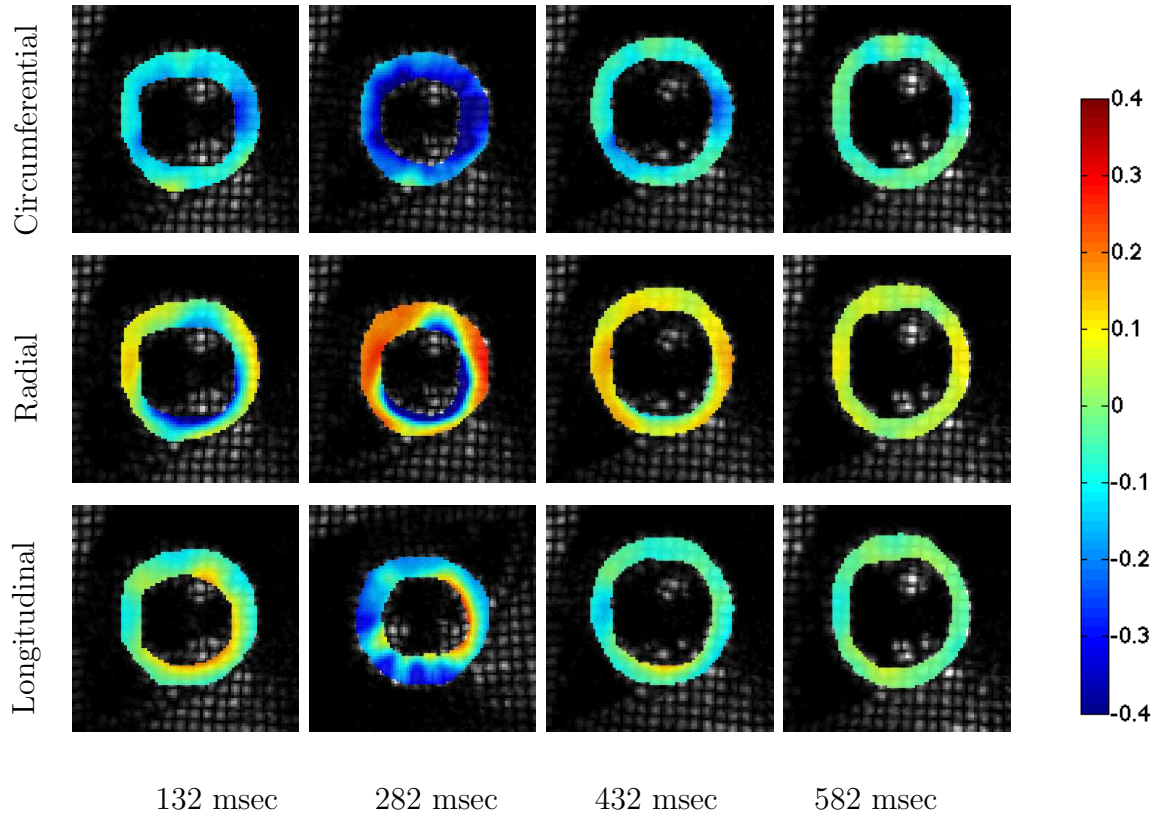
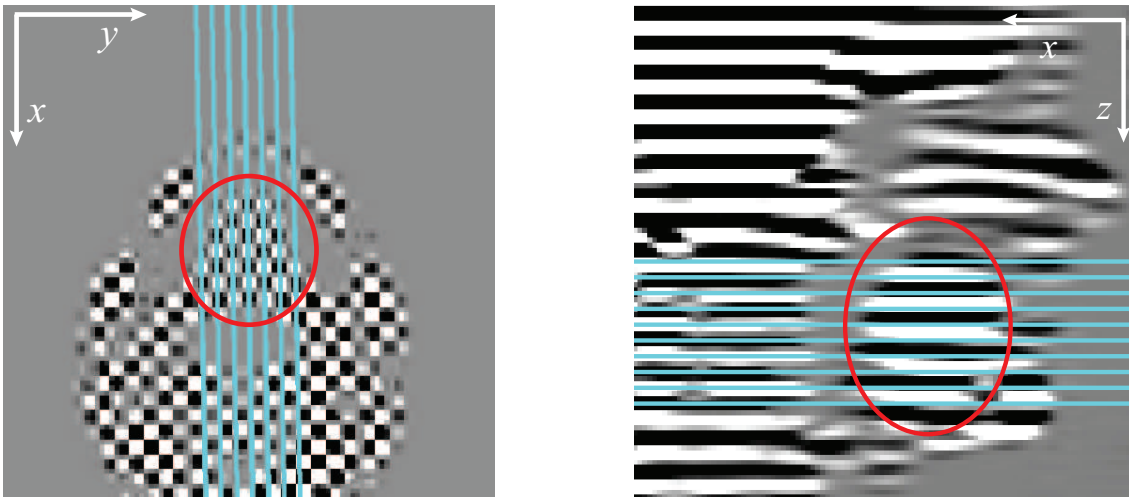


Figure 7.15: The 3D Eulerian strain in the LV region on one SA slice.

7.4.4 Tongue Motion Reconstruction

Our method was applied to reconstruct the 3D tongue motion of a normal subject during speech. The tagged MR images were acquired on a Siemens 3T Tim-Trio MRI scanner (Siemens Medical Solutions, Malvem, PA). The imaging parameters were: tag spacing = 12 mm, image matrix size = 128×128, pixel spacing = 1.875 mm, temporal resolution = 47 msec, number of time frame = 20. The image acquisition was coordinated using the triggering system as described in Chapter 5. The speech material studied was "eeoo". Tagged MR images were acquired on 10 axial slices



(a) Axial Image (b) Sagittal Image

Figure 7.16: Tagged (a) axial and (b) sagittal images. The axial image shown is the product of the separately acquired horizontal and vertical tagged images. The lines overlaying these images are the intersections of the shown images with the acquired (a) sagittal and (b) axial images. The red circles show the tongue location in the images.

and 7 sagittal slices to cover the whole tongue. On each axial slice, two images were acquired with orthogonal tag directions. On each sagittal slice, one image were acquired with tag planes parallel to the axial slices. Fig. 7.16 shows the relative locations of the image slices. The reconstructed 3D incompressible motion at all time frames are shown in Figs. 7.17 to 7.21. The three shown slices are the 1st, 4th, and 7th axial slices whose locations are shown in Fig. 7.16(b) as the 1st, 4th, and 7th lines from the top.

The reconstructed 3D motion may promote the understanding of tongue motion patterns during speech. By visually checking the images we have found that the transition from “ee” to “oo” happens mostly from the 8th to 12th time frames. It

can be seen from the reconstructed motion that, during "ee" (see Figs. 7.17 and 7.18), the top part of the tongue (see the top slice in the figures) does not move much, while the bottom part (see the bottom slice in the figures) is moving forward and a little upward. This is because the top of the tongue is touching the palate, which prevents the tongue from moving up. In the meantime, the muscle at the bottom part of the tongue is moving up and forward to push the tongue again the palate. During the transition from "ee" to "oo" (the 8th to 12th frames, shown in Figs. 7.18 and 7.19), the whole tongue starts to move downward and backward, and the top part of the tongue moves more than the bottom part. After the transition (see Figs. 7.20 and 7.21), the tongue holds its position to pronounce "oo" continuously.

We also computed the 3D Lagrangian strain tensor using the reconstructed 3D dense motion. Figs. 7.22, 7.23, 7.24, and 7.25 show the components of Eulerian strain tensors on four slices at different time frames. Though the exact relationship between 3D strain and tongue muscle activation remains a challenging problem and requires more exploration, some interesting observations can be found from these images.

Overall the top part of the tongue exhibits more strain than the bottom part. In the first slice from the top (Figs. 7.22), during the transition from "ee" to "oo", the tongue expands in the anterior-posterior (x) direction, and shortens in the head-foot (z) direction. It also shortens a little in the lateral (y) direction. This observation may be explained by the tongue behavior. To pronounce "ee", the tongue is raised up and elongated upward from its rest position so that it can touch the palate. In

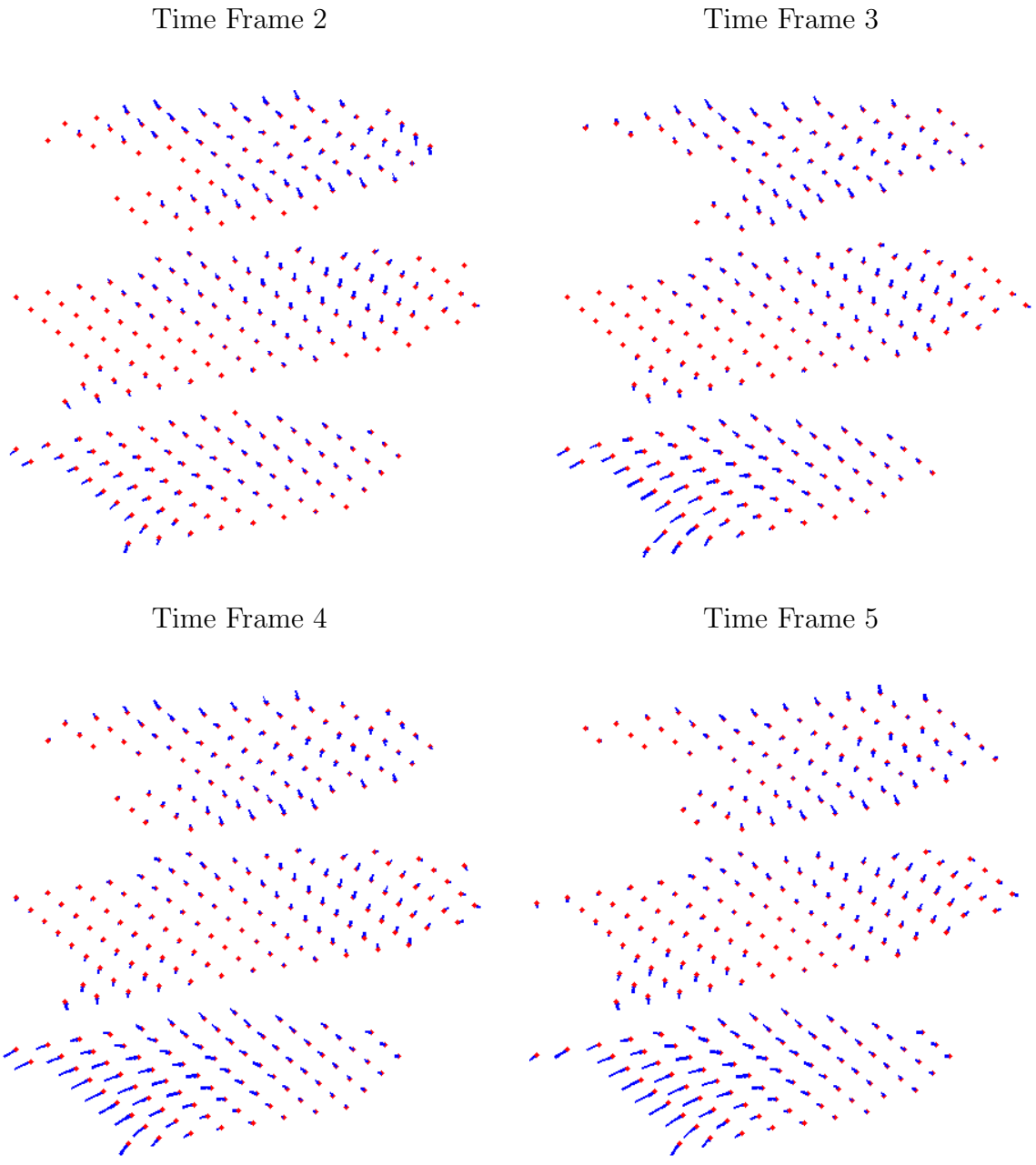


Figure 7.17: The 3D displacement fields of the tongue illustrated using three axial slices at frames 2 to 5. (The tongue tip is toward the right.)

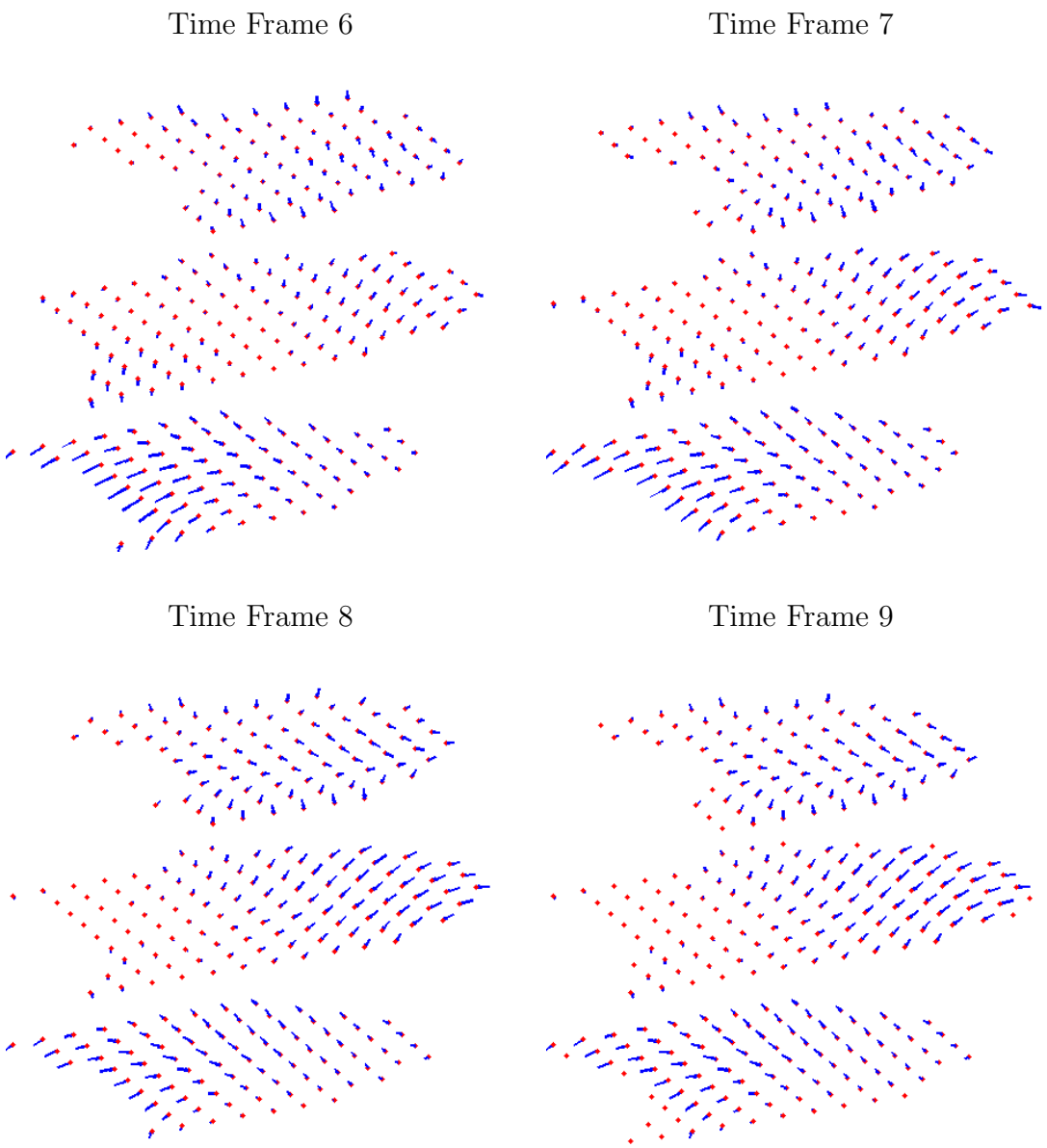


Figure 7.18: The 3D displacement fields of the tongue illustrated using three axial slices at frames 6 to 9. (The tongue tip is toward the right.)

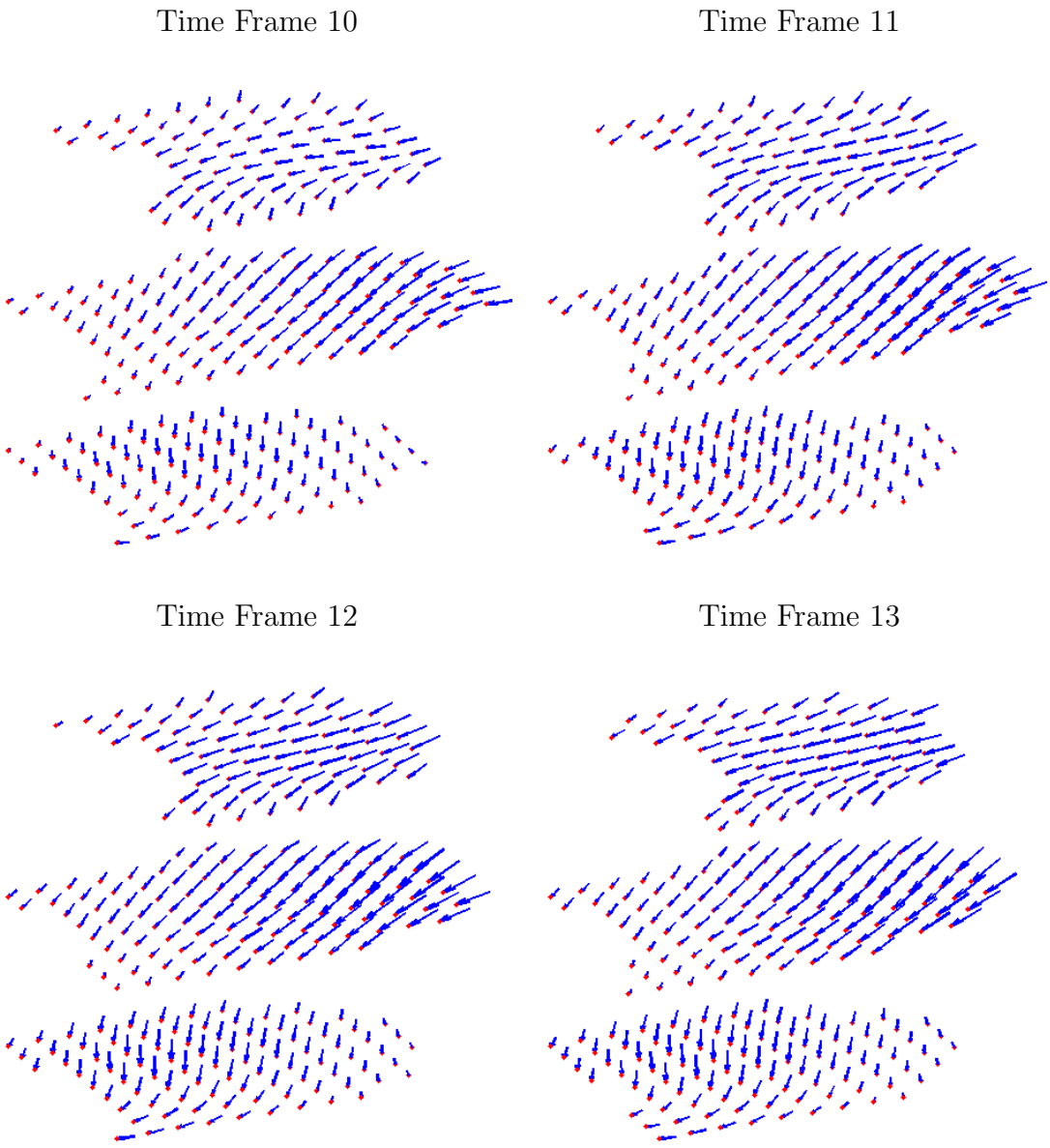


Figure 7.19: The 3D displacement fields of the tongue illustrated using three axial slices at frames 10 to 13. (The tongue tip is toward the right.)

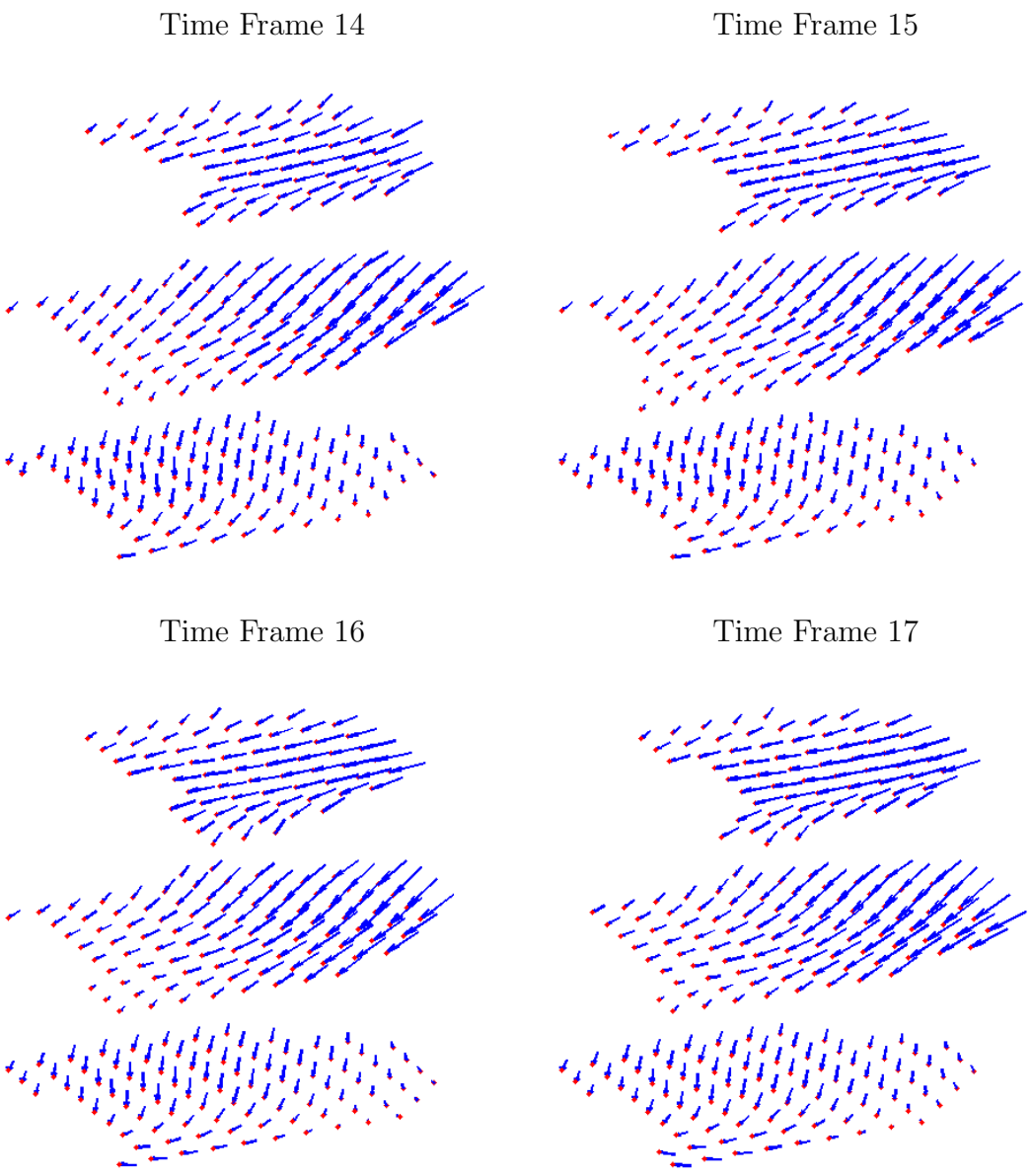


Figure 7.20: The 3D displacement fields of the tongue illustrated using three axial slices at frames 16 to 17. (The tongue tip is toward the right.)

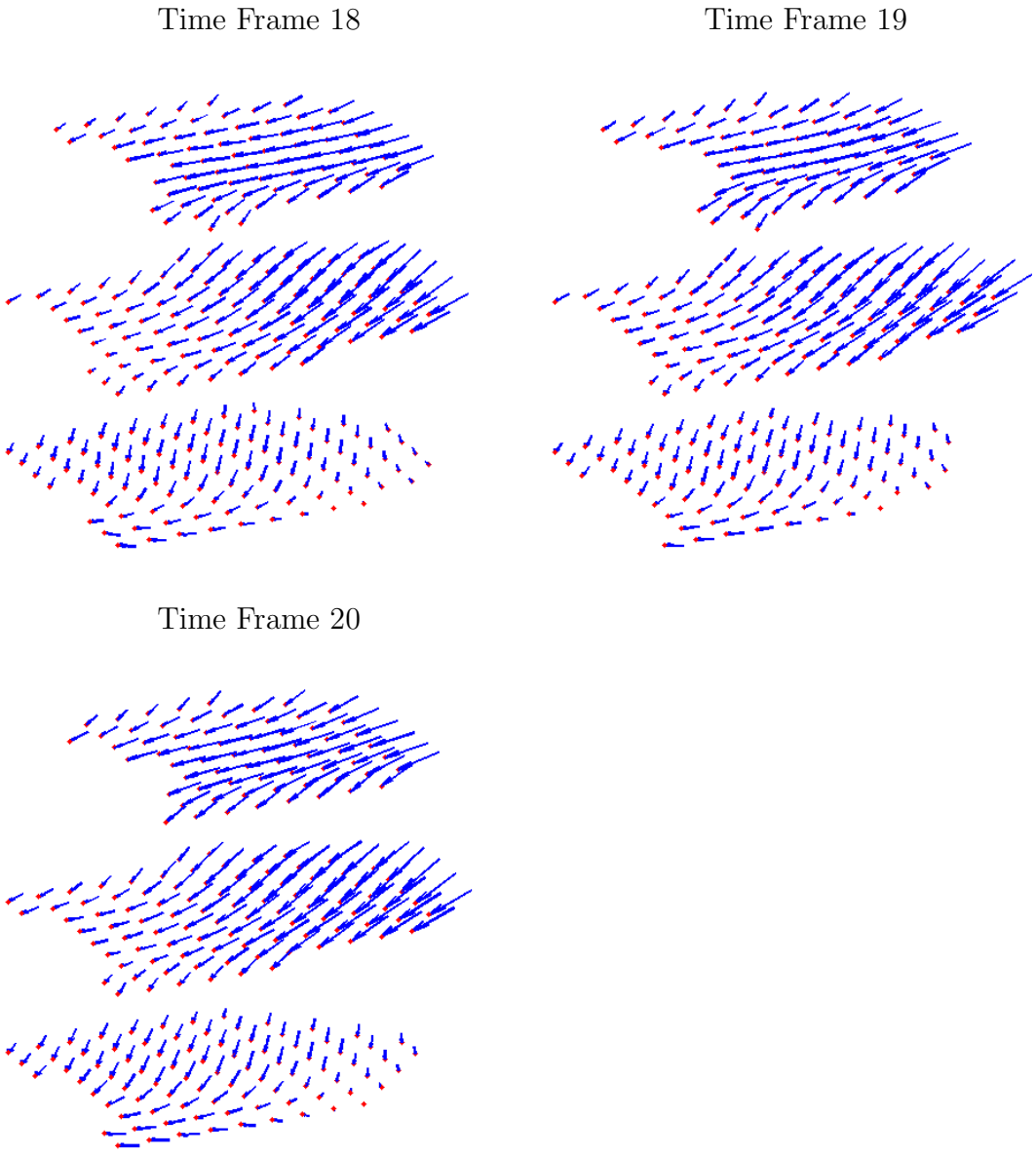


Figure 7.21: The 3D displacement fields of the tongue illustrated using three axial slices at frames 18 to 20. (The tongue tip is toward the right.)

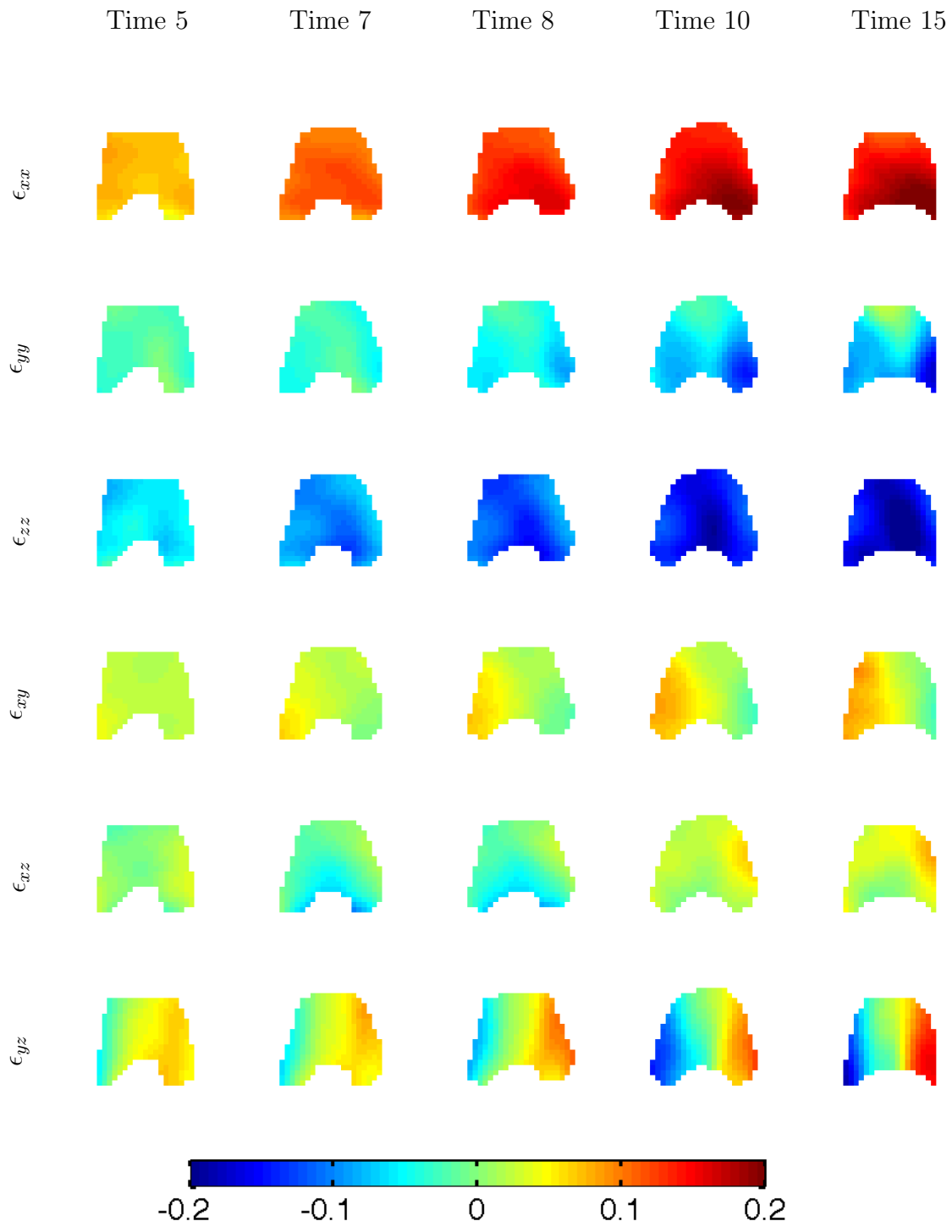


Figure 7.22: The components of 3D Eulerian strain tensor on the 1st axial slice at different time frames. The slice location is shown as the 1st line from top in Fig. 7.16(b). The x, y and z axis are as defined in Fig. 7.16 and the tongue tip is upward.

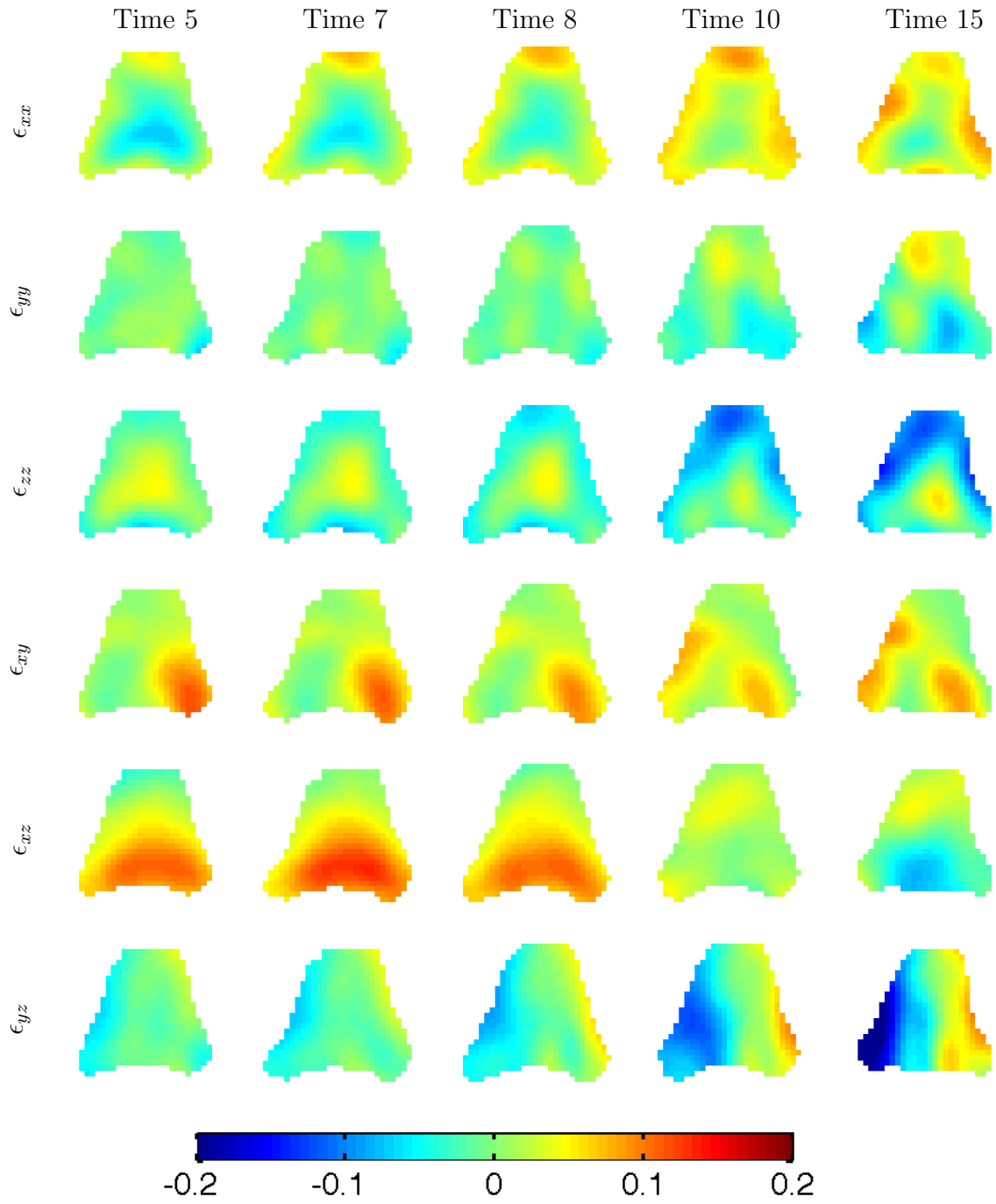


Figure 7.23: The components of 3D Eulerian strain tensor on the 3rd axial slice at different time frames. The slice location is shown as the 3rd line from top in Fig. 7.16(b). The x, y and z axis are as defined in Fig. 7.16.

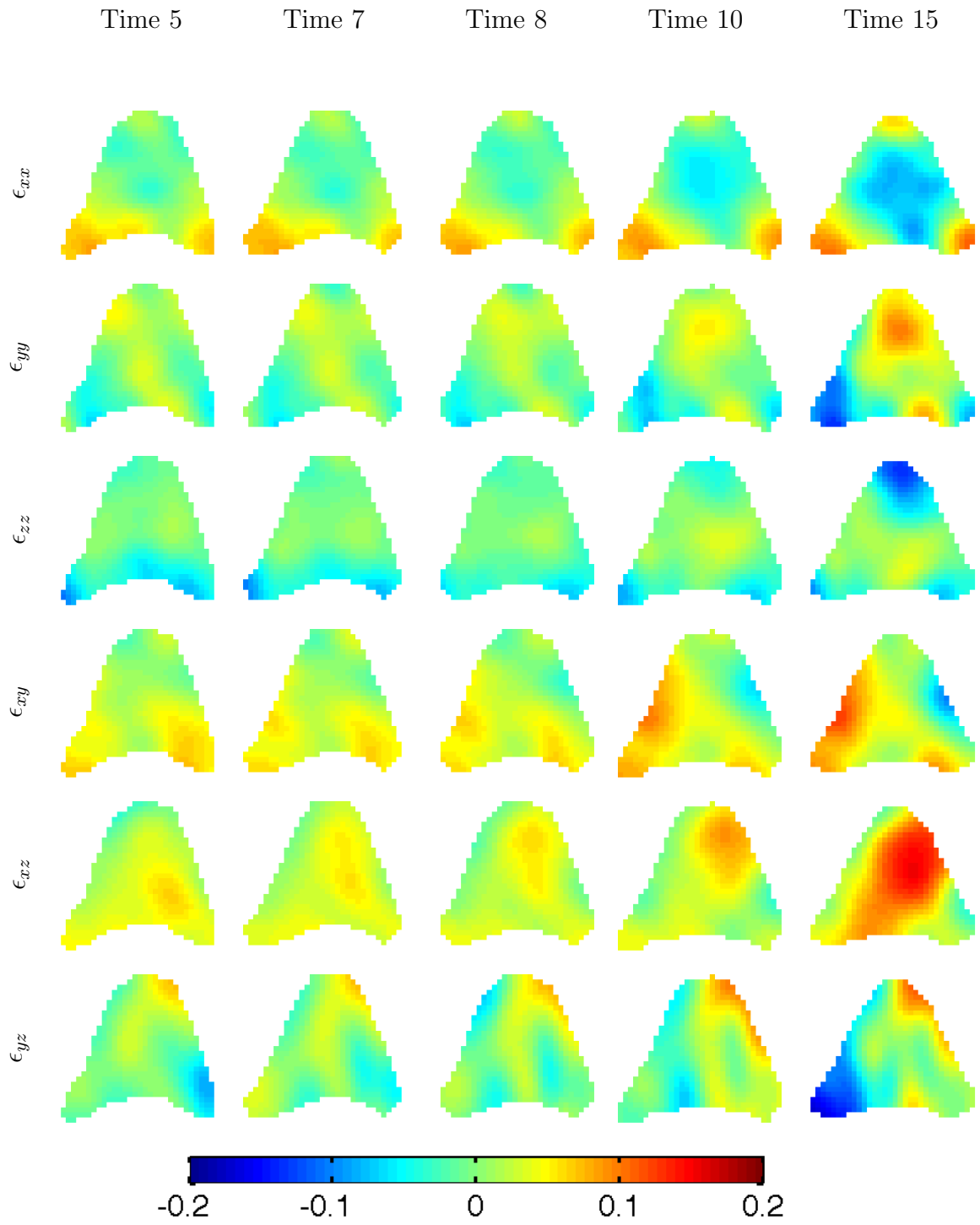


Figure 7.24: The components of 3D Eulerian strain tensor on the 5$^{\text{th}}$ axial slice at different time frames. The slice location is shown as the 5$^{\text{th}}$ line from top in Fig. 7.16(b). The x, y and z axis are as defined in Fig. 7.16.

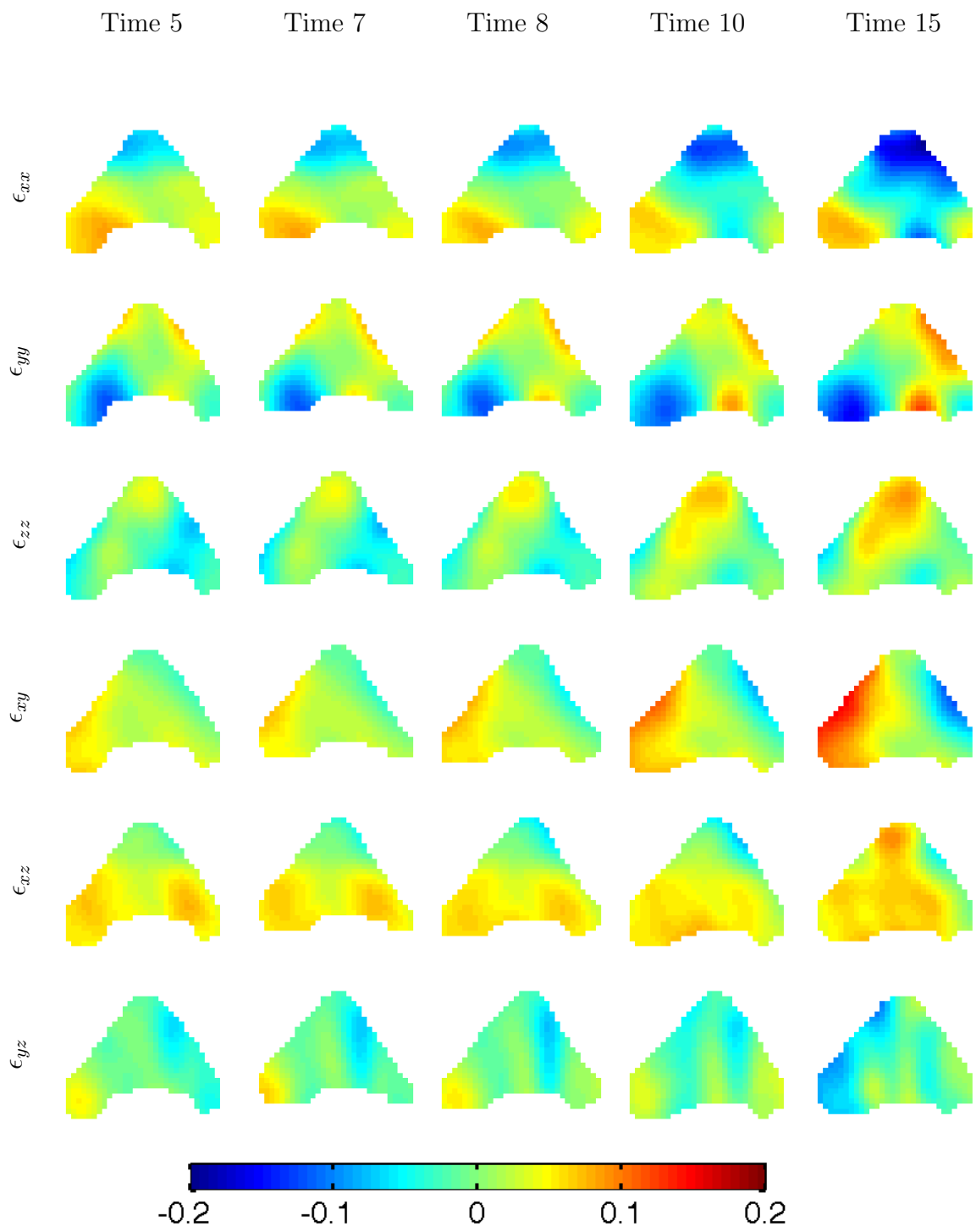


Figure 7.25: The components of 3D Eulerian strain tensor on the 7$^{\text{th}}$ axial slice at different time frames. The slice location is shown as the 7$^{\text{th}}$ line from top in Fig. 7.16(b). The x, y and z axis are as defined in Fig. 7.16.

the meantime, the tongue tip moves backward and the tongue muscle contracts in the anterior-posterior position. When the tongue is pushed against the palate, it expands in the lateral direction to preserve the volume. When transiting to "oo", the tongue relaxes and lower, and is positioned close to its rest position. Therefore in the transition, the tongue appears to be expanding in the anterior-posterior direction, shortening in the head-foot direction, and also shortening a little in the lateral direction. The tongue also shortens more laterally on the two sides than in the middle. The possible reason for this is that because of the arched shape of the palate, during "ee" the tongue is squeezed from the top more on the sides and as a result it expands more on the side. Hence during the transition it shortens laterally more on the sides of the tongue. This also explains why the middle of the tongue shortens more in the head-foot (z) direction.

7.4.5 Computation

The computation of DFVS involves the inversion of a $(N + 11) \times (N + 11)$ matrix (see Eqns. (7.5) and (7.7)) with N being the number of data samples, and its complexity is $O(N^3)$. Because of the large number of sample points from the HARP processing of tagged MR images, this matrix inversion dominates the computation of our method. For example, in the cardiac experiment, the number of samples varied from 21,585 to 22,492 at different time frames. In our implementation, 20 integration steps was adopted with three levels of subsampling of the data samples with sample

rates of 4, 2, and 1 respectively. Our algorithm was implemented in Matlab (Mathworks, Natick MA) and run on a server with 4 x 3.57 GHz processors and 32 GB RAM. In our implementation, the proposed method took about 50 minutes to reconstruct the displacement field at one time frame, while Bistoquet's method took about 25 minutes.

The computation time should not pose serious limitation for scientific study with fairly small number of data sets. For clinical use and the processing of large number of data sets, It is possible to reduce the computation time by optimizing our implementation, including the adoption of optimized matrix inversion algorithm, parallel computing, and so on.

7.5 Discussion

7.5.1 Choices of Kernel Matrix

The DFVS uses a kernel matrix with infinite support to interpolate divergence-free velocity field. There are some other works that have used different kernel matrices to interpolate or approximate divergence-free vector fields. Narcowich and Ward [103] developed the generalized Hermite interpolation using the divergence-free kernel matrix generated from Gaussian function $\phi_c(\mathbf{x}) = e^{-c||\mathbf{x}||^2} \in C^\infty$. The kernel matrix has the same form of Eqn. (7.9) except the function $h(\mathbf{x})$ is replaced by Gaussian

function, i.e.,

$$K_{\mathrm{G}}(\mathbf{x}) = \left[\triangle \mathbf{I} - \nabla\nabla^T\right] \phi_c(\mathbf{x}) = \left[(4\alpha - 4\alpha^2||\mathbf{x}||^2)\mathbf{I} + 4\alpha^2\mathbf{x}\mathbf{x}^T\right] e^{-\alpha||\mathbf{x}||^2}. \qquad (7.22)$$

This kernel matrix was used in the nearly incompressible model in Bistoquet et al. [94]. Another class of divergence-free kernel functions are derived from the compactly-supported *Wendland functions* [104]. Lowitzsch [105] applied it to certain incompressible fluid flows.

Comparing with the DFVS kernel, these kernel functions have higher order of smoothness, and have either compact support (Wendland functions) or values decreasing quickly with $\mathbf{x}$ (Gaussian functions). So the computation of spline coefficients can be more efficient because the matrix $\mathbb{K}$ in Eqns. (7.5) and (7.7) is more sparse. A downside is that they require careful parameter tuning. The parameters (α in the Gaussian function and the size of support in Wendland functions) greatly affect the interpolation results, and need to be carefully selected.

In addition, in the implementation we have found that, although for both Gaussian-based kernel and Wendland-based kernel the matrix $\mathbb{K}$ is theoretically positive definite, when the number of samples is large $\mathbb{K}$ can be close to singular to the working precision of computer and cannot be inverted. This is especially a problem for our application since the number of data samples is usually large ($\sim$20,000 samples). A way to get around is to use a large support (in Wendland functions) or a small α (in Gaussian function), but it will affect the accuracy of the interpolation.

7.5.2 A Large Deformation Diffeomorphism Approach

In our method, the strategy of estimating velocity fields at discrete time steps and integrating them together to reconstruct the displacement field is similar to that in *landmark matching via large deformation diffeomorphism* (LMLDD) [106]. Both methods compute the deformation (displacement) fields by integrating the velocity fields computed at discrete time steps based on the knowledge of displacements at sample points (in our approach) or matched landmarks (in LMLDD). The main difference is that LMLDD seeks smooth trajectories over time by directly constraining the temporal smoothness of velocity fields. In the appendix of the chapter, we extend LMLDD to incompressible motion reconstruction and demonstrate it using a 2D example, and call it iLMLDD.

ILMLDD appears to be a good solution in our application of estimating incompressible motion from tagged MRI. However, it is not feasible because of the unaffordable computation. LMLDD requires very expensive computation because it involves iteratively inverting $N \times N$ matrices, which has the complexity of $O(N^3)$ with N being the number of landmarks. Typically it takes several hours for LMLDD to run on several hundred landmarks. In our case of tagged MRI, it is common to have about 20,000 sample points or more. Therefore directly applying iLMLDD on the data may take weeks or months to run, which is not realistic.

7.6 Summary

In this chapter, we presented an approach to reconstruct a 3D, dense, incompressible displacement field from tagged MR images. Our method uses a divergence-free vector spline on incomplete and non-uniform sample data to interpolate velocity fields at discrete integration steps, and the displacement field is achieved by integrating these velocity fields. Comparing to previous methods, our method does not assume small deformations, and the reconstruction of displacement field at one time frame does not rely on the results from earlier time frames. Hence it prevents error accumulation arising from inaccurate estimation at earlier time frames caused by the small deformation assumption. Our method was validated with both numerical simulation and *in vivo* experiments. Our method has been successfully applied to reconstruct both the tongue motion during speech and the motion of the left ventricle of human heart.

Appendix — LMLDD for Incompressible Motion Reconstruction

LMLDD for Incompressible Motion

LMLDD [106] solves the large deformation landmark matching problem by reducing it to the small deformation landmark matching problem. For this, it computes the

curved trajectories through the integration of velocities instead of directly mapping the landmarks. This is achieved by minimizing an energy function defined on the velocity fields over time (in an interval of $[0, 1]$), i.e.,

$$\hat{\mathbf{v}}(\mathbf{x}, t) = \arg\min_{\mathbf{v}(\mathbf{x},t)} \int_0^1 \int_{\mathbb{R}^3} ||L\mathbf{v}(\mathbf{x}, s)||^2 d\mathbf{x}dt \,, \tag{7.23}$$

where L is a differential operator. Eqn. (7.23) can be theoretically solved by finding its Euler-Lagrangian equation, and numerically solved by iteratively computing the velocity fields on a finite number of time steps.

We demonstrate LMLDD using a simple 2D example (similar to the example in Joshi et al. [106] and the example in Section 7.4.1) and show it in Fig. 7.26. There are 6 landmarks which include the four corner points of the region, and two points in the middle (points **A** and **B** in Fig. 7.26(a)). The four corner points are fixed, and the two points **A** and **B** deform to **C** and **D**, respectively (Fig. 7.26(a)). We then reconstructed the deformation field using LMLDD with exponential kernel function:

$$\mathbf{K}_{\mathrm{g}}(\mathbf{x}) = e^{-\sqrt{(c)}||\mathbf{x}||}\mathbf{I} \,. \tag{7.24}$$

The trajectories of the landmarks over time are shown as red lines in Figs. 7.26(a) and (b). By reconstructing velocity fields at 10 discrete time steps and integrating these velocity fields, LMLDD successfully reconstructs a smooth diffeomorphic deformation field with correct topology, as shown in Figs. 7.26(c) and (d).

In the energy function of standard LMLDD, the differential operator L is diagonal, and the different components of the deformation field are considered separately. To

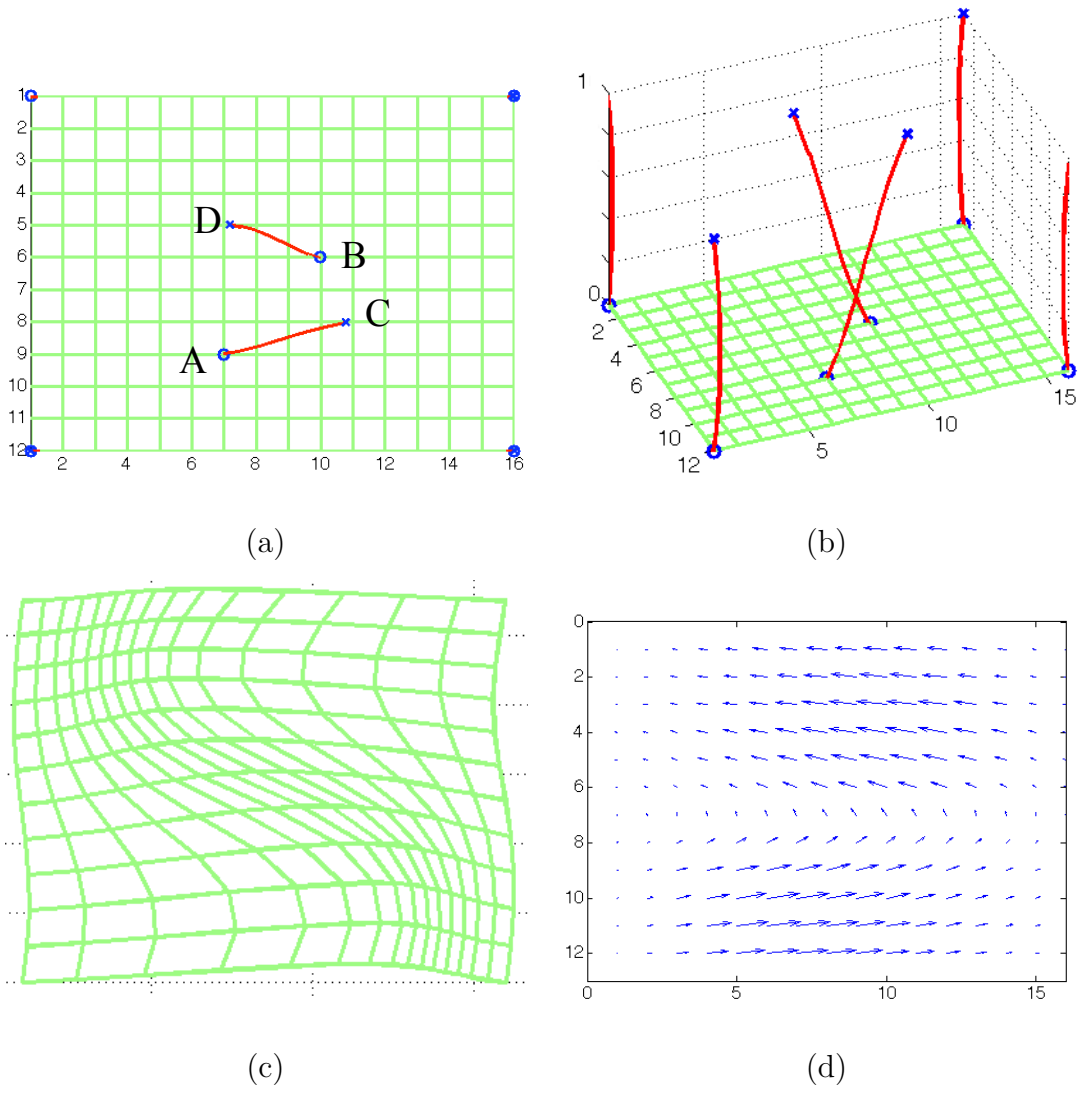


Figure 7.26: A 2D example of LMLDD. (a) The locations of landmarks and their trajectories computed using LMLDD (red lines). (b) The trajectories of landmarks over time displayed by adding the temporal axis to (a). The reconstructed 2D deformation field is represented with (c) deformed mesh and (d) vector flow respectively.

reconstruct incompressible tissue motion, we formulate the *incompressible LMLDD* (iLMLDD) landmark matching problem by enforcing the divergence-free condition,

i.e.,

$$\hat{\mathbf{v}}(\mathbf{x},t) = \arg\min_{\mathbf{v}(\mathbf{x},t)} J_{\mathrm{DF}}(\mathbf{v}) = \int_0^1 \int_{\mathbb{R}^3} \sum_{i=1}^{3} ||\nabla^k (\mathrm{rot}\mathbf{v}(\mathbf{x}))_i||^2 d\mathbf{x}\,, \tag{7.25}$$
$$\text{subject to: } \mathrm{div}\mathbf{v}(\mathbf{x}) = 0, \text{ and } \mathbf{u}(\mathbf{x}_n) = \mathbf{u}_n, n = 1, 2, ..., N.$$

with $\mathbf{u}(\mathbf{x}) = \mathbf{w}(\mathbf{x}, 1)$, and $\mathbf{w}$ is defined as in Eqn. (7.17). This can be solved in a similar way as in LMLDD [106] except the velocity at each time step is interpolated using DFVS.

We have applied iLMLDD on the same 2D example, and the results are shown in Fig. 7.27. Because of the incompressible constraint, the trajectories of the two landmarks in the center of the grid are more curved to preserve volume (Fig. 7.27(a)). The areas of all the cells on the grid are preserved so that they are equal (Fig. 7.27(c)) while in LMLDD (Fig. 7.26(c)), the cell areas are not preserved. The resulting deformation field of iLMLDD is shown in Fig. 7.27. As a preliminary work, we have successfully applied iLMLDD to 2D landmark-based prostate image matching [107].

ILMLDD (and LMLDD) can also be extended to matching incomplete data samples to be used in our application. Its solution is similar to LMLDD except the kernel matrix is projected in the same way as in DFVS from incomplete samples (see Eqn. (7.13)).

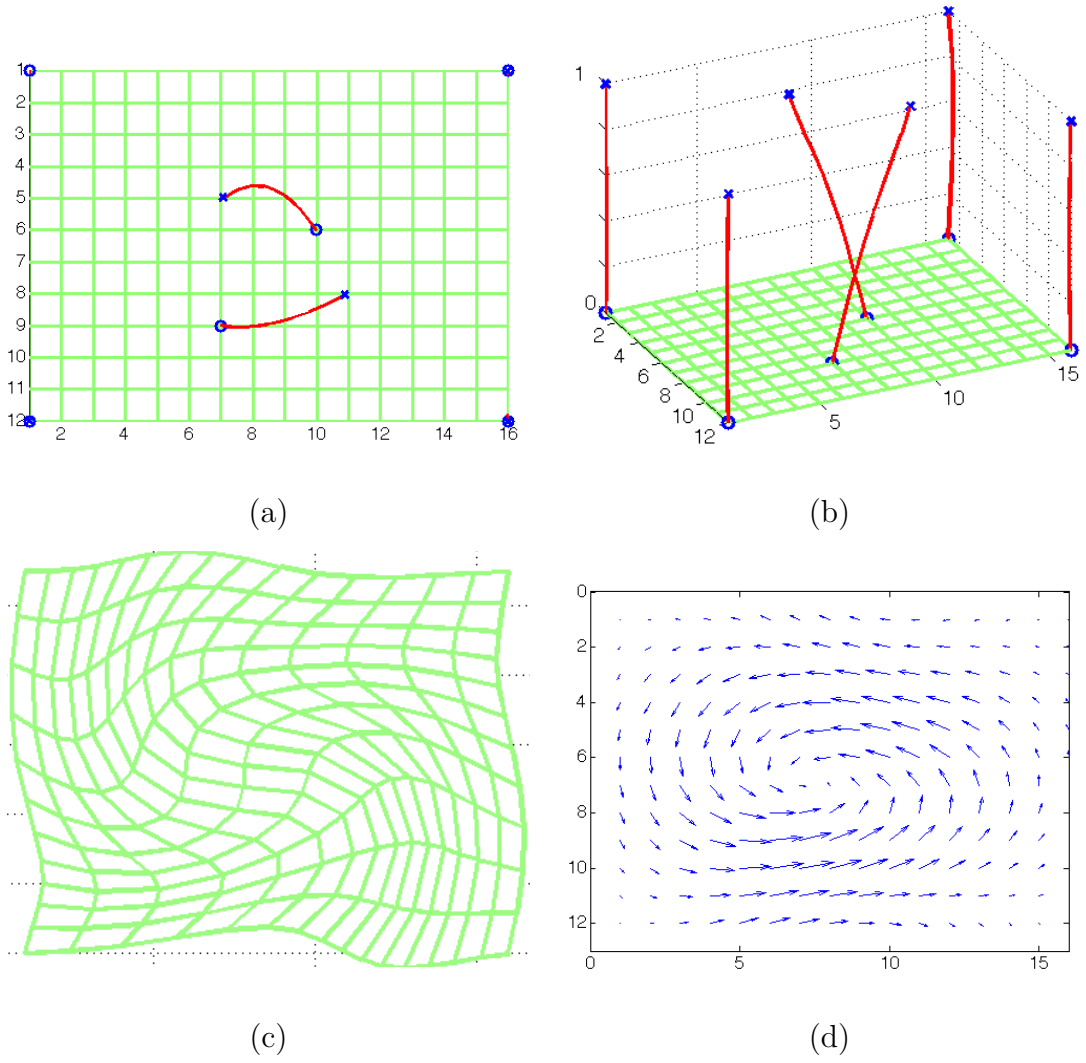


Figure 7.27: The result of iLMLDD on the same example as in Fig. 7.26. (a) The locations of landmarks and their trajectories computed using incompressible LMLDD (red lines). (b) The trajectories of landmarks over time displayed by adding the temporal axis to (a). The reconstructed 2D deformation field is represented with (c) deformed mesh and (d) vector flow respectively.

Chapter 8

Internal Tongue Motion Analysis on Glossectomy Patients: Preliminary Studies

8.1 Introduction

Glossectomy is a surgery that removes part or all of the tongue, and is an effective treatment for tongue cancer. It affects the patient's speech quality because of the changes in the motion of the post surgical tongue. In order to better interpret clinical observations and to provide data that can help predict optimal surgical outcomes, we need a better understanding of the tongue motion patterns and the underlying mechanisms of tongue muscles in glossectomy patients through careful comparison

with those in normal speakers.

One of the main factors that determine the effectiveness of tongue motion is the size of the residual tongue left after removal of the tumor [108, 109]. For fairly small tumors, variations in surgical closure procedure may affect tongue motion, including prime closure and free flap reconstruction. It is uncertain which of these methods produces better speech quality [109, 110]. The muscle mass is reduced in glossectomy, and this results in a reduction of muscle force and the need of additional supporting muscles. Free flap reconstruction adds dead weight to the remaining muscle tissue, but may provide mass for improved vocal tract shaping for speech. In addition, scar tissue may affect the tongue motion because it adds regions of rigidity. These effects together contribute to unclear speech after surgery. Even in cases where speech is unaffected, these two procedures are likely to create different motion patterns. Therefore it is required to have a better understanding of the different motion pattern of the patients with different closure procedures via comparison with normal speakers.

Tongue motion is produced through muscle activation. It has been observed that glossectomy patients tend to compensate for the morphological changes of their tongues in different ways and to different degrees. For example, the patients may move their tongue more laterally than normal speakers in order to compensate for the missing parts of the tongue on one side. In order to assist surgeons in planning and evaluating the surgery, it is desirable to understand the biomechanical and muscular mechanisms underlying the compensatory tongue motion.

In this work we performed two preliminary studies of the the tongue motion pattern and muscle mechanisms of glossectomy patients, and compared them to normal speakers. The studies were carried out at University of Maryland Medical and Dental School. Dr. Maureen Stone led and designed the studies, and performed the scientific analysis. My work provided technological support in image processing and data analysis in these studies. In the muscle mechanism study, Dr. Hideo Shinagawa extracted the muscle fiber and located them in the DTI data of the tongue. Dr. Rao Gullapalli and Ms. Jiachen Zhuo contributed in data collection. Dr. Emi Murano provided most of the "asouk" data sets and provided valuable advices.

The chapter is organized as follows. Section 8.2 describes a statistical analysis of internal tongue motion of normal and glossectomy speakers using principal component analysis (PCA) to discover their relationships. In Section 8.3, we analyze the mechanical property of inferior longitudinal muscle (IL) by combining knowledge from tagged MRI, high resolution MRI and diffusion tensor imaging. Section 8.4 provides a disscussion, and Section 8.5 summarizes the chapter.

8.2 Statistical Analysis of Internal Tongue Velocity Field

In this work, we focus on the statistical analysis of the motion patterns of internal tongue using principal component analysis (PCA). PCA is an excellent method to

extract and represent patterns in high-dimensional data for which no expectations or model is available. PCA has a wide range of applications in image analysis, including shape modelling (active shape model [111] and active appearance model [112]), predicting "average images" in a database [113], retrieving dominant modes for fast imaging [114], and so on. It has also been applied on motion or deformation field analysis; for example, it has been used to build an atlas of cardiac motion [115] and to study shape variations in the normal brain [116].

In speech research, PCA has been successfully applied to representing the tongue surface shape [117–120]. These studies discovered the main variance in vowel tongue shapes in both American English and other languages. Bressman et al. [121] studied the tongue surface shapes of glossectomy patients using PCA and 3D ultrasound, and discovered some statistical differences between patients with and without a flap.

In this preliminary study, we used PCA to examine the motion patterns of the internal muscle movements of the tongue instead of just the surface. The study compares the motion of 5 normal speakers and 5 patients saying "asouk". For each subjects, the velocity field of the mid-sagittal section of the tongue during the elevation of the tongue body was computed from tagged MR images. PCA was then performed on these velocity fields to quantitatively characterize their different motion patterns.

8.2.1 Method

8.2.1.1 Image Acquisition and Processing

We acquired both cine MR images and tagged cine-MR images on 10 subjects. For the same subject, both cine MRI and tagged MRI were acquired on a midsagittal plane with the same spatial resolution and temporal resolution on the same slice location so that the images were automatically registered. Both cine MRI and tagged MRI were acquired with a 6 mm slice thickness and 1.975 mm in-plane resolution. For tagged images, two CSPAMM image sequences in two tag orientations were acquired. The speech repetitions of the subjects were synchronized using the special MR triggering system described in Chapter 5.

All 10 subjects were native American English speakers. Five of them were normal controls (4 male, 1 female) and 5 were glossectomy patients (4 male, 1 female). Among the patients, one male had a flap reconstruction and the others had primary closure. The subjects all repeated the word "asouk" except one normal subject repeated "desouk". These words were chosen because (1) they are short so they can be easily repeated and imaged in an MRI scanner, (2) they maximize tongue deformation by engaging the jaw very little, and (3) they cover a large range of tongue positions and shapes and they contain a range of difficulties for glossectomy patients. We examined the velocity field at the onset of motion from "u" to "k", which was determined via visual inspection of the velocity patterns of the tongue. The velocity

fields were computed from tagged MR images using HARP [39, 40].

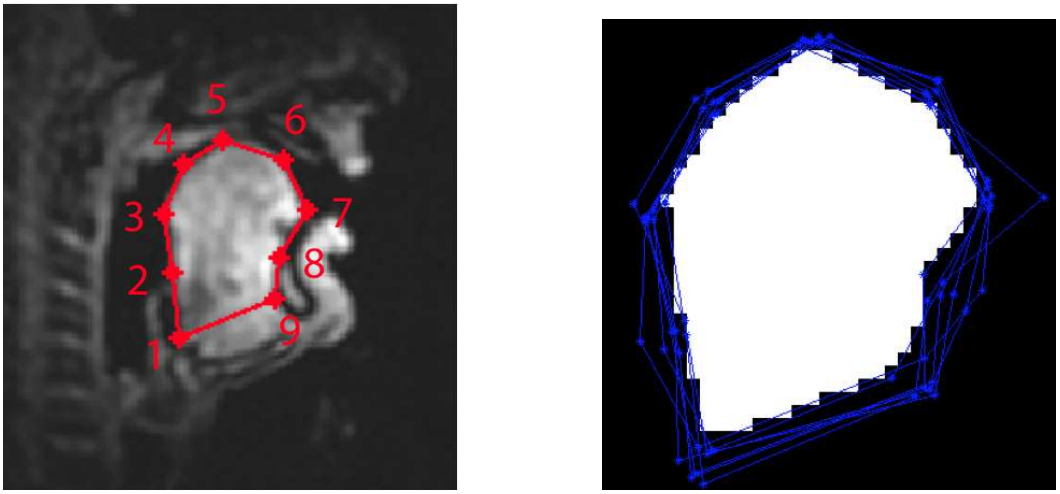
8.2.1.2 Data Pre-Processing: Alignment of Tongue Regions

The tongue regions from different subjects must be aligned before performing PCA. For this the tongue region in each data set is manually identified using nine landmark points from the cine images, as shown in Fig. 8.1(a). These points mark: the base of the valleculae, the upper tip of the eipglottis (projected on the tongue surface), the point on the tongue surface that lies between the elbow of the velum and the lower edge of the mandible, the mid palate, the tongue tip, the origin of genioglossus, and several additional points equidistant between these landmarks. We denote the i^{th} landmark on the j^{th} subject as $\mathbf{P}_{ij}$. The tongue region of each subject was defined as the area inside the polygon by connecting these landmarks.

The tongue regions in all subjects were registered using rigid transformation plus a global scaling computed from manually picked landmark points. Without loss generality, we picked the first data set as the reference coordinate to which all the other data sets are registered. The j^{th} data set was registered to the reference by computing the transformation $[s_j, \mathbf{R}_j, \mathbf{t}_j]$ that minimizes

$$E_j = \sum_{i=1}^{9} (\mathbf{P}_{i1} - (s_j \mathbf{R}_j \mathbf{P}_{ij} + \mathbf{t}_j)) \,. \tag{8.1}$$

The registered landmark points are illustrated in Fig. 8.1(b). The common region of the registered tongues were then determined (the white area in Fig. 8.1(b)), and we denote it as $\mathbb{C}$.



(a) (b)

Figure 8.1: (a) The landmark points used to align the tongues. (b) The registered landmarks.

Next, we must transform the velocity field inside the common region of each dataset to the reference coordinate. This was accomplished in three steps. For each point (pixel) $\mathbf{p}_k \in \mathbb{C}$ and the j^{th} subject, (1) compute its location $\mathbf{p}_{kj}$ in the j^{th} dataset by applying inverse transform, i.e., $\mathbf{p}_{kj} = s_j^{-1}\mathbf{R}^T(\mathbf{p}_k - \mathbf{t}_j)$; (2) compute the velocity $\mathbf{v}(\mathbf{p}_{kj}) = [u(\mathbf{p}_{kj}), v(\mathbf{p}_{kj})]^T$ at point $\mathbf{p}_{kj}$ using HARP and linear interpolation; (3) transform the velocity back to the reference coordinate using $\mathbf{v}_k^{(j)} = [u_k^{(j)}, v_k^{(j)}]^T = s_j\mathbf{R}_j\mathbf{v}(\mathbf{p}_{kj})$. These steps were executed for every pixel in $\in \mathbb{C}$ and every dataset.

8.2.1.3 Principal Component Analysis

After the tongue shapes and velocity fields were aligned, we performed PCA on the normal subjects and quantify the component motions of the midsagittal velocity

patterns. We then determined how well the motion patterns of the patients can be explained by the PCA of normal subjects.

Suppose there are N normal subjects, and M points in the common region. The velocity field of the j^{th} subject can be represented as a $2M \times 1$ vector $\mathbf{V}_j = [u_1^{(j)}, ...u_M^{(j)}, v_1^{(j)}, ...v_M^{(j)}]^T$. Through PCA, the data from any subject can be represented using a linear model

$$\mathbf{V} = \overline{\mathbf{V}} + \Phi \mathbf{b}\,, \tag{8.2}$$

where $\overline{\mathbf{V}}$ is the average velocity field for all the subjects

$$\overline{\mathbf{V}} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{V}_j\,. \tag{8.3}$$

The columns of matrix Φ represent the modes of variation of the velocity fields, and are called principal components (PCs). They are computed from the $2M \times 2M$ covariance matrix $\mathbf{S}$, given by

$$\mathbf{S} = \frac{1}{N}\sum_{j=1}^{N} (\mathbf{V}_j - \overline{\mathbf{V}})(\mathbf{V}_j - \overline{\mathbf{V}})^T\,. \tag{8.4}$$

The PCs are the eigenvectors ϕ_i of $\mathbf{S}$ with corresponding eigenvalues λ_i sorted so that $\lambda_i \geq \lambda_{i+1}$. The PC corresponding to the largest eigenvalue, i.e., ϕ_1 represents the direction of maximum variability in the velocity fields across the subjects.

To interpret the patient motion using normal subjects, each patient data $\mathbf{V}_p$ was fitted to the normal PC's by finding the coefficient vector $\mathbf{b}_p$ that minimizes the residue

$$E = ||\overline{\mathbf{V}} + \Phi \mathbf{b}_p - \mathbf{V}_p||\,. \tag{8.5}$$

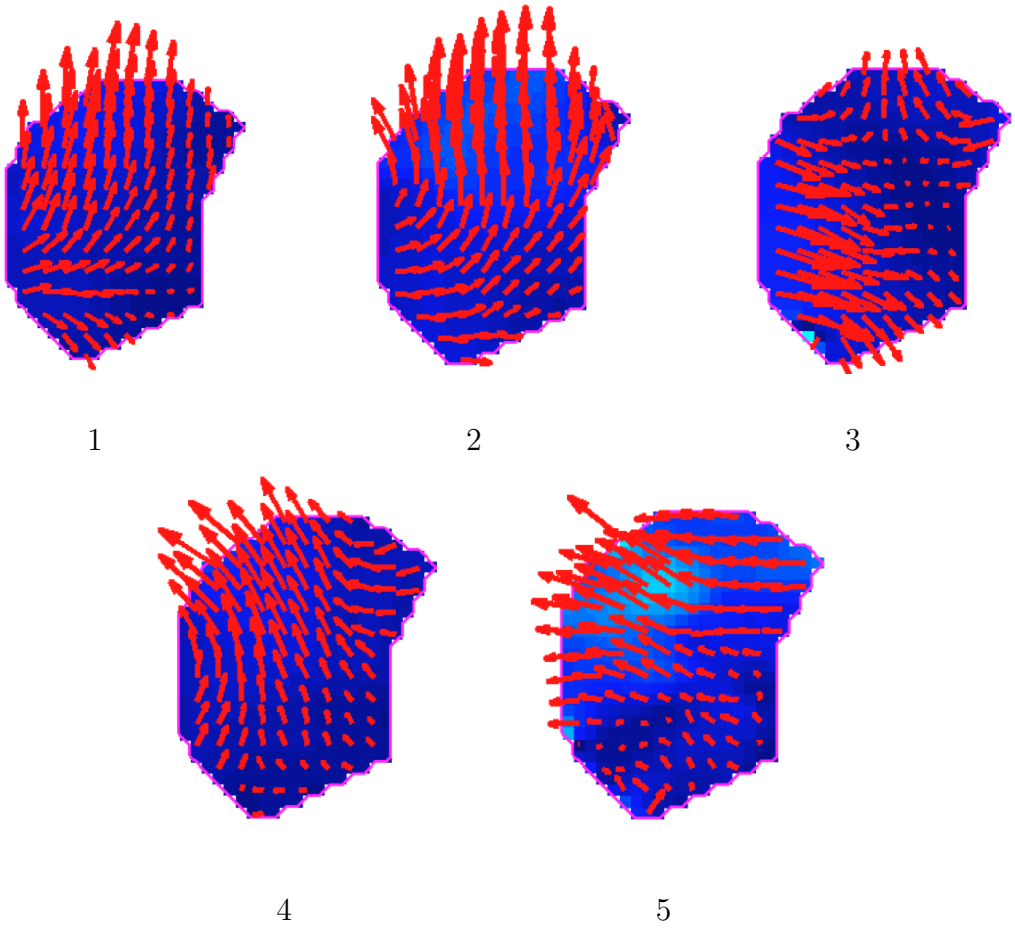


Figure 8.2: The velocity fields of 5 normal subjects at the "k" onset.

The residue represents the motion pattern of the patient that cannot be represented by the normal subjects, while $\mathbf{b}_p$ represents the amount of motion patterns that are represented by the corresponding PCs.

8.2.2 Results

The velocity fields of the 5 normal subjects after alignment are shown in Fig. 8.2. We can see there was a large amount of variability in the normal subjects. This indicated that there were a number of strategies of tongue motion to produce "k" sound from "u".

The PCA of the 5 normal subjects produced 4 PCs, and they were shown in Fig. 8.3. The eigenvalues corresponding to the 4 PCs were 372.24, 157.05, 39.27, and 11.03, respectively, and it indicated that PCs 1 and 2 accounted for most of the variance. From Fig. 8.3 it can be seen that PC1 represented primarily anterior-posterior motion, and PC2 mainly represented vertical motion.

Fig. 8.4 shows the velocity fields generated by adding PC1 and PC2 to the mean velocity. Panel 2 and 8 show the mean minus and plus one standard deviation (square root of λ_1 of PC1) respectively, and panel 4 and 6 show the mean minus and plus one standard deviation of PC2 respectively. The four corner panels show the addition of mean with both PC1 and PC2.

Fig. 8.5 shows the velocity vector fields at the time of "k"-onset for the 5 patients. Four patients' motions were very well explained by the first 2 PCs. For Patient 1, PC1 represented 61% and PC2 26% of the motion variance. PCs 3 and 4 were less than 1% each. For Patient 2, 88% of the motion variance was accounted for by PC1 and the other 3 PCs explained less than 1% each of the variance. It can be seen from Fig. 8.5 that the patient moved the tongue straight backward and upward. Patient 3

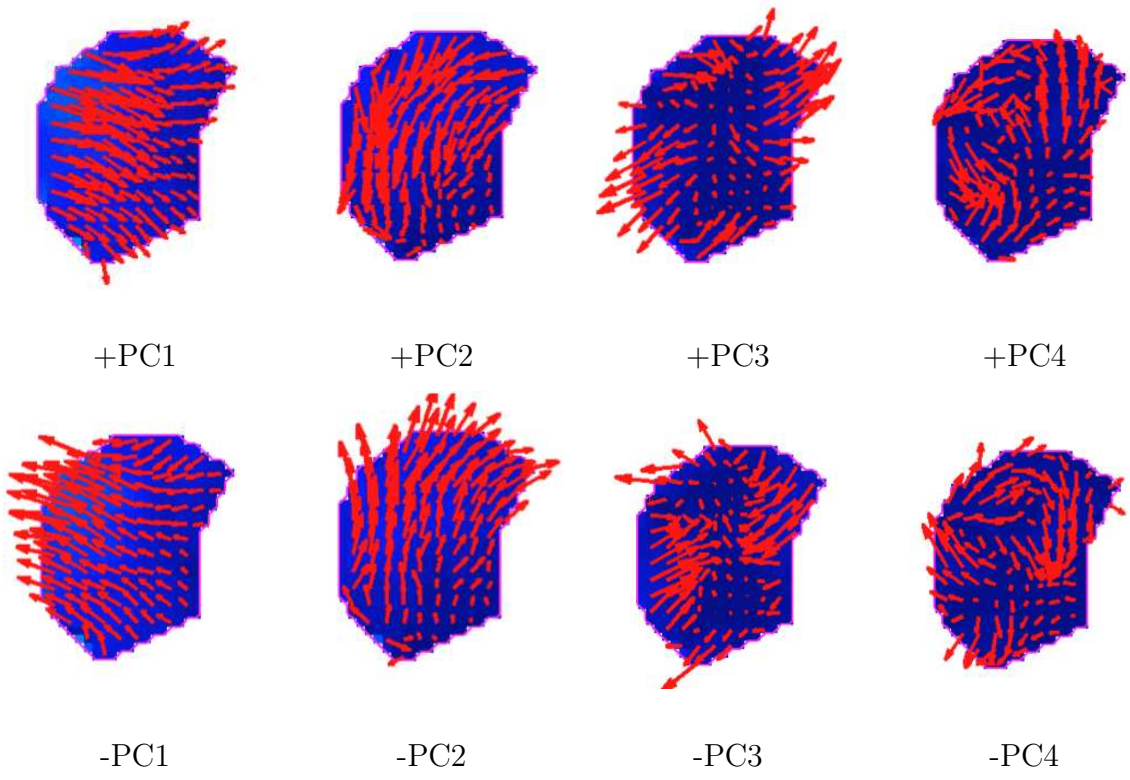


Figure 8.3: The PCs computed from 5 normal subjects.

also was fairly well explained by PC1 (44%) and PC2 (33%). Patient 4's motion was explained almost entirely by PC2 which accounted for 70% of the variance, while PC1 explained 13%. PC 4 also explained 5%. Patient 5 had 7% of his variance explained by PC1, none by PC2, 44% explained by PC3 and 5% by PC4.

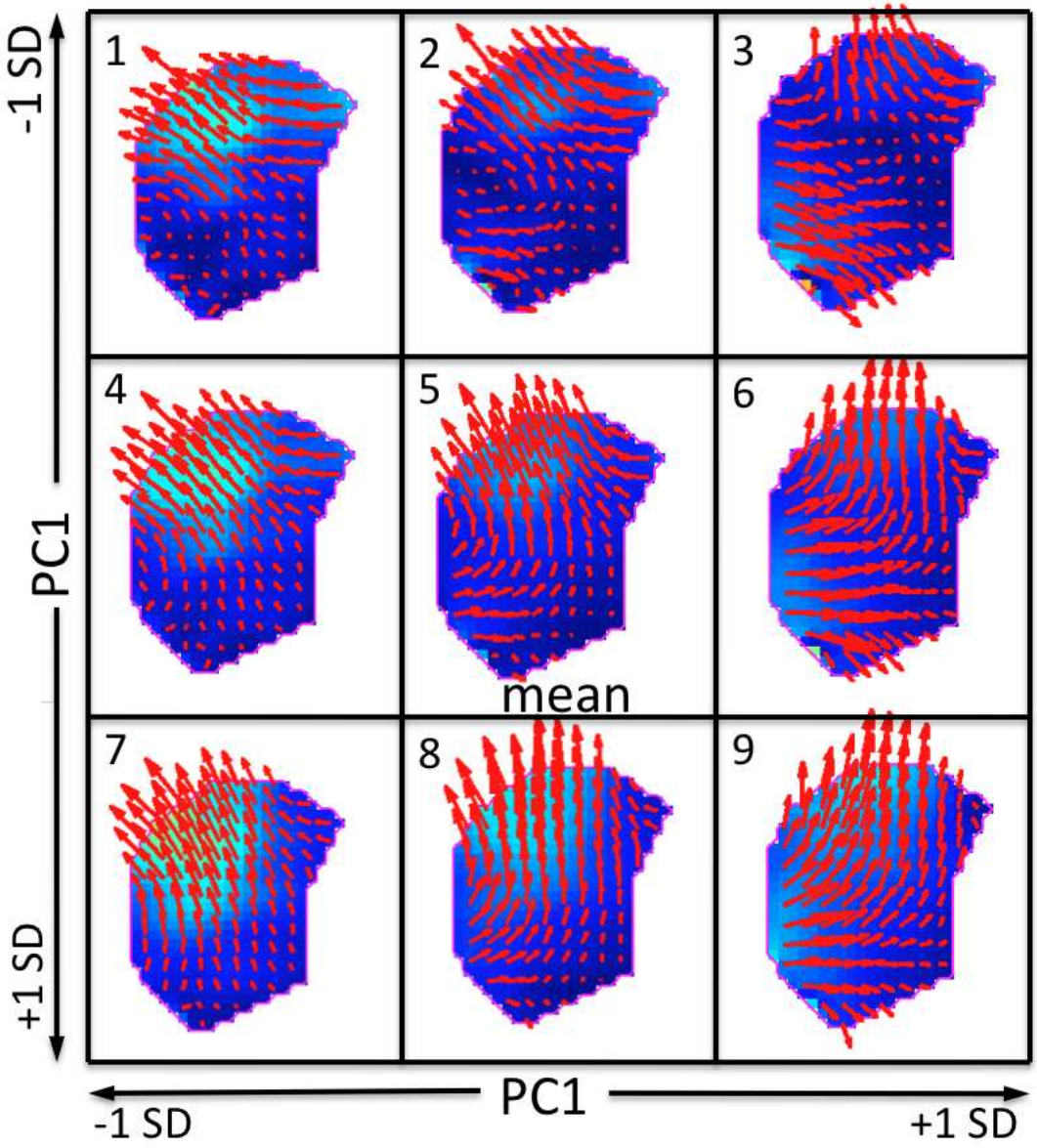


Figure 8.4: The mean velocity field and +/1 standard deviation of PC1 and PC2. The center image shows the mean velocity field. The others show the addition of mean and -/+1 standard deviation of PC1 and/or PC2.

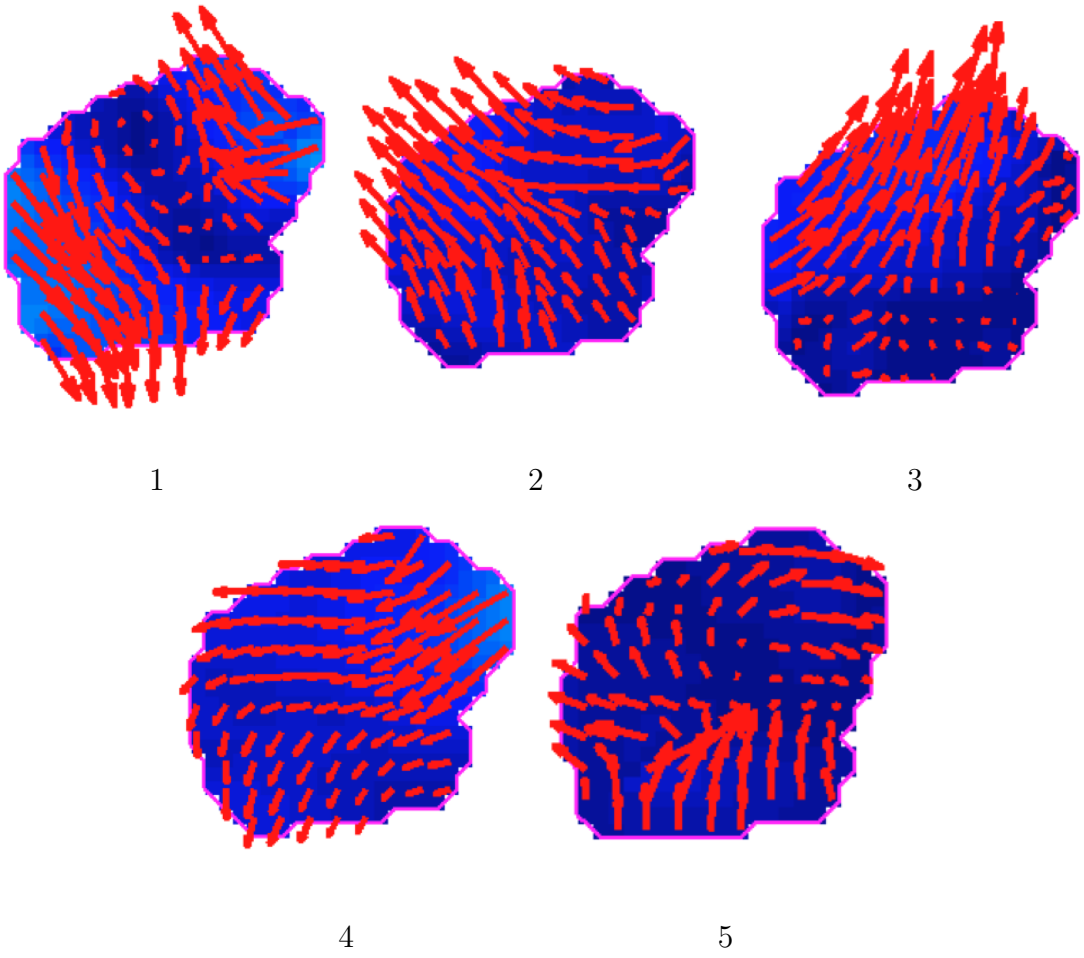


Figure 8.5: The velocity fields of the five glossectomy patients at the "k"-onset.

8.3 Tongue Muscle Mechanics

8.3.1 Introduction

In this work, we studied the mechanical property of tongue muscle during speech. We examined the motions of the inferior longitudinal muscle (IL), which is often cut during glossectomy surgery, of a normal speaker and of a glossectomy patient with a radial forearm free flap (RFFF), and analyzed whether the difference of tongue motion after surgery could be explained by the changes in muscle mechanics.

We combined knowledge from tagged MRI and diffusion tensor imaging (DTI) to compute the motion of IL muscle fiber. DTI captures muscle fiber orientation by imaging the orientation-dependent diffusion process that are associated with fibers [122], and the dominant orientation of the diffusion tensor represents the orientation of a muscle fiber passing through the voxel of interest. In speech research, DTI was initially used to image excised tissues (for example, porcine tongue [123]), but now *in vivo* human imaging in the resting state is possible [124,125]. DTI can only image the tongue in stationary position and not in a temporal sequence as would be required in speech.

In this preliminary study, both tagged MRI and DTI images were acquired on the same subject. We then extracted the muscle fiber from DTI and registered it with tagged MRI data, so that the muscle fiber can be tracked using HARP methods. Finally the tracked muscle deformations were represented using mechanical measure-

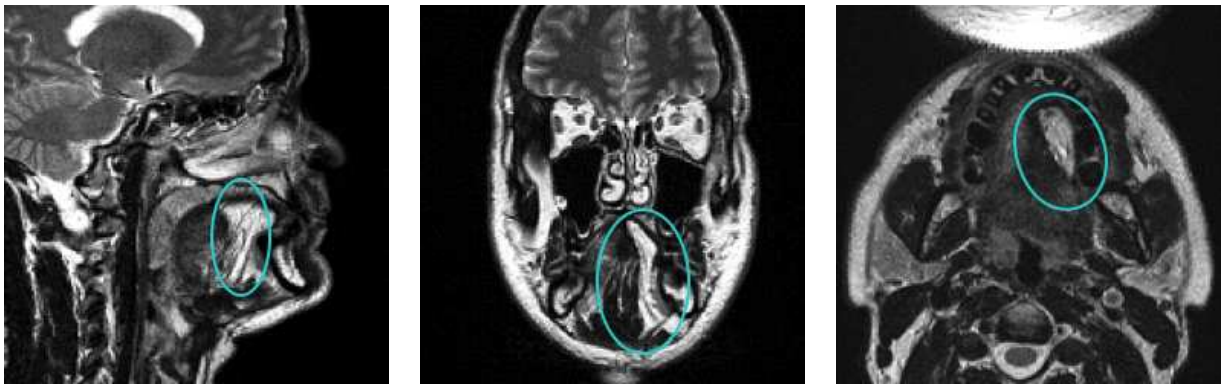
ments, and they were compared between normal and patient subjects.

8.3.2 Method

8.3.2.1 Data Acquisition and Processing

Images were acquired on a Siemens 3T MR scanner for a normal subject and a patient. The patient has a radial forearm free flap, a piece of skin tissue that was resculpted, vascularized and inserted to replace the missing tongue tissue. For each subject, DTI images were acquired with diffusion weighting along 6 directions, and a b-value of 500 s/mm^2. Other imaging parameters are: FOV=200 mm $\times$ 200 mm, slice thickness=3 mm, TR=5,000 ms, TE=66 ms, in plane resolution=3.1 mm, temporal resolution=37.5 mm. The resolution of tagged cine MRI and cine MRI was all 1.875 mm, and slice thickness was 6 mm. In the DTI acquisition, the tongue was held still in a rest position. In cine MRI and tagged MRI acquisitions, the subjects repeated each speech task multiple times and were synchronized using a specialized trigging system. For visualization purpose we also acquired high resolution MRI in a rest position. Fig. 8.6 shows the high resolution images of the patient.

The 3D location of fiber bundles within the inferior longitudinal (IL) muscle was identified using DTI data. We only look at IL at one side of the tongue, i.e., the side that was preserved for the patient. Since IL is often cut during glossectomy surgery even in partial resections, the muscle on the intact side has no pair to work with

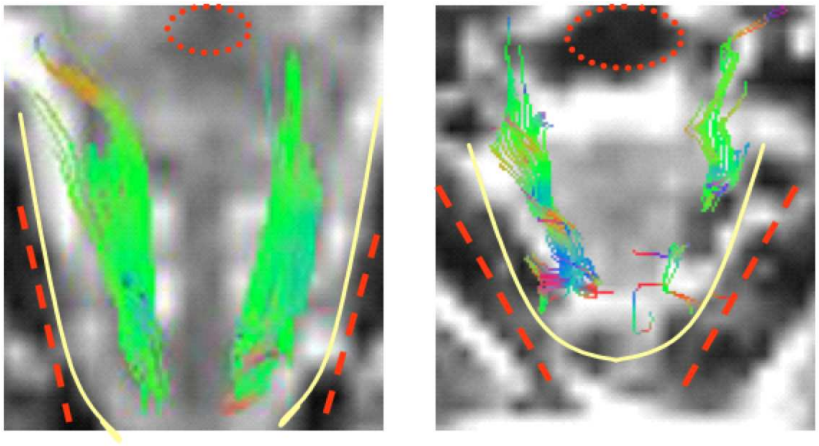


Sagittal Coronal Axial

Figure 8.6: High resolution images of the patient tongue in three planes. The flap has high intensity in the images and its position is illustrated by the circles.

and must create new motion patterns. DTI data was acquired on axial planes and reconstructed into 3D volume so that it can be re-sliced in any direction. The location of IL in the tongue was first estimated from known anatomy and visually identified on the high resolution MRI. The muscle fiber bundles were then extracted through DTI data analysis and tractography. The fiber bundle locations and orientations for both normal and patient speakers are shown in Fig. 8.7.

After IL muscle bundles were estimated, an average fiber for IL was extracted in the intact side for both subjects. This fiber was overlaid on the tagged MRI data in order to track its motion during the word "asouk". The fiber was overlaid onto the sagittal tagged MRI plane most closely corresponding to the fiber plane in the DTI data. The fiber was then automatically tracked through all time frames using HARP processing and HARP tracking.



(a) Normal (b) Patient

Figure 8.7: IL muscle fiber (green) estimated from DTI images of (a) the normal speaker and (b)the patient. The tongue is outlined in yellow, the pharynx is circled in red, and the mandibular bones are tracked in red.

8.3.2.2 Muscle Mechanics

We computed the changes of biomechanical properties of the muscle fiber between the first time frame and later time frames, and compared the patient with the normal subject. The compared properties include global rigid motion (rotation and translation), change in fiber length, and bending energy.

We used the bending energy model proposed by Duncan et al. [126] to quantitatively measure the shape change of the tongue muscle fiber. The bending energy model was originally proposed to measure the deformation of cardiac shape, and it measures the bending on a thin straight rod. It is assumed that the cross section of the rod is circular and the inertias in all directions are the same. We denote s be the length along the rod starting from one end point, and the total length of the rod is

S. The equation for its elastic energy after bending is

$$E_{\text{abs}} = \frac{1}{2} EI \int_0^S \frac{1}{R(s)^2} ds \,, \tag{8.6}$$

where $R(s)$ is the radius of curvature after bending, E is the Young's modulus, I is the inertia, and the integral is taken over the entire length of the rod. Here we assume E and I are constant, and so can be ignored from the equation. We call this measure the *absolute bending energy*. Because the muscle fiber is not straight even at the rest position, we measure its deformation using *relative bending energy*, which models the shape change of an arbitrarily-shaped thin flexible rod. The relative bending energy is expressed as

$$E_{\text{rel}} = \frac{1}{2} \int_0^S \left(\frac{1}{R_t(s)} - \frac{1}{R_0(s)} \right)^2 ds \,, \tag{8.7}$$

where $R_t(s)$ and $R_0(s)$ are the radii of curvature of the deformed and undeformed muscle fiber respectively. A nice property of the bending energy is that it is invariant to rigid transformation. Therefore it can measure the shape change of the fiber regardless of its location and orientation.

8.3.3 Results

8.3.3.1 Fiber Position and Elongation

Tables 8.1 and 8.2 present the mechanical changes in the IL fiber of the normal speaker and the patient speaker during some time-frames of "asouk". The times at which each speech sound occurred (column 2) were determined by examining the

Time Frame	Sound	x translation (mm)	y translation (mm)	Rotation (degree)	Elongation
1		0	0	0	0
3	start "s"	2.4	-0.8	0.5	5.1%
4	max "s"	2.4	-0.9	-1.3	0.5%
8		2.4	-0.1	-20.2	1.9%
11	"u"	2.7	-0.1	-20.2	-2.7%
14	"k" palate	3.3	-0.3	-18.1	-9.7%
15	"k" velum	3.3	-0.4	-19.0	-10.3%

Table 8.1: Mechanical changes in IL fiber for the normal subject.

tongue surface motion from cine MRI data. The patient started moving the tongue from the third time frame while the normal speaker started from the first time frame. The translation and rotation values represent the global motion of the fiber and the elongation represents the change of fiber length from the first time frame. x is in the posterior-anterior direction, and y is in the head-foot direction. The x translation shows that both subjects moved IL muscle forward into the "s". The normal subject then moved it backwards, whereas the patient moved it back then forward into "k". The y translation shows that the fiber moved down into "s" and then upward into "u" for both subjects. Again the patient moved downward at the end, unlike the

Time Frame	Sound	x translation (mm)	y translation (mm)	Rotation (degree)	Elongation
3		0	0	0	0
10	start "s"	3.4	-0.7	-5.3	18.9%
14	max "s"	2.8	-0.2	-9.1	17.3%
16		2.4	0.1	-6.2	14.1%
18	"u"	1.9	0.4	-4.0	7.0%
20	"k" palate	1.7	0.5	-3.3	3.3%
21	"k" velum	1.6	0.4	-1.6	5.0%

Table 8.2: Mechanical changes in IL fiber for the patient.

normal subject. The rotation shows backward rotation occurred for both subjects, with considerably less rotation for the patient. In addition the patient's muscle fiber elongated more into the "s" and more shortening thereafter, even though he never shortened to resting length (time frame 1).

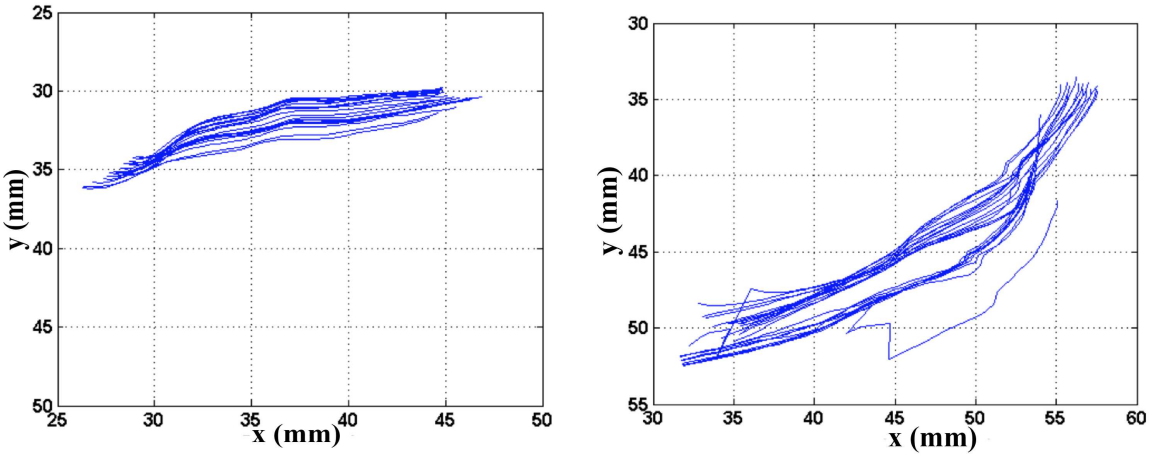
8.3.3.2 Fiber Shape and Bending Energy

Fig. 8.8 shows the overlaid position and shapes for the IL fibers at all time frames. Note for the patient two time frames were mistracked (11 and 12) because of the image quality was poor. Therefore these two time frames should not be considered for

comparison. The IL fiber shape is slight convex for the normal subject, and concave for the patient. The bending energy for the two subjects are shown in Fig. 8.9. In the figure, the mistracked time frames of the patient, 11 and 12, have been omitted. The blue line indicates a change in bending relative to the first time frame. The red line indicates a change in bending relative to a straight line. The fiber shape at the rest position for the patient is closer to a straight line than the normal subject. Therefore the absolute and relative bending energies of the patients are close. From the figures we can observe the different bending patterns of IL fiber during speech. The normal subject's largest bends occur during "k" (time frame 14), the sound with the highest tongue body, and the subsequent inhalation (time frame 17). For the patient, the largest bends occur during "u" and "k", and also "s". As bending increases, the fiber of the normal subject becomes more convex, and that of the patient becomes more concave.

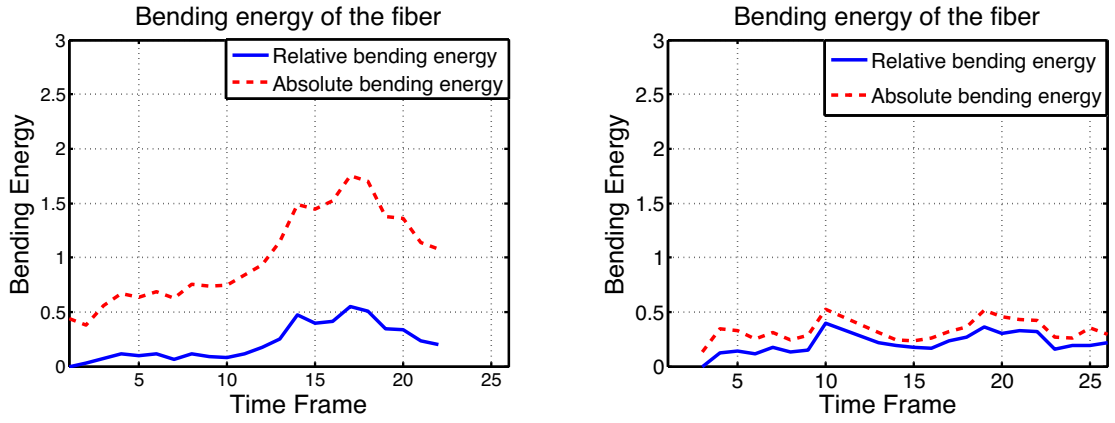
8.4 Discusion

In the PCA study, we studied the statistical property of velocity fields in the onset of the motion from "u" to "k" on five normal subjects using PCA, and tried to explain the glossectomy patients' motion using the PCs. While the PCs computed from the 5 subjects appear to be physiologically meaningful and might reasonably be made by the tongue muscles, the observations from such a small number of datasets



(a) Normal (b) Patient

Figure 8.8: Overlay of IL muscle positions at all time frames during the utterance "asouk" for (a) the normal speaker and (b) the patient.



(a) Normal (b) Patient

Figure 8.9: The absolute and relative bending energy of the IL fiber over time for (a) the normal speaker and (b) the patient.

may potentially be biased and not conclusive. More datasets are needed to achieve more statistically persuasive observations.

In the PCA study the time frame of the examined velocity field was picked manu-

ally via visual inspection through all the time frames, and the velocity at each point was estimated as the displacement from previous time frame. Though the speech repetitions were synchronized using a specialized timing device, the subjects may vary on the response time to the signal. Therefore the actual time of the velocity may be biased by up to the temporal resolution of the image acquisition. In addition, the velocity may be computed over different time interval because the temporal separations between time frames in different image acquisitions were not the same. To better align the subjects in time, we may interpolate the tongue motion in time through the image sequences. Hence, the subjects can be aligned not only spatially but also temporally so that the velocity at the interested time can be determined more accurately. Moreover, the temporal alignment of velocities also allows analysis of motion patterns through the whole utterance, which may produces more interesting and meaningful insight into the motion pattern changes of the patients after different surgical closure procedures.

In the PCA study the tongue regions were aligned using a rigid transformation plus a global scaling. This transformation assumes that the tongue shapes of different people are similar while the size is different. This simplifies the problem while still preserving important features of the tongue. The results should be improved using deformation registration to align different tongue regions so as to accommodate different tongue shapes. In this case, the velocity fields need to be transformed by the resulting deformation field using methods similar to Alexander et al. [127,128] or Cao

et al. [129].

This muscle mechanics study has observed notable differences of the IL muscle deformation during speech through comparison on the muscle mechanical properties of a glossectomy patient and a normal speaker. As a preliminary study, this work shows promising results. Studies on more data is required to achieve more conclusive observations for precise scientific study.

In the muscle mechanics study, the IL muscle from DTI was aligned with tagged MR by projecting it onto a sagittal image plane. The IL fiber is a 3D structure moving in 3D. It general does not align with any of sagittal, coronal, or axial planes on which the tagged MR images are acquired, although its orientation is approximately parallel to the sagittal plane. Therefore the 2D projection may cause a considerable amount of error in this study. In addition, fiber motion is also assumed to be 2D, i.e., inside the sagittal plane. Although this assumption is reasonable because the tongue has minimal lateral motion, it is not accurate. Better results are expected when extending this work to 3D.

8.5 Summary

In this chapter, we described preliminary studies on tongue velocity field analysis and muscle mechanical properties during speech, and discovered that the motion pattern of glossectomy patients was notably different comparing with normal speakers.

In the first study, we applied principal component analysis on the tongue velocity field during speech and compared the statistical differences between normal and patients speakers. We also studied the mechanics of IL muscle fiber by evaluating its global motion, elongation, and bending during speech. Our studies have shown interesting and promising preliminary results.

Chapter 9

Conclusions and Future Work

This dissertation focused on three-dimensional muscle motion reconstruction and analysis from tagged MR images. We concentrated on tongue motion imaging, and also worked on cardiac motion imaging. In Chapters 3 and 4, we first developed 2D HARP tracking refinement methods to prevent the mistracking in the traditional HARP method. These methods can reliably track every tissue point in 2D in tagged MR images. In Chapter 5, we measured the 3D tongue motion during speech by optimizing the zHARP method for tongue imaging, and developed a specialized MR triggering and vocal repetition method to reduce motion artifacts. In Chapter 6 we extended the 3D-HARP method in 3D cardiac motion tracking and developed the thin plate spline based 3D tongue motion tracking method to track the 3D tongue motion based on the 2D HARP method on tagged MRI. In Chapter 7, we developed a method to reconstruct 3D, dense, incompressible motion from tagged MR images

using divergence-free vector spline, and applied it to both the heart and the tongue. We also performed preliminary studies on the internal tongue motion on glossectomy patients and compared them with normal speakers, and showed them in Chapter 8. In this chapter, we summarize the key results and future research directions.

9.1 HARP Tracking Refinement

In Chapters 3 and 4, we described two HARP tracking refinement methods, region growing HARP refinement (RG-HR) and shortest path HARP refinement (SP-HR), for robust and reliable estimation of 2D motion from tagged MRI. These methods solve the mistracking problems in traditional HARP. Both methods start with a manually specified seed point that can be correctly tracked using traditional HARP tracking then carry out a region-growing process to recursively track all the points in the tissue. The RG-HR method tracks the tissue points in an order based on the local HARP phase smoothness, and may fail because points can be connected to the seed through erratic paths. To improve this, the SP-HR method explicitly resolves the optimal paths that connect tissue points to the seed by formulating a single source shortest path problem, and thus provides more robust and reliable tracking results. We have successfully applied these methods to both cardiac and tongue motion tracking.

9.1.1 Main Results

- Mistracking in 2D HARP is mainly caused by three reasons: large motion between successive time frames, through-plane motion, points close to the tissue boundary.

- The RG-HR method was applied to tongue motion tracking, and it successfully prevented mistracking in case of large tongue motion, and led to improved Lagrangian strain calculation.

- The SP-HR method was compared to traditional HARP tracking in cardiac motion tracking, and it performed better on all 18 sectors of the left ventricle, especially in the epicardium and endocardium. WIth decreased temporal resolution, the ratio of corrected tracked points with SP-HR was 98.4%, while traditional HARP was as low as 58.7%.

- In cardiac motion tracking, SP-HR was more robust than RG-HR with a success rate of 99.5% as compared to that of 93.8% in RG-HR.

- The two-step tracking strategy was adopted when tracking through a temporal sequence of images, and was shown to be more robust to through-plane motion than tracking through successive time frames.

9.1.2 Future Work

HARP refinement can be viewed as a implicit harmonic phase unwrapping process. In the future we want to investigate other phase unwrapping algorithms and adapt them to refine HARP tracking. It is also of interest to extend the proposed methods to other phase unwrapping problem and other applications, e.g., DENSE images and phase images generated with Gabor filter banks.

9.2 Measurement of 3D Tongue Motion Using zHARP

In Chapter 5, we measured 3D tongue motion during speech using zHARP. We re-implemented the zHARP sequence using a gradient echo sequence and optimized the image parameters for tongue imaging. We also used a specialized MR triggering and vocal repetition method to reduce motion artifacts. The sequence and the triggering system have been successfully applied to 3D tongue motion measurement and strain analysis.

9.2.1 Main Results

- We have successfully re-implemented and optimized zHARP for measuring 3D tongue motion, and validated it using a phantom experiment.

- Motion artifacts caused by inconsistent repetitions are the major factors that affect the zHARP image quality in tongue imaging. They are much more severe in zHARP than in standard tagging, and more severe in later time frames. To reduce motion artifacts, images are best acquired in the sagittal planes to minimize motion artifacts.

- We successfully performed *in vivo* experiments on normal speakers with optimized parameters. Each zHARP image was acquired in 4 vocal repetitions in each of the two tag directions, with an acquired k-space matrix size of 64×22, a temporal resolution of 52 msec, and a z-encoding period of 30 mm.

- We have computed 3D strain tensor based on the 3D tongue motion computed from zHARP images.

9.2.2 Future Work

- In patient studies, motion artifacts remains a major factor that lowers the zHARP image quality because the patients' tongue motion is not very consistent. Future investigation is needed to try to further suppress motion artifacts by reducing the number of vocal repetitions in zHARP imaging using fast imaging techniques such as EPI and parallel imaging.

- In our current implementation, the temporal resolution was compromised to reduce the number of repetitions required for each image. Future work is desired

to improve the temporal resolution of tongue imaging using view sharing [130].

- ZHARP made it possible to directly compute the 3D strain map for tongue functional analysis from two parallel slices. More experiments are needed to better understand the correlation between the strain map and tongue functions. This is part of our future work.

9.3 3D Tongue Motion Tracking

In Chapter 6, we developed the thin plate spline based 3D motion tracking for tongue imaging by extending the 3D-HARP method for 3D cardiac motion tracking. The method tracks a rectangular mesh placed inside the tongue in an iterative fashion. In each iteration, the method first computes the intersection points between the mesh and image planes and estimates the 2D or 1D components of the 3D motion of these points using 2D HARP tracking, then the 3D motion of the whole mesh is interpolated using a thin plate spline.

9.3.1 Main Results

- We successfully applied the method on a normal speaker during the speech of "deGoose". The tracked 3D motion was visually assessed by acquiring images with tag planes parallel to the meshes, and it was observed that the meshes conformed to the deformation of the tag lines at all time frames.

- By placing two parallel meshes closely together inside the tongue and tracking them in 3D, we were able to compute the 3D Lagrangian strain from the tracking results.

9.3.2 Future Work

The method is sensitive to incorrect 2D tracking results of points on the mesh. Future work is needed to improve the robustness of the method. Future work also includes the investigation on extending the method to the 3D motion on the whole tongue volume.

9.4 3D Incompressible Motion Reconstruction

In Chapter 7, we presented an approach to reconstruct a 3D, dense, incompressible displacement field from tagged MR images. Based on the incomplete and sparse data computed from tagged MRI, our method uses a smoothing divergence-free vector spline on incomplete data samples to interpolate the velocity fields at discrete time steps, and then integrates these velocity fields to find the incompressible displacement field. The displacement fields at different time frames are reconstructed separately to prevent error accumulation arising from inaccurate estimation at earlier time frames.

9.4.1 Main Results

- We demonstrated our method using a simple 2D example, and compared it with TPS interpolation, DFVS interpolation, and Bisotquet's method. The deformation field reconstructed using our method was incompressible with negligible error from discretization, while the deformation fields from the other three methods were not incompressible.

- Our method was validated using a cardiac motion simulator and compared to Bistoquet's method. Over the simulated 16 frames, the rms error of the reconstructed displacement field in our method was no more than 0.03 mm, while the rms error in Bistoquet's method was about 10 times greater. The incompressibility was also better reserved in our method with an average Jacobian determinant error of 0.0056 at the time frame with largest deformation, as compared to 0.052 in Bistoquet's method.

- From the phantom experiments, we have optimized the parameters in the method. We showed that the reconstructed displacement error decreases when the number of integration steps increase, which requires longer computation time. As a tradeoff we chose the number of steps to be 20. With this number of steps the reconstruction error was minimized by choosing a smoothing parameter ρ_0 of 1.15.

- Our method was applied to reconstruct the 3D incompressible motion of the

left ventricle of a human heart, and was validated by comparing the results with the 2D motion computed from HARP on validation images. Over all 20 frames, the reconstruction error of our method was about half the error of Bistoquet's method. At time frame 10, when the heart deformed the most and the reconstruction errors were the largest for both methods, the average displacement error in our method was 0.61 mm and Bistoquet's method was 1.08 mm. The Jacobian determinant errors for the two methods were 0.043 and 0.081, respectively.

- We applied our method to reconstruct the 3D tongue motion of a normal subject during speech. From the reconstructed dense 3D motion, we computed the dense 3D Eulerian strain maps which can be a useful tool for tongue functional analysis.

9.4.2 Future Work

- It is possible to adopt other other divergence-free interpolation kernels in our method, e.g., the smooth kernel matrix generated from a Gaussian function, and the class of compactly-supported kernels derived from Wendland functions. Future investigation is required to compare these kernel functions with the DFVS used in our method.

- The large deformation diffeomorphism (LMLDD) approach computes smooth

trajectories over time, and theoretically can be adapted to reconstruct the optimal incompressible displacement field. However, it requires intensive computation and is not feasible for our application because of the large number of data samples. Future investigations on reducing the computation of incompressible LMLDD on large number of data samples may open new directions for incompressible motion reconstruction. Possible strategies include using image pyramid with different levels of resolutions, reducing number of samples, and so on.

- Future work also includes the investigation of applying our 3D incompressible deformation reconstruction method to other image analysis problems, for example prostate and brain image registration.

9.5 Preliminary Studies on Internal Tongue Motion

In Chapter 8, we performed two preliminary studies on the internal tongue motion patterns of glossectomy patients through comparison with normal speakers. In the first study, we performed statistical analysis on velocity fields of normal speakers using PCA, and tried to explain the motions of patients with different surgical procedures. In the second study, we computed the mechanical properties of IL muscle fiber of

a normal speaker and a glossectomy patient with a free flap implant, and observed notable differences that can be related to the physiological changes caused by the surgery.

9.5.1 Main Results

- In the PCA study, we performed PCA on the velocity fields of five normal speakers. The eigenvalues corresponding to the four PCs were 372.24, 157.05, 39.27 and 11.03, respectively, and it indicated that PC1 and PC2 accounted for most of the variance in the five normal speakers.

- In the PCA study, we tried to explain the five patients' motions using the PCs computed from the five normal speakers, and discovered that four patients' motions were very well explained by the first two PCs, and one was very well explained by PC3.

- In the muscle mechanics study, we were able to discover notable differences in the mechanical properties of between normal and patient speakers, and the differences was used to explain the control difficulties faced by glossectomy speakers.

9.5.2 Future Work

- Both studies were performed on a limited number of patients and normal speakers. Future work is required to study more subjects for more conclusive results.
- In the PCA study, we have examined only the statistics of the 2D motion on a 2D section of the tongue. In the future we want to extend it to 3D motion field on the 3D tongue volume in order to provide more insights on the internal tongue motion.
- In the PCA study, future investigation is desired to perform PCA both across subjects and over time after aligning the tongue motion both spatially and temporally.
- In the PCA study, future work also includes aligning the subjects using deformable registration instead of the simplified transformation in our approach to better accommodate the variations of tongue shapes.
- In the muscle mechanics study, future work is required to extend the study from 2D to 3D to better reflect the facts that the fiber is a 3D structure and moves in 3D. This possibly can be done by pre-computing the 3D dense displacement fields using the method described in Chapter 7.

9.6 Overall Perspective

The dissertation aimed to improve the functional and mechanical analysis of tongue and cardiac muscles through the reconstruction of 3D motion fields from tagged MR images with minimal modeling. We have explored and provided solutions to several difficulties faced in 3D motion reconstruction, e.g., reliable 2D motion tracking, fast 3D imaging of tongue motion, incompressible motion field reconstruction, and so on. The research has provided new and beneficial tools for both scientific and clinical studies of muscle motion, especially in tongue imaging. We hope the research will also provide more insights for other image analysis tasks, including image registration and atlas-based segmentation when computing the deformation fields.

9.7 List of Publications

- Journal articles

    1. M. Stone, X. Liu, H. Chen, and J. L. Prince, "A preliminary application of principal component analysis to internal tongue deformation patterns," *Submitted to Computer Methods in Biomechanics and Biomedical Engineering.*

    2. J. Lee, X. Liu, A. K. Jain, E. C. Burdette, J. L. Prince, and G. Fitchinger, "Prostate brachytherapy seed reconstruction with Gaussian blurring and

optimal coverage cost," *IEEE Trans. Med. Imag., In Press.*

3. X. Liu and J. L. Prince, "Shortest path refinement for motion estimation from tagged MR images," *Submitted to IEEE Trans. Med. Imag.*

- Conference papers

1. M. Stone, X. Liu, H. Chen, and J. L. Prince, "A preliminary application of principal component analysis to internal tongue deformation patterns," in *1st International Workshop on Dynamic Modeling of the Oral, Pharyngeal and Laryngeal Complex for Biomedical Applications*, 2009.

2. X. Liu, K. Abd-Elmoniem, and J. L. Prince, "Incompressible cardiac motion estimation of the left ventricle using tagged MR images," in *the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2009.

3. M. Stone, X. Liu, E. Murano, J. Zhuo, R. Gullapalli, A. Salama, and J. L. Prince, "Principal component analysis of internal tongue motion in normal and glossectomy patients with primary closure and free flap," in *Proceeding of the International Symposium on Biomechanics, Healthcare and Information Science*, Kanazawa, Japan, 2009.

4. X. Liu, J. Zhuo, H. Agarwal, K. Abd-Elmoniem, E. Murano, M. Stone, R. Gullapalli, and J. L. Prince, "Quantification of three dimensional tongue motion during speech using zHARP," in *Proceeding of SPIE Medical Imag-*

ing, 2009.

5. X. Liu, Y. Bai, and J. L. Prince, "Shortest path refinement for HARP motion tracking," in *Proceeding of SPIE Medical Imaging*, 2009.

6. J. Lee, X. Liu, J. L. Prince, and G. Fitchinger, "Prostate brachytherapy seed localization with Gaussian blurring and camera self-calibration," in *the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2008.

7. J. Lee, X. Liu, A. K. Jain, J. L. Prince, and G. Fitchinger, "Tomosynthesis-based radioactive seed localization in prostate brachytherapy using modified distance map images," in *Proceeding of IEEE International Symposium on Biomedical Imaging (ISBI)*, 2008.

8. X. Liu, J. Zhuo, H. Agarwal, K. Z. Abd-Elmoniem, E. Murano, M. Stone, and J. L. Prince, "HARP MRI techniques for imaging tongue motion," in *Proceedings of the Fourth International Symposium on Biomechanics, Healthcare and Information Science*, Kanazawa, Japan, 2008.

9. M. Stone, X. Liu, S. Shinagawa, E. Murano, R. Gullapalli, J. Zhuo, and J. L. Prince, "Speech patterns in a muscular hydrostat: Normal and glossectomy tongue movement," in *Proceedings of the Fourth International Symposium on Biomechanics, Healthcare and Information Science*, Kanazawa, Japan, 2008.

10. E. Murano, M. Stone, J. Zhuo, X. Liu, R. Gullapalli, A. Salama, R. Ord, and J. L. Prince, "Functional outcomes of glossectomy surgery with and without flap reconstruction," in *Proceedings of the Fourth International Symposium on Biomechanics, Healthcare and Information Science*, Kanazawa, Japan, 2008.

11. X. Liu, A. K. Jain, and G. Fitchinger, "Prostate implant reconstruction with discrete tomography," in *the 10th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2007, pp. 734-742.

12. X. Liu, E. Murano, M. Stone, and J. L. Prince, "HARP tracking refinement using seeded region growing," in *Proceeding of IEEE International Symposium on Biomedical Imaging (ISBI)*, 2007.

13. V. Parthasarathy, X. Liu, J. L. Prince, and M. Stone, "Three dimensional tracking of tongue motion using harmonic phase (HARP) processing of tagged cine-MR images," in *Proceeding of International Symposium on Biomechanics, Healthcare and Information Science*, 2007. (Co-frst author)

14. X. Liu, M. Stone, and J. L. Prince, "Tracking tongue motion in three dimensions using tagged MR images," in *Proceeding of IEEE International Symposium on Biomedical Imaging (ISBI)*, 2006, pp. 1372–1375.

- Conference abstracts

1. X. Liu, S. Roys, J. L. Prince, and R. Gullapalli, "Landmark-based prostate MR image matching using incompressible large deformation diffeomorphism," in *Proceedings of the 18th Scientific Meeting of the ISMRM*, 2009, p. 3823.

2. K. Z. Abd-Elmoniem, M. Stuber, A. A. Youssef, T. Sasano, T. P. Abraham, X. Liu, S. Mullur, E.-J. P. Vonken, M. R. Abraham, and J. L. Prince, "Direct quantification of 3-D myocaridal principal strain orientations: A new insight into heart regional function," in *Proceedings of the 17th Scientific Meeting of the ISMRM*, 2008.

3. X. Liu and J. L. Prince, "Fast and robust HARP tracking using region growing," in *Proceedings of the 16th Scientific Meeting of the ISMRM* , 2007.

4. X. Liu, L. Pan, W. Relwald, and C. Lorenz, "Study of magnetization behavior of MR tagging with steady state free precession," in *Proceedings of the 16th Scientific Meeting of the ISMRM*, 2007.

Bibliography

[1] C. H. Shiboski, B. L. Schmidt, and R. C. Jordan, "Tongue and tonsil carcinoma: increasing trends in the U.S. population ages 20-44 years," *Cancer*, vol. 103, no. 9, pp. 1843–1849, 2005.

[2] *SEER Cancer Statistics Review, 1975-2006.* National Cancer Institute. Bethesda, MD.

[3] D. Lloyd-Jones, R. Admas, M. Carnethon, G. D. Simone, T. B. Ferguson, K. Flegal, E. Ford, K. Furie, A. Go, K. Greenlund, N. Haase, S. Hailpern, M. Ho, V. Howard, B. Kissela, S. Kittner, D. Lackland, L. Lisabeth, A. Marelli, M. McDermott, J. Meigs, D. Mozaffarian, G. Nichol, C. O'Donnell, V. Roger, W. Rosamond, R. Sacco, P. Sorlie, R. Stafford, J. Steinberger, T. Thom, S. Wasserthiel-Smoller, N. Wong, J. Wylie-Rosett, and Y. Hong, "Heart disease and stroke statistics 2009 update: a report from the American Heart Association statistics committee and stroke statistics subcommittee," *Circulation*, vol. 119, pp. e21–181, 2009.

[4] A. N. Liberman, J. L. Weiss, B. I. Jugdutt, L. C. Becker, and M. L. Weisfeldt, "Two dimensional echocardiography and infarct size: relationship of regional wall motion and thickening to the extent of myocardial infarction in the dog," *Circulation*, vol. 64, pp. 739–746, 1981.

[5] P. M. Pattynama and A. D. Roos, "MR evaluation of myocardial ischemia and infarction," *Top. Mag. Res. Imag.*, vol. 7, no. 4, pp. 218–231, 1995.

[6] E. Nagel, M. Stuber, C. Matter, M. Lakatos, P. Boesiger, and O. M. Hess, "Rotational and translational motion of the heart after myocardial infarction," *J. Cardiovasc. Pharm.*, vol. 28, no. suppl. 2, pp. S31–S35, 1996.

[7] E. Nagel, M. Stuber, M. Lakatos, M. Scheidegger, P. Boesiger, and O. M. Hess, "Cardiac rotation and relaxation after anterolateral myocardial infarction," *Coronary Artery Disease*, vol. 11, no. 3, pp. 261–267, 2000.

[8] C. M. Kramer, W. J. Rogers, T. M. Theobald, T. P. Power, G. Geskin, and N. Reichek, "Dissociation between changes in intramyocardial function and left ventricular volumes in the eight weeks after first anterior myocardial infarction," *J. Amer. Coll. Cardio.*, vol. 30, pp. 1625–1632, 1997.

[9] N. R. Clark, N. Reichek, P. Bergey, E. A. Hoffman, D. Brownson, L. Palmon, and L. Axel, "Circumferential myocardial shortening in the normal human left ventricle," *Circulation*, vol. 84, pp. 64–74, 1991.

[10] T. Buck, P. Hunold, K. U. Wentz, J. Nesser, and R. Erbel, "Tomographic three-dimensional echocardiographic determination of chamber sie and syotolic function in patients with left ventricle aneurysm: comparison to magnetic resonance imaging, cine-ventriculography, and two-diemensional echocardiography," *Circulation*, vol. 69, pp. 497–505, 1997.

[11] E. A. Zerhouni, D. M. Parish, W. J. Rogers, A. Yang, and E. P. Shapiro, "Human heart: tagging with MR imaging – a method for noninvasive assessment of myocardial motion," *Radiology*, vol. 169, pp. 59–63, 1988.

[12] L. Axel and L. Dougherty, "MR imaging of motion with spatial modulation of magnetization," *Radiology*, vol. 171, pp. 841–845, 1989.

[13] L. Axel and L. Doughety, "Heart wall motion: improved method of spatial modulation of magnetization for MR imaging," *Radiology*, vol. 172, pp. 349–350, 1989.

[14] N. J. Pelc, R. J. Herfkens, A. Shimakawa, and D. R. Enzmann, "Phase contrast cine magnetic resonance imaging," *Mag. Res. Quarterly*, vol. 7, pp. 229–254, 1991.

[15] V. J. Wedeen, "Magnetic resonance imaging of myocardial kinematics. Technique to detect, localize and quantify the strain rates of the active human myocardium," *Mag. Reson. Med.*, vol. 27, pp. 52–67, 1992.

[16] T. Baer, J. Gore, S. Boyce, and P. Nye, "Application of MRI to the analysis of speech production," *Mag. Reson. Imag.*, vol. 5, no. 5, pp. 1–7, 1987.

[17] R. Luftkin, R. Christianson, and W. Hanafee, "Normal magnetic resonance imaging anatomy of the tongue, oropharynx, hypopharnyx and larnyx," *Dysphagia*, vol. 1, pp. 119–127, 1987.

[18] K. M. McKenna, B. A. Jabour, R. Luftkin, and W. hanafee, "Magnetic resonance imaging of the tongue and oropharynx," *Top. Mag. Reson. Imag.*, vol. 2, pp. 49–59, 1990.

[19] H. Cha and B. M. Patten, "Amyotrophic lateral sclerosis abnormalities of the tongue on magnetic resonance imaging," *Am. Neurol. Assoc.*, vol. 25, pp. 468–472, 1988.

[20] S. Takashima, J. Ikezoe, K. Harada, Y. Akai, S. Hamada, J. Arisawa, S. Morimoto, N. Masaki, T. Kozuka, and H. Maeda, "Tongue cancer: correlation of imaging and sonography with pathology," *Am. J. Neuroradiol.*, vol. 10, no. 2, pp. 419–424, 1989.

[21] A. V. Lakshminarayanan, S. Le, and M. J. McCutcheon, "MR imaging of the vocal tract during vowel production," *J. Mag. Reson. Imag.*, vol. 1, pp. 71–76, 1991.

[22] A. R. Greenwood, C. C. Goodyear, and P. A. Martin, "Measurements of vocal

tract shapes using magnetic resonance imaging," *IEE Proceeding-1*, vol. 139, no. 6, pp. 553–560, 1992.

[23] C. C. Moore, W. G. O'Dell, E. R. McVeigh, and E. A. Zerhouni, "Calculation of three-dimensional left ventricular strains from biplanar tagged MR images," *J. Mag. Res. Imag.*, vol. 2, pp. 165–175, 1992.

[24] A. M. Sulter, D. G. Miller, R. F. Wolf, H. K. Schutte, H. P. Wit, and E. I. Mooyaart, "On the relation between the dimensions and resonance characteristics of the vocal-tract - a study with MRI," *Mag. Reson. Imag.*, vol. 10, no. 3, pp. 365–373, 1992.

[25] B. H. Story, I. R. Titze, and E. A.Hoffman, "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.*, vol. 100, no. 3, pp. 537–554, 1996.

[26] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area function for an adult female speake based on volumetric imaging," *J. Acoust. Soc. Am.*, vol. 104, no. 1, pp. 471–487, 1998.

[27] P. Badin, G. Bailly, L. Reveret, M. Baciu, and C. Segebarth, "Three-dimensional linear articulatory modeling of tongue lips and face, based on MRI and video images," *J. Phonetics*, vol. 30, no. 3, pp. 533–554, 2002.

[28] S. Masaki, M. K. Tiede, K. Honda, Y. Shimada, I. Fujimoto, Y. Nakamura,

and N. Ninomiya, "MRI-based speech production study using a synchronized sampling method," *J. Acoustic Science Japan*, vol. 20, pp. 375–379, 1999.

[29] C. Shadle, M. Mohammad, J. Carter, and P. Jackson, "Dynamic magnetic resonance imaging: New tools for speech research," in *Proc. 14th Int. Cong. Phon. Sci.*, 1999, pp. 623–626.

[30] M. Stone, E. P. Davis, A. S. Douglas, M. NessAiver, R. Gullapalli, W. Leving, and A. J. Lundberg, "Modeling tongue surface contours from cine-MRI images," *J. Speech, Language, and Hearing Reseach*, vol. 44, no. 5, pp. 1026–1040, 2001.

[31] D. Demolin, T. Metens, and A. Soquet, "Real time MRI and articulatory coordinations in vowels," in *Proc. of International Speech Production Seminar*, 2000.

[32] K. Mady, R. Sader, A. Zimmermann, P. Hoole, A. Beer, H.-F. Zeilofer, and C. CHannig, "Assessment of consonant articulation in glossectomee speech by dynamic MRI," in *Proc. ICSLP 2002*, 2002.

[33] O. Engwall, "A revisit to the application of MRI to the analysis of speech production - testing our assumptions," in *6th Int. Sem. Spee. Pro. (ISPS)*, 2003.

[34] S. Narayanan, K. Nayak, D. Byrd, and S. Lee, "An approach to real-time

magnetic resonance imaging for speech production," *J. Acoust. Soc. Amer.*, vol. 115, no. 4, pp. 1771–1776, 2004.

[35] M. Niitsu, M. Kumada, G. Campeau, S. Niimi, S. J. Riederer, and Y. Itai, "Tongue displacement: visualization with rapid tagged magnetization-prepared MR imaging," *Radiology*, vol. 191, pp. 578–580, 1994.

[36] V. J. Napadow, Q.Chen, V. J. Wedeen, and R. J. Gilbert, "Biomechanical basis for lingual muscular deformation during swallowing," *Am. J. Physiol.*, vol. 277, pp. G695–G701, 1999.

[37] ——, "Intramural mechanics of the human tongue in association with physiological deformations," *J. Biomech.*, vol. 322, pp. 1–12, 1999.

[38] M. Stone, E. P. Davis, A. S. Douglas, M. NessAiver, R. Gullapalli, W. Leving, and A. J. Lundberg, "Modelling the motion of the internal tongue from tagged cine-MR images," *J. Acou. Soc. Am.*, vol. 109, no. 6, pp. 2974–2982, 2001.

[39] V. Parthasarathy, J. L. Prince, M. Stone, E. Murano, and M. Nessaiver, "Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing," *J. Acoust. Soc. Am.*, vol. 121, no. 1, pp. 491–504, 2007.

[40] N. F. Osman, W. S. Kerwin, E. R. McVeigh, and J. L. Prince, "Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging," *Magn. Reson. Med.*, vol. 42, no. 6, pp. 1048–1060, 1999.

[41] N. F. Osman and J. L. Prince, "Visualizing myocardial function using HARP MRI," *Phys. Med. Biol.*, vol. 45, no. 6, pp. 1665–1682, 2000.

[42] N. F. Osman, E. R. McVeigh, and J. L. Prince, "Imaging heart motion using harmonic phase MRI," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 186–202, 2000.

[43] L. Pan, J. L. Prince, J. A. C. Lima, and N. F. Osman, "Fast tracking of cardiac motion using 3d-harp," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 8, pp. 1424–1435, 2005.

[44] S. E. Fischer, G. C. McKinnon, S. E. Maier, and P. Boesiger, "Improved myocardial tagging contrast," *Mag. Res. Med.*, vol. 30, pp. 191–200, 1993.

[45] M. NessAiver and J. L. Prince, "Magnitude image CSPAMM reconstruction (MICSR)," *Mag. Res. Med.*, vol. 50, no. 2, pp. 331–342, 2003.

[46] A. A. Amini, Y. Chen, R. W. Curwen, V. Manu, and J. Sun, "Coupled B-snake grides and constrained thin-plate splines for analysis of 2-D tissue deformations from tagged MRI," *IEEE Tran. Med. Imag.*, vol. 17, pp. 344–356, 1998.

[47] W. S. Kerwin and J. L. Prince, "Cardiac material markers from tagged MR images," *Med. Imag. Anal.*, vol. 2, no. 4, pp. 339–353, 1998.

[48] A. A. Young, D. L. Kraitchman, L. Dougherty, and L. Axel, "Tracking and

finite element analysis of stripe deformation in magnetic resonance tagging," *IEEE Trans. Med. Imag.*, vol. 14, no. 3, pp. 413–421, 1995.

[49] I. Haber, D. N. Metaxas, and L. Axel, "Three-dimensional motion reconstruction and analysis of the right ventricle using tagged MRI," *Med. Imag. Anal.*, vol. 4, no. 4, pp. 335–355, 2000.

[50] T. S. Denny and J. L. Prince, "Reconstruction of 3-D left ventricular motion from planar tagged cardiac MR images: an estimation theoretic approach," *IEEE Trans. Med. Imag.*, vol. 14, no. 4, pp. 625–635, 1995.

[51] M. A. Guttman, J. L. Prince, and E. R. McVeigh, "Tag and contour detection in tagged MR images of the left ventricle," *IEEE Trans. Med. Imag.*, vol. 13, no. 1, pp. 74–88, 1994.

[52] M. A. Guttman, E. A. Zerhouni, and E. R. MvVeigh, "Analysis of cardiac function from MR images," *IEEE Comput. Graph. Appl.*, no. 1, pp. 30–38, 1997.

[53] W. G. O'Dell, C. C. Moore, W. C. Hunter, E. A. Zerhouni, and E. R. McVeigh, "Three-dimensional myocardial deformations: calculation with displacement field fitting to tagged MR images," *Radiology*, vol. 195, pp. 829–835, 1995.

[54] T. S. Denny and E. R. McVeigh, "Model-free reconstruction of three-

dimensional myocardial strain from planar tagged MR images," *J. Magn. Reson. Imag.*, vol. 7, pp. 799–810, 1997.

[55] P. Radeva, A. A. Amini, and J. Huang, "Deformable B-Solids and implicit snakes for 3D localization and tracking of SPAMM MRI data," *Comp. Vis. Imag. Und.*, vol. 66, pp. 163–178, 1997.

[56] J. Declerck, J. Feldmar, and N. Ayache, "Definition of a four-dimensional continuous planispheric transformation for the tracking and the analysis of left-ventricle motion," *Med. Imag. Anal.*, vol. 2, no. 2, pp. 197–213, 1998.

[57] A. A. Young, "Model tags: direct 3D tracking of heart wall motion from tagged magnetic resonance images," *Med. Imag. Anal.*, vol. 3, no. 4, pp. 361–372, 1999.

[58] C. Ozturk and E. R. McVeigh, "Four-dimensional *b*-spline based motion analysis of tagged MR images: introduction and in vivo validation," *Phys. Med. Biol.*, vol. 45, no. 6, pp. 1683–1702, 2000.

[59] J. Huang, D. Abendchein, V. G. Davila-Roman, and A. A. Amini, "Spatio-temporal tracking of myocardial deformations with a 4-D B-spline model from tagged MRI," *IEEE Tran. Med. Imag.*, vol. 18, no. 10, pp. 957–972, 1999.

[60] K. Park, D. Metaxas, and L. Axel, "A finite element model for functional analysis of 4D cardiac-tagged MR images," in *MICCAI 2003, LNCS 2878*. Springer, 2003, pp. 491–498.

[61] A. Montillo, D. Metaxas, and L. Axel, "Extracting tissue deformation using Gabor filter banks," in *Proc. SPIE Med. Imag.*, vol. 5369, 2004, pp. 1–9.

[62] V. Parthasaranthy, M. Ness-Aiver, and J. L. Prince, "Tracking tongue motion from tagged magnetic resonance images using harmonic phase mri (harp-mri)," *Proc. Of 15th International Congress of Phonetic Sciences*, 2003.

[63] Y. P. Wang, Y. Chen, and A. A. Amini, "Fast LV motion estimation using subspace approximation techniques," *IEEE Trans. Med. Imag.*, vol. 20, pp. 499–513, 2001.

[64] X. Deng and T. S. Denney, "Three-dimensional myocardial strain reconstruction from tagged MRI using a cylindrical B-spline model," *IEEE Trans. Med. Imag.*, vol. 23, pp. 861–867, 2004.

[65] D. Suter and F. Chen, "Left ventricular motion reconstruction based on elastic vector splines," *IEEE Trans. Med. Imag.*, vol. 19, no. 4, pp. 295–305, 2000.

[66] S. Ryf, M. A. Spiegel, M. Gerber, and P. Boesiger, "Myocardial tagging with 3D-CSPAMM," *J. Magn. Res Imag.*, vol. 16, no. 3, pp. 320–325, 2002.

[67] W. H. Perman, L. L. Cresswell, S. G. Wyers, M. J. Moulton, and M. K. Pasque, "Hybrid DANTE and phase-contrast imaging technique for measurement of three-dimensional myocardial wall motion," *J. Magn. Reson. Imag.*, vol. 5, pp. 101–106, 1995.

[68] J. P. Kuijer, J. T. Marcus, M. J. Gotte, A. C. van Rossum, and R. M. Heethaa, "Simultaneous MRI tagging and through-plane velocity quantification: A three-dimensional myocardial motion tracking algorithm," *J. Mag. Reson. Imag.*, vol. 9, pp. 409–419, 1999.

[69] K. Z. Abd-Elmoniem, M. Stuber, N. Osman, and J. L. Prince, "ZHARP: three-dimensional motion tracking from a single image plane," in *Info. Proc. Med. Imag.*, 2005, pp. 639–651.

[70] K. Z. Abd-Elmoniem, N. F. Osman, J. L. Prince, and M. Stuber, "Three-dimensional magnetic resonance myocardial motion tracking from a single image plane," *Mag. Res. Med.*, vol. 58, pp. 92–102, 2007.

[71] D. Dick, C. Ozturk, A. Douglas, E. R. McVeigh, and M. Stone, "Three-dimensional tracking of tongue using tagged MRI," in *Proc. Intl. Soc. Magn. Reson. Med.*, 2000.

[72] M. Stone, D. Dick, A. Douglas, E. Davis, and C. Ozturk, "Modelling the internal tongue using principal strains," in *Proc. Speech Prod Seminar*, 2000, pp. 133–136.

[73] A. M. Khalifa, A. M. Youssef, and N. F. Osman, "Improved harmonic phase (HARP) method for motion tracking a tagged cardiac MR images," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2005, pp. 4298–4301.

[74] S. R. Tecelao, J. J. Zwanenburg, J. P. Kuijer, and J. T. Marcus, "Extended harmonic phase tracking of myocardial motion: Improved coverage of myocardium and its effect on strain results," *J. Magn. Reson. Imag.*, vol. 23, no. 5, pp. 682–690, 2006.

[75] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, pp. 641–647, 1994.

[76] A. H. Aletras, S. Ding, R. S. Balaban, and H. Wen, "DENSE: Displacement encoding with stimulated echoes in cardiac functional MRI," *J. Magn. Reson.*, vol. 137, pp. 247–252, 1999.

[77] B. S. Spottiswoode, X. Zhong, A. T. Hess, C. M. Kramer, E. M. Meintjes, B. M. Mayosi, and F. H. Epstein, "Tracking myocardial motion from cine DENSE images using spatiotemporal phase unwrapping and temporal fitting," *IEEE Tran. Med. Imag.*, vol. 26, no. 1, pp. 15–30, 2007.

[78] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Machematik*, vol. 1, pp. 269–271, 1959.

[79] N. Osman and J. L. Prince, "Improving HARP to produce smooth strain maps of the heart," in *Proc. Intl. Soc. Magn. Reson. Med.*, 2001, p. 605.

[80] V. Parthasarathy and J. L. Prince, "Artifact reduction in HARP strain maps," in *Proc. Intl. Soc. Magn. Reson. Med.*, 2004, p. 1785.

[81] K. Z. Abd-Elmoniem, V. Parthasarathy, and J. L. Prince, "Artifact reduction in HARP strain maps using anisotropic smoothing," in *Proc. SPIE Med. Imag.*, vol. 6143, 2006, pp. 841–849.

[82] K. Z. Abd-Elmoneim, M. Stuber, and J. L. Prince, "Direct three-dimensional myocardial strain tensor quantification and tracking using zHARP," *Med. Imag. Anal.*, vol. 12, pp. 778–786, 2008.

[83] S. E. Fischer, G. C. McKinnon, M. B. Scheidegger, W. Prins, D. Meier, and P. Boesiger, "True myocardial motion tracking," *Mag. Res. Med.*, vol. 31, pp. 401–413, 1994.

[84] M. Stuber, S. Fischer, M.B.Scheidegger, and P. Boesiger, "Toward high-resolution myocardial tagging," *Magn. Reson. Med.*, vol. 41, pp. 639–643, 1999.

[85] K. Z.Abd-Elmoniem, M. Stuber, and J. L. Prince, "Multi0slice three-dimensional myocardial strain tensor quantification using zHARP," in *Info. Proc. Med. Imag.*, 2007, pp. 62–73.

[86] T. G. Reese, D. A. Feinberg, J. Dou, and V. J. Wedeen, "Phase contrast MRI of myocardial 3D strain by encoding contiguous slices in a single shot," *Magn. Reson. Med.*, vol. 47, no. 4, pp. 665–676, 2002.

[87] E. R. McVeigh, "MRI of myocardial function: motion tracking techniques," *Magn. Reson. Imag.*, vol. 14, no. 2, pp. 137–150, 1996.

[88] F. L. Bookstein, "Principal warps: thin-plate splines and decomposition of deformations," *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 11, no. 6, pp. 567–585, 1989.

[89] F. Yin, C. Chan, and R. Judd, "Compressibility of perfused passive myocardium," *Amer. J. Physiol.-Heart Circ. Physiol.*, vol. 8, pp. 1864–1870, 1996.

[90] I. Rodriguez, D. B. Ennis, and H. Wen, "Noninvasive measurement of myocardial tissue volume change during systolic contraction and diastolic relaxation in the canine left ventricle," *Mag. Reson. Med.*, vol. 55, no. 3, pp. 484–90, 2006.

[91] S. M. Song and R. M. Leahy, "Computation of 3-D velocity fields from 3-d cine CT images of a human heart," *IEEE Trans. Med. Imag.*, vol. 10, no. 3, pp. 295–306, 1991.

[92] T. Denney and J. L. Prince, "Reconstruction of 3D left ventricular motion from planar tagged cardiac MR images: an estimation theoretic approach," *IEEE Trans. Med. Imag.*, vol. 14, no. 4, pp. 625–635, 1995.

[93] A. Bistoquet, J. Oshinski, and O. Skrinjar, "Left ventricle deformation recovery from cine MRI using an incompressible model," *IEEE Trans. Med. Imag.*, vol. 26, no. 9, pp. 1136–1153, 2007.

[94] ——, "Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model," *Med. Imag. Anal.*, vol. 12, pp. 69–85, 2008.

[95] W. M. Lai, D. Rubin, and E. Krempl, *Introduction to continuum mechanics (3rd edition)*. Butterworth-Heinemann, 1993.

[96] L. Amodei and M. N. Benbourhim, "A vector spline approximation," *J. Approx. Theo.*, vol. 67, no. 1, pp. 51–79, 1991.

[97] M. H. Davis, A. Khotanzad, D. P. Flamig, and S. E. Harms, "A physics-based coordinate transformation for 3-D image matching," *IEEE Tran. Med. Imag.*, vol. 16, no. 3, pp. 317–328, 1997.

[98] J. Kohlrausch, K. Rohr, and H. S. Stiehl, "A new class of elastic body splines for nonrigid registration of medical images," *J. Math. Imag. Vis*, vol. 23, pp. 253–280, 2005.

[99] M. Arigovindan, "Variational reconstruction of vector and scalar images from non-uniform samples," Ph.D. dissertation, École Polytechnique Federale de Lausanne, 2005.

[100] E. Waks, J. L. Prince, and A. S. Douglas, "Cardiac motion simulator for tagged MRI," in *Proc. MMBIA '96*, 1996, pp. 182–191.

[101] T. Arts, W. Hunter, A. Douglas, A. Muijtjens, and R. Reneman, "Description of deformation of the left ventricle by a kinematic model," *J. Biomechanics*, vol. 25, no. 10, pp. 1119–1127, 1992.

[102] N. F. Osman and J. L. Prince, "Improving HARP to produce smooth strain maps of the heart," in *Proc. Intl. Soc. Mag. Reson. Med.*, 2001, p. 605.

[103] F. J. Narcowich and J. D. Ward, "Generalized Hermite interpolation via matrix-valued conditionally posistive definite functions," *Math. Comp.*, vol. 64, pp. 661–687, 1994.

[104] H. Wendland, "Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree," *Adv. Comp. Math.*, vol. 4, pp. 389–396, 1995.

[105] S. Lowitzsch, "Approximation and interpolation employing divergence-free radial basis functions with applications," Ph.D. dissertation, Texas A&M University, 2002.

[106] S. C. Joshi and M. I. Miller, "Landmark matching via large deformation diffeomorphisms," *IEEE Trans. Imag. Proc.*, vol. 9, no. 8, pp. 1357–1370, 2000.

[107] X. Liu, S. Roys, J. L. Prince, and R. Gullapalli, "Landmark-based prostate MR image matching using incompressible large deformation diffeomorphism," in *Proceedings of the 18th Scientific Meeting of the ISMRM, 2009*, 2009, p. 3823.

[108] Y. Matsui, K. Ohno, Y. Yamashita, and K. Takahashi, "Factors influencing postoperative speech function of tongue cancer patients following reconstruction

with fasciocutaneous/myocutaneous flaps: a multicenter study," *Int. J. Oral and Maxillofacial Surgery*, vol. 36, no. 7, pp. 601–609, 2007.

[109] G. Nicoletti, D. S. Soutar, M. S. Jackson, A. A. Wrench, G. Robertson, and C. Robertson, "Objective assessment of speech after surgical treatment for oral cancer: experience from 196 selected cases," *Plast. Reconstr. Surg*, vol. 113, pp. 114–125, 2004.

[110] H. Terai and M. Shimahara, "Articulatory function in aptients who have undergone glossectomy with use of an artifical graft membrane," *Oral Surg. Oral Med. Ora Pathol. Oral Radiol. Endo.*, vol. 89, no. 5, pp. 560–562, 2000.

[111] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape model — their training and applications," *Comp. Vis. Imag. Under.*, vol. 61, pp. 38–59, 1995.

[112] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, 2001.

[113] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object recognition," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 19, no. 7, pp. 696–710, 1997.

[114] G. P. Zientara, L. P. Panych, and F. A. Jolesz, "Dynamically adaptive MRI with encoding by singular value decomposition," *Magn. Reson. Med.*, vol. 32, no. 2, pp. 268–274, 1994.

[115] R. CHandrashekara, A. Rao, G. I. Sanchez-Ortiz, R. H. Mohiaddin, and D. Rueckert, "Construction of a statistical model for cardiac motion analysis using nonrigid image registration," in *Info. Proc. Med. Imag.*, 2003, pp. 599–610.

[116] Z. Xue, D. Shen, and C. Davatzikos, "Statistical representation of high-dimensional deformation fields with application to statistically constrained 3d warping," *Med. Imag. Anal.*, vol. 10, pp. 740–751, 1994.

[117] R. Harshman, P. Ladeforged, and L. Goldstein, "Factor analysis of tongue shapes," *J. Acoust. Soc. Am.*, vol. 62, pp. 693–713, 1977.

[118] P. Hoole, "On the lingual organization of the German vowal system," *J. Acoust. Soc. Am.*, vol. 106, no. 2, pp. 1020–1032, 1999.

[119] M. Stone, M. Goldstein, and Y. Zhang, "Principal component analysis of cross-sectional tongue shape in vowels," *Speech Communication*, vol. 22, pp. 173–184, 1997.

[120] E. Slub, P. Smith, M. Stone, and M. Goldstein, "Principal components representation of the two-dimensional coronal tongue surface," *Phonetica*, vol. 59, pp. 108–133, 2002.

[121] T. Bressmann, E. Ackloo, C. Heng, and J. C. Irish, "Quantitative three-

dimensional ultrasound imaging of partially resected tongues," *Otolaryngology - Head and Neck Surgery*, vol. 136, no. 5, pp. 799–805, 2007.

[122] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *BIophys. J.*, vol. 66, no. 1, pp. 259–267, 1994.

[123] V. J. Napadow, Q. Chen, V. mai, P. T. So, and R. J. Gilbert, "Quantitative analysis of three-dimensional-resolved fiber architecture in heterogeneous skeletal muscle tissue using nmr and optical imaging methods," *Biophys. J.*, vol. 80, no. 6, pp. 2968–2975, 2001.

[124] R. J. Gilbert and V. J. Napadow, "Three-dimensional muscular architecture of the human tongue determined *in vivo* with diffusion tensor magnetic resonance imaging," *Dysphagia*, vol. 20, no. 1, pp. 1–7, 2005.

[125] H. Shinagawa, E. Z. Murano, J. Zhuo, B. Landman, R. P. Gullapalli, J. L. Prince, and M. Stone, "Tongue muscle fiber tracking during rest and tongue protrusion with oral appliances: a preliminary study with diffusion tensor imaging," *Acoust. Sci. Tech.*, vol. 29, no. 4, pp. 291–294, 2008.

[126] J. S. Duncan, F. A. Lee, A. W. M. Smeulders, and B. L. Zaret, "A bending energy model for measurement of cardiac shape deformity," *IEEE Trans. Med. Imag.*, vol. 10, no. 3, pp. 307–332, 1991.

[127] D. C. Alexander, J. C. Gee, and R. Bajcsy, "Strategies for data reorientation

during nonrigid warps of diffusion tensor images," in *Proc. MICCAI 1999*, 1999, pp. 463–472.

[128] D. C. Alexander, C. Piepaoli, P. J. Basser, and J. C. Gee, "Spatial transformations of diffusion tensor magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 20, no. 11, pp. 1131–1139, 2001.

[129] Y. Cao, M. I. Miller, R. L. Winslow, and L. Younes, "Large deformation diffeomorphic metric mapping of vector fields," *IEEE Trans. Med. Imag.*, vol. 24, no. 9, pp. 1216–1230, 2005.

[130] M. Markl and J. Hennig, "Phase contrast MRI with improved temporal resolution by view sharing: k-space related velocity mapping properties," *Mag. Reson. Imag.*, vol. 19, no. 5, pp. 669–676, 2001.

Vita

Xiaofeng Liu was born in Yantai, China on July 9, 1976. He received his B.E. degree in Automatic Control from University of Science and Technology of China in 1998. He received his M.E. degree in Pattern Recognition and Intelligent System from Institute of Automation, Chinese Academy of Sciences in 2001. He received his M.S. degree in Computer Science from Vanderbilt University in 2003. He is currently pursuing his PhD degree in Computer Science at the Johns Hopkins University. His research interests include medical image analysis, computer vision, motion analysis and cardiac and vocal tract imaging.