

**Three Dimensional Tissue Motion Analysis from Tagged
Magnetic Resonance Imaging**

by

Fangxu Xing

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

Baltimore, Maryland

October, 2015

© Fangxu Xing 2015

All rights reserved

Abstract

Motion estimation of soft tissues during organ deformation has been an important topic in medical imaging studies. Its application involves a variety of internal and external organs including the heart, the lung, the brain, and the tongue. Tagged magnetic resonance imaging has been used for decades to observe and quantify motion and strain of deforming tissues. It places temporary noninvasive markers—so called “tags”—in the tissue of interest that deform together with the tissue during motion, producing images that carry motion information in the deformed tagged patterns. These images can later be processed using phase-extraction algorithms to achieve motion estimation and strain computation.

In this dissertation, we study three-dimensional (3D) motion estimation and analysis using tagged magnetic resonance images with applications focused on speech studies and traumatic brain injury modeling. Novel algorithms are developed to assist tagged motion analysis. Firstly, a pipeline of methods—TMAP—is proposed to compute 3D motion from tagged and cine images of the tongue during speech. TMAP produces an estimation of motion along with a multi-subject analysis of motion pat-

ABSTRACT

tern differences between healthy control subjects and post-glossectomy patients. Secondly, an enhanced 3D motion estimation algorithm—E-IDEA—is proposed. E-IDEA tackles the incompressible motion both on the internal tissue region and the tissue boundaries, reducing the boundary errors and yielding a motion estimate that is more accurate overall. Thirdly, a novel 3D motion estimation algorithm—PVIRA—is developed. Based on image registration and tracking, PVIRA is a faster and more robust method that performs phase extraction in a novel way. Lastly, a method to reveal muscles' activity using strain in the line of action of muscle fiber directions is presented. It is a first step toward relating motion production with individual muscles and provides a new tool for future clinical and scientific use.

Primary Reader: Dr. Jerry L. Prince

Secondary Reader: Dr. Trac D. Tran, Dr. John I. Goutsias

Acknowledgments

This dissertation would not be possible without the support and help from many people. I would like to first express my gratitude to my advisor, Dr. Jerry Prince, for his patient guidance and support over the past six years. It has been a great fortune for me to study under his supervision during my graduate career. He directed me to the field of medical imaging study, guided me with great passion, and taught me how to be a better person both in conducting research and in real life.

I would like to thank Dr. Maureen Stone for her support and her guidance. She has always been a passionate user of my new algorithms and software, and has been providing me new insights from the view of a speech scientist. I would like to thank Dr. Philip Bayly and Dr. Bennett Landman for their financial support and priceless advice in our collaborations. I would like to thank Dr. Trac Tran and Dr. John Goutsias for carefully examining my thesis and being a member in my thesis committee. I would like to thank Dr. Jonghye Woo, Dr. Junghoon Lee, and Dr. A. David Gomez for our friendship and fruitful collaborations. Last but not least, I would like to thank Dr. Emi Murano, not only for her valuable teachings and

ACKNOWLEDGMENTS

feedbacks for me in our five-year collaboration time, but also for her long-standing friendship, support, and help during difficult times in my career.

I am grateful to all members from the Image Analysis and Communications Laboratory, especially every former and present member in the HARP group. We had numerous insightful discussions together which led to productive new results and papers. I want to thank Aaron Carass and Laura Granite for their technical and administrative supports. I also want to thank all my friends, who brought much wonderful time to my life during the years I spent at Johns Hopkins University.

Finally, I would like to express my deepest gratitude to my parents Tongsheng Xing and Rong Luo. This thesis is dedicated to them for their endless love, support, and encouragement.

Contents

Abstract	ii
Acknowledgments	iv
List of Tables	xi
List of Figures	xii
1 Introduction	1
1.1 Motivation	1
1.2 Anatomy of the Human Tongue and Brain	3
1.3 Tagged Magnetic Resonance Imaging	6
1.4 Motion Estimation and Challenges	7
1.5 Thesis Contributions	11
1.6 Thesis Organization	14
2 Background	16

CONTENTS

2.1	Data Acquisition	16
2.2	Harmonic Phase Algorithm	22
2.3	Incompressible Deformation Estimation Algorithm	25
2.4	Strain Tensor Calculation	27
2.5	Summary	30
3	Three Dimensional Motion Analysis Pipeline	31
3.1	Introduction	31
3.2	Phase Calibration of CSPAMM Images	32
3.3	Segmentation of the Tongue Mask	41
3.4	Tongue Motion Analysis Pipeline	44
3.5	Multi-subject Data Regularization	45
3.6	Two-step Principal Component Analysis	48
3.7	Evaluation of Angle of Motion	52
3.8	Two-step PCA of the Whole Tongue	54
3.9	Evaluation of Two-step PCA on Mirrored Hemi-tongue	54
3.10	Implementation of TMAP	58
3.11	Discussion	61
3.12	Summary	64
4	Enhanced Incompressible Motion Estimation	66
4.1	Introduction	66

CONTENTS

4.2	Measuring Tongue Surface Deformation	68
4.2.1	Enhanced IDEA	70
4.3	Visual Assessment of E-IDEA	72
4.4	Point Tracking with E-IDEA	74
4.5	Evaluating Reprojection Error	75
4.6	Discussion and Summary	75
5	Phase Vector Incompressible Registration Algorithm	77
5.1	Introduction	77
5.2	Interpolation of Tagged Slices	79
5.3	Harmonic Phase Volumes	81
5.4	Registration of Phase Volumes	82
5.5	Deformation of Phase Volumes	87
5.6	Incorporation of Incompressibility	88
5.7	Incorporation of Inverse Estimation	89
5.8	Simulation of Brain and Tongue Deformation	91
5.9	Estimation of Brain Motion in Mild Accelerations	99
5.10	Estimation of Tongue Motion in Speech	101
5.11	Discussion	103
5.12	Conclusion	106
6	Relating Motion Production with Muscle Activities	107

CONTENTS

6.1	Introduction	107
6.2	Strain in the Line of Action	111
6.3	Thresholding Strain in the Line of Action	114
6.4	Discussion	115
6.5	Summary	119
7	Conclusions and Future Work	120
7.1	3D Motion Analysis Pipeline	121
7.1.1	Main Results	121
7.1.2	Future Work	122
7.2	Enhanced Incompressible Motion Estimation	123
7.2.1	Main Results	124
7.2.2	Future Work	125
7.3	Phase Vector Incompressible Registration	125
7.3.1	Main Results	126
7.3.2	Future Work	127
7.4	Motion and Muscle Activities	127
7.4.1	Main Results	128
7.4.2	Future Work	129
7.5	Overall Perspective	130
7.6	List of Publications	130

CONTENTS

A	Continous Label Fusion of Medical Images	136
A.1	Introduction	136
A.2	EM Algorithm for Continuous Label Fusion	141
A.3	Bias Invariance Problem	144
A.4	Initialization Strategies	148
A.5	MAP Estimation for Continuous Label Fusion	150
A.6	2D Point Identification Simulations	153
A.7	Empirical Fusion: RV Insertion Points Identification in Cardiac Images	156
A.8	Empirical Fusion: Contour Identification in Cardiac Images	157
A.9	Discussion	161
A.10	Summary	163
	Bibliography	164
	Vita	182

List of Tables

3.1	Controls' and Patients' Average Elevation and Azimuth Angles of Motion (in degrees).	53
5.1	Brain Rotation Simulation Test of Three Methods	94
5.2	Tongue Protrusion Simulation Test of Three Methods	99
5.3	Brain Acceleration Motion Estimation from Three Methods	101
5.4	Tongue Motion Estimation from Three Methods	103
6.1	Percentage of compressed muscles at critical time frames during utterance "a souk".	115
A.1	RMSE (in pixels) of Estimated Truth from Generated Truth with Six Fusion Techniques in 500 Monte Carlos of 2D Simulation.	155
A.2	RMSE (in pixels) of Estimated Truth from Expert Truth with Six Fusion Techniques in 100 Monte Carlos of Real RV Insertion Points Data.	157
A.3	Dice Coefficients (in percentage) of Estimated Truth from Expert Truth with Six Fusion Techniques in 50 Monte Carlos and Discrete STAPLE for Endocardium Identification.	159
A.4	Dice Coefficients (in percentage) of Estimated Truth from Expert Truth with Two Informed CSTAPLE Methods Subject to Training Dataset Size Change.	161

List of Figures

1.1	(a) The tongue and its external structures. Source: anatomybodypics.net. (b) Internal muscles of the tongue. Source: bartleby.com. . .	4
1.2	(a) Location of the brain in the skull. (b) Basic structure of the brain. Source: wikipedia.org/wiki/Human_brain.	5
1.3	(a) Horizontally tagged mid-sagittal image slice of the tongue. (b) Vertically tagged image on the same slice location. (c) Cine image on the same slice location.	7
2.1	Illustration of the acquisition setup of tagged MR image slices for tongue motion in speech. A: anterior. P: posterior. L: left. R: right. S: superior. I: inferior. (a) Slice locations. (b) An axial tagged slice. (c) A coronal tagged slice. (d) A sagittal tagged slice.	18
2.2	Illustration of the acquisition setup of tagged MR image slices for brain motion. A: anterior. P: posterior. L: left. R: right. S: superior. I: inferior. (a) Slice locations. (b) An axial slice with horizontal tags. (c) An axial slice with vertical tags. (d) A sagittal tagged slice.	21
2.3	Processing of a tagged MR image. (a) Acquired tagged slice. (b) Harmonic peaks in the Fourier domain and a HARP filter. (c) Harmonic magnitude slice. (d) Harmonic phase slice.	22
2.4	Resulting 2D motion fields from HARP on (a) an axial slice, (b) a coronal slice, and (c) a sagittal slice of the tongue.	24
2.5	Relationship between 2D motion components and 3D IDEA estimation on (a) a slice with x and y motion components (b) a slice with z motion component. (c) shows the integration process of the estimated velocity components by IDEA (Source: Liu et al. Incompressible deformation estimation algorithm (IDEA) from tagged MR images. <i>IEEE TMI</i> , 2012.).	25
2.6	Deformation of a continuum body. Source: wikipedia.org.	28
3.1	Flowchart of the tongue motion analysis pipeline (TMAP).	33

LIST OF FIGURES

3.2	Demonstration of phase of tagged images from acquisition. (a) Phase of A and B images. (b) Demonstration of tag fading on an A image. (c) A MICS image on the first time frame. (d) A CSPAMM image on the first time frame.	35
3.3	Phase calibration result. (a) and (c) are CSPAMM images before calibration. (b) and (d) are corrected CSPAMM images.	39
3.4	Cine super volume reconstruction by the SUPERV algorithm.	40
3.5	Seed propagation in time and segmentation of temporal stack.	42
3.6	Segmentation of the supervolume.	43
3.7	3D motion estimation produce by TMAP. (a) A healthy control's HARP input at time frame /s/. (b) A healthy control's HARP input at time frame /k/. (c) A healthy control's 3D motion at time frame /s/. (d) A healthy control's 3D motion at time frame /k/. (e) A patient's 3d motion at time frame /s/. (f) A patient's 3D motion at time frame /k/. Note: cones are color-coded by motion directions as shown in the color diagram (red for left-right, blue for superior-inferior, green for anterior-posterior). Cone size is magnitude.	46
3.8	(a) Division of eight VOIs in the tongue. (b) An example using VOI-1: Average motion of 26 frames and all subjects. Horizontal line divides healthy controls (bottom) and patients (top). Vertical curves are at time frames /ə/, /s/, and /k/. (c) Interpolated motion between time frames /ə/ and /k/ from (b). Horizontal line divides healthy controls (bottom) and patients (top). Vertical lines are at time frames /ə/, /s/, and /k/.	48
3.9	Elevation and azimuth angles during motion of healthy control and patient subjects in volume of interest 1. Each curve is a subject. (a) and (b) show sixteen healthy controls and (c) and (d) show seven patients.	53
3.10	Two-step PC directions and PC weights. (a) Primary PC directions from healthy controls. (b) Secondary PC directions from patients. (c) and (d) show the PC weights for healthy controls and patients. Each dashed line is a subject.	55
3.11	First two patient-like abnormal PC weights for four volumes of interest and all subjects. The blue dot is the origin. Green dots are test controls, circles are patient glossectomy sides, and crosses are patient native sides.	57
3.12	Boxplot of the average secondary principal component weights (abnormal energy) of control test, PGS, and PNS in 12870 Monte Carlo trials. Within each box, the center bar shows the median and the circle shows the mean. Patients 12 and 13 are labeled with A and B to represent their two recordings.	58
3.13	HARP5 2D processing GUI.	59
3.14	HARP5 2D visualization GUI.	60

LIST OF FIGURES

3.15	HARP5 3D processing GUI.	61
4.1	(a) Tongue mask of a healthy control subject (sagittal view). (b) HARP field on axial and coronal slices as input for IDEA, zoomed in at the tongue back. (c) IDEA result at the tongue back. (d) Surface normal deformation component at tongue back surface. (e) Proposed method result.	67
4.2	Relationship between 2D motion components and 3D motion on (a) an axial slice, (b) a coronal slice and (c) the tongue surface.	68
4.3	Visual comparison of conventional IDEA result and E-IDEA result.	73
4.4	Comparison of IDEA and E-IDEA with manually tracked surface points. (a) Tracks of the healthy control's surface points by manual (<i>blue</i>), IDEA (<i>yellow</i>), and E-IDEA (<i>green</i>). (b) Error magnitude for the healthy control (<i>bar</i> is median and <i>circle</i> is mean). (c) Error magnitude for the patient.	74
4.5	Regularized histogram of IDEA and E-IDEA's reprojection error on internal and surface points. <i>Dotted lines</i> show the mean of four types of reprojection error.	76
5.1	Demonstration of interpolation of tagged slices on tagged images of the tongue. (a) Spatial locations of acquired samples (blue) and interpolated samples (red). (b) A tagged sagittal tongue slice from acquisition. (c) Four interpolated sagittal tongue slices with horizontal tags.	80
5.2	Linear interpolation of wrapped phase values.	88
5.3	Test of PVIRA on simulated brain rotations. (a) Simulated true motion. (b) Estimated motion. (c) Estimated inverse motion. (d) Simulated tagged images. (e) Error magnitude of the estimation. (f) Composed forward-backward motion in magnitude. (g) Jacobian determinant of motion estimate. (h) Box-plotted estimation error of all time frames.	93
5.4	Performance of three methods under noise. (a) PVIRA estimation. (b) 3D HARP estimation. (c) IDEA estimation. (d) PVIRA error. (e) 3D HARP error. (f) IDEA error. (g) Tagged images with noise. (h) Box-plotted estimation error under difference noise level. Center bar is median and circle is mean.	96
5.5	PVIRA estimation error with different regularizations and noise levels. σ is the size of the two regularizing Gaussian kernels.	98
5.6	Test of PVIRA on simulated tongue protrusion. (a) Tagged slices with three tag directions. (b) Estimation error with three methods. (c) Simulated motion and estimated motion with three methods.	100

LIST OF FIGURES

5.7	Brain mild acceleration motion estimation. (a)-(c) Left rotation estimates. (d)-(f) Right rotation estimates. (g) Examples of tagged axial and sagittal slices. (h) Example slices of estimation difference from HARP. Colorbar goes from 0 to 0.5 as in the previous figures. (i) Example slices of Jacobian determinant. (j) Histogram of difference from HARP using IDEA and PVIRA.	102
5.8	Tongue motion estimation in speech. (a)-(c) Forward motion estimates. (d)-(f) Upward Motion estimates. (g) Examples of tagged axial and sagittal slices. (h) Example slices of estimation difference from HARP. Colorbar goes from 0 to 0.5 as in the previous figures. (i) Histogram of difference from HARP using IDEA and PVIRA.	104
6.1	The three principal directions of tongue motion during speech at two critical time frames /s/ and /k/.	109
6.2	(a) A high-resolution MR mid-sagittal image slice. (b) Manual segmentation of the tongue muscles in 3D. (c) Genioglossus muscle fiber direction. (d) Transversalis muscle fiber direction.	113
6.3	Strain in the line of action of muscle fibers of a healthy control. The top row is the segmented shape of GG (green) and T (blue) muscles at different time frames. In the bottom row, local compression of genioglossus is shown in green, and compression of transverse is shown in blue for the same time frames.	116
6.4	Strain in the line of action of muscle fibers of the first post-glossectomy patient. The top row is the segmented shapes and the bottom row is the local compression of muscles.	117
6.5	Strain in the line of action of muscle fibers of the second post-glossectomy patient. The top row is the segmented shapes and the bottom row is the local compression of muscles.	117
A.1	Cardiac MR image labeling for endocardium and RV insertion points. (a) Short-axis MR image of the heart. (b) Pixel labeling of the left ventricle chamber. (c) Labeling of right ventricle insertion points. (d) Level set representation of the left ventricle chamber contour (endocardium).	139
A.2	The consequence of poor bias initialization under two continuous label fusion scenarios. (a) Six raters identify RV insertion points (dots) and their fusions (crosses) are poor because CSTAPLE is initialized in upper left corner. Six raters identify endocardium contours (three shown in (b) and three more shown in (c)). The fusion of six endocardium contours shown in (d) is poor because CSTAPLE was initialized with zeros on the entire image plane.	147

LIST OF FIGURES

A.3	CSTAPLE simulation of 2D point identification. In (a) (d) circles are generated truth and dots are rater decisions. In (b) (e) “x” markers are the fusion of zero initialization, crosses are average initialization, and dots are informed initialization. In (c) (f) “x” markers are fusion of weak prior, crosses are data-adaptive prior, and dots are informed prior.	154
A.4	Identification of RV insertion points in cardiac MRI. In (a) (b) red dots are rater decisions and green circles are an expert’s decision as ground truth. Fusions are shown in yellow, where “x” markers, crosses and dots are respectively zero, average, informed initialization in (a) and weak, data-adaptive, informed prior in (b). The error distance of all fusion points from corresponding truth points shown in (c) which compares data-adaptive prior and informed prior methods.	158
A.5	Different fusion methods for endocardium contour identification. (a) Expert decision from manual delineation regarded as truth. (b) CSTAPLE with zero initialization. (c) CSTAPLE with average initialization. (d) CSTAPLE with informed initialization. (e) Classic STAPLE fusion of discrete labels. (f) MAP-CSTAPLE with weak prior. (g) MAP-CSTAPLE with data-adaptive prior. (h) MAP-CSTAPLE with informed prior.	160

Chapter 1

Introduction

1.1 Motivation

The basic functions of many human organs and muscles are associated with various motion and deformation. Measuring motion and strain of tissues during organ deformation has been an important topic in disease diagnosis, organ function study, and surgical planning. Its application involves cardiac imaging, lung motion modeling, speech and swallowing studies, analysis of brain motion in traumatic injuries, etc. A good understanding of the organ deformation from accurate motion estimation not only benefits medical science and clinical treatment of the human body, but also directly contributes to imaging technologies and image analysis methods development.

Take the human tongue for example. The tongue is highly deformable during speech and is able to perform precise motion over short periods of time because of its

CHAPTER 1. INTRODUCTION

complex internal muscle architecture [1,2]. Studying the motor control of the tongue muscles, including their interaction and cooperation for producing speech, has been an ongoing topic for oral surgeons, linguists, and speech-language pathologists. This is because the tongue, which is a volume-preserving structure devoid of bones and joints, moves entirely by deformation to create the shapes critical to speech, chewing, and swallowing [3]. Among various subjects, the motion of post-glossectomy patients, who have received surgery to remove a malignant section of their tongue, is of special interest as these patients have altered tongue morphology and may combine unusual motor control strategies for speech [4]. The National Cancer Institute estimates that 41,000 people in the US develop oral or oropharyngeal cancer per year and that 36% of these cancers are in the tongue [5]. Tumor size and location are the two most important factors that impact speech [4,6,7]. Therefore, understanding the motion differences made by healthy subjects and glossectomy patients will assist surgical decisions as well as speech and swallowing remediation.

The study of human brain deformation is another example. A rapid deformation of brain tissue in response to fierce linear or angular accelerations of the head can cause a major health concern called traumatic brain injury (TBI). TBI is not only a direct physical impairment to the human body, but can also lead to neuro-degenerative diseases. It is known that TBI is an established risk factor for Alzheimers disease [8, 9]. It is also reported that high incidence of memory impairment and depression is likely to occur on professional athletes with histories of repeated head impacts [10].

CHAPTER 1. INTRODUCTION

According to [11], 1.4 million TBI cases are reported approximately each year in the United States and the risk is likely to increase in the near future. As a result, understanding the mechanism of brain deformation and the relationship between the magnitude, location, and direction of the mechanical impact and the resulting neuro-degeneration has become an important goal in TBI related studies, in the hope of controlling the occurrence of TBI cases as well as reducing TBI-related injuries.

1.2 Anatomy of the Human Tongue and Brain

This dissertation focuses its applications to studying tongue motion in speech and brain deformation in TBI. Thus it is necessary to introduce the anatomy of the human tongue and brain.

A vital organ for chewing, swallowing, and assisting speech, the tongue is a complex muscular-hydrostat that consists of eight muscles in the mouth [12]. The genioglossus, hyoglossus, styloglossus, and palatoglossus muscles are four extrinsic muscles. They anchor the tongue on the bones and other external structures in the mouth and are capable of alter the tongue's position for protrusion, retraction, and sideways motion. The transversalis, verticalis, superior longitudinalis, and inferior longitudinalis are four intrinsic muscles hidden inside the tongue. They are responsible for lengthening, shortening, curling, and flattening of the tongue. The intrinsic and

CHAPTER 1. INTRODUCTION

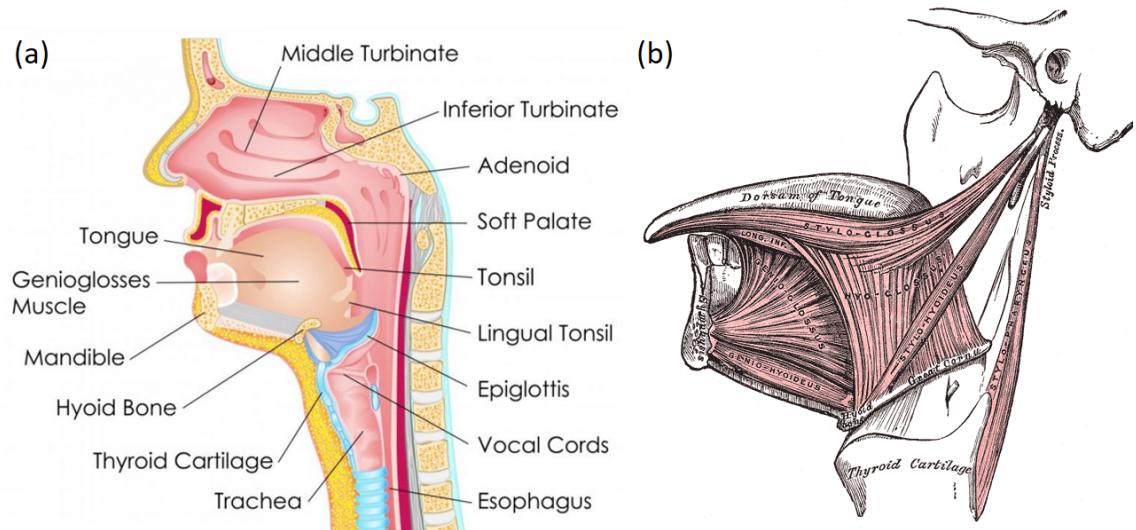


Figure 1.1: (a) The tongue and its external structures. Source: anatomybodypics.net. (b) Internal muscles of the tongue. Source: bartleby.com.

extrinsic muscles collaborate in an elaborate fashion to create various deformations during speech and swallowing. And these deformations are incompressible—meaning that any deformation of the tongue is volume-preserving [13]. Fig. 1.1 shows the anatomy of the tongue with its internal muscular structures.

The human brain is regarded as the most complex organ of the human body. Known as life’s “command center”, it is essential for maintaining the normal functionalities of the human body and has always been a common interest in medical studies. Fig. 1.2 shows the anatomy of the brain. The largest region of the brain is the cerebrum, which is divided into left and right hemispheres. The brain stem lies underneath the cerebrum and the cerebellum is located behind. The cerebral cortex is outermost layer of the cerebrum, which consists of four lobes with different func-

CHAPTER 1. INTRODUCTION

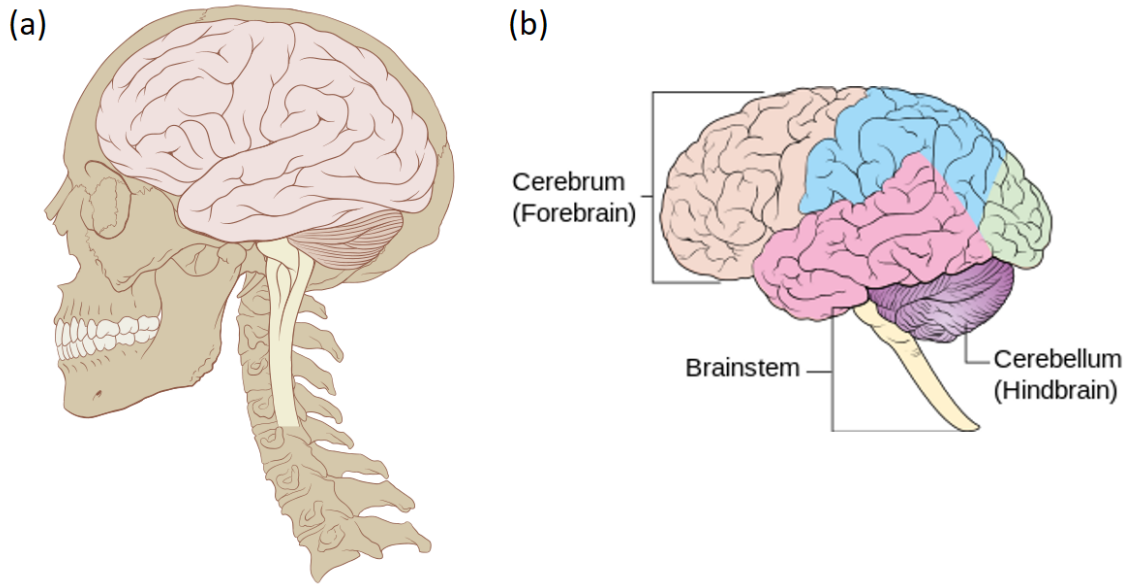


Figure 1.2: (a) Location of the brain in the skull. (b) Basic structure of the brain.
Source: [wikipedia.org/wiki/Human_brain](https://en.wikipedia.org/wiki/Human_brain).

tions [12]. In medical images, the interior of the brain consists of darker-colored areas (the grey matter) separated by lighter-colored areas (the white matter). Although the human brain is not muscular, cerebrospinal fluid (CSF) fills the brain ventricular system and the total brain volume is unlikely to change during motion caused by controlled deformations for the purpose of scientific studies. Therefore, it is safe to assume incompressibility for the imaging of brain motion under mild accelerations created inside a medical imaging scanner.

1.3 Tagged Magnetic Resonance Imaging

Motion estimation has always been a challenging task. This is because most tissue deformations happen in three dimensions, involve complex muscle structures, and are largely hidden within the internal organs. Magnetic resonance (MR) imaging is capable of revealing both anatomical structures and tissue motion. The internal motion of the tissues can be captured by tagged MR imaging [14–16]. By spatially modulating the magnetic field, magnetic tags can be placed within the tissue, and then deform with the tissue as it moves. Figs. 1.3(a) and 1.3(b) show an example of tagged images of the tongue. Such deformed tag patterns contain motion information which can be reconstructed by processing a sequence of tagged MR images acquired over time.

Although tagged MR images capture motion well, they provide only low resolution anatomical information at the tissue-air boundary. Therefore, in the present imaging protocols an additional image sequence, a cine MR sequence without tags, is collected at the same positions and same time instants during additional repetitions of the motion cycle, and these images can be used to segment the tissue region. Fig. 1.3(c) shows an example of cine image of the tongue.

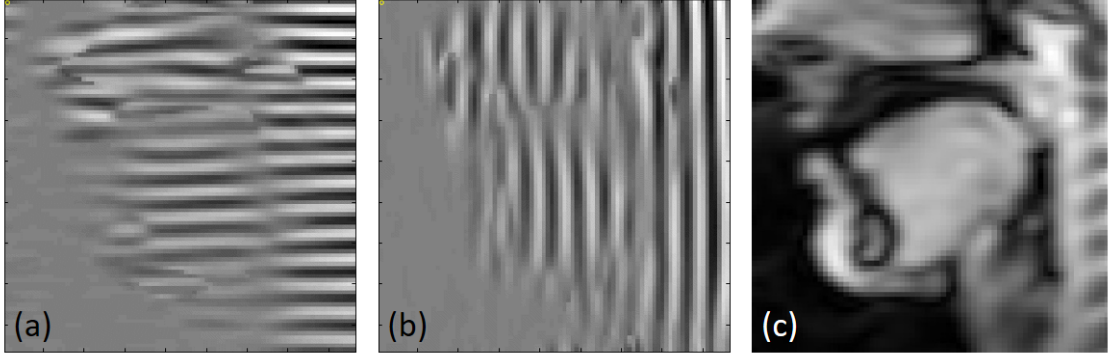


Figure 1.3: (a) Horizontally tagged mid-sagittal image slice of the tongue. (b) Vertically tagged image on the same slice location. (c) Cine image on the same slice location.

1.4 Motion Estimation and Challenges

Many methods for estimating two-dimensional (2D) tissue motion and strain from the deformed tag patterns have been proposed in the past, including tag line tracking [17, 18], tag intersection tracking [19–21], and pixel-wise tracking using harmonic phase or Gabor filters [22–25]. While measuring 2D motion alone might be sufficient in some applications, knowledge of three-dimensional (3D) motion might be necessary or highly desirable in others. Methods to measure 3D motion from densely-acquired tagged MR images have been proposed in the past. They require an acquisition of a large number of closely spaced image slices, which is equivalent to direct acquisition of a dense 3D volume. In this case, traditional 2D methods can be extended to 3D and directly applied to compute the dense motion [26–29]. However, the large amount of time that it takes to acquire these images makes this approach impractical for routine clinical or scientific use. For this reason, most of the reported 3D motion

CHAPTER 1. INTRODUCTION

estimation methods have been focused on the use of sparse collections of 2D images and 2D motion estimation followed by interpolation in order to achieve 3D motion estimation.

Incompressibility is an important consideration in the estimation of biological motions. For example, during a cardiac cycle, the change in volume of the myocardium is less than 4% [30, 31], and the change in volume of the tongue during speech and swallowing is even smaller [32]. Therefore, the motion of muscles should be assumed to be incompressible [13]. In the study of brain motion under mild accelerations, there is very little exchange of fluid in the ventricular system and the total brain volume including the ventricles remains very nearly constant [16]. Therefore, incompressibility can also be assumed in studies of small brain motions under mild accelerations created inside an MR scanner. In a word, motion estimation in many organs requires the final dense estimated motion field to be incompressible in order to represent tissue's true physical property.

Biological motions that are estimated from image data should also be inverse-consistent. That is, deformation fields estimated between two time frames should have an inverse deformation that reproduces the original image when reversing the order of the time frames. This property is not maintained, in general, by finding an inverse deformation field (i.e., the reference map) after finding the forward deformation field (i.e., the motion map). In motion estimation from tagged MR images, this is particularly problematic because the reference time frame, where the tags

CHAPTER 1. INTRODUCTION

are flat, is a unique configuration that is typically exploited in estimation of dense deformations—and its “flatness” property cannot be found in later times. Although inverse-consistency has been an important topic in image registration [33–35], it has not been properly discussed and implemented in the field of biological motion estimation. Thus numerical inversion methods are often used as a post-processing step if inverse fields are needed [36,37]. This requires extra computation time and can cause numerical errors on the boundary.

Previous works on 3D motion estimation from sparse imaging geometries are quite varied. Some use finite element or finite difference methods [38], some use tag line tracking based on 2D images [39, 40], some use spline interpolation [39, 41–43], and some use the organ’s biomechanical properties [44, 45]. More methods are summarized in the collected works of [46]. Because the computations are typically limited to the organ of interest, a 3D segmentation method is also required [38, 39, 41, 47]; this step requires either human intervention or automated segmentation algorithms [48] that increase the demand on processing time. In terms of incorporating incompressibility, because of the assumption of sparse imaging geometry, previous methods have been of the interpolation variety. The methods in [49, 50] use divergence-free radial basis functions while the methods in [43, 51] use smoothing vector splines seeking a divergence-free velocity field that yield a nearly incompressible deformation field when integrated. Unfortunately, these methods suffer from an over-complex computation due to vector interpolation, which can cause days of processing time for one single

CHAPTER 1. INTRODUCTION

subject with dozens of temporal volumes.

In contrast to the above methods, image registration algorithms based on dense imaging geometries can accommodate incompressibility in various ways. For example, the widely-adapted diffeomorphic demons algorithm [52] has been extended to include volume-preserving deformations [53, 54]. However, registration-based motion estimation methods have not been previously developed to work with the sparse imaging geometries of tagged MR images and with their unique brightness properties. In particular, the iLogDemons method was applied to the intensities of 3D tagged data [54]. But because tags fade with the T_1 time constant of the tissues, image brightness is not conserved over time and this will yield inaccurate displacement estimates [55]. Despite the number and diversity of previous approaches to process tagged data, there does not yet exist a well-accepted, high performance algorithm for estimating dense motion from the sparse imaging geometries of tagged images.

Beside the main motion estimation algorithm itself, various other aspects needed to be considered while designing a workflow of estimation methods that can be implemented in a user interface to process raw tagged data from the very beginning; these issues are all related to pre-processing or post-processing and need to be addressed with appropriate image analysis techniques involving a broad selection of topics. For example, in order to specify tissue region, manual or semi-automatic segmentation of multiple cine MR images in multiple time frames is a time-consuming task (a typical number of images is about 800 slices per subject). Efficient segmentation methods

CHAPTER 1. INTRODUCTION

must be developed. After obtaining estimates of the 3D motion from multiple subjects including healthy controls and patients, it remains challenging to analyze the resulting data in order to reveal population similarities and differences. The inconsistent tissue deforming rate between subjects is a major obstacle, so that a method is needed to provide a time alignment of the displacement data. The high dimensionality of the time sequences of 3D displacements is another obstacle; a method is needed to perform dimensionality reduction so that population analysis can be readily carried out. To analyze the relationship between motion and muscle activity, strain derived from the estimated motion needs to be studied. However, 3D strain has been hard to interpret, because internal muscles are mostly interdigitated and activate in different patterns to create a deformation, resulting in complex principal strain directions that are difficult to visually and quantitatively assess. This problem needs to be solved in order to trace the production of speech and motion patterns back to the activation of specific muscle groups.

1.5 Thesis Contributions

In this dissertation, we aim to develop efficient methods for 3D motion estimation and analysis using tagged MR images. The five major contributions of the thesis are summarized as follows.

1. A workflow of algorithms is developed to compute the 3D motion from tagged

CHAPTER 1. INTRODUCTION

and cine images of the tongue during speech. Named as the *tongue motion analysis pipeline* (TMAP), TMAP produces a sequence of estimated motion fields along with a multi-subject analysis of motion pattern differences between healthy control subjects and post-glossectomy patients using a novel two-step principal component analysis strategy. TMAP is based on a series of existing algorithms. It incorporates all of them as steps in a carefully optimized semi-automatic pipeline which also includes steps for time alignment and a population analysis. TMAP provides a usable tool for clinical studies and shows a capability in capturing healthy control and patient subjects' unique motion patterns. Its efficacy is demonstrated and the potential for quantitative motion analysis of the tongue is revealed.

2. An updated 3D motion estimation algorithm that is an enhanced version of the previously reported incompressible deformation estimation algorithm (IDEA) [43] is developed. The method, named E-IDEA, tackles the incompressible motion computation both on the internal tissue regions and tissue boundaries, reducing boundary errors and yielding a motion estimate that is more accurate on the whole. E-IDEA combines data from tagged images with surface deformation information derived from cine MR images to achieve a dramatical improvement. Deformable image registration techniques are applied to help extracting the motion on the boundary. Quantitative evaluations achieved by comparing with 2D harmonic phase results show that E-IDEA efficiently handles the well-known

CHAPTER 1. INTRODUCTION

boundary effects from traditional harmonic phase motion estimation.

3. A novel 3D motion estimation algorithm is proposed as a substitute for the previous methods. Based entirely on the imaging registration framework, the proposed method, called *phase vector incompressible registration algorithm* (PVIRA), is a faster and more robust method that performs phase tracking in a novel way. PVIRA uses several concepts from diffeomorphic registration with a key novelty that defines a symmetric similarity metric enabling the simultaneous use of multiple phase volumes. The new metric solves the aperture problem of optical flow by exploiting the 3D material properties of harmonic phase. Evaluations show a large reduction in computation time over E-IDEA and better robustness against noise. PVIRA is demonstrated to yield a nearly incompressible result and to produce motion and inverse motion fields that are very nearly inverse-consistent.
4. A method to reveal muscle's activities using strain in the line of action along the fiber directions is proposed. Both tagged MR images and static high-resolution MR images are used to obtain motion and muscle anatomy, respectively. Strain tensors and local tissue compression along the muscle fiber directions are computed in order to reveal their shortening pattern, which is then linked with clinical data to suggest a possible pattern of muscle activation. The method is capable of relating motion to particular muscle functions, which is a first step in

CHAPTER 1. INTRODUCTION

learning individual muscle activity from tagged images and may provide novel information for future clinical and scientific studies.

5. In statistical label fusion of medical images, labels of different regions and structures are often specified by a number of different human raters or automated algorithms. In this contribution, label fusion in the continuous domain is studied, and a bias problem in the continuous *simultaneous truth and performance level estimation* (STAPLE) algorithm is revealed. We prove that bias and variance used to represent rater performance yield a maximum likelihood solution in which bias is indeterminate. We analyze the cause of the deficiency and resolve this ambiguity by proposing two classes of solutions, one that estimates the bias as part of the algorithm initialization and the other that uses a maximum a posteriori criterion with a priori probabilities on the rater bias. Since this contribution is not directly related to motion estimation—the main theme of the dissertation, we put its description in the appendix.

1.6 Thesis Organization

This thesis is organized as follows. Chapter 2 introduces background knowledge on MR data acquisition, the harmonic phase method, the IDEA method, and strain tensor calculation. In Chapter 3, we describe the tongue motion analysis pipeline TMAP used to establish a first step to 3D motion estimation. In Chapter 4, we

CHAPTER 1. INTRODUCTION

describe the enhance IDEA method that addresses the boundary problem in the previous pipeline and IDEA estimation. In Chapter 5, we present the novel method PVIRA for fast and robust motion estimation. In Chapter 6, we describe the method for revealing the relationship between motion and muscle activities. Finally, Chapter 7 concludes the thesis and discusses future work. In the Appendix, we present the investigation on continuous label fusion.

Chapter 2

Background

In this chapter, we provide a general overview of MR data acquisition and a brief introduction to related algorithms. In Section 2.1, we introduce the scanner setup and imaging parameters of the tongue and brain data used in this dissertation. Section 2.2 describes the standard harmonic phase method. In Section 2.3, we review the incompressible deformation estimation algorithm. Section 2.4 reviews the definition of deformation gradient tensors and strain tensors. Finally, Section 2.5 summarizes this chapter.

2.1 Data Acquisition

To study tongue motion in speech, a dataset consisting of both healthy control subjects and post-glossectomy patients was used. The subjects who participated in

CHAPTER 2. BACKGROUND

the study were consented under IRB approved protocols. Some patients had small (T1) tumors and other patients had mid-size (T2) tumors. The T1 tumors were removed with glossectomy and the wound closed by sutures. The T2 tumor was closed by adding external tissue, a radial-forearm free flap, to replace the resected mass. The speech material was designed into a phrase “a souk”. Using international phonetic alphabet (IPA), the phrase is pronounced /ə’suk/, which was used to elicit specific tongue motions. The phrase started with /ə/, a centralized tongue position, moved into /s/, where forward tongue motion was expected to be prominent, and ended with /k/, where upward tongue motion was expected to be prominent. Because subjects needed to repeat the speech materials to a metronome in the MR scanner, prior to entering the scanner, they were trained to speak to the same metronome beat for about 15 minutes. The 4 beat metronome sequence represented the two syllables of the speech task, plus an inhalation and exhalation. Thus, every motion of the oral cavity was timed as precisely as possible.

Fig. 2.1 shows the imaging geometry for slice acquisition of the tongue motion in speech [15]. MR scanning was performed on a Siemens 3.0T Tim Trio system (Siemens Medical Solutions, Malvern, PA) with a 12-channel head coil and a 4-channel neck coil using a segmented gradient echo sequence. The field of view was $240\text{ mm} \times 240\text{ mm}$ with an in plane resolution of $1.87\text{ mm} \times 1.87\text{ mm}$ and a slice thickness of 6 mm. Each data set contained a sagittal, coronal, and axial stack of images encompassing the tongue and surrounding structures. The image sequence was obtained into 26

CHAPTER 2. BACKGROUND

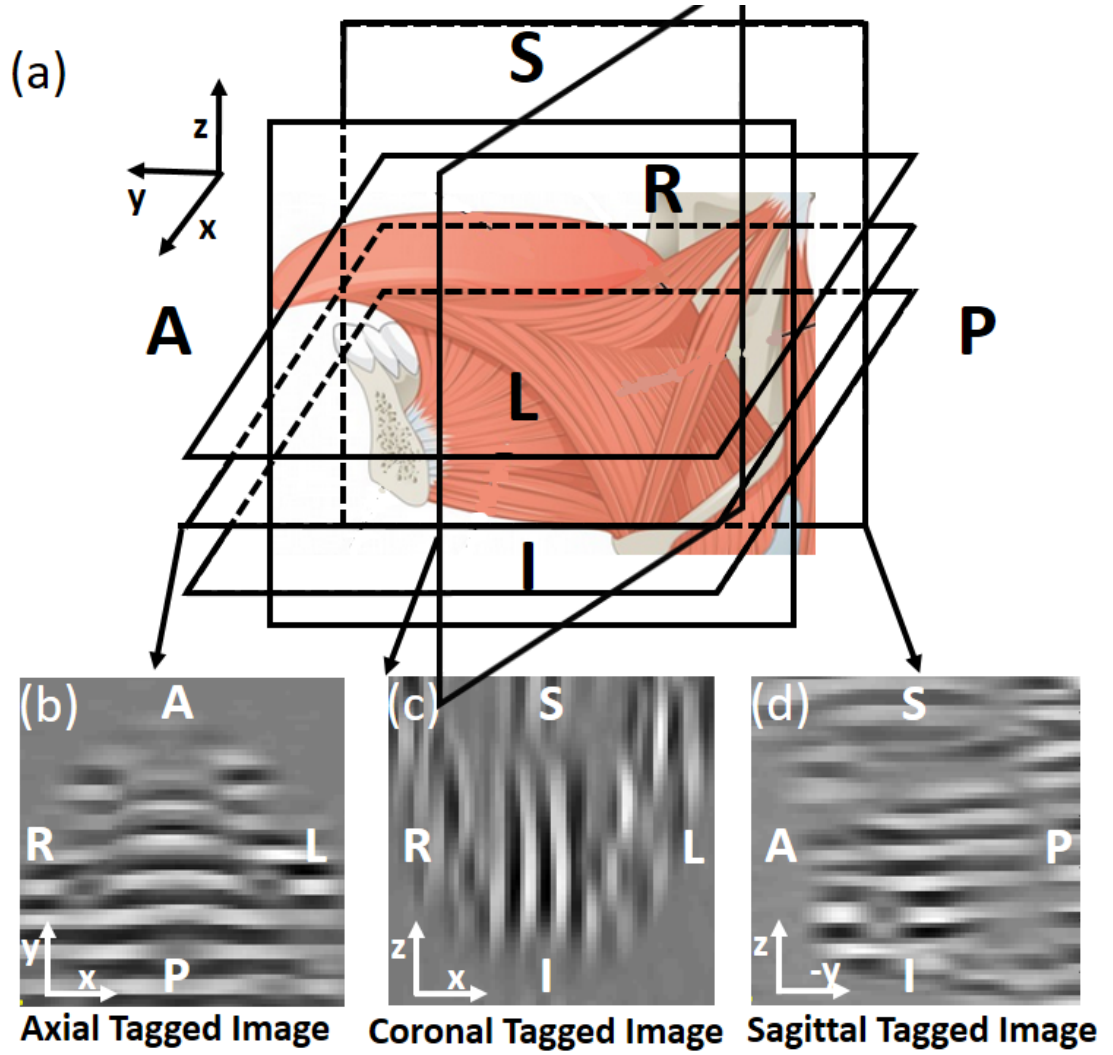


Figure 2.1: Illustration of the acquisition setup of tagged MR image slices for tongue motion in speech. A: anterior. P: posterior. L: left. R: right. S: superior. I: inferior. (a) Slice locations. (b) An axial tagged slice. (c) A coronal tagged slice. (d) A sagittal tagged slice.

CHAPTER 2. BACKGROUND

time frames that spanned for about one second. The complete MR recording session lasted for about an hour. The MR imaging machine collected a very weak signal, namely the amount of hydrogen protons in each unit of tissue. Therefore, multiple repetitions of the speech utterance needed to be collected and summed to generate a single time-series showing tongue motion. To sum the cine MR data collection, five repetitions were needed per slice. For the tagged data collection, three repetitions were needed per slice. In addition, data for each tag direction (superior–inferior, anterior–posterior, left–right) were collected twice, once for a sinusoidal tag pattern and once for another sinusoidal pattern on a 180 degree phase shift [56]. The number of slices depended on the size of the subject’s tongue and ranged as follows: sagittal—5 to 9 slices; coronal—10 to 14 slices; axial—10 to 14 slices. Pauses were allowed after each set of slices, so that consecutive acquisitions contained 15 to 42 repetitions for tagged acquisitions and 25 to 70 repetitions for cine acquisitions [15].

To study the brain deformation, we imitate an acceleration about 10–15% during a soccer player’s heading ball motion. A mild angular acceleration was generated from healthy subjects who participated in the study. The subjects were consented under IRB approved protocols. Each subject, lying down in a Siemens 3.0T mMR Biograph scanner (Siemens, Munich, Germany), was constrained with a head rotation device that accelerates the head towards the left shoulder direction [16]. The head rotated around the body’s longitudinal axis (from head to foot), generating a brain deformation in a “shaking head” movement. In repeated motions, tagged images were

CHAPTER 2. BACKGROUND

acquired with a SPAMM pulse sequence into sparse parallel slices covering the brain region and spanning across 12 time frames (resolution: 1.5 mm in-plane and 8.00 mm through-plane).

Fig. 2.2 shows the imaging geometry for slice acquisition of the brain during mild acceleration [16]. The slices are taken to cover the whole brain. Similar to the tongue application, due to time constraints the slices are sparsely positioned. In each axial image plane, two sets of images are taken in the same location, one with horizontal tags and the other with vertical tags. In this way, the motion in the x and y directions can be observed in these two sets. In order to record the z motion, sagittal slices with horizontal tags are also acquired. For each slice position, MR tags are placed at a reference time and a sequence of images are acquired over time in order to reveal the motion as a deforming tag pattern.

Because of the acquisition geometry, the available motion data in both scenarios are sparse in the through-plane direction. As a result, estimation of dense 3D motion from these data can be summarized as interpolating information gleaned from sparsely acquired image data. In particular, motion features that are observed in the acquired image planes must be interpolated in some way in order to observe dense, 3D motion.

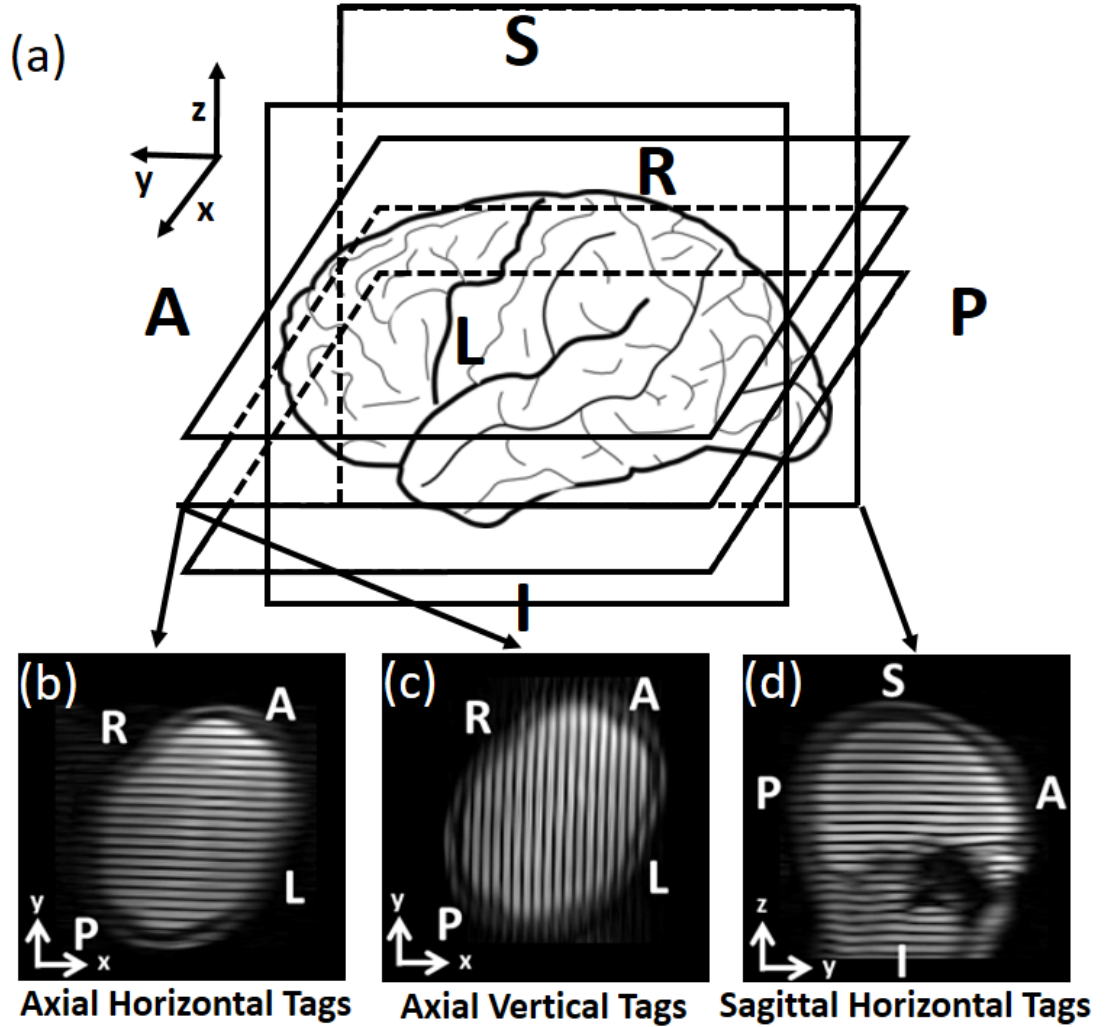


Figure 2.2: Illustration of the acquisition setup of tagged MR image slices for brain motion. A: anterior. P: posterior. L: left. R: right. S: superior. I: inferior. (a) Slice locations. (b) An axial slice with horizontal tags. (c) An axial slice with vertical tags. (d) A sagittal tagged slice.

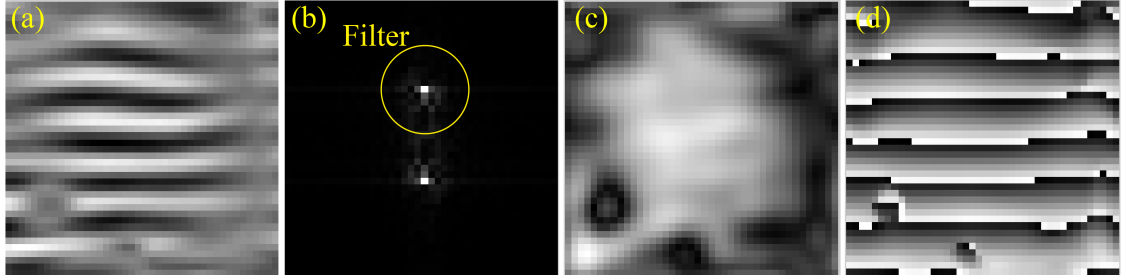


Figure 2.3: Processing of a tagged MR image. (a) Acquired tagged slice. (b) Harmonic peaks in the Fourier domain and a HARP filter. (c) Harmonic magnitude slice. (d) Harmonic phase slice.

2.2 Harmonic Phase Algorithm

The harmonic phase (HARP) algorithm is a standard, well-known method to process and analyze tagged MR images [23]. It has been applied in applications for the heart [14], the brain [15], the tongue [16], and other areas of the body [57–59], and has become an established tool for scientific use and clinical protocols. The reason that HARP is often preferred over other methods is because it is fast and accurate with an estimation error less than about a third of a pixel’s size. HARP is normally considered to include two steps: a HARP filtering step and a HARP tracking step. HARP filtering operates in the Fourier domain and aims at extracting a harmonic phase image carrying motion information, and HARP tracking estimates motion based on the deformation of the harmonic phase image.

Fig. 2.3 shows an tagged image processing by HARP. In particular, one of the major harmonic peaks in the Fourier domain of a tagged image slice can be filtered (so-called HARP filtering) to yield a complex-valued image, where motion information

CHAPTER 2. BACKGROUND

is contained in the phase part (HARP phase) and anatomical information is contained in the magnitude part (HARP magnitude). Mathematically, if we denote the complex image after HARP filtering as $J(\mathbf{x})$, it can be written as

$$J(\mathbf{x}) = M(\mathbf{x})e^{j\Phi(\mathbf{x})}, \quad (2.1)$$

where $M(\mathbf{x})$ is the HARP magnitude slice and $\Phi(\mathbf{x})$ is the HARP phase slice.

To extract the 2D motion contained in the phase images, two slices $\Phi_h(\mathbf{x})$ and $\Phi_v(\mathbf{x})$ with horizontal tags and vertical tags are used; each of them records motion in the normal direction of its tag plane. Denote $\Theta(\mathbf{x}) = [\Phi_h(\mathbf{x}), \Phi_v(\mathbf{x})]^T$, HARP tracking follows such an assumption that for a material point located at $\mathbf{x}(t_i)$ at time frame t_i , its HARP phase value Θ stays the same over time. As a result, at the next time frame t_{i+1} , $\mathbf{x}(t_{i+1})$ can be found by

$$\Theta(\mathbf{x}(t_i), t_i) = \Theta(\mathbf{x}(t_{i+1}), t_{i+1}). \quad (2.2)$$

In practice, this equation can be solved iteratively using the Newton-Raphson method [60] as follows:

$$\mathbf{x}^{(n+1)}(t_{i+1}) = \mathbf{x}^{(n)}(t_{i+1}) - \nabla \Theta(\mathbf{x}^{(n)}(t_i), t_i)^{-1} (\Theta(\mathbf{x}^{(n)}(t_i), t_i) - \Theta(\mathbf{x}(t_i), t_i)). \quad (2.3)$$

HARP processes each pixel independently, and the result is a sequence of tracked

CHAPTER 2. BACKGROUND

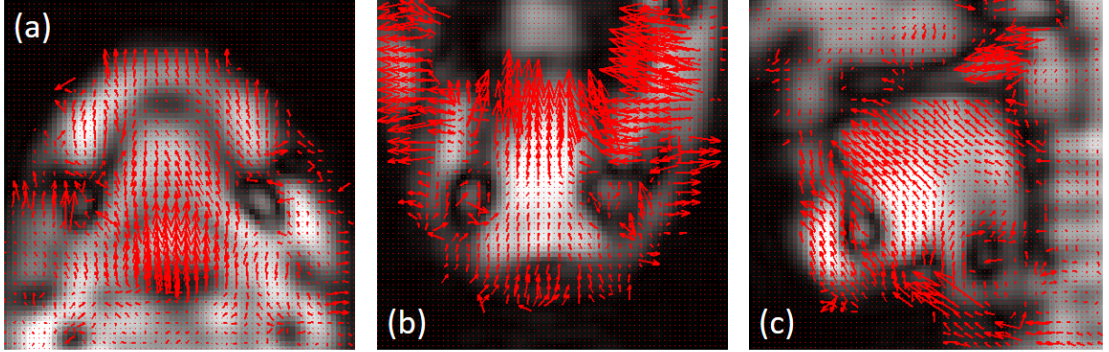


Figure 2.4: Resulting 2D motion fields from HARP on (a) an axial slice, (b) a coronal slice, and (c) a sagittal slice of the tongue.

points over all time frames (see Fig. 2.4). Their displacements can be computed by subtraction between neighboring point locations and a sequence of 2D in-plane displacement fields are achieved. Due to phase wrapping [23, 61], HARP may fail by tracking a “jumped” tag when a tissue point makes large movements commonly seen in organs like the tongue. A HARP refinement can be used to prevent this problem. In the works for this dissertation, we use the shortest-path HARP refinement method [61] to reduce this type of error.

While originally developed to analyze 2D images, the HARP framework is valid for 3D images, and has been previously used for 3D images which were acquired on a dense grid [26]. The only changes are to switch the 2D circular HARP filter into a 3D ball-shaped HARP filter and to use three tag directions in Eqn. 2.3.

CHAPTER 2. BACKGROUND

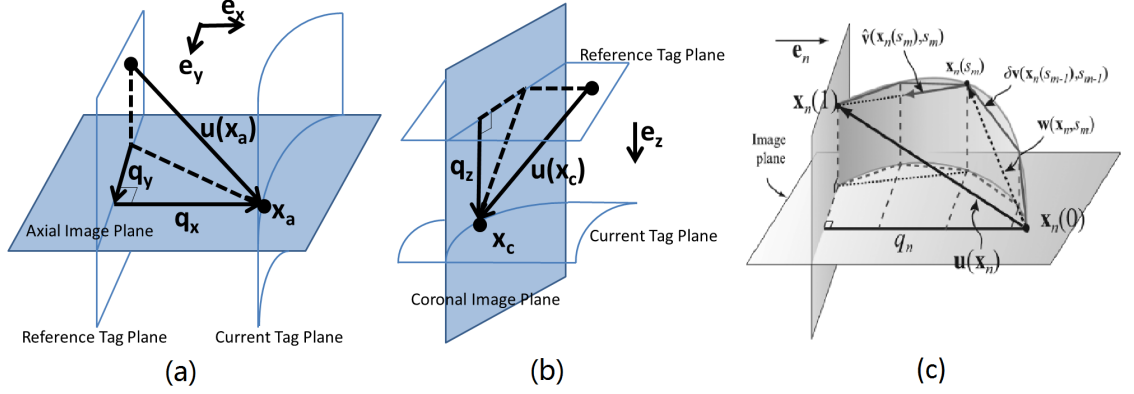


Figure 2.5: Relationship between 2D motion components and 3D IDEA estimation on (a) a slice with x and y motion components (b) a slice with z motion component. (c) shows the integration process of the estimated velocity components by IDEA (Source: Liu et al. Incompressible deformation estimation algorithm (IDEA) from tagged MR images. *IEEE TMI*, 2012.).

2.3 Incompressible Deformation Estimation Algorithm

The 2D motion on slices obtained from tagged data can be viewed as multiple observations about the underlying 3D motion. However, each observation is an in-plane projection of the true 3D motion and it is spatially sparse due to the low through-plane resolution. The incompressible deformation estimation algorithm (IDEA) incorporates these sparse and incomplete projections as well as the incompressibility constraint by using a divergence-free vector spline to interpolate a 3D dense motion estimate [43]. Specifically, it reconstructs a sequence of divergence-free velocity fields over small time steps so that the integrated 3D velocities yield the observed 2D HARP displacements to good approximation.

CHAPTER 2. BACKGROUND

Fig. 2.5 illustrates how 2D in-plane motions are processed by IDEA. The undeformed tissue at the first time frame has undeformed reference tag planes. At a current time frame, the tag planes have deformed along with the tissue. Suppose axial slices contain x and y motion components and coronal slices contain z motion component from 2D HARP, to each point (pixel location) \mathbf{x}_a on an axial image such as Fig. 2.5(a), HARP produces two vectors representing components of displacement:

$$\begin{cases} \mathbf{q}_x = q_x \mathbf{e}_x , \\ \mathbf{q}_y = q_y \mathbf{e}_y , \end{cases} \quad (2.4)$$

where \mathbf{e}_x and \mathbf{e}_y are unit vectors in the x and y directions and \mathbf{q}_x and \mathbf{q}_y are the projections of the 3D motion $\mathbf{u}(\mathbf{x}_a)$ on the current axial plane. Similarly, for each point \mathbf{x}_c on a coronal image such as Fig. 2.5(b), HARP yields the displacement component vector

$$\mathbf{q}_z = q_z \mathbf{e}_z , \quad (2.5)$$

where \mathbf{e}_z is the unit vector in the z direction. IDEA takes such data on all pixels $\{\mathbf{x}_a, \mathbf{q}_x(\mathbf{x}_a), \mathbf{x}_a, \mathbf{q}_y(\mathbf{x}_a), \mathbf{x}_c, \mathbf{q}_z(\mathbf{x}_c)\}$ as input, and estimates an incompressible deformation field $\mathbf{u}(\mathbf{x})$ on a high-resolution grid within the tissue mask. We note two important aspects. First, IDEA is carried out as a series of smoothing splines, each of which seeks a divergence-free velocity field yielding the deformation field only when integrated (as shown in Fig. 2.5(c)). Thus the final field $\mathbf{u}(\mathbf{x})$ is nearly incompress-

CHAPTER 2. BACKGROUND

ible and its re-projected components at all input points approximately agree with the input measurements. Second, the inputs are observed components of displacements that can arise at any physical position and in any direction of motion.

In Chapter 3, IDEA’s framework is used as the basis of a proposed 3D motion estimation pipeline. In Chapter 4, we will extend IDEA to incorporate boundary information, and in Chapter 5, we will propose a new motion estimation method. IDEA is used as a baseline method to compare in both of these chapters.

2.4 Strain Tensor Calculation

The strain tensor is a commonly used metric to measure a motion field’s local feature and to better characterize deformation and tissue function. It is capable of quantifying how much a deformation differs locally from a rigid-body displacement and computing the amount of stretch and compression along any specific motion direction. The strain tensor is closely related to the deformation gradient tensor. We briefly review their definitions in this section.

Fig. 2.6 shows the deformation of a continuum body. Denote a material point as \mathbf{X} , which deforms to location \mathbf{x} under a displacement $\mathbf{u}(\mathbf{X})$. Thus $\mathbf{x} = \mathbf{X} + \mathbf{u}(\mathbf{X})$. The deformation gradient tensor \mathbf{F} is defined as

$$\mathbf{F} = \frac{d\mathbf{x}}{d\mathbf{X}} = \mathbf{I} + \frac{d\mathbf{u}}{d\mathbf{X}}. \quad (2.6)$$

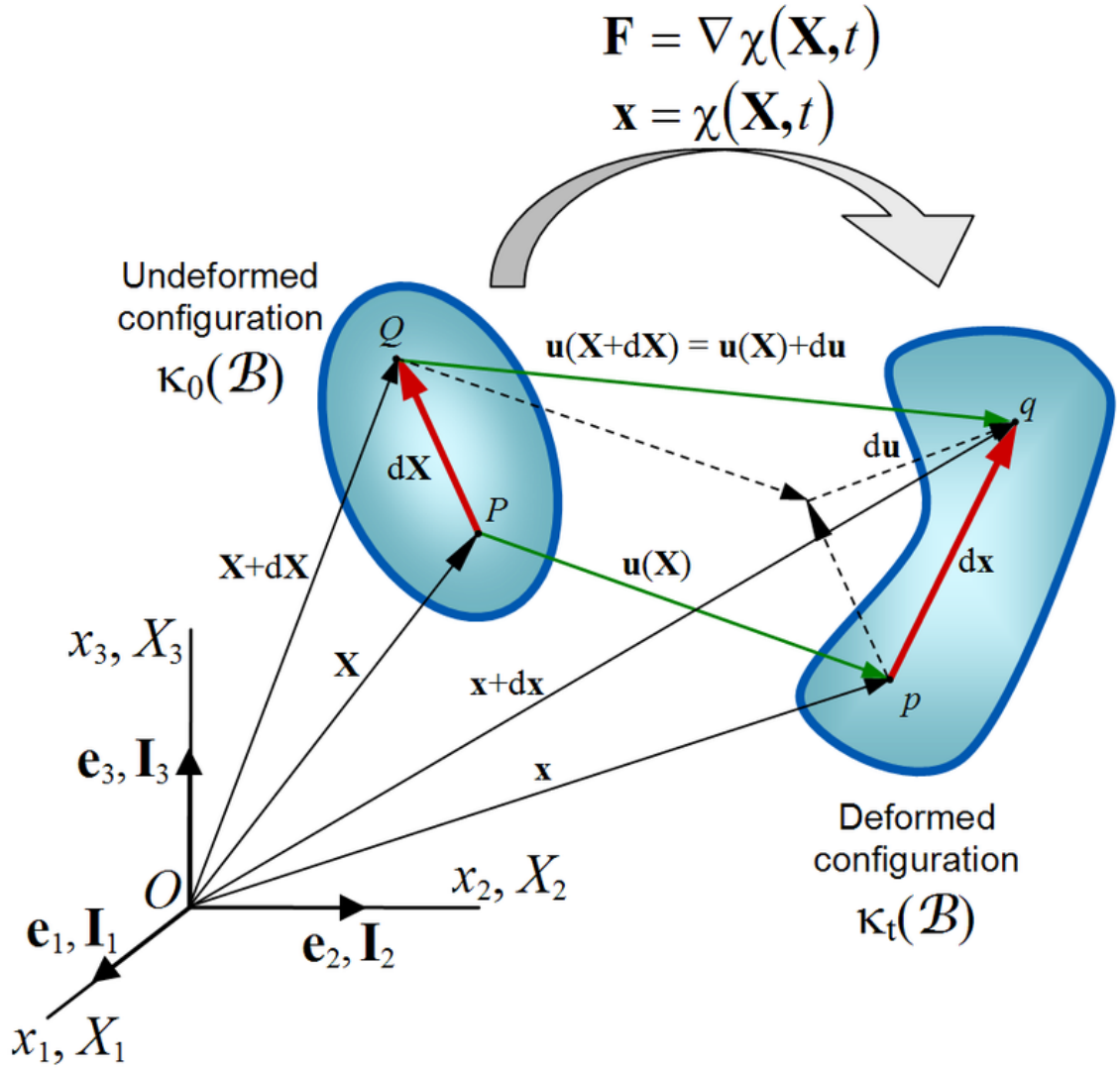


Figure 2.6: Deformation of a continuum body. Source: wikipedia.org.

CHAPTER 2. BACKGROUND

This tensor measures the local ratio of stretch or compression in all directions. Note that the explicit form of a tensor is a matrix. In this case, F is of dimension 2×2 for 2D data and 3×3 for 3D data.

Since the deformation gradient tensor involves a transformation between two coordinate systems, namely the material domain and the deformed domain, a rotation is accounted for in the transformation of coordinates. However, in a deformable body, a pure rotation is not capable of inducing any stresses or strains, and it is convenient to use other tensors independent of rotation to measure the deformation. Thus, the strain tensors are defined. In the material coordinate system \mathbf{X} , the *Lagrangian strain tensor* is defined by

$$E = \frac{1}{2}(F^T F - I) = \frac{1}{2}\left(\frac{d\mathbf{u}}{d\mathbf{X}} + \left(\frac{d\mathbf{u}}{d\mathbf{X}}\right)^T + \left(\frac{d\mathbf{u}}{d\mathbf{X}}\right)^T \left(\frac{d\mathbf{u}}{d\mathbf{X}}\right)\right). \quad (2.7)$$

In the deformed coordinate system \mathbf{x} , the *Eulerian strain tensor* is defined by

$$e = \frac{1}{2}(I - F^{-T} F^{-1}) = \frac{1}{2}\left(\frac{d\mathbf{u}}{d\mathbf{x}} + \left(\frac{d\mathbf{u}}{d\mathbf{x}}\right)^T - \left(\frac{d\mathbf{u}}{d\mathbf{x}}\right)^T \left(\frac{d\mathbf{u}}{d\mathbf{x}}\right)\right). \quad (2.8)$$

The two strains are used to characterize deformations in the corresponding coordinated system of interest. In practice, these tensors can also be projected onto the directions of interest to compute the deformation along a certain direction, such as the circumferential strain and radial strain in cardiac motion studies [62].

2.5 Summary

In this chapter, we briefly introduced some background knowledge on the tagged data acquisition, the HARP method that computes 2D motion, and the IDEA method for 3D incompressible motion estimation. We also reviewed the definition of strain tensor and the equations to compute it from motion fields.

Chapter 3

Three Dimensional Motion

Analysis Pipeline

3.1 Introduction

As a first step, we aim to develop a straightforward pipeline of algorithms to achieve motion estimation of the tongue data in speech and target a few immediate difficulties mentioned in Section 1.4. We focus on reducing the labor of manually segmenting hundreds of cine MR slices in multiple time frames and on resolving the difficulty of providing a time alignment of the displacements to regularize multi-subject data. We also aim to develop a method to perform dimensionality reduction on healthy control groups and patient groups to analyze the resulting data in order to reveal population similarities and differences.

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

We base the pipeline on some well-established algorithms: the HARP algorithm [23] for computing 2D motion from tagged data (see Section 2.2), the IDEA method [43] for estimating 3D motion from the HARP result (see Section 2.3), the super-resolution methods [63] for building 3D tongue volumes, segmentation algorithms for providing automatic or semi-automatic labeling of the tongue region [47], and principal components analysis (PCA) for dimensionality reduction of populations of vector fields [64]. The proposed method, called TMAP for *Tongue Motion Analysis Pipeline*, incorporates all of these steps in a carefully optimized semi-automatic pipeline which also includes steps for time alignment and a population analysis. TMAP is evaluated with a dataset consists of sixteen healthy controls and post-glossectomy patients in a specific speech task. It is intended to be used as a tool and feasible for speech analysts to process their data with.

As illustrated in Fig. 3.1, the input to TMAP is tagged and cine MR data. The output is the 3D motion field and a multi-subject statistical analysis result. The critical steps we have developed are described in the following sections.

3.2 Phase Calibration of CSPAMM Images

Before any motion extraction can be performed, the first problem to be solved is to find a way to correctly read the raw tagged images. As mentioned in Section 2.1,

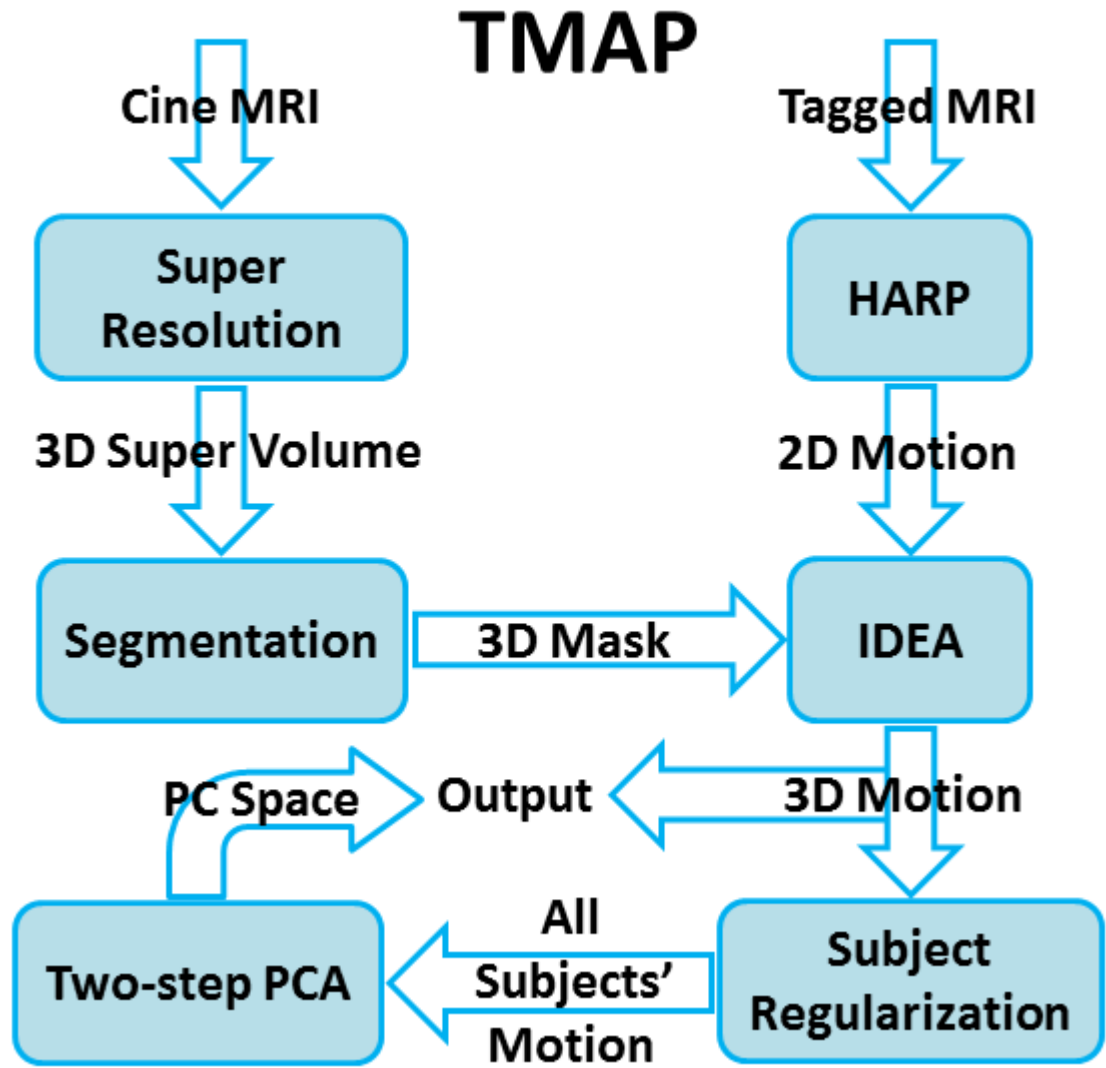


Figure 3.1: Flowchart of the tongue motion analysis pipeline (TMAP).

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

in order to eliminate DC components, the tagged tongue data is acquired at a pair of sinusoidal tag patterns with a 180 degree phase difference. The two images in the pair, denoted as $A(x, t)$ and $B(x, t)$, are modeled as a function of spatial location x and time t by

$$\begin{aligned} A(x, t) &= M_0(1 - (1 - \cos \frac{2\pi x}{P})e^{-t/T_1}), \\ B(x, t) &= M_0(1 - (1 - \cos(\frac{2\pi x}{P} - \pi))e^{-t/T_1}), \end{aligned} \quad (3.1)$$

where P is the tag period, M_0 is the magnetization magnitude, and T_1 is the relaxation time constant [56]. Fig. 3.2(a) shows the tag's wave patterns in the A and B images. In practice, since the acquired data is stored in two parts—the magnitude part and the phase part—the actual images directly from the MR scanner are complex representations of A and B denoted as

$$\begin{aligned} A_c(x, t) &= |M_0(1 - (1 - \cos \frac{2\pi x}{P})e^{-t/T_1})| \cdot e^{j\beta_A(x, t)}, \\ B_c(x, t) &= |M_0(1 - (1 - \cos(\frac{2\pi x}{P} - \pi))e^{-t/T_1})| \cdot e^{j\beta_B(x, t)}, \end{aligned} \quad (3.2)$$

where $|M_0(1 - (1 - \cos(2\pi x/P))e^{-t/T_1})|$ and $|M_0(1 - (1 - \cos(2\pi x/P - \pi))e^{-t/T_1})|$ are the acquired magnitude parts of A and B images and $\beta_A(x, t)$ and $\beta_B(x, t)$ are the acquired phase parts. Here we use the symbol β in order to distinguish the acquired phase from HARP phase. In an ideal situation, β_A and β_B are constants independent of spatial position x and time t . However, whether constant or not, they can be used to recover the original A and B images from A_c and B_c , as we now discuss.

It is known that tags fade with the T_1 time constant in a very short period af-

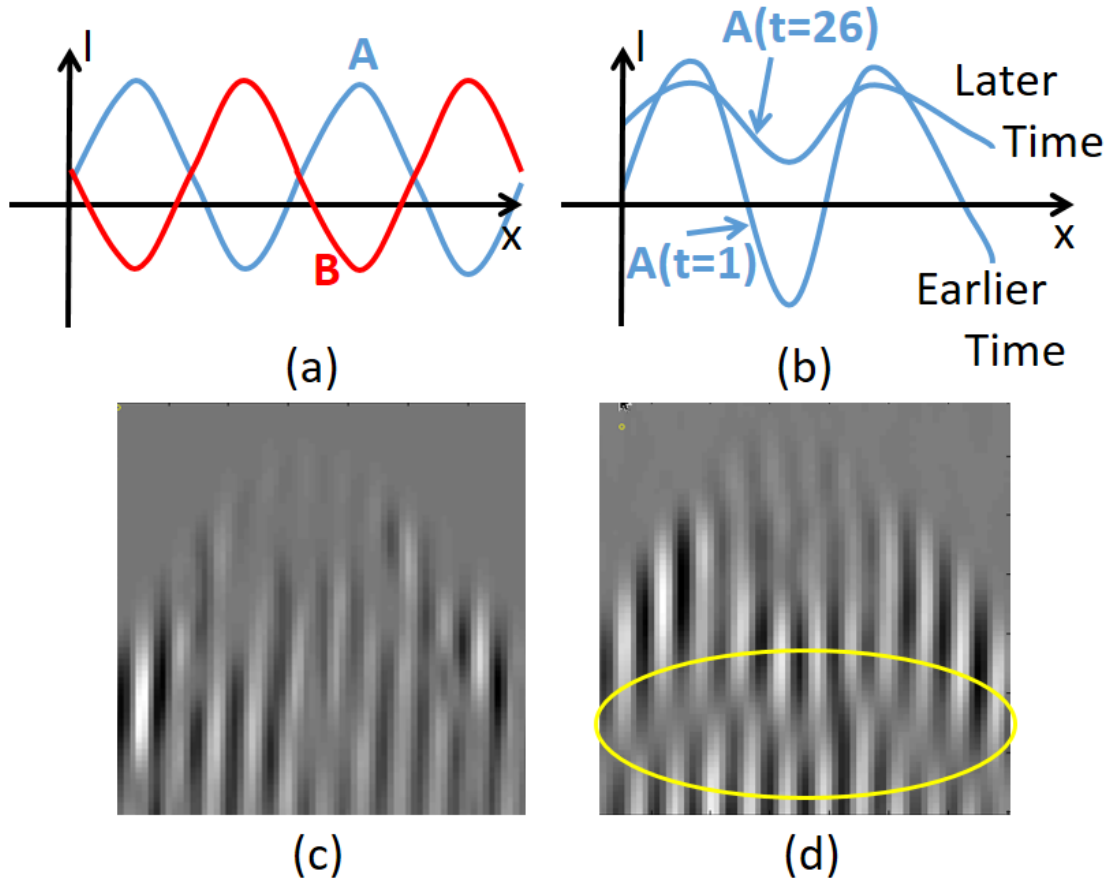


Figure 3.2: Demonstration of phase of tagged images from acquisition. (a) Phase of A and B images. (b) Demonstration of tag fading on an A image. (c) A MICSR image on the first time frame. (d) A CSPAMM image on the first time frame.

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

ter they are initially applied, i.e., their amplitude becomes smaller over time (see Fig. 3.2(b)). To make time for capturing more frames during the tongue's motion in the entire one-second speech task, a 180-degree MR tip angle is used for the tags during the scan session. This causes a problem because the tip angle crosses 90 degrees and the magnitude part of A_c and B_c images in Eqn. 3.2 becomes negative. To recover the original A and B images from A_c and B_c , we need to use the phase values β_A and β_B to find the locations where the magnitude of A_c and B_c should be flipped with a negative sign. At a pixel where its tag value is negative, its phase value β_A or β_B makes a sudden jump of π , which is a criterion that can be used to pinpoint the location. Therefore, considering Eqns. 3.1 and 3.2, A and B images can be recovered by

$$A(x, t) = \begin{cases} |A_c(x, t)|, & \text{if } -\frac{\pi}{2} \leq \beta_A(x, t) < \frac{\pi}{2}, \\ -|A_c(x, t)|, & \text{otherwise.} \end{cases} \quad (3.3)$$

and

$$B(x, t) = \begin{cases} |B_c(x, t)|, & \text{if } -\frac{\pi}{2} \leq \beta_B(x, t) < \frac{\pi}{2}, \\ -|B_c(x, t)|, & \text{otherwise.} \end{cases} \quad (3.4)$$

There are three methods to read A and B images for motion estimation [56]. The SPAMM method directly uses only one image of A and B . However, SPAMM's DC component in the Fourier domain limits the radius of the followed HARP filter, causing a loss of high-frequency information. Thus the CSPAMM method or the MICSRR method can be used to eliminate the DC component by combining A and B

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

images to yield a new tagged image. The definitions of CSPAMM and MICSR images are

$$\text{CSPAMM} = A(x, t) - B(x, t) = 2M_0 e^{-t/T_1} \cos \frac{2\pi x}{P} \quad (3.5)$$

and

$$\text{MICSR} = |A(x, t)|^2 - |B(x, t)|^2 = 4M_0^2 e^{-2t/T_1} (1 - e^{-t/T_1}) \cos \frac{2\pi x}{P}. \quad (3.6)$$

Apparently, MICSR is preferred in this case, because Eqn. 3.6 only uses the magnitude part of A and B images, which can be directly replaced by the magnitude of A_c and B_c images because $|A| = |A_c|$ and $|B| = |B_c|$ according to Eqns. 3.3 and 3.4. Thus, MICSR saves the step of recovering A and B using Eqns. 3.3 and 3.4. Fig. 3.2(c) shows a MICSR image. Although MICSR provides a feasible result, from Eqn. 3.6 we can see that MICSR suffers from a low signal at the initial time ($t = 0$); in fact, its magnitude is close to zero with very low contrast. Due to the characteristics of TMAP, all motions need to be calculated with respect to the first time frame. As a result, CSPAMM with its stronger signal and higher contrast in the first time frame should be used (Fig. 3.2(d)).

In order to use CSPAMM, the recovery of A and B using Eqns. 3.3 and 3.4, which require a check on the values of β_A and β_B , is essential. However, instead of being constant, the actual β_A and β_B values are inconsistent; they shift overtime and are not spatially homogeneous. If used directly, they can yield broken tags that have a sudden flip of sign in some regions, as demonstrated within the ellipse in Fig. 3.2(d).

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

To eliminate this inconsistency, we introduce a *phase calibration* preprocessing step.

We note that due to tag fading, in the last time frame of data acquisition, the tags have lasted for about one second and their amplitude has reduced to a very small value. The tip angle no longer crosses 90 degrees and the magnitude of all pixels stays positive (see Fig. 3.2(b)). As a result, the phase values $\beta_A(x, 26)$ and $\beta_B(x, 26)$ at the last (26th) time frame can be considered as the bias of all previous time frames due to phase inconsistency. Removal of this bias corrects all previous phases. Thus, for all pixels and all time frames the phase calibration is performed by

$$\begin{aligned}\beta'_A(x, t) &= W(\beta_A(x, t) - \beta_A(x, 26)), \\ \beta'_B(x, t) &= W(\beta_B(x, t) - \beta_B(x, 26)),\end{aligned}\tag{3.7}$$

where $W(\cdot)$ is a wrapping operator defined by $W(\beta) = \text{mod}(\beta + \pi, 2\pi) - \pi$ to wrap the phase back into the range of $[-\pi, \pi)$. The calibrated phases β'_A and β'_B are used in the following phase calibration operation:

$$A'(x, t) = \begin{cases} |A_c(x, t)|, & \text{if } -\frac{\pi}{2} \leq \beta'_A(x, t) < \frac{\pi}{2}, \\ -|A_c(x, t)|, & \text{otherwise.} \end{cases}\tag{3.8}$$

and

$$B'(x, t) = \begin{cases} |B_c(x, t)|, & \text{if } -\frac{\pi}{2} \leq \beta'_B(x, t) < \frac{\pi}{2}, \\ -|B_c(x, t)|, & \text{otherwise.} \end{cases}\tag{3.9}$$

In this way, a correct CSPAMM combination of A' and B' images can be found

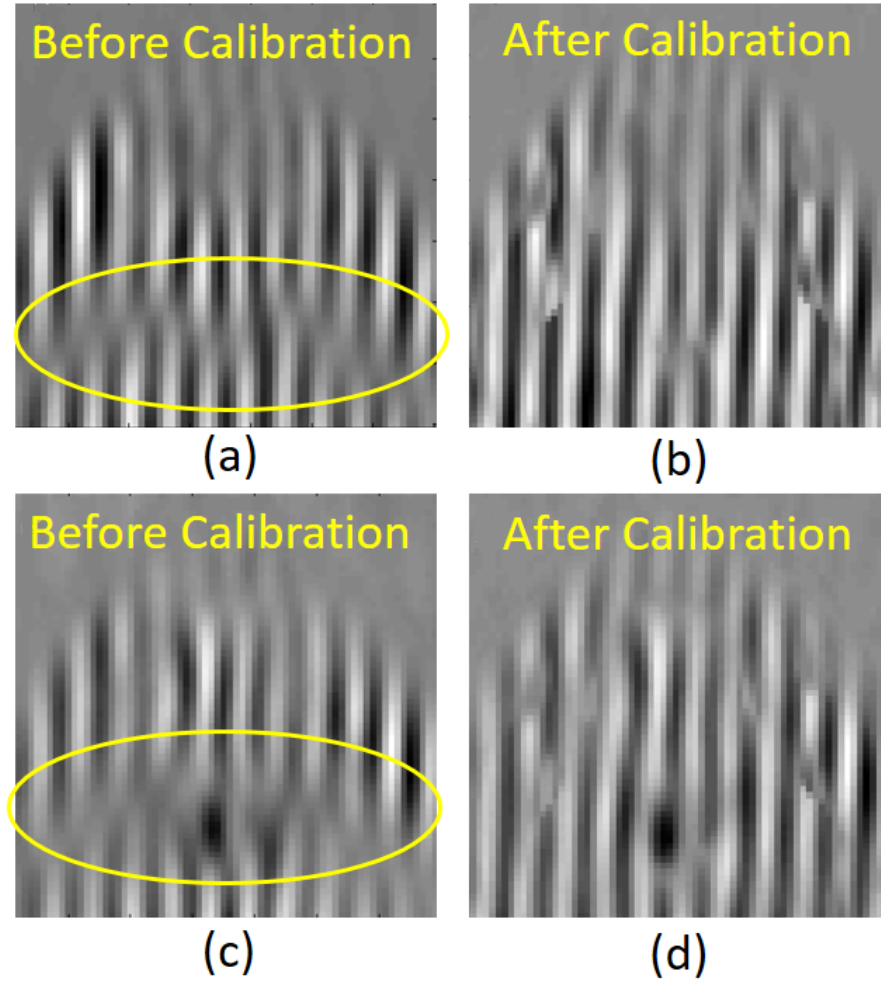


Figure 3.3: Phase calibration result. (a) and (c) are CSPAMM images before calibration. (b) and (d) are corrected CSPAMM images.

using $A'(x, t) - B'(x, t)$. Fig. 3.3 shows two examples of the corrected CSPAMM computation. The broken tags have been fixed and a high-contrast image at the first time frame is obtained that can be used in the following TMAP processing.

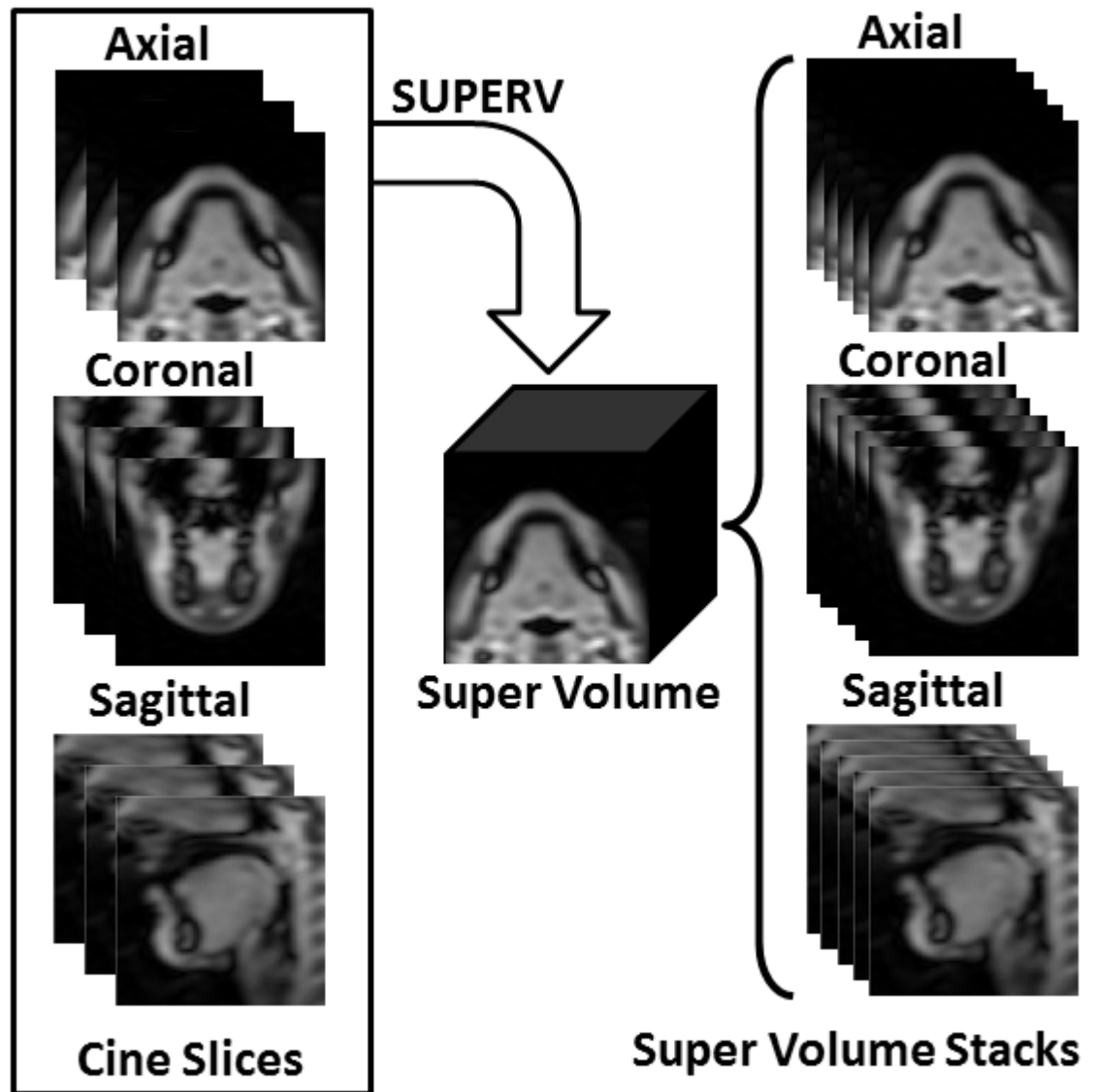


Figure 3.4: Cine super volume reconstruction by the SUPERV algorithm.

3.3 Segmentation of the Tongue Mask

Cine MR images are used to provide anatomical information for tongue segmentation. Because of the need to acquire data rapidly during speech and to maintain high overall signal-to-noise ratio (SNR), cine images are acquired in the same positions with the same relatively large slice thickness as the tagged MR data. Any single stack of cine MR images cannot be used for high resolution segmentation because of the poor through-plane resolution. Therefore, the entire collection (axial, sagittal, and coronal) is combined using super-resolution methods into a single image on a 3D grid whose voxel resolution is the same as the original 2D in-plane resolution. Specifically, we use the SUPERV algorithm [63] to obtain one *supervolume* at each time frame (see Fig. 3.4).

In order to constrain our analysis to the tongue region only, we carry out a segmentation of the tongue using the supervolume images. Despite the improved quality of the supervolume over the cine images, we have found that it is helpful to have manual guidance in this stage. We use the random walker (RW) algorithm [65], which is a graph-based algorithm to find a global optimal probabilistic solution for multi-label image segmentation. In practice, a user specifies (draws) a small number of pixels as seeds within pre-defined structures (labels), such as the tongue and the background. Each unlabeled pixel is then assigned to the label with the greatest probability in such a way that a random walker starting at this pixel will reach one of the seeds with this label. In TMAP, a human user is required to input seeds on a few slices (6 to 9

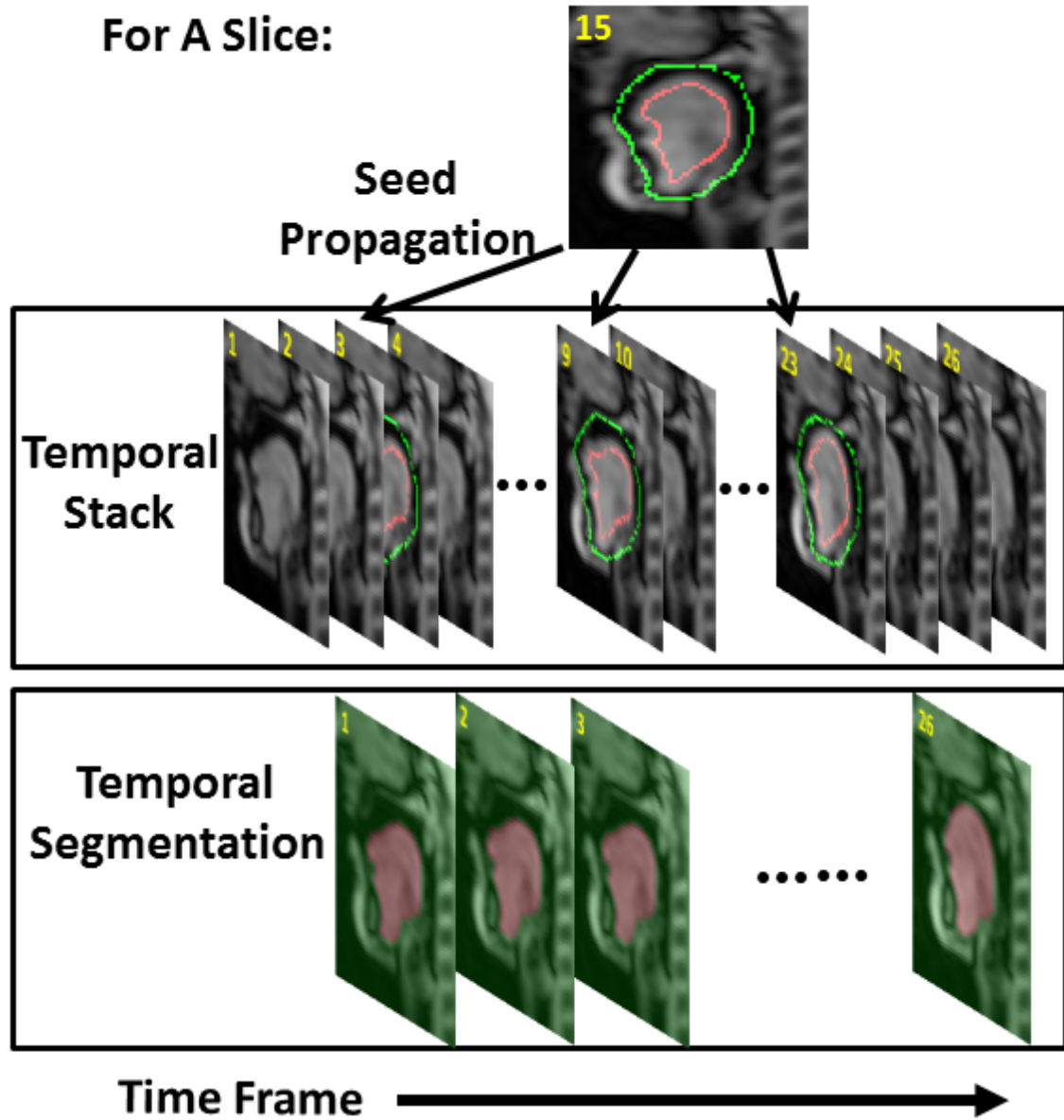


Figure 3.5: Seed propagation in time and segmentation of temporal stack.

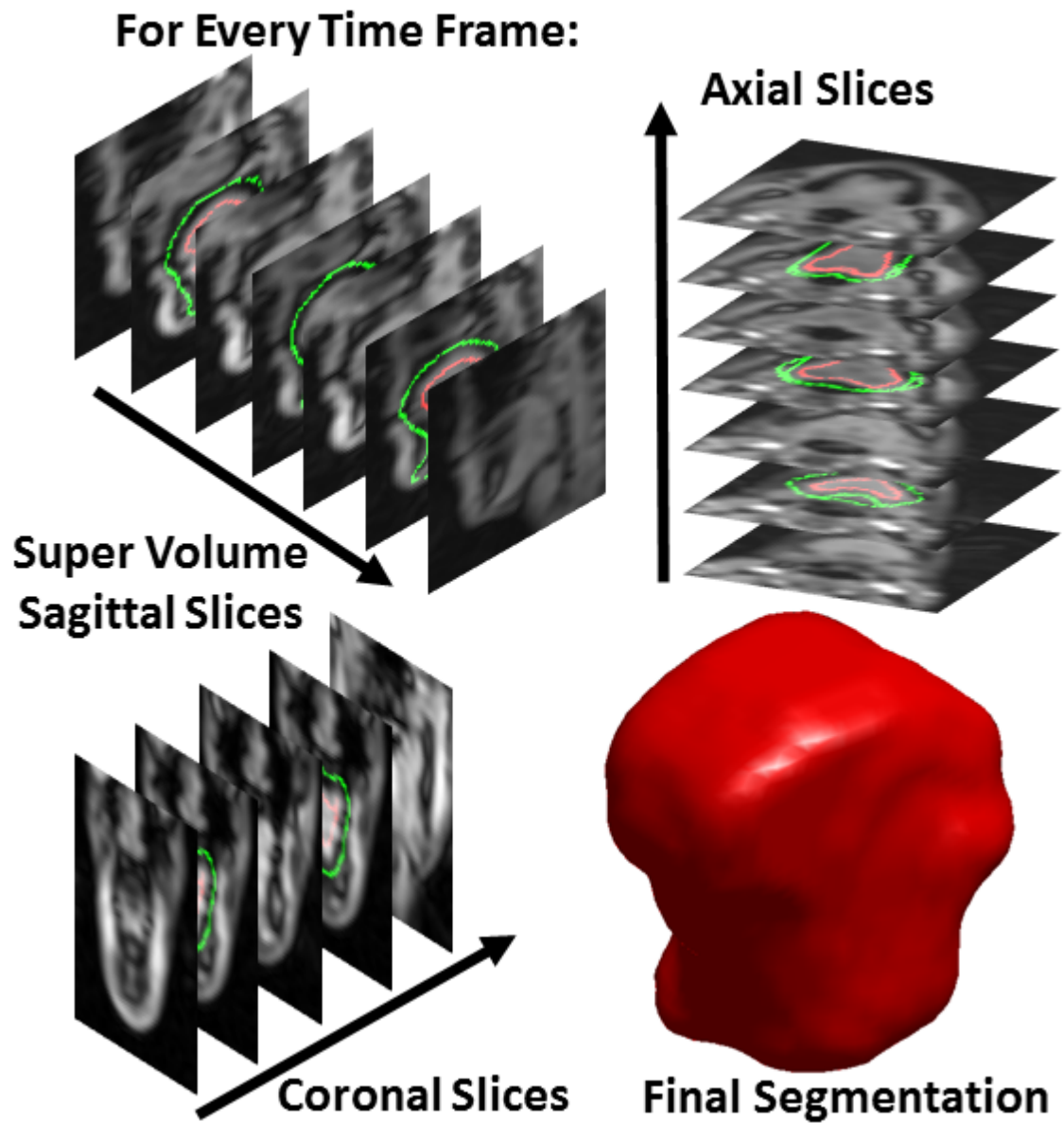


Figure 3.6: Segmentation of the supervolume.

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

slices in our case) of the cine images at one time frame. The user-given seeds are then propagated by deformable registration [52] to additional user-determined time frames (in our case, four time frames uniformly distributed over 26 time frames) at the same slice location. For the rest time frames, seeds are automatically generated by: 1) segmenting a 3D temporal stack using RW; 2) using the skeleton [66] of the temporal segmentation as seeds (Fig. 3.5). After seeds are found for all time frames at this slice location, they are exported to the 3D supervolume space and RW is computed, yielding the final segmentation (Fig. 3.6). During this process, the user is allowed to validate and correct the propagated and automatically generated seeds. Finally, the 3D tongue masks can be used to cut the 2D in-plane motion at the positions where it intersects the slice plane, leaving the 2D motion only on the tongue region as input for IDEA.

3.4 Tongue Motion Analysis Pipeline

As introduced in Section 2.2, HARP yields a collection of 2D vector-valued images, each representing the 2D projection of the 3D motion occurring from the current time frame to the initial time frame when the tags were applied. The segmented 2D motion slices from tagged data can be viewed as multiple observations about the underlying 3D motion (Figs. 3.7(a) and 3.7(b)). With the semi-automatic segmentation results from RW, IDEA can be applied on the HARP results and yield a 3D motion estimate.

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

Since IDEA is computationally demanding, the segmented 3D tongue mask alleviates the problem by letting it compute only on the tongue region, yielding a desired 3D motion field (Figs. 3.7(c) and 3.7(d)). We note that in this dissertation, *cones* are used to visualize 3D motion. They are color-coded by motion directions in classic diffusion tensor imaging (DTI) color scheme (red for left–right direction, blue for superior–inferior direction, green for anterior–posterior direction), and the cone size reflects magnitude of motion. This completes the estimation part of TMAP.

3.5 Multi-subject Data Regularization

Due to varying speaking rates among different subjects, before comparing their motion, displacement fields relative to a common position must be computed from the result of IDEA. Since the first time frame is normally an unpredictable pre-speech position of the tongue, and since we expect to observe the forward motion into /s/ and then upward motion into /k/ during the utterance “a souk”, the mid-central vowel /ə/ should be used as the common reference frame to compare motion across subjects [67]. Therefore we must switch the reference frame from 1 to the maximum /ə/ position.

For each subject, we denote the sequence of 3D vector fields as $\{\mathbf{v}_{1,1}(\mathbf{X}_1), \mathbf{v}_{1,2}(\mathbf{X}_1), \dots, \mathbf{v}_{1,26}(\mathbf{X}_1)\}$, obtained from the output of IDEA. Each vector field $\mathbf{v}_{1,t}(\mathbf{X}_1)$, as visualized in Fig. 3.7, shows the displacement from time frame 1 (default reference

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

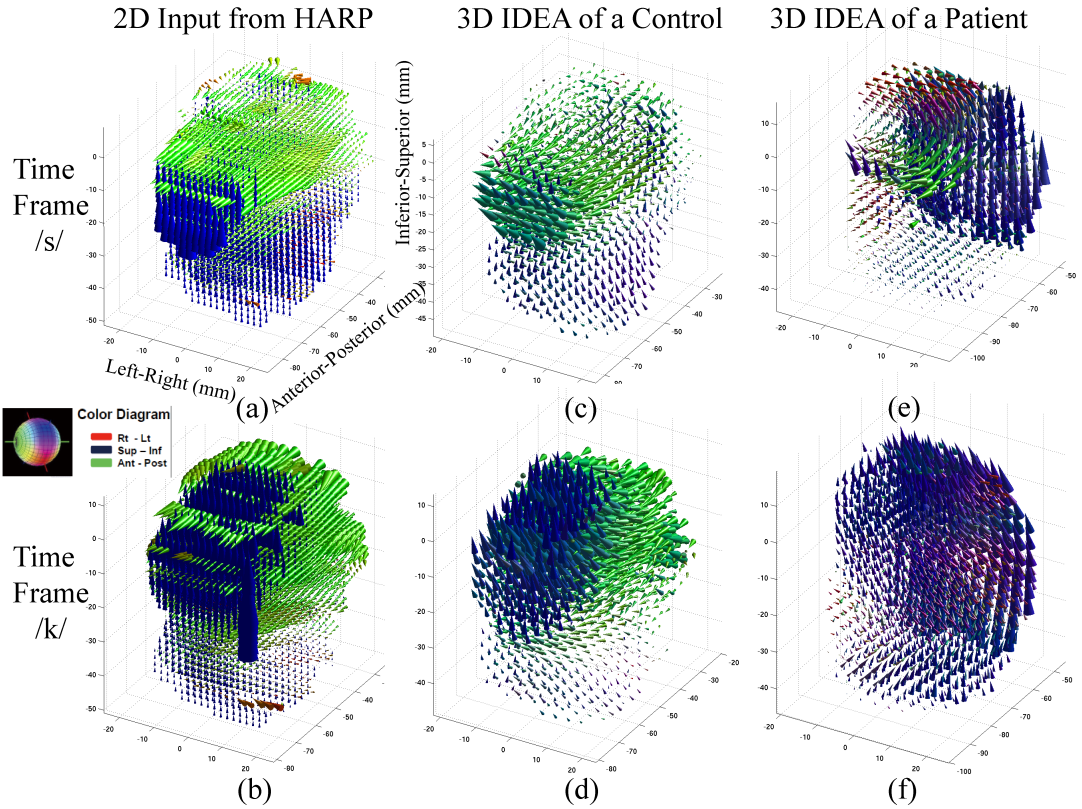


Figure 3.7: 3D motion estimation produce by TMAP. (a) A healthy control’s HARP input at time frame /s/. (b) A healthy control’s HARP input at time frame /k/. (c) A healthy control’s 3D motion at time frame /s/. (d) A healthy control’s 3D motion at time frame /k/. (e) A patient’s 3d motion at time frame /s/. (f) A patient’s 3D motion at time frame /k/. Note: cones are color-coded by motion directions as shown in the color diagram (red for left–right, blue for superior–inferior, green for anterior–posterior). Cone size is magnitude.

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

frame) to the current time frame t . \mathbf{X}_1 is the 3D grid located at time frame 1 (here we use a Lagrangian representation). As a result, if we consider the vector field $\mathbf{v}_{1,t}(\mathbf{X}_1)$ as arrows, they grow from the grid locations \mathbf{X}_1 at time frame 1 and end up pointing at the non-grid locations in the current frame t .

Suppose maximum $\partial/\partial t$ happens at time frame r . At arbitrary time frame t , related motion fields are $\mathbf{v}_{1,r}(\mathbf{X}_1)$ and $\mathbf{v}_{1,t}(\mathbf{X}_1)$. If we are able to find the inverse field of $\mathbf{v}_{1,r}(\mathbf{X}_1)$, namely $\mathbf{v}_{r,1}(\mathbf{X}_r)$, the following field is what we want

$$\mathbf{v}_{r,t}(\mathbf{X}_r) = \mathbf{v}_{r,1}(\mathbf{X}_r) + \mathbf{v}_{1,t}(\mathbf{X}_r + \mathbf{v}_{r,1}(\mathbf{X}_r)), \quad (3.10)$$

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

Because the field $\mathbf{v}_{1,r}(\mathbf{X}_1)$ is discrete, from the definition of inverse field we know that at time frame r , $\mathbf{v}_{r,1}(\mathbf{X}_1 + \mathbf{v}_{1,r}(\mathbf{X}_1)) = -\mathbf{v}_{1,r}(\mathbf{X}_1)$. To find the value of $\mathbf{v}_{r,1}$ at \mathbf{X}_r , we apply a fixed-point method [36] by iteratively solving the following equation

$$\mathbf{v}_{r,1}^{(n)}(\mathbf{X}_r) = -\mathbf{v}_{1,r}(\mathbf{X}_r + \mathbf{v}_{r,1}^{(n-1)}(\mathbf{X}_r)). \quad (3.11)$$

Finally, by substituting the converged result of Eqn. 3.11 into Eqn. 3.10 and repeating for every time frame, we get a new sequence of displacement fields $\{\mathbf{v}_{r,1}(\mathbf{X}_r), \mathbf{v}_{r,2}(\mathbf{X}_r), \dots, \mathbf{v}_{r,26}(\mathbf{X}_r)\}$ for every subject starting at time frame $\partial/\partial t$.

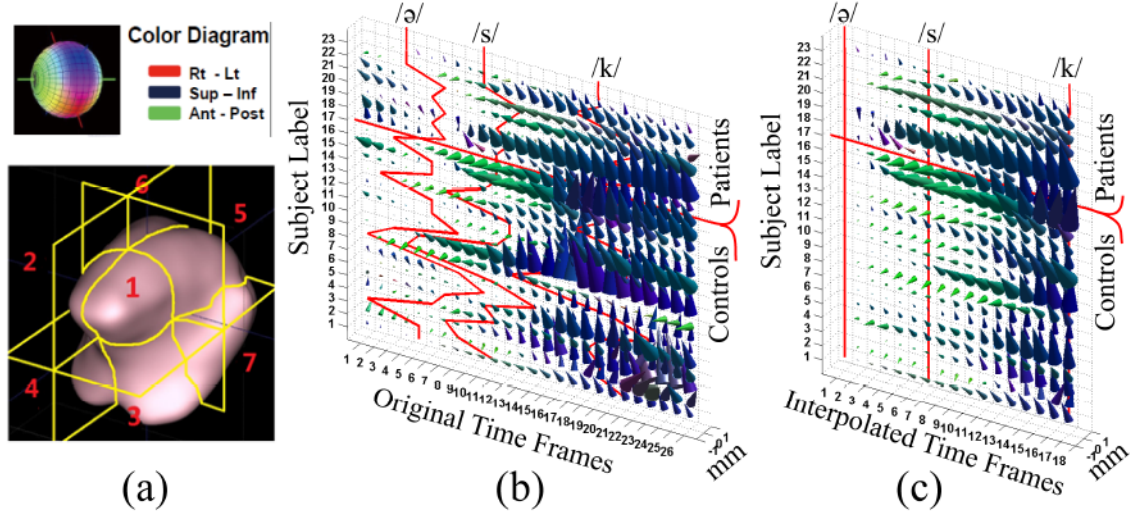


Figure 3.8: (a) Division of eight VOIs in the tongue. (b) An example using VOI-1: Average motion of 26 frames and all subjects. Horizontal line divides healthy controls (bottom) and patients (top). Vertical curves are at time frames /ə/, /s/, and /k/. (c) Interpolated motion between time frames /ə/ and /k/ from (b). Horizontal line divides healthy controls (bottom) and patients (top). Vertical lines are at time frames /ə/, /s/, and /k/.

3.6 Two-step Principal Component Analysis

Multi-subject PCA requires a certain tongue motion quantity to be in the same frame of reference. Although we have made all displacement fields to be with respect to /ə/, different subjects' tongue shapes vary extensively. Therefore, we divide the tongue into eight volumes of interests (VOIs) as shown in Fig. 3.8(a). We average the motion field inside each VOI to get one vector to represent its general motion (Fig. 3.8(b)), denoted by $\{\mathbf{d}_{r,1}, \mathbf{d}_{r,2}, \dots, \mathbf{d}_{r,26}\}_v$, where v is the VOI number from 1 to 8. We treat these VOIs independently in the following processing.

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

Since we are only interested in the periods from /ə/ to /s/ to /k/, we create a common time interval by taking the average motion between these two periods, and using a cubic spline [68] (denoted as “cspline” in Eqn. 3.12) to interpolate them into 17 time frames for all subjects, where /ə/ is time-frame 1, /s/ is 7, and /k/ is 17. Denoting the time frame number of maximum /ə/, /s/, and /k/ as a , s , and k , we have

$$\{\hat{\mathbf{d}}_{1,1}, \dots, \hat{\mathbf{d}}_{1,7}, \dots, \hat{\mathbf{d}}_{1,17}\}_v = \text{cspline}(\{\mathbf{d}_{a,a}, \dots, \mathbf{d}_{a,s}, \dots, \mathbf{d}_{a,k}\}_v) \quad (3.12)$$

For any VOI, $\hat{\mathbf{d}}_{1,t}$ is the interpolated mean motion we are interested in, which puts all subjects’ motions in the same framework and ready for PCA (Fig. 3.8(c)). Labeling the subject number by i , we stack the mean motion of all 17 frames as one vector

$$\hat{\mathbf{d}}^i = [\hat{\mathbf{d}}_{1,1}^i; \dots; \hat{\mathbf{d}}_{1,7}^i; \dots; \hat{\mathbf{d}}_{1,17}^i], \quad (3.13)$$

which lies in a $3 \times 17 = 51$ -dimensional space. The physical meaning of $\hat{\mathbf{d}}^i$ is all motion both in space and in time of subject i when performing the complete speech task of “a souk”. Note that by doing so we have avoided treating each time frame independently. Instead, we consider the entire speech task as an evaluation of the subject’s speech function.

Suppose the number of healthy controls is C . If we want to perform a conventional PCA on these healthy controls, the following steps are required: 1) subtract the mean

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

of healthy control motion $\hat{\mathbf{s}}^i = \hat{\mathbf{d}}^i - \text{mean}\{\hat{\mathbf{d}}^1, \dots, \hat{\mathbf{d}}^i, \dots, \hat{\mathbf{d}}^C\}$, 2) compute the covariance matrix of the subtracted motion $\text{COV} = [\hat{\mathbf{s}}^1, \dots, \hat{\mathbf{s}}^i, \dots, \hat{\mathbf{s}}^C][\hat{\mathbf{s}}^1, \dots, \hat{\mathbf{s}}^i, \dots, \hat{\mathbf{s}}^C]^T$, and 3) find the eigendecomposition of COV to get $C - 1$ principal component (PC) directions $\{\mathbf{e}^1, \dots, \mathbf{e}^{C-1}\}$ and principal values $\{P^1, \dots, P^{C-1}\}$. To evaluate this PC space to see if it is able to distinguish patients' motion from healthy controls', we project a test group of healthy controls' and a group of patients' motion onto these principal directions. Unsatisfyingly, after carrying this out we observed that the patients' PC scores are similar to those of the test controls and the two groups are hardly distinguishable.

To handle this situation, we introduce a PCA step involving the patients after performing the first PCA on healthy controls. Since the first PC space has a rank of $C - 1$, the remaining $51 - (C - 1)$ principal directions can be vectors generated by any feasible orthogonalization method (e.g., the Gram-Schmidt process). And this remaining $51 - (C - 1)$ dimensional space contains only the motion of the patients, because ideally the healthy controls should project a zero PC score in this space. Suppose the number of patients is P . We take their motions indexed by $j, 1 \leq j \leq P$, subtract the control mean, and then compute their "normal part" by projection onto the control PC space; that is,

$$\hat{\mathbf{s}}^j = \hat{\mathbf{d}}^j - \text{mean}\{\hat{\mathbf{d}}^1, \dots, \hat{\mathbf{d}}^i, \dots, \hat{\mathbf{d}}^C\}, \quad (3.14)$$

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

$$\hat{\mathbf{s}}_{normal}^j = ((\hat{\mathbf{s}}^j)^T \cdot \mathbf{e}^1)\mathbf{e}^1 + \dots + ((\hat{\mathbf{s}}^j)^T \cdot \mathbf{e}^{C-1})\mathbf{e}^{C-1}. \quad (3.15)$$

The remaining motion is considered abnormal and is given by

$$\hat{\mathbf{s}}_{abnormal}^j = \hat{\mathbf{s}}^j - \hat{\mathbf{s}}_{normal}^j. \quad (3.16)$$

We then compute the covariance matrix of $\hat{\mathbf{s}}_{abnormal}^j$ and find its eigen-decomposition to get P more vectors as the PC directions for abnormal motion $\{\mathbf{u}^1, \dots, \mathbf{u}^P\}$. Taken together, $\{\mathbf{e}^1, \dots, \mathbf{e}^{C-1}, \mathbf{u}^1, \dots, \mathbf{u}^P\}$ come from a two-step PCA to represent the normal and abnormal motion parts, which we refer to as “primary” and “secondary” PC spaces. We note that the reason that the rank of the secondary PC space is P instead of $P - 1$ is because the healthy controls’ (not patients’) mean motion is subtracted in Eqn. 3.14 so that patient motion is not zero-centered. We emphasize that the entire purpose of building the secondary PC space is to use it to contain the patient’s abnormal motion part and to separate it from its normal motion part. After the two-step PCA, we can take any subject’s motion (patient or healthy control), project it onto the space $\{\mathbf{e}^1, \dots, \mathbf{e}^{C-1}, \mathbf{u}^1, \dots, \mathbf{u}^P\}$, and look at the “primary PC scores” and “secondary PC scores”.

3.7 Evaluation of Angle of Motion

We used TMAP to process all twenty-three subjects' speech data, obtaining their motion fields. A typical healthy control's motion and a typical patient's motion at critical time frames /s/ and /k/ are shown in Figs. 3.7(c), 3.7(d), 3.7(e), and 3.7(f). In the motion of a healthy control, both forward and upward motions are expected to be prominent, while a patient's motion is expected to contain more left-right motion. After multi-subject data regularization, we took all subjects' interpolated average motion and computed two angles of the motion vector at every time frame (Fig. 3.9). We began this data analysis at time frame 4 because the magnitude of motion in the first three time frames was nearly zero, so that noise became a major interruption in angle visualization. The first angle, shown in Figs. 3.9(a) and 3.9(c), is the elevation angle of the motion vector upward from the anterior direction. It shows the degree that the motion vector tilts up and is denoted as θ . The second angle, shown in Figs. 3.9(b) and 3.9(d), is the azimuth angle of the motion vector sideways from the anterior direction (to the left or to the right of the head). It shows the degree that the motion vector tilted left or right and is denoted as ϕ . Fig. 3.9 shows that for healthy controls, θ grows from small values to around 90 degrees, while the ϕ value stays small. For patients, θ values are inconsistent throughout the speech task and some ϕ values are large. Thus, ϕ distinguishes normal speech, which contains less left-right tongue motion, from patient motion, which has evidence of left-right asymmetrical motion.

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

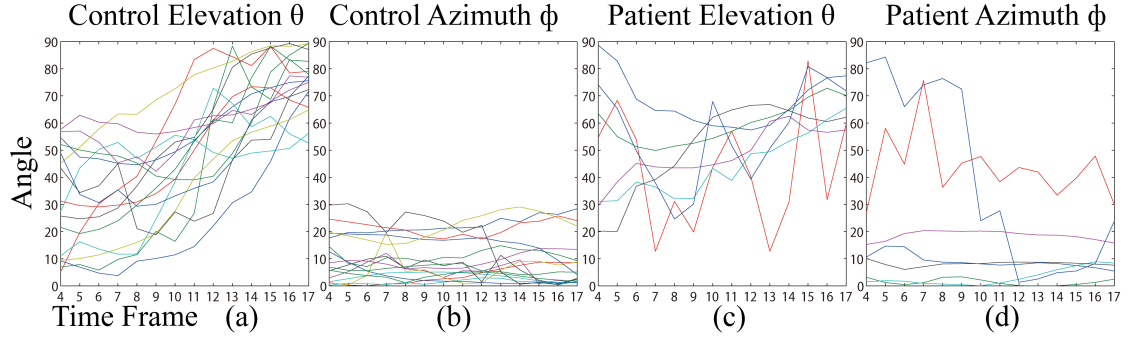


Figure 3.9: Elevation and azimuth angles during motion of healthy control and patient subjects in volume of interest 1. Each curve is a subject. (a) and (b) show sixteen healthy controls and (c) and (d) show seven patients.

Table 3.1: Controls' and Patients' Average Elevation and Azimuth Angles of Motion (in degrees).

	VOI-1	VOI-2	VOI-3	VOI-4	VOI-5	VOI-6	VOI-7	VOI-8
Control θ	49.7	49.1	43.7	44.0	34.1	30.5	42.7	44.1
Patient θ	53.0	54.0	58.6	60.8	62.4	57.4	62.9	61.5
Control ϕ	7.2	6.4	9.9	10.1	8.5	9.3	11.3	11.9
Patient ϕ	13.3	14.6	17.6	18.6	11.4	9.2	11.7	13.5

Table 3.1 shows the average of the two angle values across all time for all eight VOIs. Both angles show larger values for the patients. The higher θ values for patients reflect an earlier upward angle of motion than the healthy controls in all tongue regions. The value of ϕ is larger in patients than healthy controls ($p < 0.01$, student t-test statistic -3.39 , effect size $r^2 = 0.51$), showing larger left-right motion in patients versus healthy controls in all eight tongue regions.

3.8 Two-step PCA of the Whole Tongue

We applied our two-step PCA strategy on this dataset of sixteen healthy controls and seven patients, obtaining fifteen primary PC directions and seven secondary PC directions. In Fig. 3.10, we only show the PC directions and PC weights of VOI-1 as an example for visualization. Other VOIs possess the same pattern of the PC directions and PC weights. We note that the focus of interpreting PC directions should be on the first few primary PCs starting from number 1 (Fig. 3.10(a)) and the first few secondary PCs starting from number 16 (Fig. 3.10(b)). The primary directions show little lateral motion until PC number 7, indicating a consistent control motion featuring mostly anterior–posterior and superior–inferior. The secondary directions contain all abnormal motion patterns showing many red glyphs that correspond to left–right motion. As shown in Figs. 3.10(c) and 3.10(d), healthy controls load zero weight in secondary space while patients load in both spaces.

3.9 Evaluation of Two-step PCA on Mirrored Hemi-tongue

To test the reliability of our two-step PCA method, we measured its capability of distinguishing patient motion from control motion by mirroring the tongue’s right side to the left. Since the sixteen healthy controls’ tongue motions are generally

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

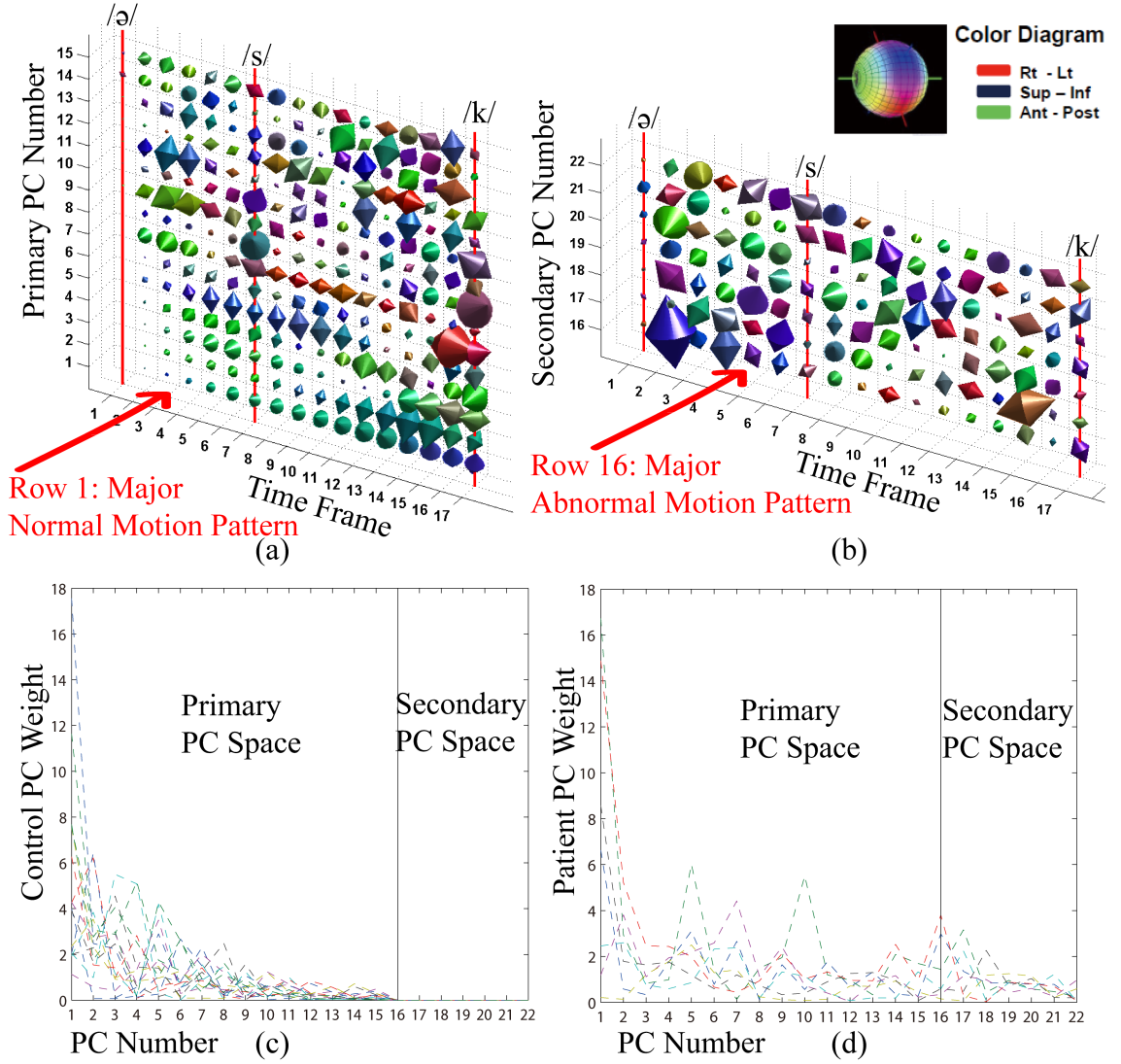


Figure 3.10: Two-step PC directions and PC weights. (a) Primary PC directions from healthy controls. (b) Secondary PC directions from patients. (c) and (d) show the PC weights for healthy controls and patients. Each dashed line is a subject.

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

symmetric, we overlaid the mirrored right hemi-tongue on the left hemi-tongue and averaged their motion at each time frame. This yielded sixteen healthy controls with only the left VOIs: 1, 3, 5, and 7. For the seven patients, since their glossectomy was only performed on one side (either left or right) of the tongue, their tongues are no longer symmetric and their data should not be averaged. Therefore, after we mirrored their right hemi-tongue to the left, the result is seven resected hemi-tongues and seven native hemi-tongues both on the left. We named the former “Patient Glossectomy Side (PGS)” and the latter “Patient Native Side (PNS)”.

With this modified data on hemi-tongues, we then performed four two-step PCAs, one for each of the four VOIs, using eight of the healthy controls as training data to build the primary PC space, leaving eight controls as test data. The seven PGS subjects were then used to create the secondary PC space. Following this procedure, we obtained seven primary PC directions and seven secondary PC directions for each VOI. We projected three subject groups onto the secondary PC directions: eight test controls, seven PGSs, and seven PNSs. The first two secondary PC weights of all three subject groups for all four VOIs are shown in Fig. 3.11. The solid blue dot represents no abnormal motion. The green dots, which are the test controls, show smaller loading values. The patients have higher loadings in both the PNS group (crosses) and the PGS group (circles), although these two groups are not particularly distinguished from each other in this representation.

Finally, we randomized the choice of training controls from all sixteen healthy

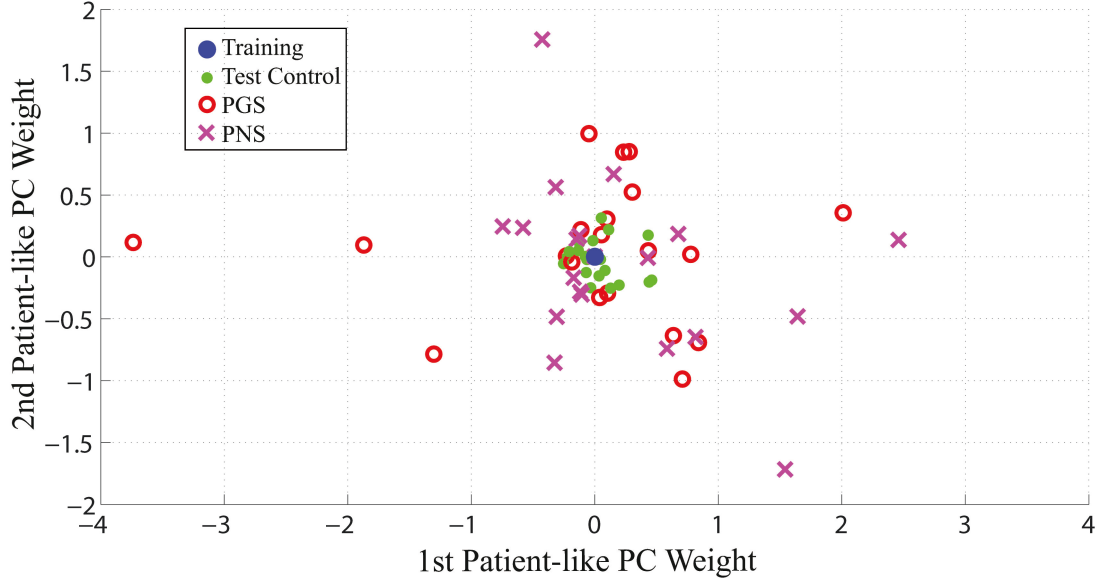


Figure 3.11: First two patient-like abnormal PC weights for four volumes of interest and all subjects. The blue dot is the origin. Green dots are test controls, circles are patient glossectomy sides, and crosses are patient native sides.

controls (16 choose 8 is 12870 possibilities), thus randomizing the construction of the primary and the secondary PC spaces. We repeated the above hemi-tongue experiment in 12870 Monte Carlo trials for all possible forms of the two-step PC spaces, while keeping the eight controls as test group to compare with PGS group and PNS group. After every trial, the three groups' motions were projected onto the secondary PC directions to compute their secondary (abnormal) PC weights. The average abnormal PC weights of all trials are box-plotted in Fig. 3.12. In all four VOIs, the mean of the test controls' abnormal energy is lower than both PGS and PNS from all 12870 cases ($p < 0.01$, student t-test statistic -8.12 , effect size $r^2 = 0.49$). Despite the small amount of training data, this analysis is capable of distinguishing

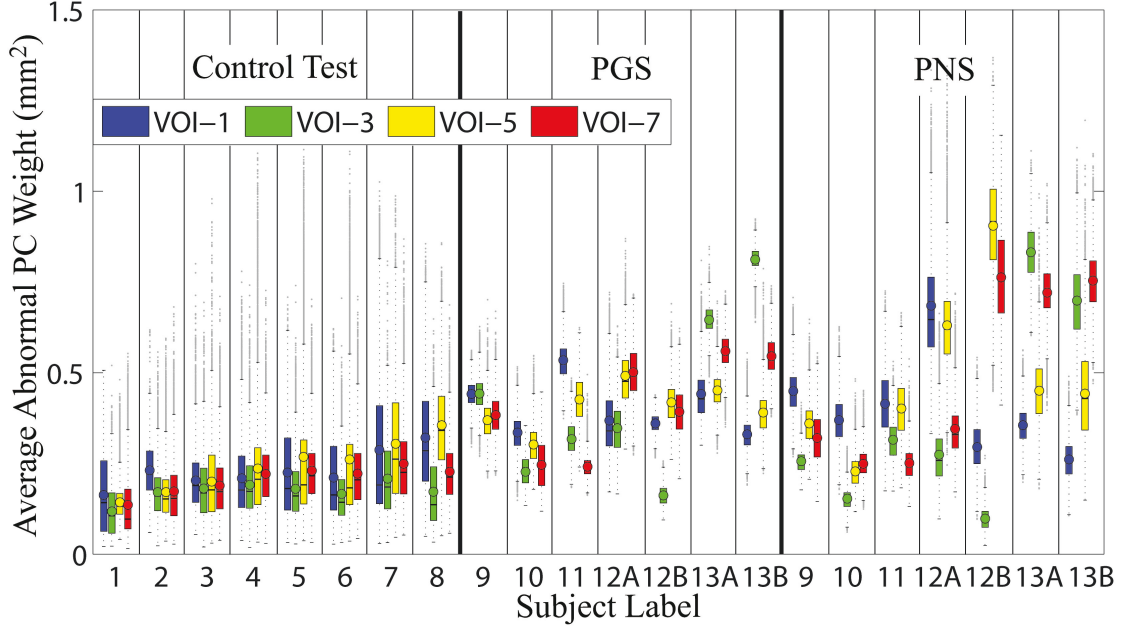


Figure 3.12: Boxplot of the average secondary principal component weights (abnormal energy) of control test, PGS, and PNS in 12870 Monte Carlo trials. Within each box, the center bar shows the median and the circle shows the mean. Patients 12 and 13 are labeled with A and B to represent their two recordings.

control motion from the two groups of patient motion and the difference is statistically significant.

3.10 Implementation of TMAP

TMAP has been implemented in an in-house software called “HARP5” based on Matlab (Mathworks, Natick, MA). HARP5 is the most recent version from generations of “HARP series” software until 2015. HARP5 has three major graphical user interfaces (GUIs), as demonstrated in Figs. 3.13, 3.14, and 3.15. The first GUI is related to 2D processing functions using the HARP algorithm. It provides major

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

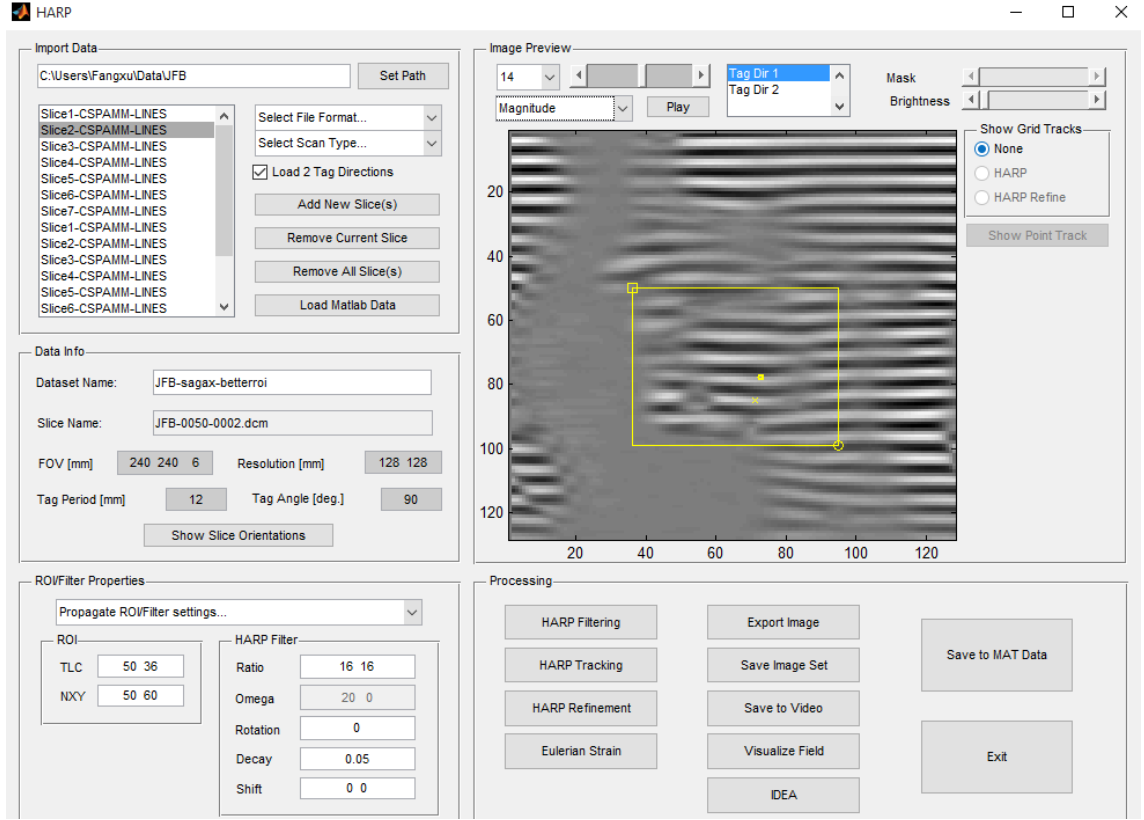


Figure 3.13: HARP5 2D processing GUI.

reading and visualizing functions of the raw tagged MR data. Users can manually set parameters such as region of interest and HARP filter properties. The second GUI provides a basic visualization of the estimated 2D motion fields. The third GUI is the main “workhorse” for 3D motion processing. It includes IDEA computation, segmentation and 3D mask specification, and strain analysis. Together, HARP5 provides all the basic functions in TMAP, and has been used in current speech studies for years. A simplified version of HARP5 has been released to public, and it can be downloaded from the Image Analysis and Communication Laboratory webpage (<https://iacl.ece.jhu.edu>).

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

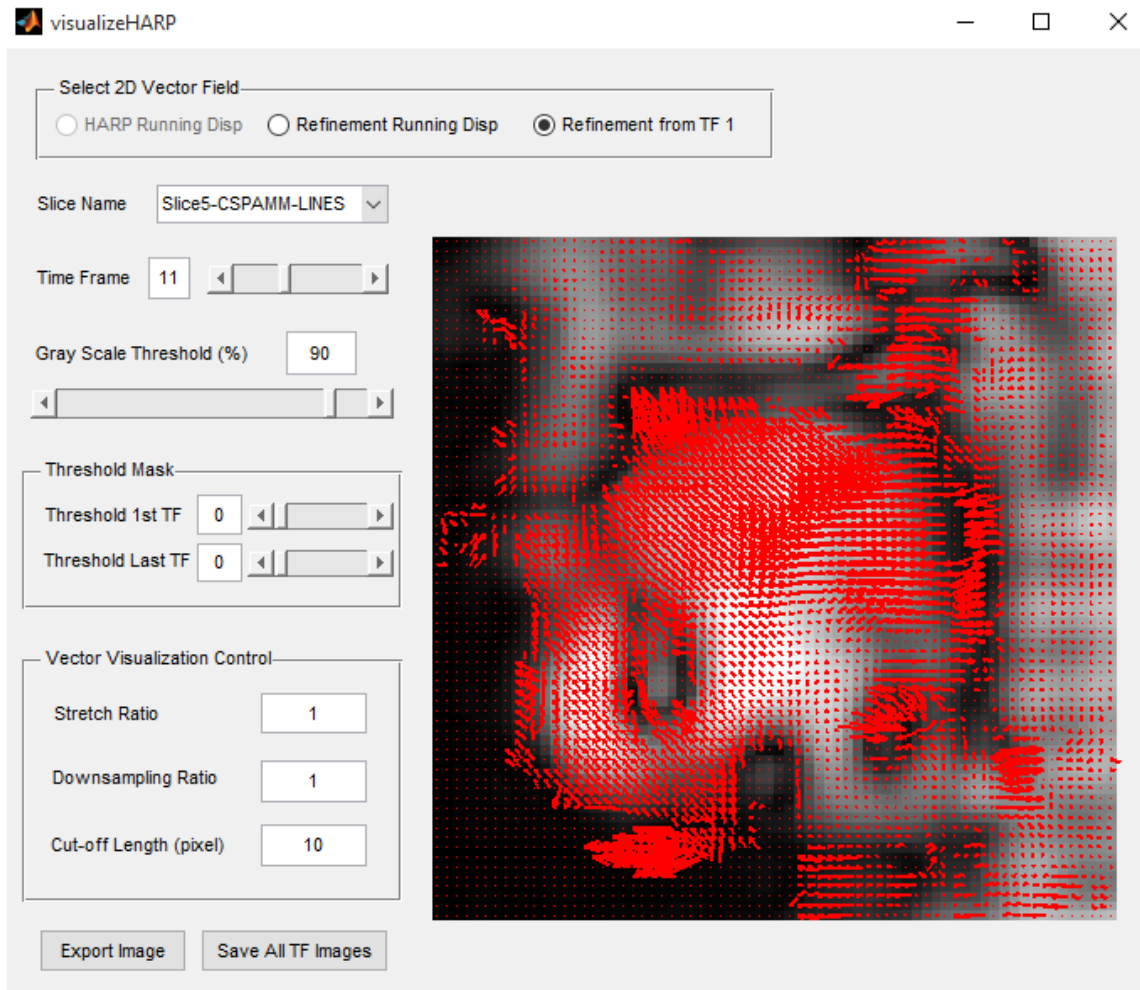


Figure 3.14: HARP5 2D visualization GUI.

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

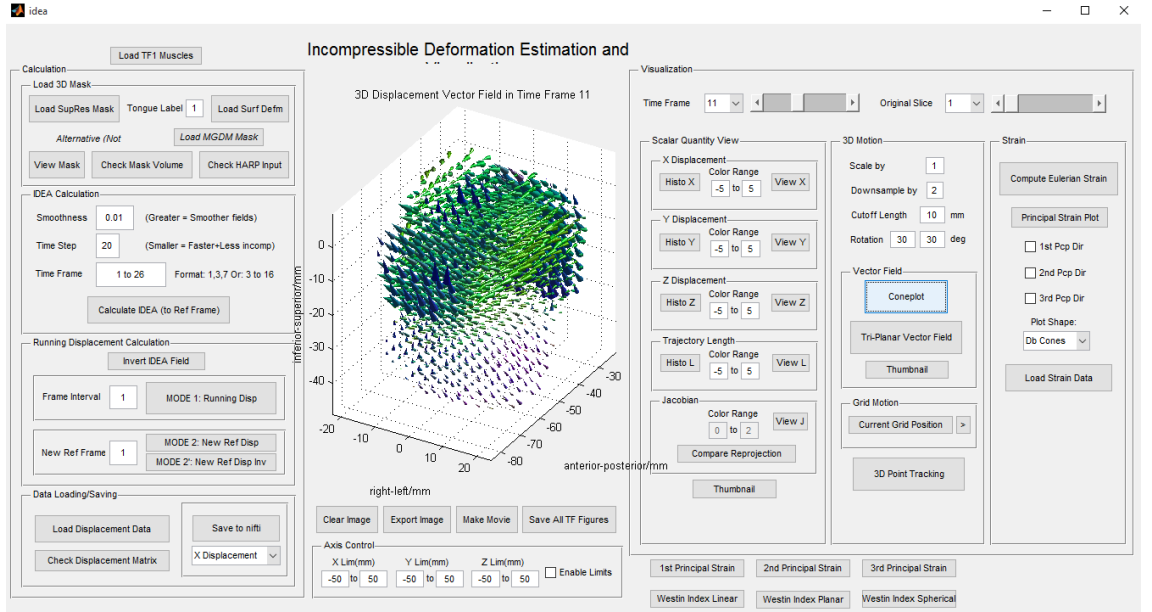


Figure 3.15: HARP5 3D processing GUI.

3.11 Discussion

In the multi-subject analysis of the selected speech dataset, the angle of motion experiment was intended to reveal general trends in the motion fields. The results agreed with the expected “a souk” motion. Overall, the healthy controls’ tongues went forward to produce /s/ then upward to produce /k/ while maintaining left–right symmetry. The patients’ tongues showed a similar trend but were noisier and less predictable while producing a large amount of left–right asymmetry. In the temporally averaged motion in all eight VOIs (Table 3.1), the azimuth angle of the patient tongues always had greater values than the control, especially at VOIs 1, 2, 3, and 4 in the anterior tongue. Since our data used only subjects with unilateral tumors behind the tongue tip, the motor control of the tongue tip was reduced. Therefore,

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

the anterior VOIs (1, 2, 3, and 4) were expected to be affected more strongly than the back part of the tongue. On the other hand, the elevation angle was greater for the patients possibly due to their compensation strategies in speech after partial tongue resection.

Secondly, the two-step PCA experiment further revealed the patients' abnormal motion pattern. The primary PCs showed little lateral motion until PC number 7 (Fig. 3.10(a)), consistent with normal speech. The secondary PCs had more lateral motion, consistent with abnormal motion, and also showed less predictability of patterning. The plot of the PC weights showed zero load of the healthy controls on the secondary PC space (Fig. 3.10(c)) and non-zero weight of patients on the secondary PC space (Fig. 3.10(d)), which was a direct mathematical result of the two-step PCA. The subsequent experiment using the mirrored hemi-tongue data additionally showed that both sides of the patient tongue, native and resected, loaded more highly than the control tongues on the secondary PC space.

Lastly, 12870 Monte Carlo trials tested the efficacy of TMAP and confirmed that the healthy controls' motions were more consistent and stable than the patients. The mean of the control abnormal energy is lower than both PGS and PNS in all VOIs, and TMAP is capable of distinguishing the subtle patient motion variation pattern from the healthy controls. However, PGS and PNS motions are not well distinguished by our current approach, suggesting that compensation may be occurring on both sides of the tongue. In previous work [69], the patients studied moved their glossectomy side to

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

a greater extent than the healthy controls, while the native side moved similarly to the healthy controls. In this work, however, the patients loaded on the secondary PCs in both PGS and PNS, suggesting that both sides of the tongue used different strategies from the healthy controls. And the higher PNS loading suggested patients used the native side in a more unusual manner than the glossectomy side for compensation.

A closer look at the result of the Monte Carlo trials (Fig. 3.12) with reference to clinical recordings can provide us more information on the potential compensation strategies used by the patients. Although seven patient datasets were acquired, five patients participated in this study. The first four subjects (9 to 12) had small (T1) tumors and the fifth, subject 13, had a mid-size (T2) tumor. The small tumors were removed with glossectomy and the wound closed by sutures. The moderate tumor was closed by adding external tissue, a radial-forearm free flap, to replace the resected mass. Thus, subject 13 had different anatomy from the other four patients. In addition, two subjects, 12 and 13, were recorded twice. Specifically, subject 12's recordings were 25 and 73 months post-surgery and subject 13's recordings were 14 and 24 months post-surgery.

Fig. 3.12 shows high loadings for subjects 9, 10, and 11 on the secondary PCs of VOI-1 in both PGS and PNS, which means they moved the tongue tip in an unusual manor compared to the healthy controls. Subject 12A followed this pattern in PNS. However, in his PGS, and for both sides during the second visit (12B), this subject used the back of the tongue (VOIs 5 and 7). This may be evidence of an incomplete

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

adaptation in the first visit moving toward a fully realized pattern of motion in the second visit. Subject 13's motion pattern differed from all the other subjects. He used the floor of the mouth muscles (VOIs 3 and 7) on both sides of the tongue during both visits. Subject 13's first recording was much later post-surgically than subject 12 and his adaptation was likely to be more complete. In addition, his closure procedure was unique in this group. Thus, his differences were reasonable. The floor muscles helped move the upper tongue appropriately to shape the vocal tract and produce good quality speech, as seen by the low loadings on the secondary PCs of VOIs 1 and 5 on the upper tongue. The importance of this result is that the secondary PCs allow an explanation of the degree and nature of the abnormal movement in a way that was not possible before.

3.12 Summary

We proposed a pipeline of algorithms, TMAP, for processing speech MR image data to obtain a reliable estimate of the motion field. It is semi-automatic and easy to operate. We proposed methods to achieve effective segmentation, multi-subject data regularization, and a two-step principal component analysis to reveal subtle abnormal motion patterns. We tested TMAP using a various collection of subjects in multiple Monte Carlo trials and the result showed promising value in current speech studies and medical application. It is important for the progress of biological research into the

CHAPTER 3. 3D MOTION ANALYSIS PIPELINE

human body to have a pipeline method suitable for any user, which can process tagged and cine MR data from the raw images to the final data analysis and interpretation.

Chapter 4

Enhanced Incompressible Motion Estimation

4.1 Introduction

In the previous chapter, we developed TMAP to provide an initial step for motion estimation. However, the use of HARP and IDEA methods in TMAP can result in an increase of estimation errors on the tissue boundary. Since HARP uses a bandpass filter to extract the harmonic images, object boundaries are blurred and motion estimates near the anatomical surfaces are inaccurate [23, 61]. To make matters worse, HARP measurements near the boundaries are sparse because of the sparseness of image plane acquisition. These two problems severely affect 3D motion estimation near anatomical surfaces, as shown in Fig. 4.1. Zooming in on the back of the tongue (see

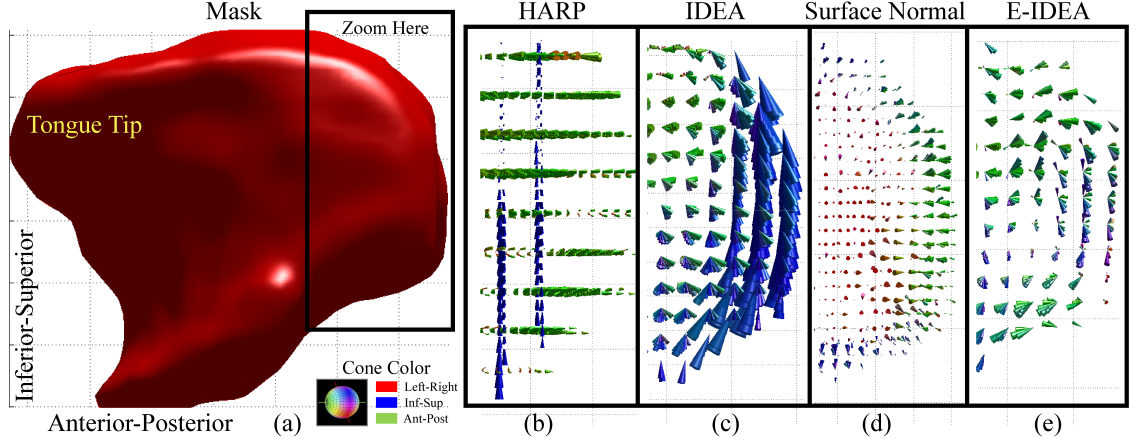


Figure 4.1: (a) Tongue mask of a healthy control subject (sagittal view). (b) HARP field on axial and coronal slices as input for IDEA, zoomed in at the tongue back. (c) IDEA result at the tongue back. (d) Surface normal deformation component at tongue back surface. (e) Proposed method result.

Fig. 4.1(a)), Fig. 4.1(b) shows the sparse 2D motion components from HARP and 4.1(c) is the IDEA reconstruction of 3D motion that shows inaccurate large motion.

In this chapter, we develop a novel approach that combines data from tagged images with surface deformation information derived from cine MR images to dramatically improve 3D tongue motion estimation. At every time frame, the tongue is segmented to achieve a 3D mask, and the deformation between the reference mask at the resting position and the deformed mask is computed using deformable registration. The normal components of surface deformation are then used to augment the HARP measurements within the IDEA estimation framework. Fig. 4.1(d) shows the additional input and Fig. 4.1(e) shows the result of proposed method. Comparing with Fig. 4.1(c), this result is more sensible from a qualitative point of view. Quantitative evaluations provided below also show that this method achieves a more

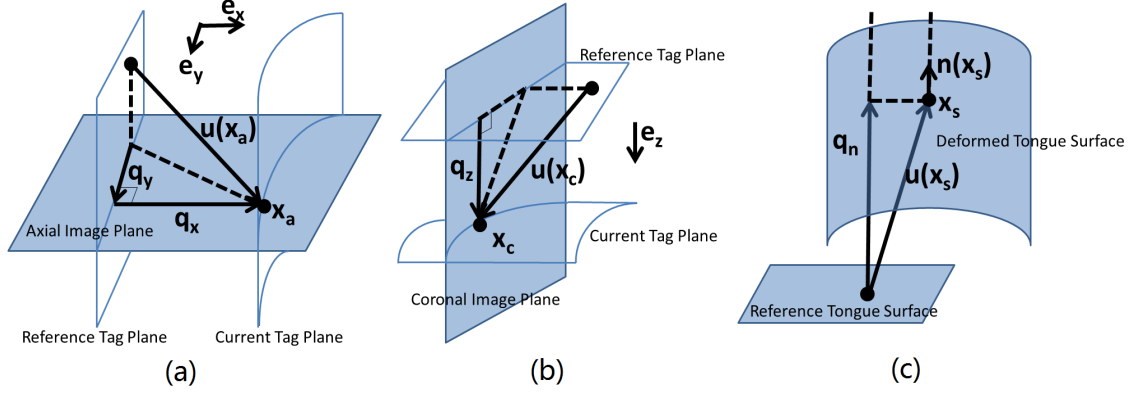


Figure 4.2: Relationship between 2D motion components and 3D motion on (a) an axial slice, (b) a coronal slice and (c) the tongue surface.

accurate estimate of the whole tongue motion.

4.2 Measuring Tongue Surface Deformation

Figs. 4.2(a) and 4.2(b) illustrate how HARP data are processed in IDEA [43]. As explained in Section 2.3, on each pixel \mathbf{x}_a from an axial image such as Fig. 4.2(a), HARP produces two vectors \mathbf{q}_x and \mathbf{q}_y . On each pixel \mathbf{x}_c from a coronal image such as Fig. 4.2(b), HARP produces a vector \mathbf{q}_z . IDEA takes such data on all pixels $\{\mathbf{x}_a, \mathbf{q}_x(\mathbf{x}_a), \mathbf{x}_a, \mathbf{q}_y(\mathbf{x}_a), \mathbf{x}_c, \mathbf{q}_z(\mathbf{x}_c)\}$ as input, and estimates the dense incompressible 3D field $\mathbf{u}(\mathbf{x})$ within the tongue mask. Since the inputs for IDEA are observed components of displacements that can arise at any physical position and in any direction of motion, it provides a key to utilization of surface deformation measurements within

CHAPTER 4. ENHANCED INCOMPRESSIBLE MOTION ESTIMATION

the IDEA framework. In particular, as shown in Fig. 4.2(c), the tongue surface may deform between time frames, and a point \mathbf{x}_s on the surface at current time frame can be associated with a point on the reference tongue surface. However, like the traditional aperture problem in optical flow [70, 71], we should not assume to know any tangential information about the surface displacement. This leads to a perfect analogy with HARP data: observations about surface normal deformation, if available, can be used in 3D reconstruction.

IDEA requires segmentation of the tongue volume in order to limit the tissue region that is assumed to be incompressible. In TMAP, cine MR images are used to construct a super-resolution volume [63] at each time frame, which is then segmented for the tongue surface mask. We note that these 3D masks can also be used for deformable registration in order to provide surface deformation information. The diffeomorphic demons method [52] is applied to the pair of masks between the two time frames where motion is to be computed. Denoting the reference mask at time frame 1 as $I_1 : \Omega_1 \subset \mathbb{R}^3 \rightarrow \{0, 1\}$ and the current deformed mask as $I_t : \Omega_t \subset \mathbb{R}^3 \rightarrow \{0, 1\}$ defined on the open and bounded domains Ω_1 and Ω_t , the deformation field is found and denoted by the mapping $\mathbf{d} : \Omega_t \mapsto \Omega_1$. The estimated displacement field at a point \mathbf{x}_s on the surface of the tongue in current time frame can be denoted as

$$\mathbf{u}(\mathbf{x}_s) = -\mathbf{d}(\mathbf{x}_s) . \quad (4.1)$$

CHAPTER 4. ENHANCED INCOMPRESSIBLE MOTION ESTIMATION

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

$$\mathbf{n}(\mathbf{x}_s) = \frac{\nabla\phi(\mathbf{x}_s)}{|\nabla\phi(\mathbf{x}_s)|} . \quad (4.2)$$

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

$$\mathbf{q}_n(\mathbf{x}_s) = (\mathbf{u}(\mathbf{x}_s) \cdot \mathbf{n}(\mathbf{x}_s))\mathbf{n}(\mathbf{x}_s) . \quad (4.3)$$

An example of such a field is shown in Fig. 4.1(d).

4.2.1 Enhanced IDEA

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

Algorithm 4.1. Enhanced Incompressible Deformation Estimation Algorithm (E-IDEA)

CHAPTER 4. ENHANCED INCOMPRESSIBLE MOTION ESTIMATION

1. Set $\mathbf{u}(\mathbf{x}_i) = \mathbf{0}$ and $\mathbf{u}(\mathbf{x}_s) = \mathbf{0}$.
2. Set M time steps, **for** $m = 1$ to M **do**
3. Project currently computed displacement onto input directions by $p_x(\mathbf{x}_a) = \mathbf{u}(\mathbf{x}_a) \cdot \mathbf{e}_x$, $p_y(\mathbf{x}_a) = \mathbf{u}(\mathbf{x}_a) \cdot \mathbf{e}_y$, $p_z(\mathbf{x}_c) = \mathbf{u}(\mathbf{x}_c) \cdot \mathbf{e}_z$, $p_n(\mathbf{x}_s) = \mathbf{u}(\mathbf{x}_s) \cdot \mathbf{n}(\mathbf{x}_s)$.
4. Compute remaining motion projection by $r_x(\mathbf{x}_a) = q_x(\mathbf{x}_a) - p_x(\mathbf{x}_a)$, $r_y(\mathbf{x}_a) = q_y(\mathbf{x}_a) - p_y(\mathbf{x}_a)$, $r_z(\mathbf{x}_c) = q_z(\mathbf{x}_c) - p_z(\mathbf{x}_c)$, $r_n(\mathbf{x}_s) = q_n(\mathbf{x}_s) - p_n(\mathbf{x}_s)$.
5. Use part of the remaining motion to approximate velocity: $v_x(\mathbf{x}_a) = r_x(\mathbf{x}_a)/(M - m + 1)$, $v_y(\mathbf{x}_a) = r_y(\mathbf{x}_a)/(M - m + 1)$, $v_z(\mathbf{x}_c) = r_z(\mathbf{x}_c)/(M - m + 1)$, $v_n(\mathbf{x}_s) = r_n(\mathbf{x}_s)/(M - m + 1)$.
6. Update estimation: $\mathbf{u}(\mathbf{x}_i) = \mathbf{u}(\mathbf{x}_i) + \text{DFVS}\{v_x(\mathbf{x}_a), v_y(\mathbf{x}_a), v_z(\mathbf{x}_c), v_n(\mathbf{x}_s)\}$, $\mathbf{u}(\mathbf{x}_s) = \mathbf{u}(\mathbf{x}_s) + \text{DFVS}\{v_x(\mathbf{x}_a), v_y(\mathbf{x}_a), v_z(\mathbf{x}_c), v_n(\mathbf{x}_s)\}$.
7. **end for**

Here DFVS stands for divergence-free vector spline, which is also the key algorithm “workhorse” of IDEA [43]. M is typically set to 20 which provides a proper trade-off between accuracy and computation time. Enhanced IDEA, which we refer to as E-IDEA below, typically takes about 5 hours to process a subject with 26 time frames.

4.3 Visual Assessment of E-IDEA

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

Firstly, we visually assessed the motion fields. The results of both subjects are shown in Figs. 4.1 and 4.3 on two critical time frames: at the /s/, when forward motion is prominent, and at the /k/, when upward motion is prominent (Fig. 4.1 is for healthy control at time frame /s/). Knowing that the internal muscular structure of tongue prevents its back from performing either too large or zero motion [13], at tongue’s back, we see E-IDEA has reduced the erroneous large motions for the healthy control, and has captured those small motions where IDEA mistakenly interpolates as zero for the patient. We also see E-IDEA can straighten up the motion at the top of the tongue to better estimate the displacement when the tongue hits the palate vertically (Figs. 4.3(a) and 4.3(d)). In general, the boundary estimation agrees more with tongue physical mechanics [13].

CHAPTER 4. ENHANCED INCOMPRESSIBLE MOTION ESTIMATION

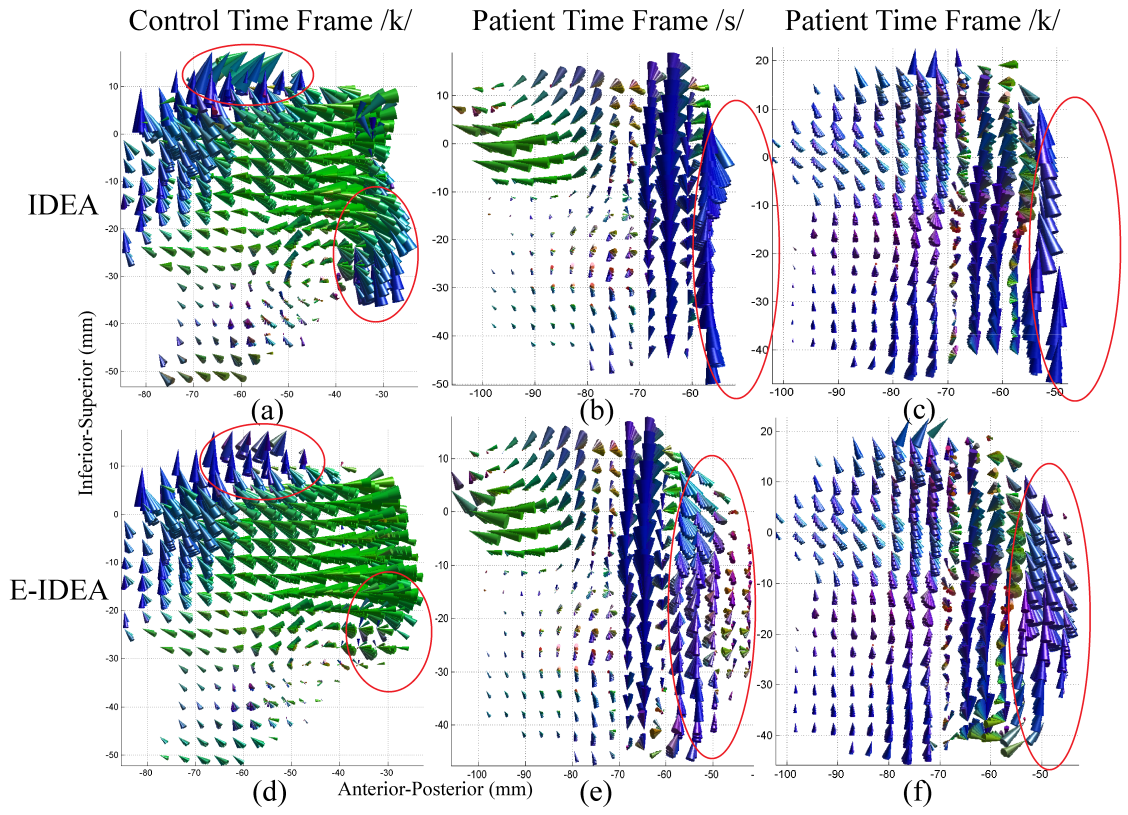


Figure 4.3: Visual comparison of conventional IDEA result and E-IDEA result.

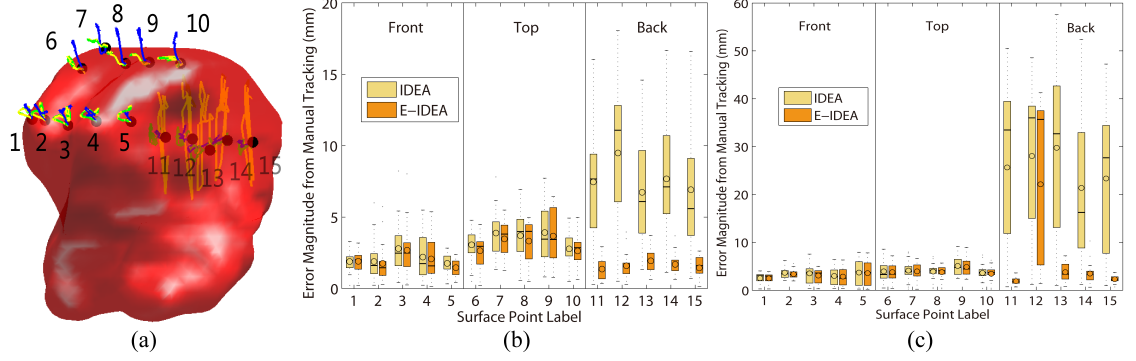


Figure 4.4: Comparison of IDEA and E-IDEA with manually tracked surface points. (a) Tracks of the healthy control’s surface points by manual (*blue*), IDEA (*yellow*), and E-IDEA (*green*). (b) Error magnitude for the healthy control (*bar* is median and *circle* is mean). (c) Error magnitude for the patient.

4.4 Point Tracking with E-IDEA

To obtain a numerical comparison, we manually tracked the motions of 15 surface points distributed 5 each on the front, top, and back parts of the tongue (labeled in Fig. 4.4(a)). We then computed their trajectories with IDEA and E-IDEA motion fields. The tracks of three methods are shown in Fig. 4.4(a) and errors from manual tracking at each point are shown in Figs. 4.4(b) and 4.4(c), boxplotted across all time frames. The error magnitude has been reduced by E-IDEA, especially on the back part of the tongue. Also, the mean error (circles in boxes) is reduced by E-IDEA at all 15 points. The improvement is significant (student t-test $p < 0.01$).

4.5 Evaluating Reprojection Error

We took the estimated 3D motions at input sample locations and reprojected them onto input directions using Eqns. 2.4, 2.5, and 4.3. We then computed a *reprojection error* that gives the error in distance in the input directions between the estimated sample components and the input sample components. This measure assumes input motion components (HARP and surface normal motions) are the truth. We compare four types of reprojection errors in histograms of Fig. 4.5: on IDEA internal points, on E-IDEA internal points, on E-IDEA boundary points, and on IDEA boundary points as indicated in the legend. For the healthy control, on a total of 105455 internal points and 108853 boundary points, the mean of the four errors are: 0.32 mm, 0.35 mm, 0.65 mm, and 1.33 mm, respectively. The boundary error has been reduced by 0.68 mm and the internal error has been raised by 0.03 mm. For the patient, on 133302 internal points and 100523 boundary points, the mean of the four errors are: 0.22 mm, 0.24 mm, 0.96 mm and 3.11 mm. The boundary error has been reduced by 2.15 mm and the internal error has been raised by 0.02 mm.

4.6 Discussion and Summary

In this chapter, we have proposed a novel algorithm for estimating the tongue's motion field in 3D. The major innovation is in the incorporation of surface motion as additional information, which compensates for the well-known deficiencies of HARP

CHAPTER 4. ENHANCED INCOMPRESSIBLE MOTION ESTIMATION

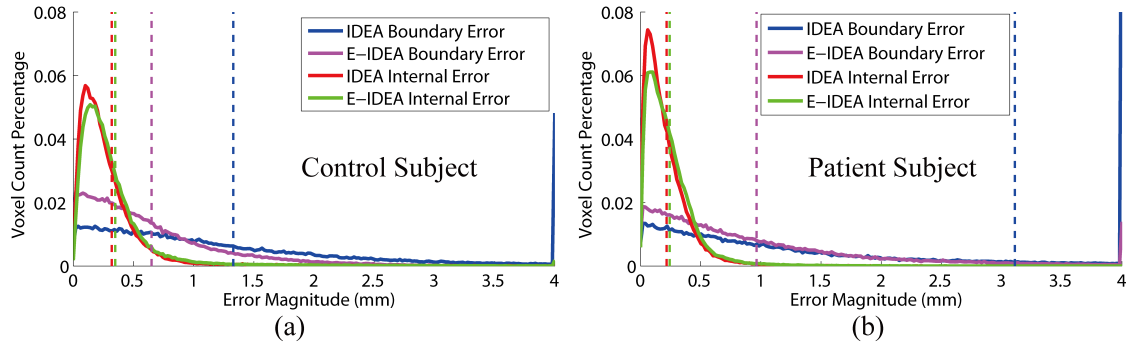


Figure 4.5: Regularized histogram of IDEA and E-IDEA's reprojection error on internal and surface points. *Dotted lines* show the mean of four types of reprojection error.

in estimating boundary motions. Both qualitative and quantitative improvements are evident using two independent metrics. Especially, from reprojection error, we see that boundary error is substantially reduced while internal error is only minimally increased.

Chapter 5

Phase Vector Incompressible Registration Algorithm

5.1 Introduction

Chapters 3 and 4 focused on building and improving a 3D motion analysis pipeline, whose core estimation algorithm is IDEA or E-IDEA. From previous experiments, we have observed that TMAP suffers from an over-complexity problem caused by multiple preprocessing steps and the IDEA framework’s multi-step divergence-free vector spline computation, which results in days of processing time for a single subject.

Stepping backwards to the very essence of estimation of a dense 3D motion field, since the acquired data is inherently sparse, interpolation is a natural and necessary step for achieving a desirable estimate. From existing literature summarized in Sec-

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

tion 1.4, all proposed 3D algorithms in the past have used interpolation of 2D motion estimated from sparse slices, which requires a combination of vector-valued components. On the other hand, the reversed process—interpolation of sparse scalar-valued image slices to a dense image volume followed by direct estimation of 3D motion—has never been explored. This forms the motivation of the novel method described in this chapter.

In this chapter, we propose a new method to compute 3D motion fields from sparse tagged MR images. Rather than interpolating the motion fields derived from the sparse images onto a denser grid, we turn the process around and first interpolate the raw tagged images onto a denser grid. We then apply the HARP method [23] on the interpolated tagged images to yield 3D harmonic phase volumes. Using the fact that harmonic phase is a material property of the tissue, we use these images in a multichannel image registration algorithm to track tissue points throughout the image time frames. The image registration algorithm is based on iLogDemons [54], but it contains several key differences. First, to drive the registration process we developed a new symmetric harmonic phase vector similarity metric. Second, to compensate for harmonic phase wrapping, we introduce a new phase interpolation method. Third, to enforce incompressibility only within tissue regions, we incorporate the harmonic magnitude image along with the divergence-free velocity constraint of iLogDemons. Finally, to simultaneously compute a consistent inverse motion, we incorporated the method of equivalent perturbation [72].

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

We call the proposed method PVIRA, which stands for *Phase Vector Incompressible Registration Algorithm*. PVIRA was evaluated using simulated tagged images, real data from a human brain in mild accelerations, and real data from the human tongue in speech. We compared its result against 3D HARP tracking and IDEA. The evaluations show a large reduction in computation time over IDEA and better robustness against noise as compared to both methods. PVIRA is demonstrated to yield a nearly incompressible result and to produce motion and inverse motion fields that are very nearly inverse-consistent.

5.2 Interpolation of Tagged Slices

PVIRA starts by using interpolation to create a dense grid of tagged images. We justify the use of interpolation on tagged images as follows. First, due to the nature of tagged imaging, the tag period (spacing between two consecutive tag lines) is usually set at a value much higher than the image resolution, e.g. 1.88 mm image resolution with a 12.00 mm tag period, which ensures that the tag pattern itself has a low frequency contribution to the image [14]. Second, when the tagged images are processed later with a bandpass filter (HARP filter), they also lose high frequency information before motion estimation [23]. Therefore, it is reasonable to interpolate the missing samples and form more tagged slices between the acquired sparse data.

We use tricubic b-spline interpolation [73] to produce an arbitrarily dense 3D grid

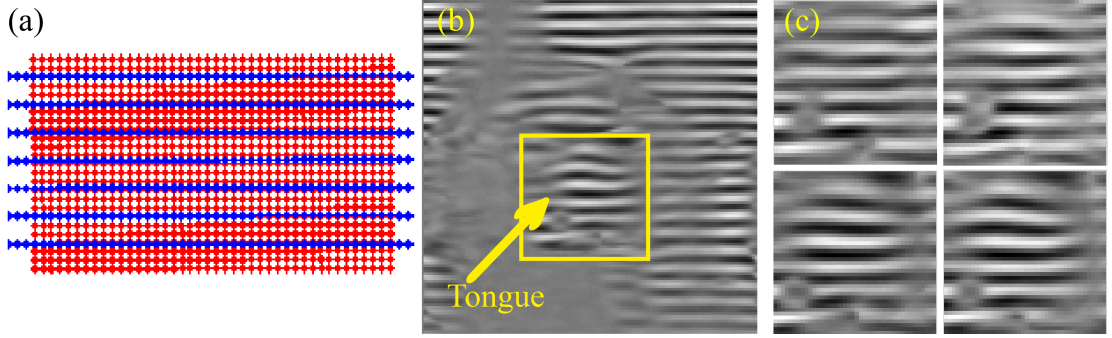


Figure 5.1: Demonstration of interpolation of tagged slices on tagged images of the tongue. (a) Spatial locations of acquired samples (blue) and interpolated samples (red). (b) A tagged sagittal tongue slice from acquisition. (c) Four interpolated sagittal tongue slices with horizontal tags.

of tagged images, as illustrated in Fig. 5.1. In this example, we let \mathbf{x}_s denote the locations of the sparsely acquired sagittal images $I_a(\mathbf{x}_s)$ having horizontal tag planes. These locations are shown as blue dots in Fig. 5.1(a). The tag planes in this sagittal acquisition are oriented in the axial direction, which is the reason that subscript a is used in the notation I_a . We denote the sample points in the dense 3D grid by \mathbf{x} , as shown using red dots in Fig. 5.1(a). Tricubic b-spline interpolation then finds the values $I_a(\mathbf{x})$. i.e.,

$$I_a(\mathbf{x}) = \sum_{\mathbf{x}_s} c(\mathbf{x}_s) \beta^3(\mathbf{x} - \mathbf{x}_s), \quad (5.1)$$

where $\beta^3(\mathbf{x})$ is the cubic b-spline interpolation kernel and $c(\mathbf{x}_s)$ are the interpolation coefficients computed from I_a [74]. Four interpolated horizontally-tagged sagittal images zoomed in on the tongue region are shown in Fig. 5.1(c).

In a similar fashion, the two axial acquisitions containing horizontal and vertical tags can be interpolated onto the same dense 3D grid to create the image $I_c(\mathbf{x})$ having

coronal tag planes and the image $I_s(\mathbf{x})$ having sagittal tag planes, respectively. The result of this interpolation step is a dense 3D volume in which each point \mathbf{x} has samples of three tagged images where the tags (prior to motion) are oriented in the three cardinal directions. This process is carried out independently at each acquired time frame, yielding a sequence of such 3D, vector-valued tagged images.

5.3 Harmonic Phase Volumes

While originally developed to analyze 2D images, the HARP framework is valid for 3D images, and has been previously used for 3D images that were acquired on a dense grid [26]. Following this strategy, our method performs HARP filtering on the three interpolated tag volumes I_a , I_c , and I_s . For example, for the volume $I_a(\mathbf{x})$, the complex image after HARP filtering can be denoted as

$$J_a(\mathbf{x}) = M_a(\mathbf{x})e^{j\Phi_a(\mathbf{x})}, \quad (5.2)$$

where $M_a(\mathbf{x})$ is the HARP magnitude volume and $\Phi_a(\mathbf{x})$ is the HARP phase volume. The same notation applies for coronally and sagittally tagged volumes, yielding $M_c(\mathbf{x})$, $\Phi_c(\mathbf{x})$, $M_s(\mathbf{x})$, and $\Phi_s(\mathbf{x})$.

5.4 Registration of Phase Volumes

Since harmonic phase is a material property, it can be used instead of brightness to associate voxels between image time frames. Furthermore, since there are three harmonic phase values at each time frame (and these arise from linearly independent directions), tracking all three phases solves the aperture problem of optical flow [70]. The approach of Ryf et al. [26] carried out such a 3D HARP tracking approach, based on the 2D fixed point HARP tracking method proposed in Osman et al. [23], to compute dense displacement fields in 3D tagged MR data. Since we have established such a dense volume by interpolation, we can use the same approach. This approach, which we call 3D HARP tracking, is used as a comparison algorithm in the following experiments. However, 3D HARP tracking does not incorporate global regularization, incompressibility, and inverse consistency, properties that we deem necessary for most applications in biological motion estimation. Therefore, we now develop a registration framework, which incorporates these desirable additional properties, to process the harmonic phase volumes.

We use the iLogDemons approach [54] as a starting point for PVIRA. Summarized in Algorithm 5.1, iLogDemons is an iterative method alternating between force-driven stepwise update and deformation field regularization. The result is an invertible motion field and incompressibility is applied in the process. Suppose that $\mathbf{u}(\mathbf{x})$ is the deformation field to be found to align the moving image $I_t(\mathbf{x})$ to the fixed image $I_0(\mathbf{x})$. The method iLogDemons maps the displacement field $\mathbf{u}(\mathbf{x})$ to its “log domain”

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

by using a single stationary velocity field $\mathbf{v}(\mathbf{x})$ to represent $\mathbf{u}(\mathbf{x})$. The relationship between $\mathbf{u}(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$ is that $\mathbf{u}(\mathbf{x})$ is the exponential map [75] of $\mathbf{v}(\mathbf{x})$, i.e., $\mathbf{u}(\mathbf{x}) = \exp(\mathbf{v}(\mathbf{x}))$. By minimizing the demon energy $\|I_0(\mathbf{x}) - I_t \circ \mathbf{d}(\mathbf{x})\|^2 + K\|\mathbf{d}\|^2$, an optimal update $\mathbf{d}(\mathbf{x})$ to the velocity field $\mathbf{v}(\mathbf{x})$ can be found. According to [76], the symmetric form of $\mathbf{d}(\mathbf{x})$ is given by

$$\mathbf{d}(\mathbf{x}) = \frac{2(I_0(\mathbf{x}) - I_t(\mathbf{x}))(\nabla I_0(\mathbf{x}) + \nabla I_t(\mathbf{x}))}{\|\nabla I_0(\mathbf{x}) + \nabla I_t(\mathbf{x})\|^2 + K(I_0(\mathbf{x}) - I_t(\mathbf{x}))^2}, \quad (5.3)$$

where K is a normalization factor. This equation is a variant of the classical result from optical flow motion tracking [77].

Algorithm 5.1. iLogDemons Registration Algorithm

1. Set the initial stationary velocity field $\mathbf{v}(\mathbf{x})$ to zero.
2. Compute the update velocity $\mathbf{d}(\mathbf{x})$ using Eqn. 5.3.
3. Fluid-like regularization: regularize $\mathbf{d}(\mathbf{x})$ with a Gaussian kernel.
4. Update $\mathbf{v}(\mathbf{x})$ with $\mathbf{d}(\mathbf{x})$ using the Baker-Campbell-Hausdorff (BCH) formula [54] and set the result to be the new velocity estimate $\mathbf{v}(\mathbf{x})$.
5. Elastic-like regularization: regularize $\mathbf{v}(\mathbf{x})$ with a Gaussian kernel.
6. Enforce incompressibility by solving Poisson's equation $\nabla^2 p = \nabla \cdot \mathbf{v}(\mathbf{x})$ and setting $\mathbf{v}(\mathbf{x}) - \nabla p$ to be the new $\mathbf{v}(\mathbf{x})$.
7. Go to Step 2. Repeat until convergence.

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

8. Compute the final displacement field using $\mathbf{u}(\mathbf{x}) = \exp(\mathbf{v}(\mathbf{x}))$.

Since iLogDemons works with regular intensity images, we modified its framework to use wrapped phase data. We note that a HARP phase volume $\Phi(\mathbf{x})$ is in fact a true phase volume $\Theta(\mathbf{x})$ wrapped into the range of $[-\pi, \pi)$, i.e.,

$$\Phi(\mathbf{x}) = W(\Theta(\mathbf{x})), \quad (5.4)$$

where

$$W(\Theta) = \text{mod}(\Theta + \pi, 2\pi) - \pi. \quad (5.5)$$

This is illustrated in Fig. 5.2. Although phase unwrapping [78–80] could potentially recover the true unwrapped phase, it is known that global phase unwrapping of images is problematic [81]. Therefore, rather than attempting to carry out global phase unwrapping we reformulate the required mathematical operations on HARP phases as follows.

The first operation of interest is the calculation of phase difference. Consider the two (wrapped) phases $\Phi_1 = W(\Theta_1)$ and $\Phi_2 = W(\Theta_2)$. If the underlying true phases satisfy $|\Theta_1 - \Theta_2| < \pi$, then

$$\Theta_1 - \Theta_2 = W(\Phi_1 - \Phi_2). \quad (5.6)$$

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

This implies that carrying out subtraction on observed harmonic phases followed by an explicit rewrapping operation will recover the true phase difference when it is small enough. This can be satisfied for the tissue motion when tags do not deform more than half a period, a condition that normally holds for tagged MR acquisitions in the brain, tongue, and heart. We call this a *small motion condition*.

The second operation of interest is the calculation of the spatial gradient of phase. The true gradient can be recovered from the wrapped phase with a mathematical trick denoted as ∇^* [23]:

$$\begin{aligned} \nabla\Theta(\mathbf{x}) &= \nabla^*\Phi(\mathbf{x}) \\ &= \begin{cases} \nabla\Phi(\mathbf{x}), & \text{if } |\nabla\Phi(\mathbf{x})| \leq |\nabla W(\Phi(\mathbf{x}) + \pi)|, \\ \nabla W(\Phi(\mathbf{x}) + \pi), & \text{otherwise.} \end{cases} \end{aligned} \quad (5.7)$$

This procedure computes the gradient on both the original phase function and on the phase that has been shifted by π and then re-wrapped. The correct gradient is the one with smaller magnitude.

For the three HARP phase volumes, we redefine the demon energy as

$$\begin{aligned} E_t(\mathbf{d}(\mathbf{x})) &= ||\Phi_{a0}(\mathbf{x}) - \Phi_{at} \circ \mathbf{d}(\mathbf{x})||^2 + ||\Phi_{c0}(\mathbf{x}) - \Phi_{ct} \circ \mathbf{d}(\mathbf{x})||^2 \\ &\quad + ||\Phi_{s0}(\mathbf{x}) - \Phi_{st} \circ \mathbf{d}(\mathbf{x})||^2 + K||\mathbf{d}||^2 \end{aligned} \quad (5.8)$$

and minimize it in a similar fashion as Eqn. 5.3. The subscripts 0 and t denote time frame numbers. Note that all three pairs of phase volumes with three tag directions

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

are used simultaneously because each pair provides a main motion component in x , y , and z . We assume equal weights of the three pairs because they contribute equally from three directions. We follow an analogous strategy as that in [52] to find the $\mathbf{d}(\mathbf{x})$ that minimizes E_t . It yields the proposed update velocity field in the demons framework:

$$\begin{aligned}
 \mathbf{d}(\mathbf{x}) &= 2\mathbf{d}_0(\mathbf{x})/(d_1(\mathbf{x}) + d_2(\mathbf{x})/K) , \text{ where} \\
 \mathbf{d}_0(\mathbf{x}) &= W(\Phi_{a0}(\mathbf{x}) - \Phi_{at}(\mathbf{x}))(\nabla^* \Phi_{a0}(\mathbf{x}) + \nabla^* \Phi_{at}(\mathbf{x})) \\
 &\quad + W(\Phi_{s0}(\mathbf{x}) - \Phi_{st}(\mathbf{x}))(\nabla^* \Phi_{s0}(\mathbf{x}) + \nabla^* \Phi_{st}(\mathbf{x})) \\
 &\quad + W(\Phi_{c0}(\mathbf{x}) - \Phi_{ct}(\mathbf{x}))(\nabla^* \Phi_{c0}(\mathbf{x}) + \nabla^* \Phi_{ct}(\mathbf{x})) , \\
 d_1(\mathbf{x}) &= \|\nabla^* \Phi_{a0}(\mathbf{x}) + \nabla^* \Phi_{at}(\mathbf{x})\|^2 \\
 &\quad + \|\nabla^* \Phi_{s0}(\mathbf{x}) + \nabla^* \Phi_{st}(\mathbf{x})\|^2 \\
 &\quad + \|\nabla^* \Phi_{c0}(\mathbf{x}) + \nabla^* \Phi_{ct}(\mathbf{x})\|^2 , \\
 d_2(\mathbf{x}) &= W(\Phi_{a0}(\mathbf{x}) - \Phi_{at}(\mathbf{x}))^2 + W(\Phi_{s0}(\mathbf{x}) - \Phi_{st}(\mathbf{x}))^2 \\
 &\quad + W(\Phi_{c0}(\mathbf{x}) - \Phi_{ct}(\mathbf{x}))^2 .
 \end{aligned} \tag{5.9}$$

This metric is similar to previous multi-channel image registration metrics [82], but it is specifically adapted to wrapped HARP phase. In particular, the difference and gradient operations have been modified to deal with phase wrapping.

5.5 Deformation of Phase Volumes

At each demons iteration step, the current field estimate is used to deform the moving phase volumes $\Phi_{at}(\mathbf{x})$, $\Phi_{ct}(\mathbf{x})$, and $\Phi_{st}(\mathbf{x})$ to $\Phi_{at}(\mathbf{u}(\mathbf{x}))$, $\Phi_{ct}(\mathbf{u}(\mathbf{x}))$, and $\Phi_{st}(\mathbf{u}(\mathbf{x}))$. During this deformation, interpolation is needed to find sub-voxel values of the deformed phase and the effect of phase wrapping must be considered. For example, as shown in Fig. 5.2, suppose the phase value at sub-voxel location x_0 needs to be found from the known phases $\Phi(x_1)$ and $\Phi(x_2)$ at two neighbor voxels x_1 and x_2 , linear interpolation gives $\hat{\Phi}(x_0) = (\Phi(x_2) - \Phi(x_1))(x_0 - x_1)/(x_2 - x_1)$. And in this case the value $\hat{\Phi}(x_0)$ equals to the real phase $\Theta(x_0)$ because no wrapping is involved. On the other hand, at location y_0 , since $\Phi(y_2)$ has been wrapped from the true value $\Theta(y_2)$, $(\Phi(y_2) - \Phi(y_1))(y_0 - y_1)/(y_2 - y_1)$ yields a wrong result, while the correct value is computed by unwrapping $\Phi(y_2)$ to $\Phi(y_2) + 2\pi$.

We note that $(\Phi(y_2) + 2\pi) - \Phi(y_1) = W(\Phi(y_2) - \Phi(y_1))$. Thus, whenever an abnormal phase difference greater than 2π is caused by wrapping, under the small motion condition (discussed in the previous section), the abnormal difference must be a jump of $\pm 2\pi$. Therefore, it can be removed by wrapping the difference. In summary, when deforming phase volumes and using linear interpolation to find a sub-voxel value, every difference operation needs to be computed with wrapping, i.e.,

$$\hat{\Phi}(y_0) = W\left(\frac{y_0 - y_1}{y_2 - y_1} \cdot W(\Phi(y_2) - \Phi(y_1))\right). \quad (5.10)$$

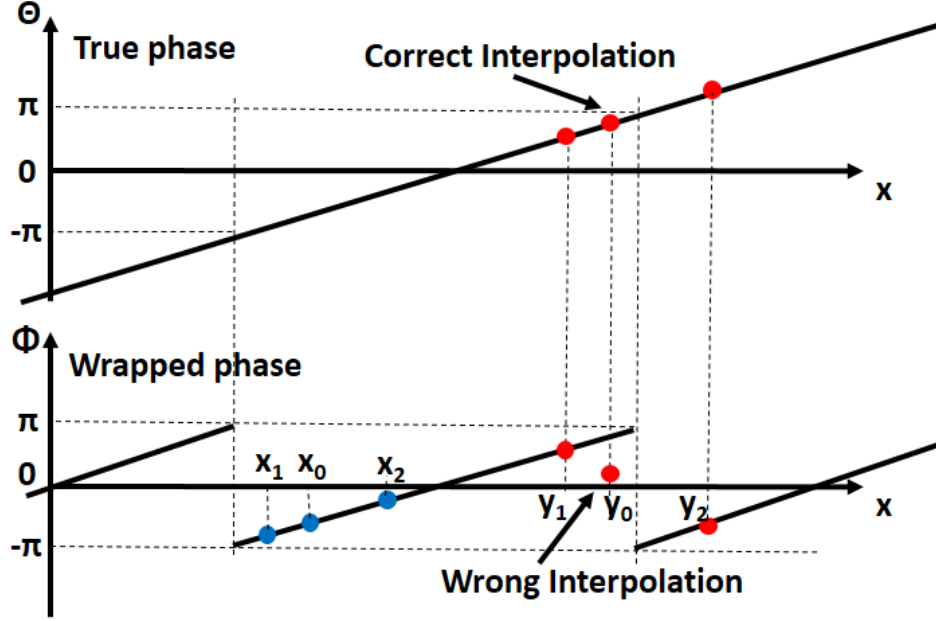


Figure 5.2: Linear interpolation of wrapped phase values.

This is also the reason that linear interpolation is used to deform the phase volumes, because either bilinear or trilinear interpolation can be separated by two or three consecutive one-dimensional operations using Eqn. 5.10 in the corresponding directions.

5.6 Incorporation of Incompressibility

In iLogDemons [54], incompressibility is enforced at every iteration by computing the “divergence part” of the estimate and removing it, i.e., solve Poisson’s equation $\nabla^2 p = \nabla \cdot \mathbf{v}(\mathbf{x})$ to find the divergence motion $\mathbf{v}_d(\mathbf{x}) = \nabla p$ and compute the incompressible motion by $\mathbf{v}(\mathbf{x}) - \mathbf{v}_d(\mathbf{x})$. In PVIRA, since only the tissue region is incom-

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

pressible, we regularize the previously generated HARP magnitude volumes $M_a(\mathbf{x})$, $M_c(\mathbf{x})$, and $M_s(\mathbf{x})$ to the range of $[0,1]$ and use their mean $M(\mathbf{x})$ as a weighted mask to specify the region of incompressibility. Thus $\mathbf{v}(\mathbf{x})$ is updated by

$$\mathbf{v}(\mathbf{x}) \longleftarrow \mathbf{v}(\mathbf{x}) - M(\mathbf{x})\mathbf{v}_d(\mathbf{x}) . \quad (5.11)$$

Since HARP magnitude is computed at the same time as HARP phase, this strategy removes the requirement of a manual or automated segmentation step, as often required in other tag tracking approaches. Only at the location where M has a high value indicating tissue existence is incompressibility automatically enforced.

5.7 Incorporation of Inverse Estimation

Since the estimated motion is a diffeomorphism, an inverse-consistent backward field can be found. The method of equivalent perturbation minimizes a symmetry cost function to guarantee inverse consistency [72]. It gradually builds the inverse field $\mathbf{u}_{-1}(\mathbf{x})$ from every iteration by computing the additive change $\boldsymbol{\xi}(\mathbf{x})$ to the forward field estimate $\mathbf{u}(\mathbf{x})$ and regards $\boldsymbol{\xi}(\mathbf{x})$ as an infinitesimal perturbation. Suppose the forward field at the n -th iteration is $\mathbf{u}^{(n)}(\mathbf{x})$; then the backward field $\mathbf{u}_{-1}^{(n)}(\mathbf{x})$ at the

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

n -th iteration is computed by

$$\begin{aligned}\boldsymbol{\xi}(\mathbf{x}) &= \mathbf{u}^{(n)}(\mathbf{x}) - \mathbf{u}^{(n-1)}(\mathbf{x}), \\ \mathbf{u}_{-1}^{(n)}(\mathbf{x}) &= \mathbf{u}_{-1}^{(n-1)}(\mathbf{x}) - D(\mathbf{x} + \mathbf{u}_{-1}^{(n-1)}(\mathbf{x})) \cdot \boldsymbol{\xi}(\mathbf{x} + \mathbf{u}_{-1}^{(n-1)}(\mathbf{x})),\end{aligned}\tag{5.12}$$

where $D(\mathbf{x})$ is the Jacobian matrix of \mathbf{x} . This is a direct application from [72], thus we omit the derivation. We integrate this step within every iteration to simultaneously compute the inverse motion.

PVIRA is summarized in Algorithm 5.2.

Algorithm 5.2. Phase Vector Incompressible Registration Algorithm (PVIRA)

1. Interpolate the three sets of sparsely acquired tagged images to yield three dense tagged volumes using Eqn. 5.1.
2. Apply HARP filtering using Eqn. 5.2 on every tagged volume to get HARP phase and HARP magnitude volumes.
3. Set the initial stationary velocity field $\mathbf{v}(\mathbf{x})$ to zero.
4. Compute the update velocity $\mathbf{d}(\mathbf{x})$ using Eqn. 5.9. Difference and gradient computations must follow Eqns. 5.6 and 5.7.
5. Fluid-like regularization: regularize $\mathbf{d}(\mathbf{x})$ with a Gaussian kernel.
6. Update $\mathbf{v}(\mathbf{x})$ with $\mathbf{d}(\mathbf{x})$ using the BCH formula and set the result to be the new velocity estimate $\mathbf{v}(\mathbf{x})$.
7. Elastic-like regularization: regularize $\mathbf{v}(\mathbf{x})$ with a Gaussian kernel.

8. Enforce incompressibility using Eqn. 5.11.
 9. Compute the current motion estimate using $\mathbf{u}(\mathbf{x}) = \exp(\mathbf{v}(\mathbf{x}))$. Then use Eqn. 5.12 to update the inverse field.
 10. Go to Step 4. Repeat until convergence.
 11. Compute the final motion estimate using $\mathbf{u}(\mathbf{x}) = \exp(\mathbf{v}(\mathbf{x}))$.
-

5.8 Simulation of Brain and Tongue Deformation

First, we evaluated PVIRA with simulated tagged images. In current studies, since true motion of real data is difficult to learn beforehand, simulation plays a very important role in evaluating a new algorithm. We simulated both the brain deformation in a mild rotation and the tongue deformation in a protruding motion.

For the brain, a $64 \times 64 \times 64$ volume was generated with 1.0 mm voxel resolution and 10.0 mm tag period. The tissue exists on a cylinder-shaped region with a circular cross-section in the x - y plane and is isometric in the z direction. Synthetic displacement fields were generated by a finite element simulation (COMSOL v4.3, COMSOL Multiphysics, Burlington, MA) of a nearly-incompressible soft material (11.2 cm diameter, 18 cm long, Youngs modulus $E = 5000$, Poissons ratio $\nu = 0.49$). The outer boundary of the cylinder was subjected to a half-sinusoidal angular acceleration pulse.

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

Simulated displacements were x - y in-plane rotations around the center, as shown in Fig. 5.3(a). This increasing rotation lasted for 18 time frames and yielded a 4.8 mm maximum displacement, satisfying the small motion condition. At every time frame, the motion field was used to deform horizontally and vertically tagged synthetic volumes. Fig. 5.3(d) shows the x - y cross section of the simulated tagged volumes before and after deformation at one time frame. These volumes were processed with PVIRA, yielding the motion estimate shown in Fig. 5.3(b), and its error magnitude from truth was shown in Fig. 5.3(e). It can be observed that the error in most voxels is below 0.2 mm. Fig. 5.3(c) shows the simultaneously generated inverse field and Fig. 5.3(f) is the result when the inverse field was composed with the forward field. We expect the error to be close to 0 to indicate a good inverse estimation. Moreover, Fig. 5.3(g) shows the Jacobian determinant on this cross section. The Jacobian values are close to 1, which indicates incompressibility except some boundary effects. Besides this particular time frame, the estimation error from the truth of all 18 time frames are box-plotted in Fig. 5.3(h), where the center bar indicates the median with a 25 and 75 percentile box. The medians are all lower than 0.1 voxel. It is observed that all outliers are errors coming from the boundaries as indicated by the red spots in Fig. 5.3(e).

We also used this simulation to test two other methods: 3D HARP and IDEA. Note that IDEA is designed specifically to process sparse data instead of dense volumes. Thus we deleted five middle slices out of every six slices from the dense volume

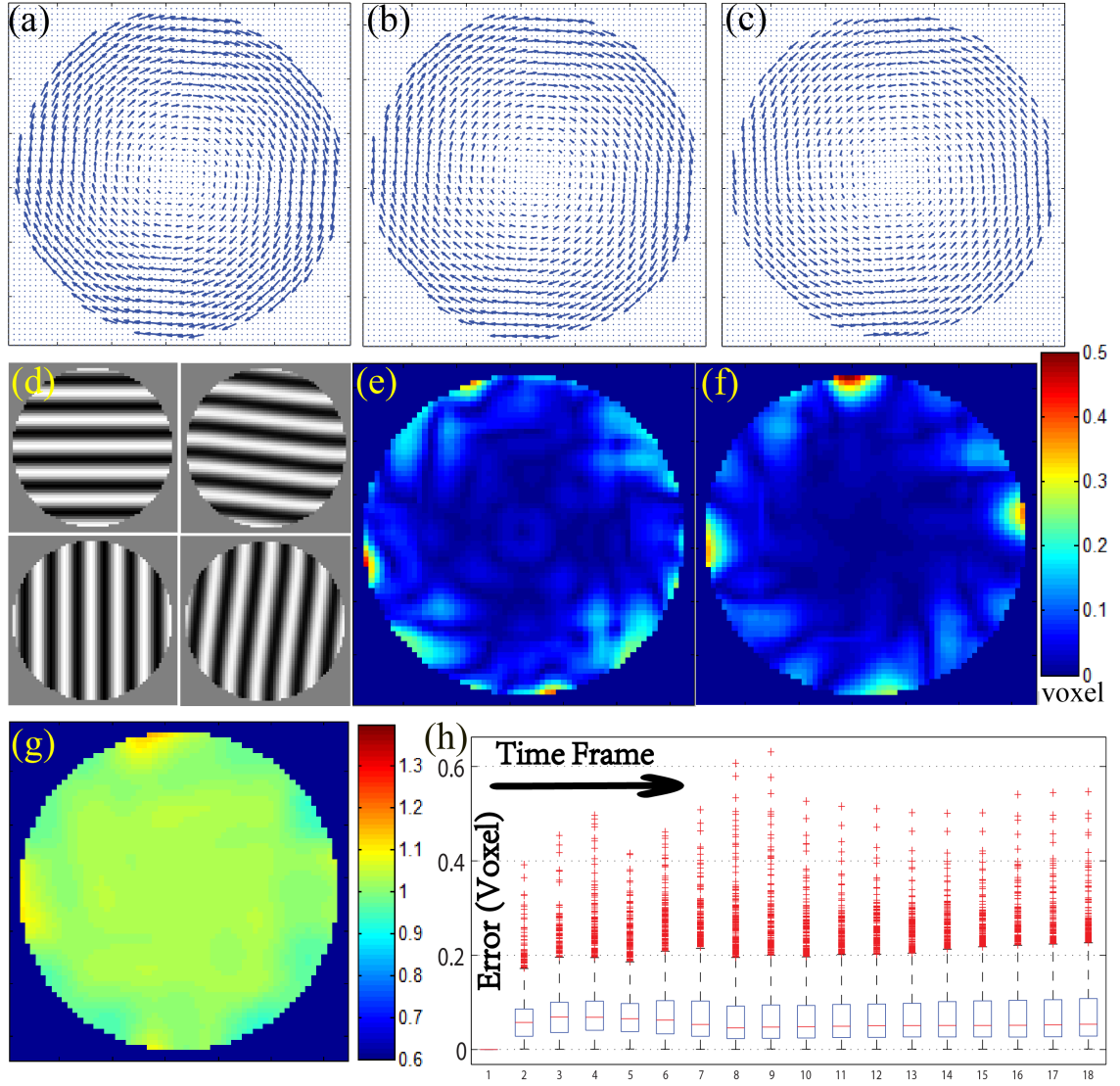


Figure 5.3: Test of PVIRA on simulated brain rotations. (a) Simulated true motion. (b) Estimated motion. (c) Estimated inverse motion. (d) Simulated tagged images. (e) Error magnitude of the estimation. (f) Composed forward-backward motion in magnitude. (g) Jacobian determinant of motion estimate. (h) Box-plotted estimation error of all time frames.

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

Table 5.1: Brain Rotation Simulation Test of Three Methods

	PVIRA	3D HARP	IDEA
Estimation Error (Voxel)	0.07 ± 0.06	0.06 ± 0.06	0.08 ± 0.07
Inversion Error (Voxel)	0.13 ± 0.14	0.11 ± 0.13	0.12 ± 0.13
Jacobian Determinant	1.01 ± 0.03	0.98 ± 0.14	1.00 ± 0.05

to create sparse data for IDEA processing. Both 3D HARP and IDEA provided good results that were visually similar to the truth. We computed the estimation error magnitude, composed inversion error magnitude, and Jacobian determinant of all three methods and listed their statistics over all time frames in Table 5.1. The mean estimation error of IDEA was slightly higher than the other two methods, but all were under a tenth of a voxel. Moreover, when computing the inverse motion, since only PVIRA automatically provides an inverse field, we used an extra step of a fixed point method [36] to numerically compute the inverse of 3D HARP and IDEA. The three composition errors are around the same level (Table 5.1). Lastly, although only PVIRA and IDEA are incompressible methods, the mean Jacobian of 3D HARP is also close to 1 because the simulated rotation field was essentially incompressible and 3D HARP provided an estimation close to the simulation. But its local incompressibility fluctuated, causing a larger standard deviation. Overall, we can roughly conclude that under the simulated brain deformation, PVIRA contains all the desirable properties of HARP and IDEA while maintaining an accuracy close to both. We note that on the same computer with an 8 GB memory and a 64-bit Windows system, PVIRA took on average 89 seconds to process one volume, while 3D HARP took 144 seconds

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

and IDEA took 527 seconds.

Next, we added normally distributed Gaussian noise to the simulation. We tested ten scenarios when the noise energy was raised step-wisely from 0.1 to 1.0 (by increasing Gaussian variance) so that the signal to noise ratio (SNR) decreased from 20 to 0 (see Fig. 5.4(h)). An example of tagged volumes at SNR 8 are shown in Fig. 5.4(g) and the corresponding estimated motion fields from three methods are shown in Figs. 5.4(a), 5.4(b), and 5.4(c). Visually, from these motion estimates and their error magnitudes shown in Figs. 5.4(d), 5.4(e), and 5.4(f), the PVIRA result is closer to the truth shown in Fig. 5.3(a) due to the contribution from its internal regularization. The error magnitudes at different SNR levels are boxed-plotted in Fig. 5.4(h). The error grows rapidly as noise increases. However, the PVIRA error stays lower than both 3D HARP and IDEA in all circumstances: both its mean and median are lower than the other two methods and it has fewer number of outlier voxels with larger errors. In this simulation, PVIRA is more accurate than both 3D HARP and IDEA (a student t-test on all noise levels indicated $p < 0.01$).

Since regularization of PVIRA plays an important role, the parameter selection of its smoothing Gaussian kernel needs to be studied. Apparently regularization helps in reducing the effect of noise, but over-smoothing may cause problems as well. Since the current algorithm does not contain an automatic parameter optimization process, the smoothing parameter needs to be manually specified. We tested the previous ten noise levels when the SNR ranges from 0 to 20. With each noise level, different

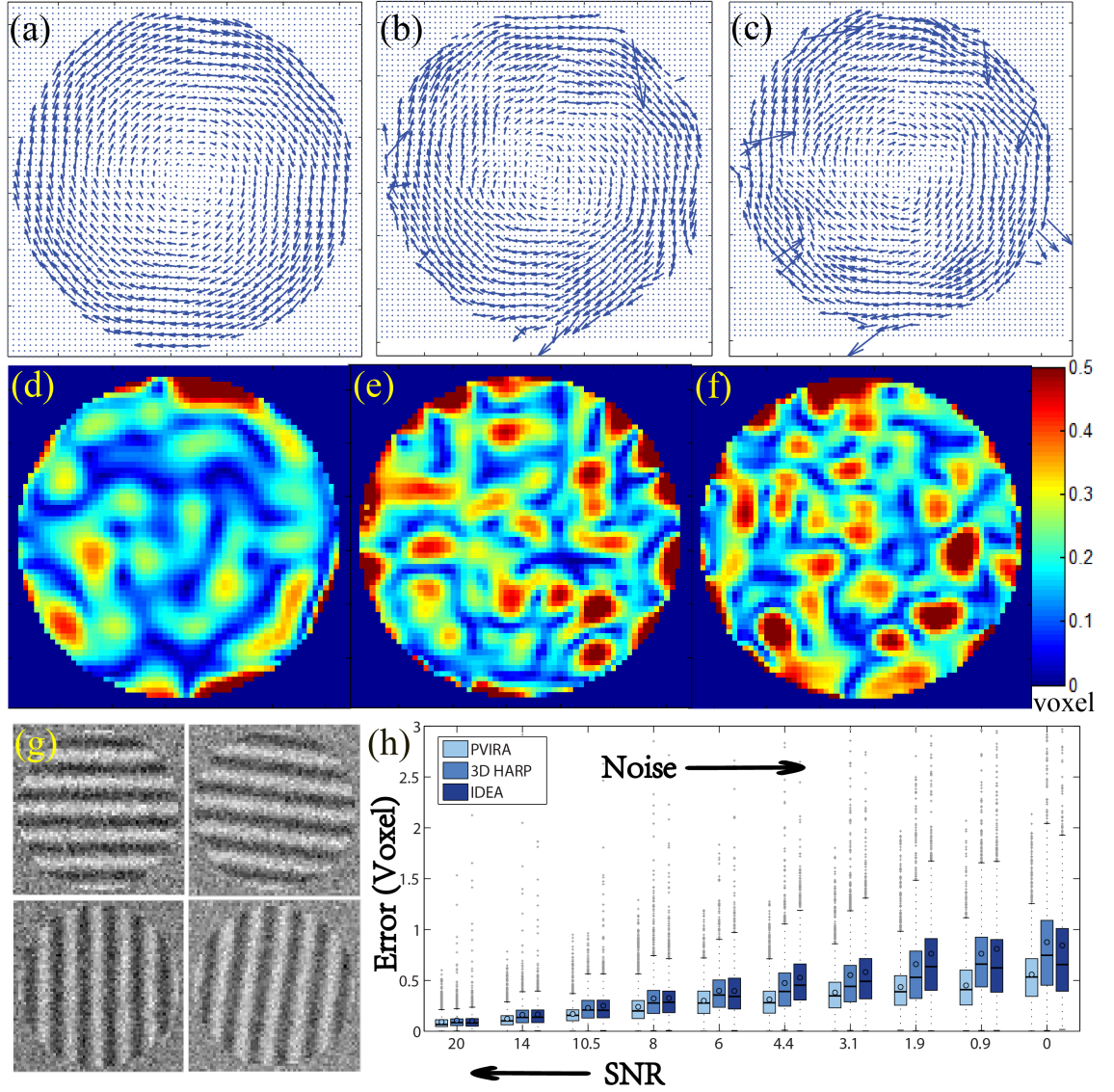


Figure 5.4: Performance of three methods under noise. (a) PVIRA estimation. (b) 3D HARP estimation. (c) IDEA estimation. (d) PVIRA error. (e) 3D HARP error. (f) IDEA error. (g) Tagged images with noise. (h) Box-plotted estimation error under difference noise level. Center bar is median and circle is mean.

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

regularization parameters were selected. In this test, for single parameter control we used only one level of regularization (with zero elastic-like regularization). We tuned the fluid-like regularization where the variance of the Gaussian kernel was set at 0, 2, 4, 6, 8, 10, and 12—from no regularization to strong smoothing. PVIRA was performed with all parameters and all noise levels and the mean estimation error was computed and plotted in Fig. 5.5. Note that for each noise level, since the generation of noise is random, in order to increase sample size, the entire experiment was repeated twenty times, each with a new random noise generation, and Fig. 5.5 shows the mean of all repetitions. Apparently, when the noise level is high, the estimation error is reduced with stronger regularization ($\sigma = 12$). However, this stronger regularization also seems to be over-smoothing when the noise level is low, causing a reduced accuracy. A σ value of 6 or 8 appears to be well-balanced, being able to reduce error under high noise while keeping a lower error under low noise. In practice, we selected $\sigma = 6$ for all experiments.

Since all simulations above focused on the estimation of x - y motion and were homogeneous in the z direction, the next test was performed on simulated tongue images containing larger amount of motion in all three directions. We simulated a forwardly-protruding tongue motion by interpolating nodal displacement results from a forward finite element simulation under muscular activation onto a $60 \times 40 \times 50$ lattice in ten time frames. The model geometry was derived from the ArtiSynth biomechanical modeling toolkit [83], and was solved using the FEBio software suite [84]. Material co-

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

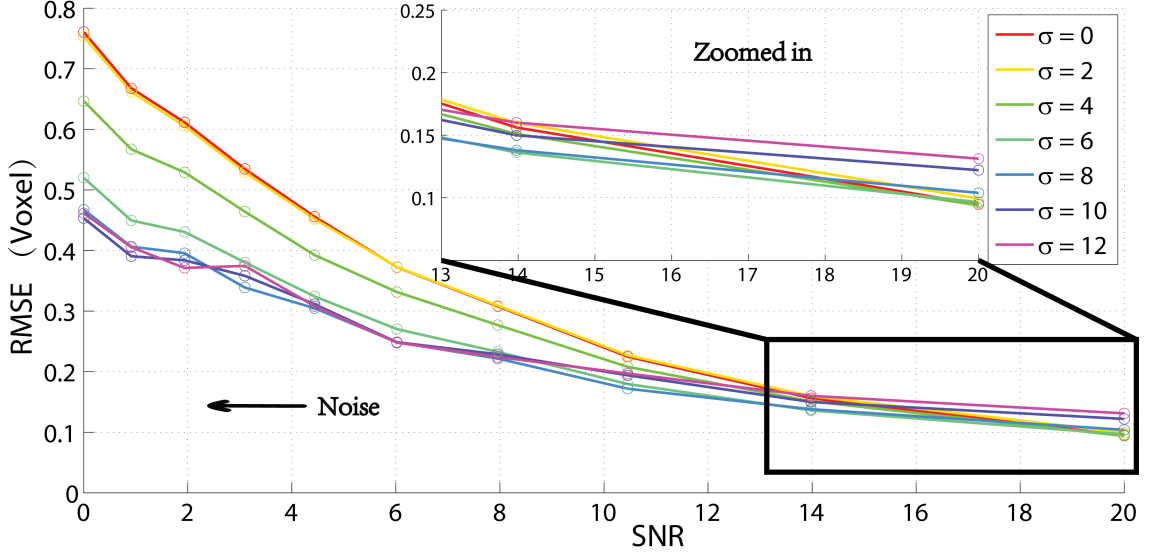


Figure 5.5: PVIRA estimation error with different regularizations and noise levels. σ is the size of the two regularizing Gaussian kernels.

efficients were manually adjusted to generate displacements similar to the magnitude expected in a live subject and to enforce incompressibility. We performed the same tests as in the previous simulation by using all three methods first and then computed their estimation error, inversion error, and Jacobian determinant. An example of one axial slice and one sagittal slice is shown in Fig. 5.6. In this simple motion when the tongue tip was protruding forward and downward, the tongue was using x and z motion to expand forward while using y motion to compress from left and right sides. Fig. 5.6(c) shows that all three methods give good estimates within the body of the tongue, but 3D HARP and IDEA suffer from stronger boundary effects. From Fig. 5.6(b), we can see that all three methods show large errors on the boundary, and IDEA clearly shows some planar artifacts due to the fact that it works with sparse sample slices.

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

Table 5.2: Tongue Protrusion Simulation Test of Three Methods

	PVIRA	3D HARP	IDEA
Estimation Error (Voxel)	0.13 ± 0.17	0.19 ± 0.81	0.21 ± 0.69
Inversion Error (Voxel)	0.12 ± 0.12	0.15 ± 0.26	0.14 ± 0.23
Jacobian Determinant	1.00 ± 0.07	1.05 ± 0.55	1.00 ± 0.11

Table 5.2 lists the average errors of the three methods over all time frames. PVIRA is more accurate on average and has less variance. IDEA is less accurate due to the artifacts from sparse data. However, IDEA and PVIRA are both more incompressible than 3D HARP. With the same computer mentioned above, PVIRA took on average 107 seconds to process one volume, while 3D HARP took 112 seconds and IDEA took 577 seconds.

5.9 Estimation of Brain Motion in Mild Accelerations

We used 3D HARP, IDEA, and PVIRA to estimate the 3D motion of the subjects in the mild brain acceleration study. Two examples of a subject's motion estimate at time frame 3 (strongest left rotation) and time frame 7 (strongest right rotation) are shown in Figs. 5.7(a) to 5.7(f). Visually, the three estimates are similar, but 3D HARP has worse boundary effects. We also computed the Jacobian determinant to check incompressibility (see Fig. 5.7(i) for example). Since the true motion was unknown, to compare the difference between the two incompressible methods (PVIRA

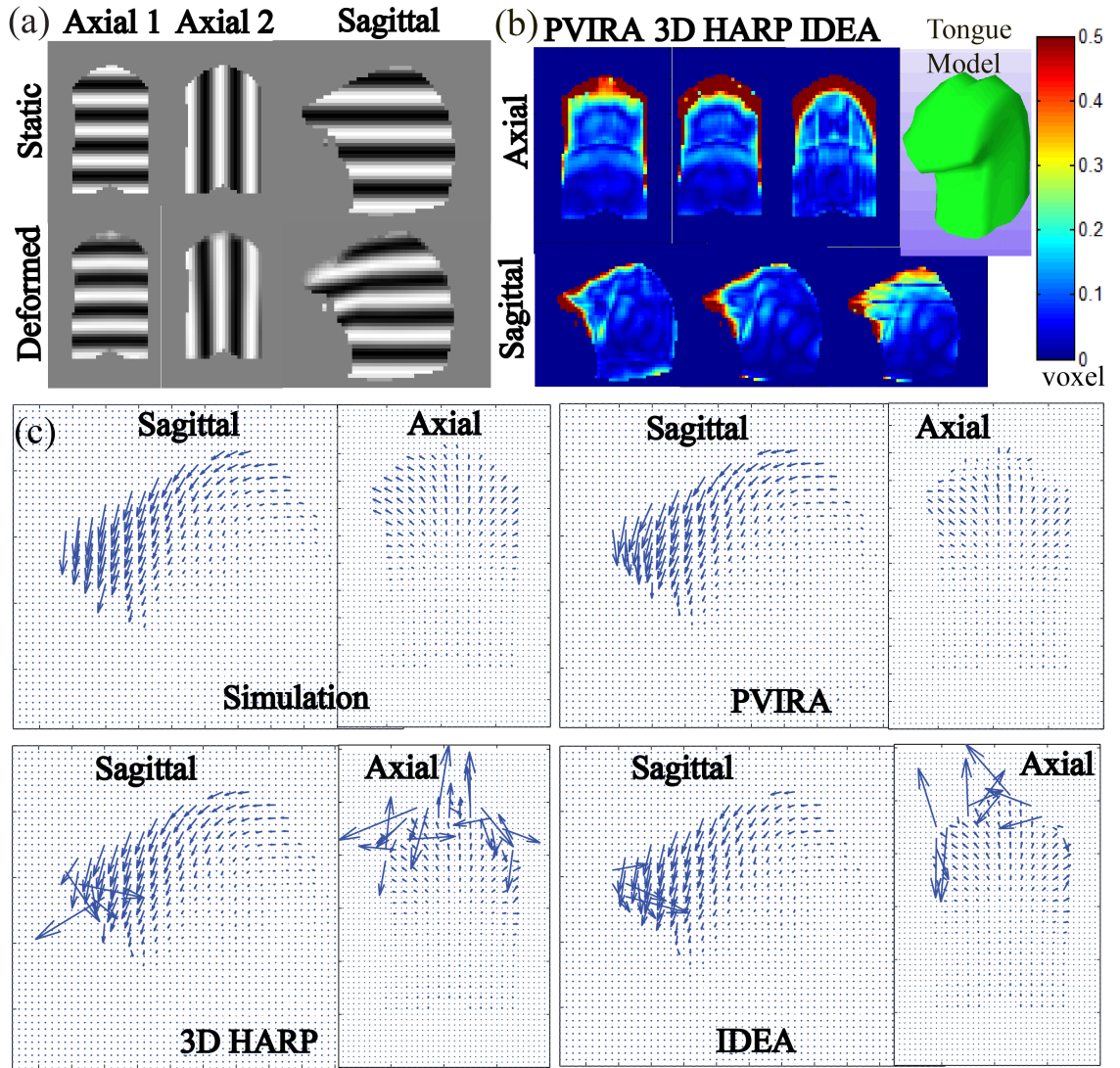


Figure 5.6: Test of PVIRA on simulated tongue protrusion. (a) Tagged slices with three tag directions. (b) Estimation error with three methods. (c) Simulated motion and estimated motion with three methods.

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

Table 5.3: Brain Acceleration Motion Estimation from Three Methods

	3D HARP	IDEA	PVIRA
HARP Difference (mm)	0	0.84 ± 1.80	0.62 ± 1.24
Jacobian Determinant	1.11 ± 0.82	1.01 ± 0.27	1.01 ± 0.26
Computing Time (min)	4.9	141.6	4.7

and IDEA), we used the 3D HARP result as a reference and computed the other two methods' differences from it (see Fig. 5.7(h) for example). We plotted the normalized histogram of the two differences in Fig. 5.7(j). PVIRA has more voxels that have smaller difference and its mean difference is also smaller than that of IDEA. Especially, this was the case for all 33 volumes of the three human subjects. We conclude that PVIRA result is closer to HARP than IDEA (student t-test on all volumes indicated $p < 0.05$). The mean difference, Jacobian determinant, and average computing time for all 33 volumes are shown in Table 5.3.

5.10 Estimation of Tongue Motion in Speech

We used 3D HARP, IDEA, and PVIRA to estimate the 3D motion of two subjects from the tongue motion study. Excluding time frame 1, a total of 50 volumes was evaluated. Two examples of a subject's motion estimate at time frame 8 (strongest forward motion) and time frame 18 (strongest upward motion) are shown in Figs. 5.8(a) to 5.8(f). In the tongue motion, the boundary effect of 3D HARP was stronger than that of the brain. We also computed the difference from HARP using

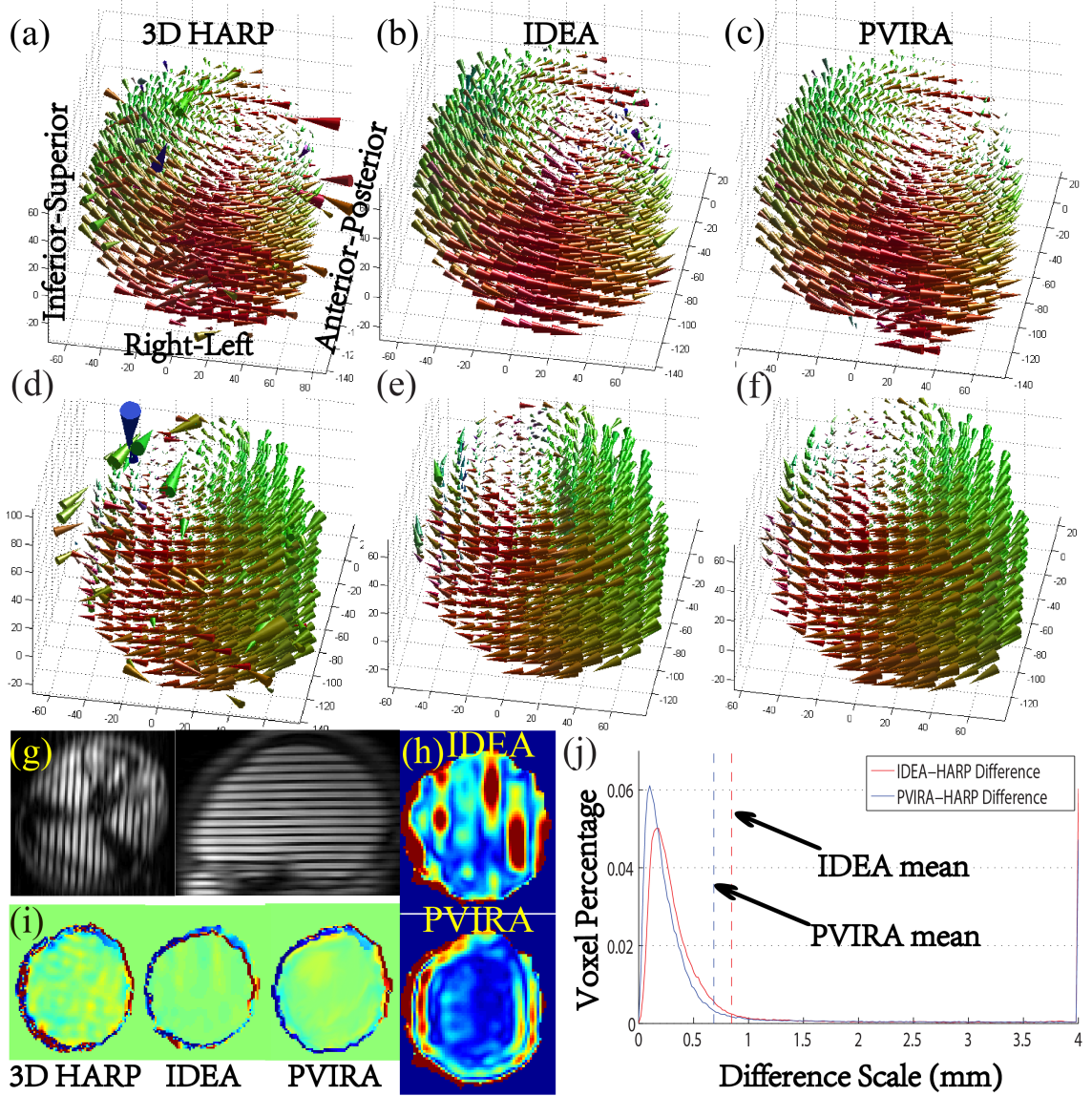


Figure 5.7: Brain mild acceleration motion estimation. (a)-(c) Left rotation estimates. (d)-(f) Right rotation estimates. (g) Examples of tagged axial and sagittal slices. (h) Example slices of estimation difference from HARP. Colorbar goes from 0 to 0.5 as in the previous figures. (i) Example slices of Jacobian determinant. (j) Histogram of difference from HARP using IDEA and PVIRA.

Table 5.4: Tongue Motion Estimation from Three Methods

	3D HARP	IDEA	PVIRA
HARP Difference (mm)	0	1.63 ± 1.70	0.97 ± 1.48
Jacobian Determinant	1.21 ± 1.83	1.00 ± 0.37	1.00 ± 0.21
Computing Time (min)	0.3	9.1	0.3

the other two methods (see Fig. 5.8(h) for example) and its histogram is plotted in Fig. 5.8(j). PVIRA result is also closer to HARP than IDEA for all 50 volumes from two subjects (student t-test on all volumes indicated $p < 0.05$). The mean difference, Jacobian determinant, and average one-volume computing time for 50 volumes are shown in Table 5.4.

5.11 Discussion

In all tests, PVIRA demonstrated comparable accuracy to HARP and IDEA although it was computed by a fundamentally different process. However, 3D HARP is not preferred over the other two methods because of its lack of physical constraints, especially incompressibility. In terms of IDEA, to achieve the final estimation, 2D HARP tracking, 3D super-resolution reconstruction, and 3D image segmentation are needed that require multiple preprocessing steps. In our tests, these preprocessing steps took about 3 hours per subject on average. Furthermore, the execution time of IDEA was much slower than PVIRA due to its multi-step divergence-free vector spline process. IDEA took days to process all subjects while PVIRA took hours. We

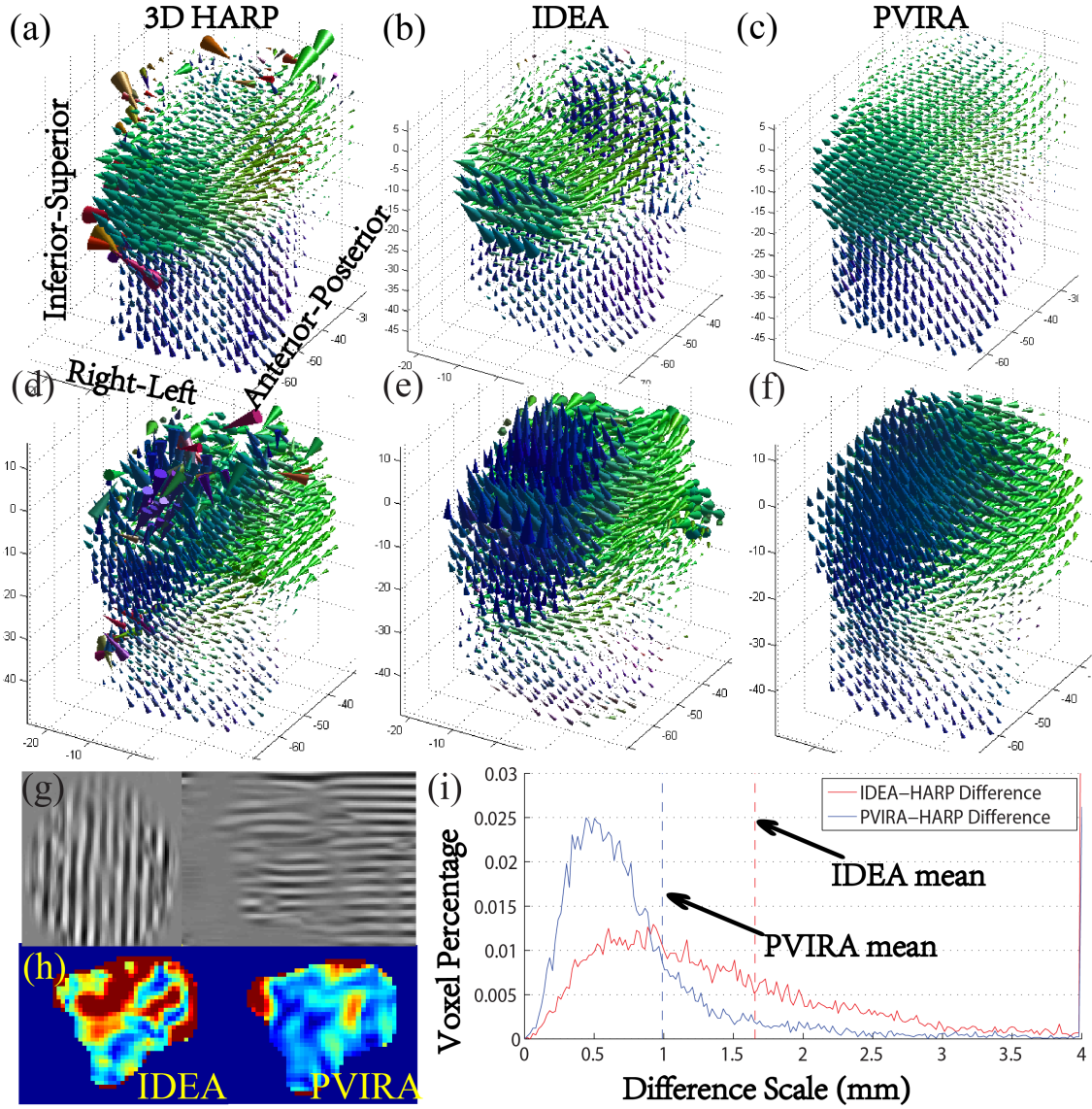


Figure 5.8: Tongue motion estimation in speech. (a)-(c) Forward motion estimates. (d)-(f) Upward Motion estimates. (g) Examples of tagged axial and sagittal slices. (h) Example slices of estimation difference from HARP. Colorbar goes from 0 to 0.5 as in the previous figures. (i) Histogram of difference from HARP using IDEA and PVIRA.

CHAPTER 5. PHASE VECTOR INCOMPRESSIBLE REGISTRATION

also note that PVIRA has the advantage to estimate motion on all tissue regions covered by tags instead of being restricted to the segmented region as is required by IDEA. This provides potential for further motion studies on other muscles or organs.

The simulation tests showed PVIRA’s robustness to noise. Helped by its internal regularization, PVIRA results on the real data were visually smoother. From Figs. 5.7(h) and 5.8(h), planar-shaped artifacts can be observed in IDEA estimation due to its use of sparse slices. PVIRA starts the estimation with a dense interpolation and does not suffer from these artifacts. This also explains the reason that PVIRA is closer to HARP than IDEA because both PVIRA and HARP are volume-processing methods.

There remains problems to be explored. PVIRA suffers from boundary effects—although they are not as severed as in the other methods. This problem could be solved with extra boundary deformations added into the estimation [51]. Besides, although PVIRA simultaneously provides an inverse field and saves an extra post-processing step, the accuracy of this inverse field is at the same level as numerically inverting the HARP and IDEA results. Potential methods to improve the inversion accuracy could be explored. Lastly, the effect of regularization needs to be further studied to prevent over-smoothing. And the effects of both regularization coefficients must be explored. Automatic local regularization with prior knowledge could be a potential direction for the improvement.

5.12 Conclusion

We proposed a novel method, PVIRA, based on an image registration framework to adapt to tagged images for motion estimation. A new similarity metric specifically designed to interpolated phase volumes was discussed, yielding an estimate that is diffeomorphic, incompressible, and simultaneously provides an inverse field. PVIRA was tested with simple motion simulations and real brain and tongue motion data. Compared to previous methods—HARP and IDEA—PVIRA demonstrated comparable accuracy, showed strong robustness against noise, greatly simplified the processing workflow, and was much faster to execute. More importantly, PVIRA looked into the tagged motion estimation problem from a registration point of view and demonstrated its capability, which was not a well-explored area in previous works.

Chapter 6

Relating Motion Production with Muscle Activities

6.1 Introduction

For surgeons, neurologists, and speech pathologists, having an accurate motion estimation is not sufficient. Understanding the relationship between the production of a tongue shape and the functioning of specific muscle groups to create such deformation is a more desired goal. This is an essential step toward establishing an informative tool for the study of motor control. However, it has been a challenging task because the tongue is composed of orthogonal interdigitated muscles that interact with each other to create a myriad of shapes, as described in Section 1.2. Especially, when comparing post-glossectomy patients with healthy control subjects, their abnormal

CHAPTER 6. MOTION AND MUSCLE ACTIVITIES

anatomical structure may result in defective muscle coordination and deficient shape creation [4, 7]. We are therefore interested in linking the activity of each internal tongue muscle to various overall tongue motions. We note that no brain data is used in this application because the brain does not contain major muscular structures.

Since tagged and cine MR images provide relatively low resolution anatomical information, tongue muscle groups are difficult to specify from these images. Therefore, to reveal the internal tongue muscles and their fiber direction, static high-resolution MR images are acquired in the same slice locations when the tongue is in a resting state. These images have double the resolution of cine images and can be used to manually segment tongue muscles and show their anatomical structure.

Previous chapters have focused on computing the overall motion of the tongue (Chapters 3–5). The major goal has been to increase estimation accuracy while respecting the incompressibility constraint. Meanwhile, strain derived from the estimated motion has been hard to visually interpret. For example, Fig. 6.1 shows the three principal strain directions computed from a normal subject’s motion during time frames /s/ and /k/. We can only roughly observe that during time frame /s/ for forward motion, the first principal direction, showing the directions of stretching, is mainly green near the tongue tip, indicating an anterior–posterior stretching to extend the tongue forward. While the third principal direction, showing the directions of compression, is mainly blue, indicating a superior–inferior compression to squeeze the tongue forward (Figs. 6.1(a) and 6.1(d)). On the other hand, during time

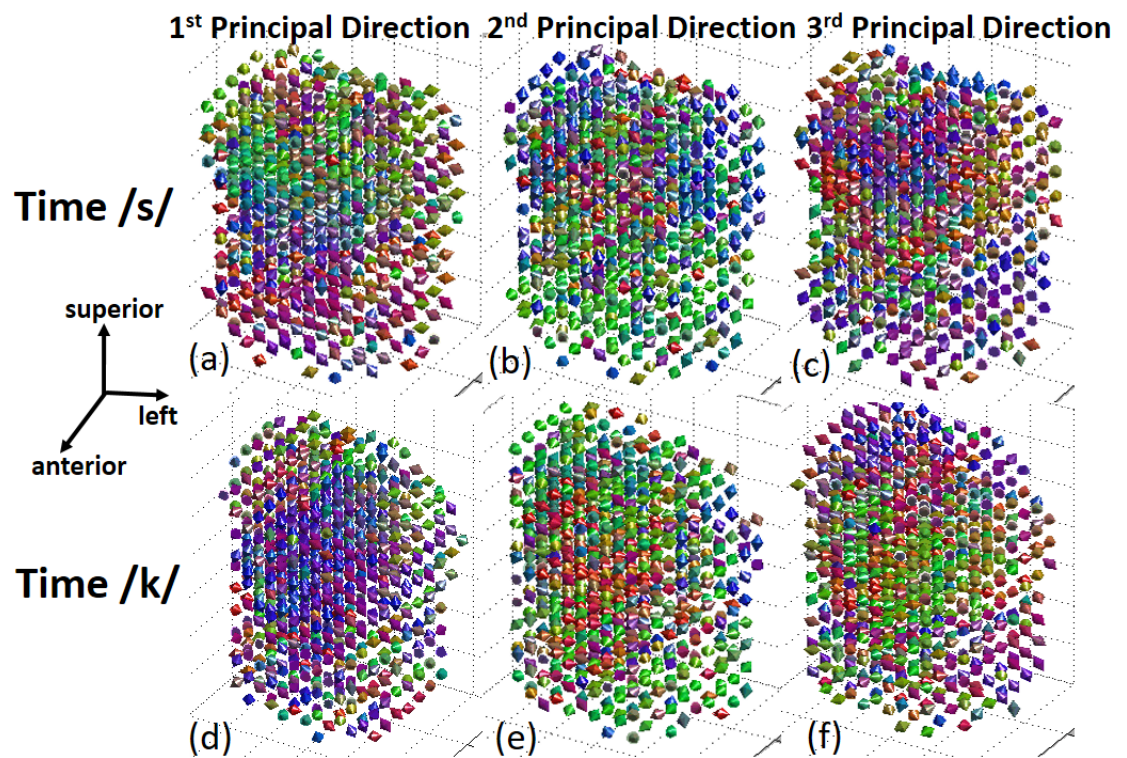


Figure 6.1: The three principal directions of tongue motion during speech at two critical time frames /s/ and /k/.

CHAPTER 6. MOTION AND MUSCLE ACTIVITIES

frame /k/ for upward motion, the first principal direction is mainly blue, indicating a superior–inferior stretching to extend the tongue upward. While the third principal direction is mainly green, indicating an anterior–posterior compression to squeeze the tongue upward (Figs. 6.1(c) and 6.1(f)). The second principal direction is orthogonal to the other two and does not provide direct visual information. The reason that visual interpretation is difficult in these results is that various tongue muscles activate in different patterns to create a certain deformation, resulting in complex principal strain directions that are over-complicated for the human eyes to assess. This even proves to be an obstacle for quantitative assessment, as inferential information is hard to be extracted.

In this chapter, we tackle the problem by grouping the estimated tongue motion using a fine segmentation of the internal muscle structures from high-resolution images. Strain along the muscle fiber directions can be computed after motion is related to individual muscles, yielding a result that enables visual and quantitative assessment. The final goal is to trace the production of speech and motion patterns back to the compression of specific muscles. An appropriate computation of projected strain tensors in the line of action of muscle fibers is a key step to achieve this goal.

6.2 Strain in the Line of Action

After tagged image slices are processed, a sequence of dense 3D motion fields at every time frame is obtained. We denote the displacement field at time t as $\mathbf{u}(\mathbf{x}, t)$. Hereafter in this chapter, we omit the variable t since we process each deformed time frame independently. As shown in Fig. 6.2(a), for every subject’s tongue, high-resolution MR images are acquired in a static state. The super-resolution algorithm SUPERV [63] that we used for creating supervolumes from the cine images in TMAP is applied on the high-resolution slices to generate dense 3D high-resolution supervolumes. Manual segmentation is performed on this volume to label all muscles of the tongue (Fig. 6.2(b)). With reference to the anatomy of the tongue introduced in Section 1.2, we now study the genioglossus (GG) muscle, which has fan shaped fibers coursing from its origin at the inner aspect of the mandible to its insertion along the entire upper surface of the midline tongue, from front to back. We also study a second muscle, the transversalis (T) muscle (also known as transverse muscle), which courses from left to right across the tongue, with two halves divided by the midline septum of the tongue. T muscle interdigitates with GG for about 1 cm on either side of the midline of the tongue.

The reason we study these two muscles as a first step toward understanding motor control is stated as follows. First, GG and T muscles form a major part of the human tongue. Their fibers include all anterior–posterior, inferior–superior, and left–right directions that can be used to form major tongue deformations in 3D. Second,

CHAPTER 6. MOTION AND MUSCLE ACTIVITIES

the intersected region of GG and T muscles is large (about 2 cm thick) so that collaborative functions of these two muscles can be studied in this region. Third, due to the larger size of these two muscles comparing to other smaller muscles, it is easier for a human user to manually delineate their fiber directions using a graphical user interface. We use an open source software package “3D Slicer” to delineate the fiber directions (<http://www.slicer.org/>). In particular, the muscle fiber directions are specified on every GG and T voxel (Figs. 6.2(c) and 6.2(d)). We denote these directions by $\mathbf{d}_{GG}(\mathbf{X})$ and $\mathbf{d}_T(\mathbf{X})$, respectively. Here we use \mathbf{X} for the coordinates of the undeformed static time frame. According to [85] summarized in Section 2.4, the deformation gradient tensor can be computed by $F(\mathbf{x}) = d\mathbf{x}/d\mathbf{X} = (I - d\mathbf{u}/d\mathbf{X})^{-1}$ from the 3D motion field $\mathbf{u}(\mathbf{x})$. The muscle fiber direction at each deformed time frame can then be propagated by

$$\begin{aligned}\mathbf{d}_{GG}(\mathbf{x}) &= F(\mathbf{x})\mathbf{d}_{GG}(\mathbf{X}(\mathbf{x})), \\ \mathbf{d}_T(\mathbf{x}) &= F(\mathbf{x})\mathbf{d}_T(\mathbf{X}(\mathbf{x})).\end{aligned}\tag{6.1}$$

The strain tensor is often computed to reveal a local stretch or compression caused by tissue deformation. The Eulerian strain tensor, defined as $E(\mathbf{x}) = \frac{1}{2}(I - F(\mathbf{x})^{-T}F(\mathbf{x})^{-1})$, can be computed in this case because the direct TMAP motion estimate is rooted in the deformed time frame. Its three principal directions and eigenvalues show the direction and type of deformation, where positive values indi-

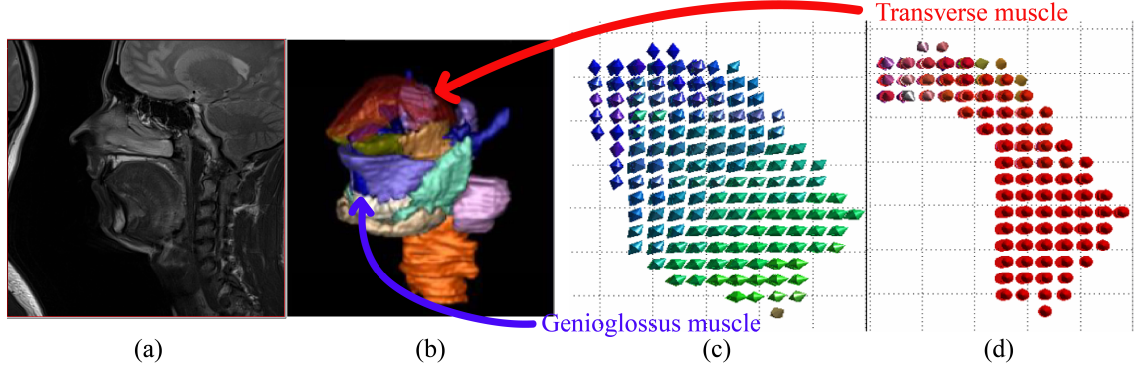


Figure 6.2: (a) A high-resolution MR mid-sagittal image slice. (b) Manual segmentation of the tongue muscles in 3D. (c) Genioglossus muscle fiber direction. (d) Transversalis muscle fiber direction.

cate stretch and negative values indicate compression. In order to see the deformation along the muscles, a projection of Eulerian strain to the fiber directions followed by computing of the 2-norm is necessary, i.e., $\|E(\mathbf{x})\mathbf{d}_{GG}(\mathbf{x})\|_2$ and $\|E(\mathbf{x})\mathbf{d}_T(\mathbf{x})\|_2$. However, any negative value will be converted to positive by the 2-norm operation, losing its indication of compression.

To solve this problem, we notice the relationship between the deformation gradient tensor $F(\mathbf{x})$ and the Eulerian strain tensor $E(\mathbf{x})$. Instead of using $E(\mathbf{x})$, we project $F(\mathbf{x})$ to the fiber directions, i.e.,

$$\begin{aligned} e_{GG}(\mathbf{x}) &= \|F(\mathbf{x})\mathbf{d}_{GG}(\mathbf{x})\|_2, \\ e_T(\mathbf{x}) &= \|F(\mathbf{x})\mathbf{d}_T(\mathbf{x})\|_2. \end{aligned} \tag{6.2}$$

At each voxel, the result is a scalar between the first and the third singular value of $F(\mathbf{x})$, indicating the ratio that a local tissue point deforms along a muscle fiber

CHAPTER 6. MOTION AND MUSCLE ACTIVITIES

against its original length. We define this quantity as *strain in the line of action*. A value greater than 1 shows stretch and a value less than 1 shows compression. To show compression only (which could be an indication of muscle activation), we threshold the ratio with a value $T, 0 < T < 1$. The result is a sequence of 3D masks on the corresponding muscle that shows its potentially activated region at each time frame, i.e.,

$$\begin{aligned} M_{GG}(\mathbf{x}) &= \begin{cases} 1, & \text{if } e_{GG}(\mathbf{x}) < T, \\ 0, & \text{otherwise.} \end{cases} \\ M_T(\mathbf{x}) &= \begin{cases} 1, & \text{if } e_T(\mathbf{x}) < T, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{6.3}$$

6.3 Thresholding Strain in the Line of Action

For each subject in the speech study, we computed strain in the line of action, and considered more than 2% compression to be above the level of noise. The muscle masks after thresholding are shown in the bottom rows of Figs. 6.3, 6.4, and 6.5 for a healthy control and two post-glossectomy patients, respectively, while the top rows show their original muscle masks for comparison. These figures visualize the parts of the GG and T muscles that are compressed during different tongue deformations.

CHAPTER 6. MOTION AND MUSCLE ACTIVITIES

Table 6.1: Percentage of compressed muscles at critical time frames during utterance “a souk”.

	GG			T		
	Time /ə/	Time /s/	Time /k/	Time /ə/	Time /s/	Time /k/
Control	10.1	45.3	20.4	32.4	41.3	52.3
Patient 1	24.5	25.3	3.7	6.4	27.3	49.7
Patient 2	13.4	26.1	9.7	4.0	9.9	16.4

For example, GG shortens anteriorly during /ə/, medially during /s/, and posteriorly during /k/ for the healthy control and the first patient, but the pattern for the second patient is different. In addition, for numerical assessment, we computed the percentage of the compressed muscle volumes relative to the whole volume. This value at critical time frames is shown in Table 6.1 and reflects the portion of the GG and T muscle that is under more than 2% compression.

6.4 Discussion

Results on a healthy control and two post-glossectomy patients show great correspondence between tongue muscle compression and speech production. In the healthy control, the activated parts of GG and T muscles at different time frames can be clearly interpreted. At time frame /ə/ where the tongue is at a speech-ready position, a small portion of anterior GG and T are both compressed during the slight retraction of the tip of the tongue. Then the tip is moved forward to pronounce /s/, compressing a larger region of anterior GG and a small portion of T. During /s/, we

CHAPTER 6. MOTION AND MUSCLE ACTIVITIES

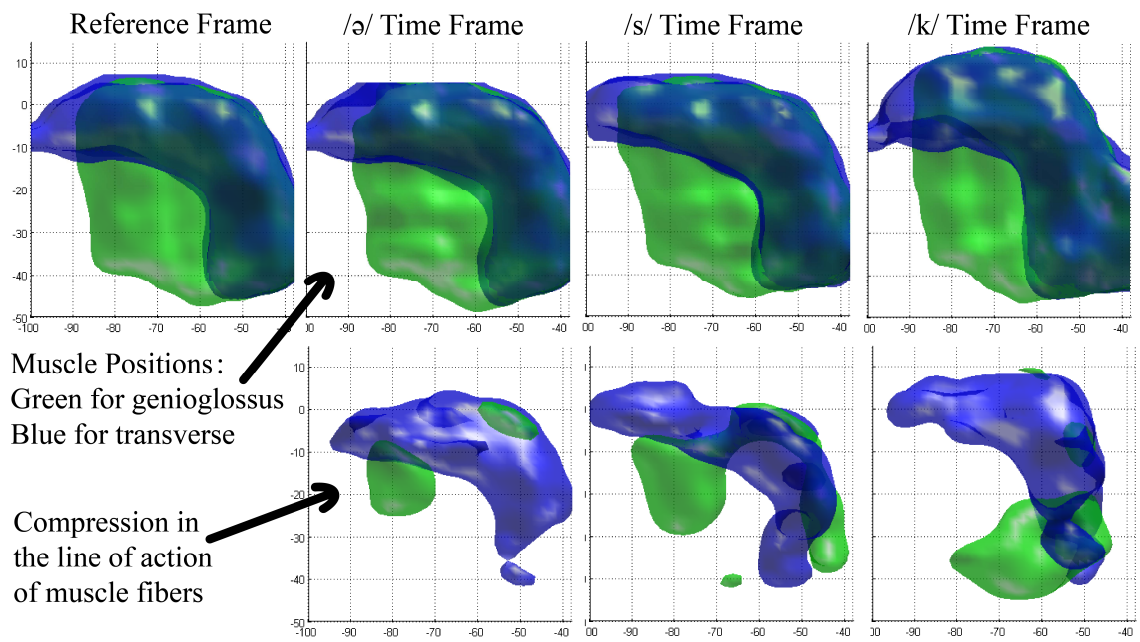


Figure 6.3: Strain in the line of action of muscle fibers of a healthy control. The top row is the segmented shape of GG (green) and T (blue) muscles at different time frames. In the bottom row, local compression of genioglossus is shown in green, and compression of transverse is shown in blue for the same time frames.

CHAPTER 6. MOTION AND MUSCLE ACTIVITIES

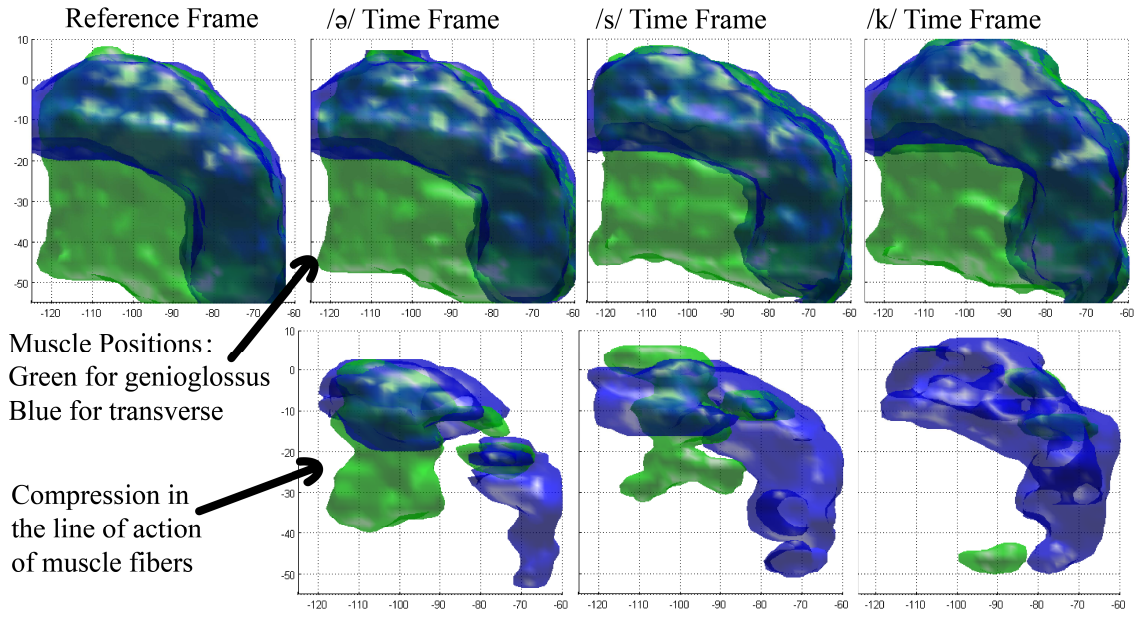


Figure 6.4: Strain in the line of action of muscle fibers of the first post-glossectomy patient. The top row is the segmented shapes and the bottom row is the local compression of muscles.

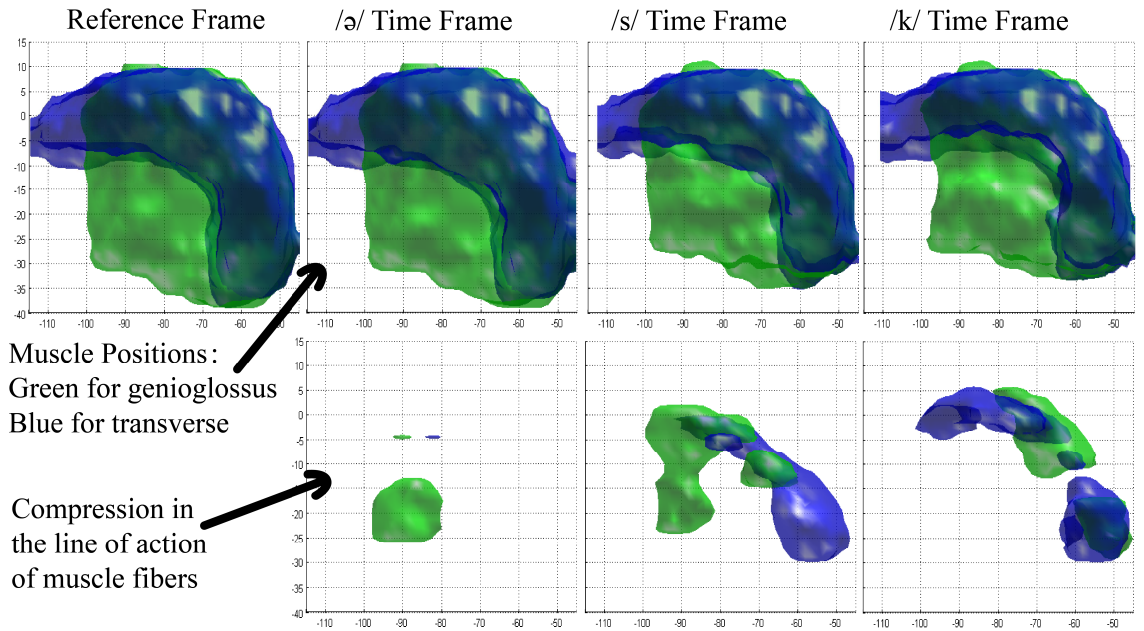


Figure 6.5: Strain in the line of action of muscle fibers of the second post-glossectomy patient. The top row is the segmented shapes and the bottom row is the local compression of muscles.

CHAPTER 6. MOTION AND MUSCLE ACTIVITIES

also see that the middle portion of GG compresses in order to flatten the mid-sagittal part of the tongue and create the center groove necessary to funnel the air across the top of the tongue. Lastly, at the /k/ sound, anterior GG stops compressing, while posterior GG and posterior T overlap in a highly compressed region at the bottom back part of the tongue. This strong joint compression in this region is consistent with squeezing the tongue both left–right and forward–backward, making it elongate upwards, which is necessary to touch the palate and produce the /k/ sound.

On the other hand, the patients’ muscle activities are less predictable than the healthy control subject due to their resected tongues, which impair their regular motor functions. In the first patient, the pronunciation of the /ə/ sound and the /s/ sound appears normal with similar muscle behavior to the healthy control. But during the /k/ sound, posterior GG does not show strong activity to compress the bottom of the tongue in forward and backward directions. The T muscle only is responsible for pushing the tongue upward. This could cause improper /k/ pronunciation or muscle fatigue in normal speech. In the second patient, the compressed muscle part is less than the other two subjects by a large amount. The resection of the tongue may be impeding its ordinary contracting abilities. Table 6.1 also provides similar interpretation. We note that although the patients’ tongues show special behaviors, their general trend of motion is the same as that of the healthy control, with strong GG activity at /s/ and increasing T activity at /k/. This enables all three subjects to be able to legibly pronounce the utterance “a souk”.

6.5 Summary

In this chapter, we described a method for analysis of motor control of internal tongue muscles. The method combined both sources of information acquired from tagged and high-resolution MR images, and used strain in the line of action along the muscle fibers to reveal tissue compression consistent with muscle activation. Preliminary results on a small number of subjects were obtained, which showed high correspondence between muscle activities and speech production, demonstrating the method's potential to aid future motor control studies.

Chapter 7

Conclusions and Future Work

This dissertation focused on resolving challenges and problems in three-dimensional tissue motion estimation and analysis using tagged MR images. The applications used tongue data in speech and brain data in mild head accelerations. In Chapter 3, we developed a 3D motion analysis pipeline that enables the implementation of a graphical user interface that was used in practical research. In Chapter 4, we improved the 3D motion estimation method in the pipeline by enhancing the boundary motion estimation. This provided a more accurate overall motion estimate on the whole tissue region. In Chapter 5, we developed a novel 3D motion estimation algorithm using a different perspective—the image registration framework. The new algorithm greatly reduces the preprocessing steps required in the previous workflow. It alleviates the burden of manual intervention and greatly improves computation speed. In Chapter 6, we introduced a method to reveal internal muscles’ activity. It related muscle

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

compression with the production of specific deformation and was an initial step to look at the motor control of individual muscles from tagged imaging. In this chapter, we summarize the key results and provide a discussion on future research directions.

7.1 3D Motion Analysis Pipeline

In Chapter 3, we proposed a workflow of algorithms, TMAP, for processing speech MR image data to obtain a reliable estimate of the motion field. TMAP is semi-automatic and easy to operate. We proposed methods to achieve effective segmentation, multi-subject data regularization, and a two-step principal component analysis to reveal subtle abnormal motion patterns. We tested TMAP using a collection of subjects in multiple Monte Carlo trials and the result showed promise of use in current speech studies and medical applications.

7.1.1 Main Results

- Phase calibration enabled a correct reading of the raw tagged image. It combined A and B images correctly using a phase shift, solved the problem that resulted in broken tags, and provided an image at the first time frame with high contrast.
- TMAP was implemented in graphical user interfaces that enabled practical use and direct operation for speech scientists.

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

- The result from the angle of motion experiment satisfied the expectation of the designed motion. Overall, the healthy controls' tongues went forward to produce /s/ then upward to produce /k/ while maintaining left-right symmetry. The patients' tongues showed a similar trend but were noisier and less predictable while producing a large amount of left-right asymmetry.
- The two-step PCA experiment revealed the patients' abnormal motion pattern. The primary PCs showed little lateral motion consistent with normal speech. The secondary PCs had more lateral motion consistent with abnormal motion, and also showed less predictability of patterning.
- Monte Carlo trials tested the efficacy of TMAP and confirmed that the healthy controls' motions were more consistent and stable than the patients. The mean of the control abnormal energy is lower than both PGS and PNS in all VOIs, and TMAP is capable of distinguishing the subtle patient motion variation pattern from the healthy controls.

7.1.2 Future Work

The ability to convert tissue-point motion from multiple 2D orientations into a 3D dataset is challenging, and further improvement of the detailed methods in TMAP are being studied and made. TMAP's system involves registration, segmentation, motion analysis, interpolation, and other problems of high interest in medical image

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

analysis. The construction of the supervolume is essentially an image registration and fusion process. The selection of registration method and the weighting of various image sources could all affect the quality of the supervolume. TMAP used the random walker algorithm for segmentation, and more sophisticated methods can be applied. In terms of motion reconstruction, we have improved the IDEA step by replacing it with E-IDEA. In summary, an upgrade to any TMAP step could lead to an improved system and better results.

7.2 Enhanced Incompressible Motion Estimation

In Chapter 4, we proposed an enhanced algorithm for estimating the tongue’s motion field in 3D. The major innovation is in the incorporation of surface motion as additional information, which compensates for the well-known deficiencies of HARP in estimating boundary motions. Both qualitative and quantitative improvements are evident using two independent metrics. Boundary error was substantially reduced while internal error is only minimally increased.

7.2.1 Main Results

- Visual assessment of the motion fields from IDEA and E-IDEA revealed that E-IDEA reduces the erroneous large motions on the boundary, and captures smaller motions where IDEA mistakenly extrapolates as zero. E-IDEA also straightens up the motion at the top of the tongue to better estimate the displacement when the tongue hits the palate vertically. In general, the boundary estimation agrees more with the tongue’s physical mechanics.
- In terms of quantitative analysis, the comparison of manual tracking, IDEA tracking, and E-IDEA tracking of the surface points on the tongue mask shows the trajectory of the points over time. The error magnitude is smaller for E-IDEA, especially on the back of the tongue. The mean error is also significantly smaller for E-IDEA at all surface points ($p < 0.01$).
- The reprojection error against 2D HARP gives the error in distance in the input directions between the estimated sample components and the input sample components. Four types of reprojection errors were compared in histograms. They are errors on IDEA internal points, E-IDEA internal points, E-IDEA boundary points, and IDEA boundary points. For the control, the boundary error was reduced by 0.68 mm and the internal error was raised by 0.03 mm. For the patient, the boundary error was reduced by 2.15 mm and the internal error was raised by 0.02 mm. In summary, the boundary error was greatly reduced while

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

the internal error was slightly increased. The overall estimation was greatly improved.

7.2.2 Future Work

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

7.3 Phase Vector Incompressible Registration

In Chapter 5, we proposed a novel method PVIRA based on an image registration framework to adapt to tagged images for motion estimation. A new similarity metric specifically designed to interpolated phase volumes is incorporated, yielding an estimate that is diffeomorphic, incompressible, and simultaneously provides a consistent inverse field. PVIRA was tested with simulations and real brain and tongue motion data. Compared against HARP and IDEA, PVIRA demonstrates comparable accu-

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

racy, shows strong robustness against noise, greatly simplifies the processing workflow, and is much faster to execute.

7.3.1 Main Results

- PVIRA shows comparable accuracy to HARP and IDEA although it is computed using a fundamentally different process. 3D HARP is not preferred over the other two methods because of its lack of incompressibility constraint. IDEA requires multiple preprocessing steps that took hours for each subject, and the execution time of IDEA is much slower than PVIRA due to its stepwise divergence-free vector spline process. Eventually, IDEA can take days to process all subjects in total while PVIRA takes hours.
- PVIRA shows its advantage in estimating motion on all tissue regions covered by tags instead of being restricted only on the segmented region as IDEA is. This provides potential for motion studies on other muscles or organs in the future.
- The simulation tests showed PVIRA's robustness to noise. Helped by its internal demons regularization, PVIRA results on the real data are visually smoother. While planar-shaped artifacts can be observed in IDEA estimation due to the usage of sparse slices in IDEA processing, PVIRA begins its estimation with a dense interpolation and does not suffer from these artifacts.

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

- PVIRA simultaneously provides an inverse field and saves an extra post-processing step required by most previous algorithms. The accuracy of inversion was similar to the post-processing inversion of 3D HARP and IDEA.

7.3.2 Future Work

Firstly, PVIRA could suffer the same boundary effect as HARP and IDEA. This problem could be solved by incorporating extra boundary deformation such as in E-IDEA. Secondly, the effect of regularization needs to be further explored to prevent potential over-smoothing. Automatic regularization parameter selection is to be developed. This could lead to a final solution to the problem. Thirdly, more complex simulations are needed to reveal the full properties of the PVIRA algorithm. Lastly, use of PVIRA in other applications such as cardiac motion estimation is yet to be studied. It can provide further tests to the robustness of the method and reveal its potential features that can be enhanced.

7.4 Motion and Muscle Activities

In Chapter 6, we proposed a method for analysis of motor control of internal tongue muscles. The method combines both sources of information acquired from tagged and high-resolution MR images, and uses strain in the line of action along the muscle fibers to reveal tissue compression consistent with muscle activation. Pre-

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

liminary results on a small number of subjects were obtained, which showed high correspondence between muscle activities and speech production, demonstrating the method’s potential to aid future motor control studies.

7.4.1 Main Results

- Visual observation showed that the muscles’ behavior agrees with expectation to produce certain deformation. During time frame /s/ for forward motion, the first principal direction indicated an anterior–posterior stretching to extend the tongue forward, while the third principal direction indicated a superior–inferior compression to squeeze the tongue forward. During time frame /k/ for upward motion, the first principal direction indicated a superior–inferior stretching to extend the tongue upward, while the third principal direction indicated an anterior–posterior compression to squeeze the tongue upward.
- Results of thresholded strain in the line of action showed great correspondence between tongue muscle compression and speech production. At time frame /ə/ where the tongue was at a speech-ready position, a small portion of anterior GG and T were both compressed during the slight retraction of the tip. The tongue tip moved forward to pronounce /s/ and the middle portion of GG compressed in order to flatten the mid-sagittal part of the tongue and create the center groove necessary to funnel the air across the top of the tongue. At the /k/

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

sound, posterior GG and posterior T overlapped in a highly compressed region at the bottom back part of the tongue. This strong joint compression elongated the tongue upwards, which was necessary to touch the palate and produce the /k/ sound.

- Patient’s muscle activities are less predictable than healthy controls due to their resected tongues, which can impair their regular motor functions. During the /k/ sound, patients’ posterior GG did not show strong activity to compress the bottom of the tongue in forward and backward directions. The T muscle only was responsible for pushing the tongue upward. This could cause improper /k/ pronunciation or muscle fatigue in normal speech. Furthermore, the compressed muscle part could be less than normal subjects. The resection of the tongue could be impeding its ordinary contracting abilities.
- Although patients’ tongues showed special behaviors, their general trend of motion was the same as healthy control’s, with strong GG activity at /s/ and increasing T activity at /k/. This enabled patients to also be able to legibly pronounce the utterance “a souk”.

7.4.2 Future Work

The method of strain analysis can be further developed. In the future, manual generation of fiber directions should be replaced by the result from diffusion imaging

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

of the tongue. Manual segmentation of the two GG and T muscles should be replaced by a complete set of internal tongue muscles obtained from registration of a tongue atlas. Moreover, the behavior of the complete eight muscle groups in the tongue should be studied with more subjects in various speech patterns.

7.5 Overall Perspective

This dissertation aimed to provide efficient algorithmic tools to help the functional analysis of tissue motion by exploring the estimation of 3D motion fields from tagged MR images. We developed algorithmic solutions to several key challenges in 3D motion analysis including efficient motion tracking, accurate motion estimation, incompressible motion registration, and muscle function quantification. In practice, the combined results from this research have provided new and efficient implementations that enabled scientific and clinical use. Currently it is being applied in tongue imaging and brain motion study. Besides motion analysis, this research also involves a wide range of other topics in the field of image processing and analysis, including image registration, tissue segmentation, atlas construction, label fusion, etc. We hope this work can provide more insights for future medical imaging studies.

7.6 List of Publications

Journal articles:

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

- Landman, B. A., Asman, A. J., Scoggins, A. G., Bogovic, J. A., Xing, F., & Prince, J. L. (2012). Robust statistical fusion of image labels. *Medical Imaging, IEEE Transactions on*, 31(2), 512-522.
- Knutsen, A. K., Magrath, E., McEntee, J. E., Xing, F., Prince, J. L., Bayly, P. V., ... & Pham, D. L. (2014). Improved measurement of brain deformation during mild head acceleration using a novel tagged MRI sequence. *Journal of Biomechanics*, 47(14), 3475-3481.
- Lee, J., Woo, J., Xing, F., Murano, E. Z., Stone, M., & Prince, J. L. (2014). Semi-automatic segmentation for 3D motion analysis of the tongue with dynamic MRI. *Computerized Medical Imaging and Graphics*, 38(8), 714-724.
- Woo, J., Lee, J., Murano, E. Z., Xing, F., Al-Talib, M., Stone, M., & Prince, J. L. (2015). A high-resolution atlas and statistical model of the vocal tract from structural MRI. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 3(1), 47-60.
- Xing, F., Prince, J. L., & Landman, B. A. (2015). Investigation of bias in continuous medical image label fusion. Submitted to *PLOS ONE*.
- Xing, F., Woo, J., Lee, J., Murano, E. Z., Stone, M., & Prince, J. L. (2015). Analysis of 3D tongue motion from tagged and cine MR images. Submitted to *Journal of Speech, Language, and Hearing Research*.

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

- Xing, F., Woo, J., Gomez, A. D., Knutsen, A. K., Pham, D. L., Bayly, P. V., Stone, M., & Prince, J. L. (2015). Phase vector incompressible registration algorithm (PVIRA) for motion estimation from tagged magnetic resonance images. In preparation for submission.

Conference papers:

- Xing, F., Soleimanifard, S., Prince, J. L., & Landman, B. A. (2011). Statistical fusion of continuous labels: identification of cardiac landmarks. In *SPIE Medical Imaging*, pp. 796206-796206. International Society for Optics and Photonics.
- Xing, F., Asman, A. J., Prince, J. L., & Landman, B. A. (2012). Finding seeds for segmentation using statistical fusion. In *SPIE Medical Imaging*, pp. 831430-831430. International Society for Optics and Photonics.
- Lee, J., Woo, J., Xing, F., Murano, E. Z., Stone, M., & Prince, J. L. (2013). Semi-automatic segmentation of the tongue for 3D motion analysis with dynamic MRI. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pp. 1465-1468. IEEE.
- Xing, F., Murano, E. Z., Lee, J., Woo, J., Stone, M., & Prince, J. L. (2013). MRI analysis of 3D normal and post-glossectomy tongue motion in speech. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pp. 816-819. IEEE.

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

- Xing, F., Woo, J., Murano, E. Z., Lee, J., Stone, M., & Prince, J. L. (2013). 3D tongue motion from tagged and cine MR images. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013*, pp. 41-48. Springer Berlin Heidelberg.
- Woo, J., Xing, F., Lee, J., Stone, M., & Prince, J. L. (2014). Determining functional units of tongue motion via graph-regularized sparse non-negative matrix factorization. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2014*, pp. 146-153. Springer International Publishing.
- Woo, J., Xing, F., Lee, J., Stone, M., & Prince, J. L. (2015). Construction of an unbiased spatio-temporal atlas of the tongue during speech. In *Information Processing in Medical Imaging (IPMI)*, pp. 723-732. Springer International Publishing.
- Xing, F., Ye, C., Woo, J., Stone, M., & Prince, J. (2015). Relating speech production to tongue muscle compressions using tagged and high-resolution magnetic resonance imaging. In *SPIE Medical Imaging*, pp. 94131L-94131L. International Society for Optics and Photonics. (Robert F. Wagner Best Student Paper Award Finalist.)

Conference abstracts:

- Murano, E. Z., Xing, F., Evitts, P., Woo, J., Lee, J., Zhuo, J., Gullapalli, R., Stone, M., Ord, R., & Prince, J. L. (2012). Tongue muscle adaptation measured

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

with 4D dynamic MRI—glossectomy with flap vs. no-flap. *Joint Meeting of the American Academy of Maxillofacial Prosthetics (AAMP) and International Society for Maxillofacial Rehabilitation (ISMR)*. Baltimore, MD.

- Murano, E. Z., Suo, Y., Xing, F., Roy, S., Bogovic, J., Stone, M., & Prince, J. L. (2011). Three dimensional digital articulator labeled atlas. *The Journal of the Acoustical Society of America*, 129(4), 2455-2455.
- Murano, E. Z., Xing, F., Zhuo, J., Woo, J., Gullapalli, R., Stone, M., & Prince, J. L. (2012). 4D Dynamic-MRI as Glossectomy Speech Assessment Tool. *Otolaryngology–Head and Neck Surgery*, 147(2 suppl), P143-P144.
- Murano, E. Z., Xing, F., Soleimanifard, S., Woo, J., Stone, M., & Prince, J. L. (2012). Measuring 3D tongue deformation based on the muscular hydrostat model. *Conference on Motor Speech*. Santa Rosa, CA.
- Xing, F., Lee, J., Woo, J., Murano, E. Z., Stone, M., & Prince, J. L. (2012). Estimating 3D tongue motion with MR images. In *Asilomar Conference on Signals, Systems, and Computers*, Monterey, CA.
- Murano, E. Z., Xing, F., Woo, J., Lee, J., Zhuo, J., Gullapalli, R., Ord, R. A., Stone, M., & Prince, J. L. (2013). Primary compensatory speech strategy in post-glossectomy patients—A 4D MRI study. *Proceedings of the 10th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research*. Cincinnati, OH.

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

- Pedersen, A. D., Hwang, J., Woo, J., Xing, F., Prince, J. L., & Stone, M. (2013). An MRI comparison of /s/ production in four subject conditions. *The Journal of the Acoustical Society of America*, 134(5), 4168-4168.
- Woo, J., Lee, J., Bogovic, J., Murano, E. Z., Xing, F., Stone, M., & Prince, J. L. (2013). Multi-subject atlas built from structural tongue magnetic resonance images. In *Proceedings of Meetings on Acoustics*, Vol. 19, No. 1, p. 060194. Acoustical Society of America.
- Woo, J., Xing, F., Stone, M., & Prince, J. L. (2014). Determining functional units of tongue motion from magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 135(4), 2196-2196.
- Xing, F., Woo, J., Ziemba, J. K., Stone, M., & Prince, J. L. (2014). Tracking four dimensional muscle mechanics from high-resolution and tagged magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 135(4), 2196-2197.
- Ziemba, J. K., Stone, M., Pedersen, A. D., Woo, J., Xing, F., & Prince, J. L. (2014). Internal three dimensional tongue motion during “s” and “sh” from tagged magnetic resonance imaging; control and glossectomy motion. *The Journal of the Acoustical Society of America*, 135(4), 2389-2389.

Appendix A

Continuous Label Fusion of Medical Images

A.1 Introduction

Characterization of the morphometric features of human organs—e.g., their size and shape—requires their delineation and labeling within medical images. This can be accomplished either by automated segmentation algorithms, manual delineation, or a combination of both efforts. For example, cardiac imaging studies commonly use either human raters or algorithms to 1) delineate the epicardium (the outer contour of the left ventricle), 2) delineate the endocardium (the inner contour of the left ventricle), and 3) identify the two right ventricle (RV) insertion points where the right and left ventricles connect [86]. These features are typically identified on short

APPENDIX A. APPENDIX

axis images showing the cross section of the heart that is perpendicular to long axis connecting the heart's apex and base (Fig. A.1(a)). In this process, the raters will introduce errors, generate ambiguous interpretation of structures, and occasionally make careless mistakes. Hence, it is adequate to employ more than one rater to label each image and enhance accuracy using statistical label fusion methods [87].

The *simultaneous truth and performance level estimation* (STAPLE) algorithm is a popular method for fusing labeled datasets [88]. STAPLE iteratively constructs estimates of both the true labels as well as the performance parameters of the raters using the E-step and M-step, respectively, of the expectation maximization (EM) algorithm [89, 90]. In the discrete case in which there are a finite number of labels to assign, rater performance is characterized by the sensitivity and specificity values for binary labels or the confusion matrix for multi-labels, both of which characterize the likelihood that raters assign the correct labels to the corresponding voxels. In the continuous case, raters select labels that are characterized by continuous values that lie in an uncountably infinite set. For example, the spatial locations for the RV insertion points cannot be characterized by discrete labels because their positions are defined by continuous-valued vectors in a 2D space indicating (potentially sub-voxel) locations. Another example of continuous labels is the levelset method for representing shapes [91], which is distinguished from voxel labeling of shapes because it has the capability to represent shapes with sub-voxel resolution. Fig. A.1 shows the labeling of a typical cardiac MR image slice. Shapes like endocardium contour

APPENDIX A. APPENDIX

can be labeled either by discrete volumetric labels (Fig. A.1(b)) or by its continuous signed distance function (Fig. A.1(d)), while the two RV insertion points must be labeled by continuous 2D vectors (Fig. A.1(c)). When multiple raters are used in the scenarios depicted in Figs. A.1(c) and A.1(d), continuous fusion must be used; this is the general framework and problem considered in this chapter.

In the continuous version of STAPLE (CSTAPLE), a Gaussian mixture model is used where rater performance can be represented by the bias and variance of the rater’s ability to locate the true value [92, 93]. CSTAPLE uses an analogous approach to the discrete STAPLE in that the truth, bias, and variance parameters are estimated simultaneously using maximum likelihood. However, we prove below that CSTAPLE yields an equal likelihood for any bias parameter, which implies that bias is indeterminate and this approach cannot fully evaluate rater performance. We will show that bias estimate in CSTAPLE is completely determined by its initialization, and this value—however it may have been specified—can strongly influence the continuous label estimate. One could ignore this problem by tweaking the initialization. However, the major contribution is to point out that adequate bias estimation is needed because the core algorithm does not estimate bias, which is a fundamental flaw of the theory.

This chapter is based on previous work of [94]. Recent developments have continued to improve statistical fusion, including robustness enhancement [95], introducing spatially-varying statistical models [96], and applying continuous label fusion to

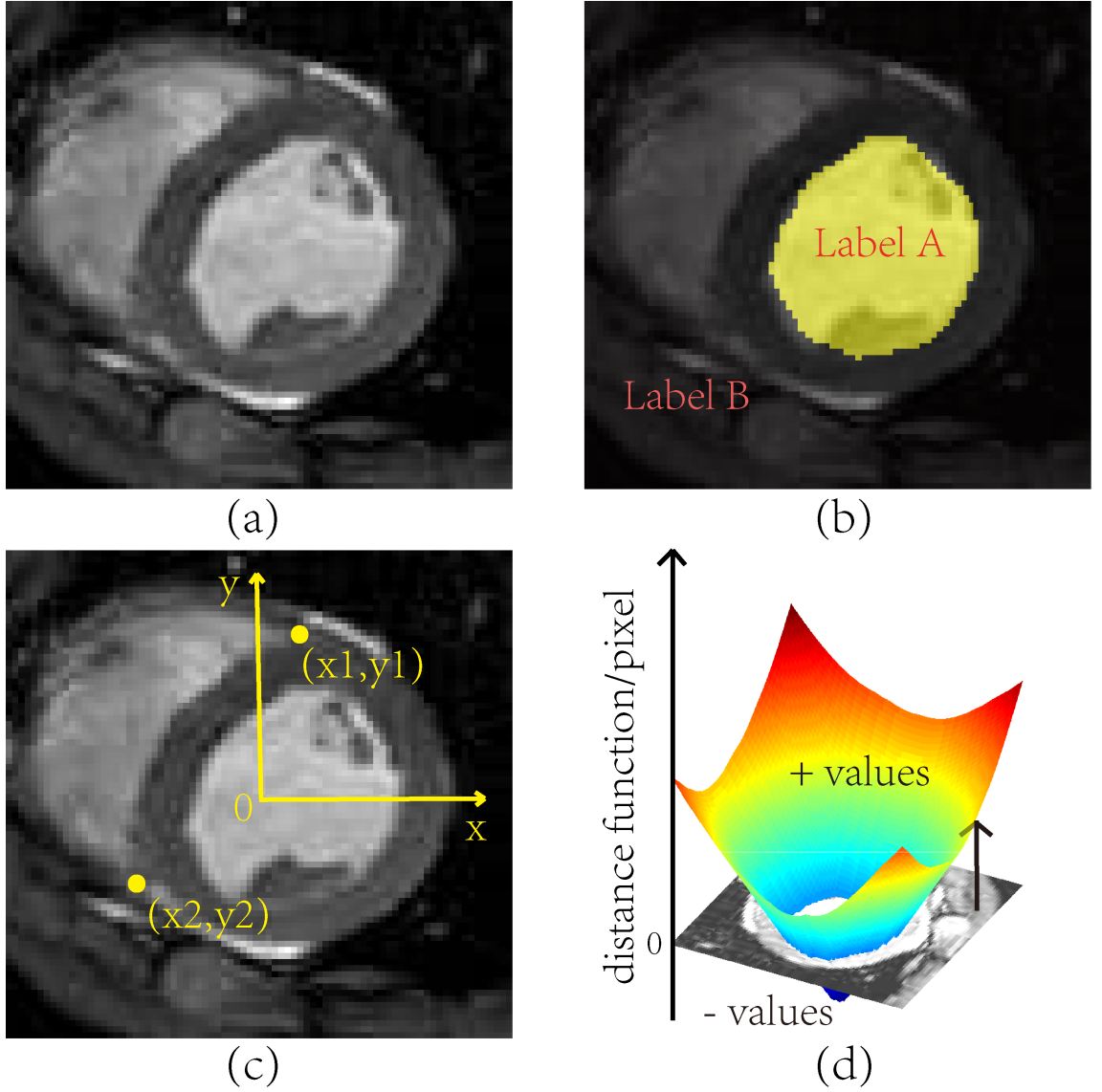


Figure A.1: Cardiac MR image labeling for endocardium and RV insertion points. (a) Short-axis MR image of the heart. (b) Pixel labeling of the left ventricle chamber. (c) Labeling of right ventricle insertion points. (d) Level set representation of the left ventricle chamber contour (endocardium).

APPENDIX A. APPENDIX

correct the bias in the application of brain imaging [97]. An evaluation on the performance of all STAPLE-related works and their variants has been proposed in [98]. On the other hand, new ways of modeling the fusion problem has been explored, such as regression-based models [99], and a generative model for segmentation based on label fusion [100]. Other methods such as shape-based averaging [101] have also tried to tackle the label fusion problem from a non-STAPLE point of view. Moreover, the problem of automated cardiac ventricular segmentation has been studied from a collective point of view using collaborative resources to build consensus [102]. Although these works have been moving forward in new applications with novel approaches, CSTAPLE is still a common reference of study and its bias problem has not been adequately analyzed and solved. Clearly, a deeper look at the cause of bias indeterminacy and a comprehensive solution inside the CSTAPLE framework is necessary.

In this chapter, we present two solutions for continuous label fusion, one that estimates the bias as part of the initialization and the other that uses a maximum *a posteriori* criterion with *a priori* probabilities on the rater bias. While re-deriving the mathematics of EM iteration to appreciate these new approaches, we also describe the method of determining initialization or prior parameters in both cases when only current observed data is available and when extra information is provided by previous knowledge.

A.2 EM Algorithm for Continuous Label Fusion

In K -dimensions, the goal is to identify N continuous vectors \mathbf{t}_i , the collection of which can be represented by the truth matrix

$$T = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_i^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1K} \\ \vdots & \vdots & \cdots & \vdots \\ t_{i1} & t_{i2} & \cdots & t_{iK} \\ \vdots & \vdots & \cdots & \vdots \\ t_{N1} & t_{N2} & \cdots & t_{NK} \end{bmatrix}_{N \times K}, t_{ik} \in \mathbb{R}. \quad (\text{A.1})$$

Consider R raters specifying all N vectors, each exactly once. Then the collection of all observations can be represented by the observation matrices

$$D_j = \begin{bmatrix} \mathbf{d}_{j1}^T \\ \vdots \\ \mathbf{d}_{ji}^T \\ \vdots \\ \mathbf{d}_{jN}^T \end{bmatrix} = \begin{bmatrix} d_{j11} & d_{j12} & \cdots & d_{j1K} \\ \vdots & \vdots & \cdots & \vdots \\ d_{ji1} & d_{ji2} & \cdots & d_{jiK} \\ \vdots & \vdots & \cdots & \vdots \\ d_{jN1} & d_{jN2} & \cdots & d_{jNK} \end{bmatrix}_{N \times K}, d_{jik} \in \mathbb{R}, j = 1, \dots, R. \quad (\text{A.2})$$

We assume that each rater j has the same performance parameters, a $K \times 1$ bias vector $\boldsymbol{\mu}_j$ and a $K \times K$ covariance matrix Σ_j , which characterize the rater's ability

APPENDIX A. APPENDIX

to specify any vector, and these parameters are deterministic and unknown. Under a Gaussian model, the probability density of rater j 's decision for vector i is

$$f(\mathbf{d}_{ji}|\mathbf{t}_i, \boldsymbol{\mu}_j, \Sigma_j) = \frac{1}{(2\pi)^{K/2} \sqrt{\det(\Sigma_j)}} e^{-\frac{1}{2}(\mathbf{d}_{ji} - (\mathbf{t}_i + \boldsymbol{\mu}_j))^T \Sigma_j^{-1} (\mathbf{d}_{ji} - (\mathbf{t}_i + \boldsymbol{\mu}_j))}. \quad (\text{A.3})$$

Our goal is to estimate $\theta = \{\theta_1, \dots, \theta_R\}$ where $\theta_j = \{\boldsymbol{\mu}_j, \Sigma_j\}$ using maximum likelihood. By viewing T as hidden data, the EM algorithm can be used to simultaneously estimate both θ and T . The expectation of the log likelihood function, i.e.,

$$E[\ln f(D, T|\theta) | D, \theta^{(n)}] = \int_{\mathbb{R}^{N \times K}} \ln f(D, T|\theta) f(T|D, \theta^{(n)}) dT \quad (\text{A.4})$$

is to be maximized by an appropriate $\theta = \hat{\theta}$. Assuming the distribution of truth is independent of performance, i.e., $f(T|\theta) = f(T)$, it is the same as maximizing

$$\int_{\mathbb{R}^{N \times K}} \ln f(D|T, \theta) f(T|D, \theta^{(n)}) dT. \quad (\text{A.5})$$

The first term in the integrand of Eqn. A.5 is the logarithm of the Gaussian density in Eqn. A.3. Assuming independence among different raters and among different vector points and assuming a constant $f(T)$. The second term by Bayes' theorem is

$$f(T|D, \theta^{(n)}) = \frac{f(D|T, \theta^{(n)}) f(T)}{\int_{\mathbb{R}^{N \times K}} f(D|T', \theta^{(n)}) f(T') dT'} = \prod_i \frac{\prod_j f(\mathbf{d}_{ji}|\mathbf{t}_i, \theta_j^{(n)})}{\int_{\mathbb{R}^K} \prod_j f(\mathbf{d}_{ji}|\mathbf{t}'_i, \theta_j^{(n)}) d\mathbf{t}'_i}. \quad (\text{A.6})$$

APPENDIX A. APPENDIX

Thus the weight of each vector point can be defined as

$$W_i^{(n)}(\mathbf{t}_i) = \frac{\prod_j f(\mathbf{d}_{ji}|\mathbf{t}_i, \theta_j^{(n)})}{\int_{\mathbb{R}^K} \prod_j f(\mathbf{d}_{ji}|\mathbf{t}'_i, \theta_j^{(n)}) d\mathbf{t}'_i} = \frac{e^{-\frac{1}{2}(\mathbf{t}_i - A^{(n)}\mathbf{b}_i^{(n)})^T (A^{(n)})^{-1} (\mathbf{t}_i - A^{(n)}\mathbf{b}_i^{(n)})}}{(2\pi)^{K/2} \sqrt{\det(A^{(n)})}}. \quad (\text{A.7})$$

where $A^{(n)} = (\sum_j (\Sigma_j^{(n)})^{-1})^{-1}$ and $\mathbf{b}_i^{(n)} = \sum_j (\Sigma_j^{(n)})^{-1} (\mathbf{d}_{ji} - \boldsymbol{\mu}_j^{(n)})$. After a sufficient number of iterations, $A^{(n)}\mathbf{b}_i^{(n)} \rightarrow A^{(\infty)}\mathbf{b}_i^{(\infty)} =: \mathbf{t}_i^{(\infty)}$, which is the estimated true position of vector point i .

Eqn. A.7 completes the derivation of the E-step. For the M-step, we need to update the performance parameters $\boldsymbol{\mu}_j^{(n)}$ and $\Sigma_j^{(n)}$ in each iteration. For each rater, from Eqn. A.5 we have

$$\{\boldsymbol{\mu}_j^{(n+1)}, \Sigma_j^{(n+1)}\} = \arg \max \sum_i \int_{\mathbb{R}^K} \ln f(\mathbf{d}_{ji}|\mathbf{t}_i, \theta_j) W_i^{(n)}(\mathbf{t}_i) d\mathbf{t}_i =: \arg \max F_j^{(n)}. \quad (\text{A.8})$$

To find the maximum point of $F_j^{(n)}$, we take its partial derivatives and set them to zero, i.e. $\partial F_j^{(n)} / \partial \boldsymbol{\mu}_j = 0, \partial F_j^{(n)} / \partial \Sigma_j = 0$, which yields

$$\begin{aligned} \boldsymbol{\mu}_j^{(n+1)} &= \frac{1}{N} \sum_i (\mathbf{d}_{ji} - A^{(n)}\mathbf{b}_i^{(n)}), \\ \Sigma_j^{(n+1)} &= \frac{1}{N} \sum_i [A^{(n)} + (\mathbf{d}_{ji} - \boldsymbol{\mu}_j^{(n+1)} - A^{(n)}\mathbf{b}_i^{(n)}) (\mathbf{d}_{ji} - \boldsymbol{\mu}_j^{(n+1)} - A^{(n)}\mathbf{b}_i^{(n)})^T]. \end{aligned} \quad (\text{A.9})$$

Eqn. A.9 completes the derivation of the M-step. These updated parameters are used in the E-step of the next iteration to compute a new estimate of the truth, which is

then used to calculate new updated parameters, and so on. Convergence is guaranteed by the nature of EM algorithm [103].

A.3 Bias Invariance Problem

From Eqn. A.3, the density of rater decision can be regarded equivalently as a function of \mathbf{t}_i or as a function of $\boldsymbol{\mu}_j$; thus the overall estimation of \mathbf{t}_i is closely related to the estimation of $\boldsymbol{\mu}_j$. Eqn. A.9 can be algebraically manipulated to reveal the fact that the bias does not change after the first calculation from initialization. First, we note that

$$\boldsymbol{\mu}_j^{(n+1)} = \frac{1}{N} \sum_i \left(\mathbf{d}_{ji} - A^{(n)} \mathbf{b}_i^{(n)} \right) = \frac{1}{N} \sum_i \left(\mathbf{d}_{ji} - \mathbf{t}_i^{(n)} \right). \quad (\text{A.10})$$

Moving the summation of $\mathbf{t}_i^{(n)}$ to the left hand side and $\boldsymbol{\mu}_j^{(n+1)}$ to the right yields

$$\frac{1}{N} \sum_i \mathbf{t}_i^{(n)} = \frac{1}{N} \sum_i \left(\mathbf{d}_{ji} - \boldsymbol{\mu}_j^{(n+1)} \right). \quad (\text{A.11})$$

While the right-hand side appears to be related to j , the left-hand side is independent of j , which means that regardless of different raters, this quantity stays the same as iteration goes on. We should also note that $A^{(n)}$ does not depend on i . As a result, by substituting both Eqn. A.11 and the definitions of $A^{(n)}$ and $\mathbf{b}_i^{(n)}$ into Eqn. A.10,

APPENDIX A. APPENDIX

we can make the following manipulations

$$\begin{aligned}
\boldsymbol{\mu}_j^{(n+1)} &= \frac{1}{N} \sum_i \left(\mathbf{d}_{ji} - \mathbf{t}_i^{(n)} \right) \\
&= \frac{1}{N} \sum_i \mathbf{d}_{ji} - \frac{A^{(n)}}{N} \sum_{i,j} (\Sigma_j^{(n)})^{-1} \left(\mathbf{d}_{ji} - \boldsymbol{\mu}_j^{(n)} \right) \\
&= \frac{1}{N} \sum_i \mathbf{d}_{ji} - \frac{A^{(n)}}{N} \sum_j (\Sigma_j^{(n)})^{-1} \sum_i \mathbf{t}_i^{(n-1)} \\
&= \frac{1}{N} \sum_i \mathbf{d}_{ji} - \frac{A^{(n)}}{N} (A^{(n)})^{-1} \sum_i \mathbf{t}_i^{(n-1)} \\
&= \frac{1}{N} \sum_i \left(\mathbf{d}_{ji} - \mathbf{t}_i^{(n-1)} \right) \\
&= \dots \\
&= \frac{1}{N} \sum_i \left(\mathbf{d}_{ji} - \mathbf{t}_i^{(0)} \right) \\
&= \boldsymbol{\mu}_j^{(1)}.
\end{aligned} \tag{A.12}$$

The computed rater bias at any iteration is equal to the initial bias. Fundamentally, although the value of $\mathbf{t}_i^{(n)}$ changes at each iteration, their summation over all points i stays the same, which causes the bias invariance problem. Although the EM algorithm is guaranteed to converge to a local optimum [104], the local optimum is independent of the bias in this case, which indicates an irrelevant relationship between the bias parameter and the likelihood function. This result has two key implications. First, the CSTAPLE algorithm does not actually estimate rater bias, which is one of the rater performance measures. Instead the bias is indeterminate from the maximum likelihood estimation framework. Second, if the initial bias is specified to be far

APPENDIX A. APPENDIX

from the true bias, the estimate of the true label could also be negatively affected. Therefore, rather than the expected situation in which the EM algorithm uses the observed data to optimally estimate both rater performances and the true continuous label, we find ourselves facing a situation in which initialization is crucial in fact, it is “the whole game”.

Fig. A.2 illustrates the consequences of poor bias initialization. In the identification of RV insertion points, the decisions of six raters are denoted by dots in Fig. A.2(a). They are then fused by CSTAPLE with the truth estimate initialized at the image origin (top left pixel), which results in the first calculation of rater bias to be very large and the final estimated truth denoted by “+” in Fig. A.2(a) to be far away from the correct position. In the distance transform approach to calculate fusion of the endocardium, the true distance function estimate can be initialized with zeros on the entire image plane. The fusion of six raters’ distance functions (whose zero level sets are shown as colored contours in Figs. A.2(b) and A.2(c)) is calculated and its zero position is extracted as the estimated endocardium contour (shown in Fig. A.2(d)). Because of this initialization the fusion result is clearly wrong, yet it is nevertheless optimal from a maximum likelihood perspective. This demonstrates that naive initialization may lead to inaccurate interpretation of the bias, thereby degrading the final truth estimate.

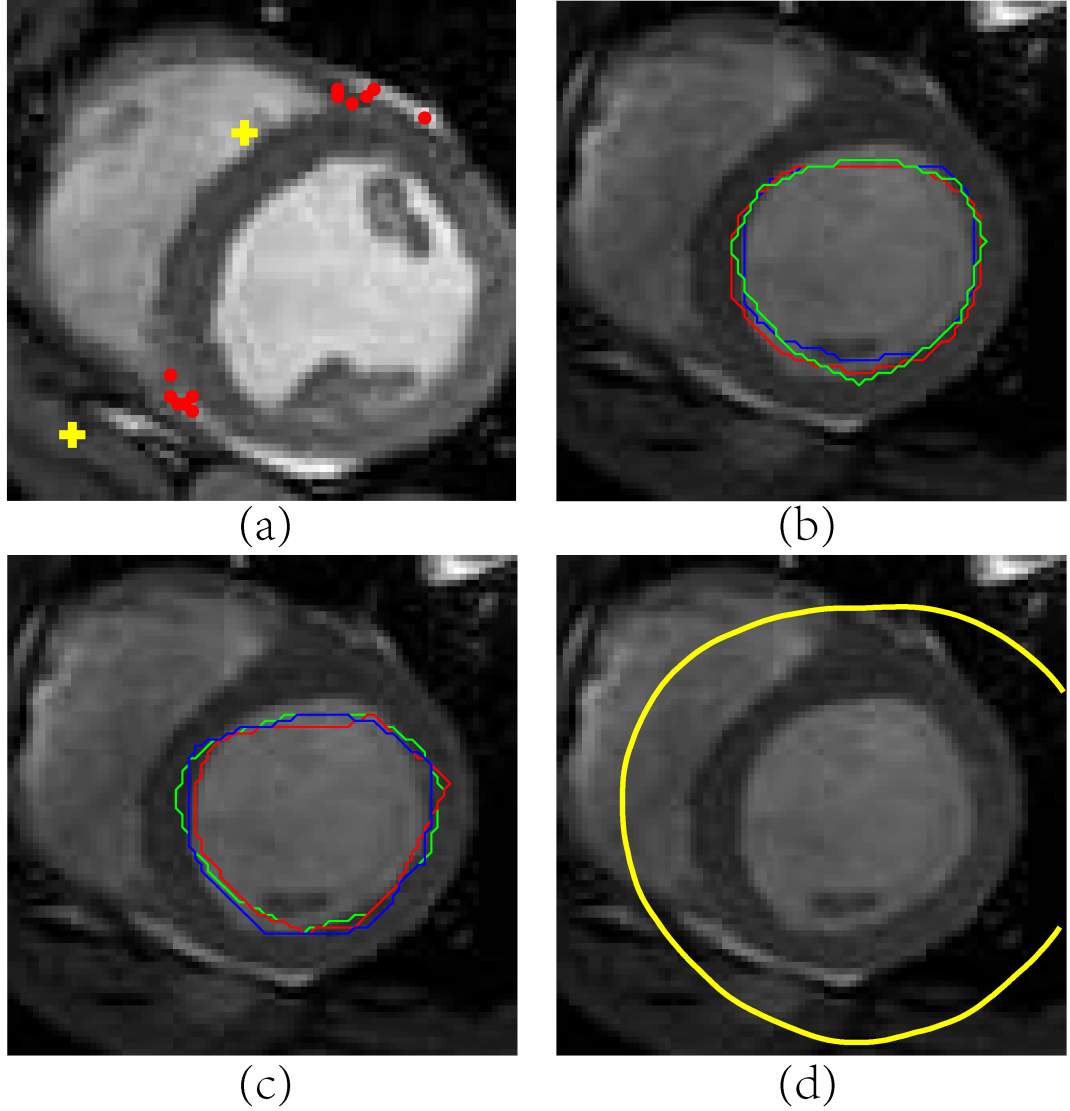


Figure A.2: The consequence of poor bias initialization under two continuous label fusion scenarios. (a) Six raters identify RV insertion points (dots) and their fusions (crosses) are poor because CSTAPLE is initialized in upper left corner. Six raters identify endocardium contours (three shown in (b) and three more shown in (c)). The fusion of six endocardium contours shown in (d) is poor because CSTAPLE was initialized with zeros on the entire image plane.

A.4 Initialization Strategies

Since use of incorrect rater biases yields poor fusion results and since any bias is optimal given a maximum likelihood criterion, it is necessary to consider alternate ways to handle rater bias. Since CSTAPLE does not change bias, one approach is to estimate rater bias in advance and use that estimate to initialize CSTAPLE. Here we state and comment on three possible strategies.

- *Zero initialization:* The EM iterations in discrete STAPLE are often started using a confusion matrix with diagonal values close to one (commonly observed performance parameters). Similarly, in the continuous case we can start with all rater biases as zero vectors. There will be cases, however, where this choice is wrong. For example, if one class of raters (i.e., novices) made systematic mistakes with large biases relative to a set of other (i.e., experienced) raters, then the larger bias of the first class relative to the second would never be estimated and used in fusion. Although this initialization would seem to be “fair” in that it makes no particular prior assumption about the rater bias, the fact that it never adapts to the fusion result that emerges is counter to the well-founded and elegant principles of the STAPLE approach.
- *Average initialization:* In this approach, we first calculate the mean $\mathbf{t}_i^{(0)}$ of all rater decisions, which also serves as an initial estimate of the truth. Each rater’s bias $\boldsymbol{\mu}_j^{(1)}$ is then calculated as the average deviation from this initial estimate

APPENDIX A. APPENDIX

of the truth. In equation form

$$\mathbf{t}_i^{(0)} = \frac{1}{R} \sum_j \mathbf{d}_{ji}, \quad i = 1, \dots, N \quad (\text{A.13})$$

$$\boldsymbol{\mu}_j^{(1)} = \frac{1}{N} \sum_i \left(\mathbf{d}_{ji} - \mathbf{t}_i^{(0)} \right), \quad j = 1, \dots, R. \quad (\text{A.14})$$

Without prior information of the acquired data, the mean location is already an appropriate fusion (unbiased estimator), although not necessarily optimal under the STAPLE framework. Using it to achieve an ML solution can be viewed as a coarse-to-fine strategy. Furthermore, various averaging strategies can be considered to adapt different cases, e.g., if majority's decisions are more trusted, a robust weighted mean can be calculated to reduce impacts of outlier, where the weight can be the inverse square distance of \mathbf{d}_{ji} from $\mathbf{t}_i^{(0)}$.

- *Informed initialization:* If prior knowledge of rater bias is available from previous experience or from a training dataset, it can be used in an informed initialization strategy. One could subtract off the prior bias $\boldsymbol{\mu}_{\mu_j}$ from corresponding rater's decisions and then use average initialization, i.e., for every j , prune all rater decisions by

$$\mathbf{d}_{ji, \text{new}} = \mathbf{d}_{ji} - \boldsymbol{\mu}_{\mu_j}, \quad \forall i, j \quad (\text{A.15})$$

and then update Eqn. A.13 and Eqn. A.14 with $\mathbf{d}_{ji, \text{new}}$. When the prior knowl-

APPENDIX A. APPENDIX

edge is reliable, this strategy sidesteps the problem of bias estimation and proves to be the most accurate one and effective in distinguishing bad raters even if they are the majority. To be effective, the bias prior $\boldsymbol{\mu}_{\mu_j}$ must be correctly learned, and this in itself may not be an easy task.

In general, although these initialization strategies are either based on the current data or obtained from previous experience, they are mathematically equivalent in that they do not affect the numerical value of the maximum likelihood optimum. A common concern for all these methods is that the bias is estimated separately from the truth estimation process.

A.5 MAP Estimation for Continuous Label Fusion

As an alternative to the pre-estimation of bias, we may apply soft constraints on the bias parameter in the form of a maximum *a priori* (MAP) optimization, so that bias can be estimated simultaneously with the truth levels. We refer to this approach as MAP-CSTAPLE. In line with previous work on Gaussian mixture models [105–107], we use the following prior on bias

$$f(\theta_j) = f(\boldsymbol{\mu}_j) = \frac{1}{(2\pi)^{K/2} \sqrt{\det(\Sigma_{\boldsymbol{\mu}_j})}} e^{-\frac{1}{2}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{\boldsymbol{\mu}_j})^T \Sigma_{\boldsymbol{\mu}_j}^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_{\boldsymbol{\mu}_j})}, \quad (\text{A.16})$$

APPENDIX A. APPENDIX

where $\boldsymbol{\mu}_{\mu_j}$ and Σ_{μ_j} are the mean and covariance of rater j 's bias $\boldsymbol{\mu}_j$.

Comparing to Eqn. A.4, now we seek to maximize the logarithm of the *a posteriori* distribution

$$E[(\ln f(D, T|\theta) + \ln f(\theta))|D, \theta^{(n)}] = E[\ln f(D, T|\theta)|D, \theta^{(n)}] + \ln f(\theta). \quad (\text{A.17})$$

Consequently, in Eqn. A.8 function $F_j^{(n)}$ now becomes

$$F_j^{(n)} = \sum_i \int_{\mathbb{R}^K} \ln f(\mathbf{d}_{ji}|\mathbf{t}_i, \theta_j) W_i^{(n)}(\mathbf{t}_i) d\mathbf{t}_i + \ln f(\theta_j). \quad (\text{A.18})$$

The E-step is the same as before by Eqn. A.7 but the M-step becomes

$$\begin{aligned} \boldsymbol{\mu}_j^{(n+1)} &= (I + \frac{1}{N} \Sigma_j^{(n+1)} \Sigma_{\mu_j}^{-1})^{-1} (\frac{1}{N} \sum_i (\mathbf{d}_{ji} - A^{(n)} \mathbf{b}_i^{(n)}) + \frac{1}{N} \Sigma_j^{(n+1)} \Sigma_{\mu_j}^{-1} \boldsymbol{\mu}_{\mu_j}), \\ \Sigma_j^{(n+1)} &= \frac{1}{N} \sum_i [A^{(n)} + (\mathbf{d}_{ji} - \boldsymbol{\mu}_j^{(n+1)} - A^{(n)} \mathbf{b}_i^{(n)}) (\mathbf{d}_{ji} - \boldsymbol{\mu}_j^{(n+1)} - A^{(n)} \mathbf{b}_i^{(n)})^T]. \end{aligned} \quad (\text{A.19})$$

With these modifications, the bias is updated in the EM steps and convergence is achieved.

To implement this algorithm, $\boldsymbol{\mu}_{\mu_j}$ and Σ_{μ_j} must be determined or specified in advance. Similar to three initialization strategies, we present three possible ways to determine these quantities.

- *Weak prior*: If the raters are not known to have bias, one can let $\boldsymbol{\mu}_{\mu_j}$ be zero

APPENDIX A. APPENDIX

and Σ_{μ_j} be large (e.g., 10 voxels for RV identification). Because of the existence of Σ_{μ_j} , the estimation process is able to account for unusually large deviations from the truth in order to compensate for the assumed prior zero bias and therefore achieve more stable results than zero initialization.

- *Data adaptive prior*: Here, we apply the average initialization strategy and then take $\mu_j^{(1)}$ in Eqn. A.14 as the bias prior μ_{μ_j} , and Σ_{μ_j} will be the covariance of \mathbf{d}_{ji} for all i . This strategy uses the current data to estimate a more restrictive bias prior than the weak prior strategy.
- *Informed prior*: If prior knowledge (mean and covariance) of the rater bias is available, we can use it directly by setting it as μ_{μ_j} and Σ_{μ_j} . It is similar to informed initialization except that estimated bias is allowed to vary around the prior mean according to the deviation specified by the prior variance.

If CSTAPLE is used with a correct bias initialization, then it is optimal. MAP-CSTAPLE provides a degree of “protection” against improper bias initialization, which may be useful in some cases. We now present experiments that demonstrate both the utility and pitfalls of the two classes of proposed approaches and their variations.

A.6 2D Point Identification Simulations

Six raters were simulated with manually assigned biases and variances in a 2D point identification problem. Each rater evaluated 10 randomly generated points in a 100×100 region of interest (ROI) according to the two models, instances of which are shown in Figs. A.3(a) and A.3(d). We evaluated all six methods.

In the first model, we assigned each rater to have a bias that is uniformly sampled from $(0, 5]$ pixels in length and with random direction. Rater covariance matrices were set to random positive definite matrices whose diagonal values are around 9 pixel^2 . An instance of this is shown in Fig. A.3(a). Assuming no prior information of the bias was known, CSTAPLE with zero initialization and average initialization and MAP-CSTAPLE with weak and data-adaptive prior were evaluated. Then assuming prior information of the bias was available (using generated bias and variance), CSTAPLE with informed initialization and MAP-CSTAPLE with informed prior were evaluated. For each of the six fusion techniques, the entire experiment was repeated in 500 Monte Carlo trials. A typical model instance is shown in Fig. A.3(a) and its estimates are shown in Figs. A.3(b) and A.3(c).

In the second model, we changed the generation of rater bias to let 3 out of 6 raters make similar mistakes, deviating 10 ± 2 pixels in length toward the upper right direction. An instance of this model is shown in Fig. A.3(d) and the estimation results are shown in Figs. A.3(e) and A.3(f). As before, we repeated the experiment in 500 Monte Carlo trials. Table A.1 compares the RMSE (root mean squared error)

APPENDIX A. APPENDIX

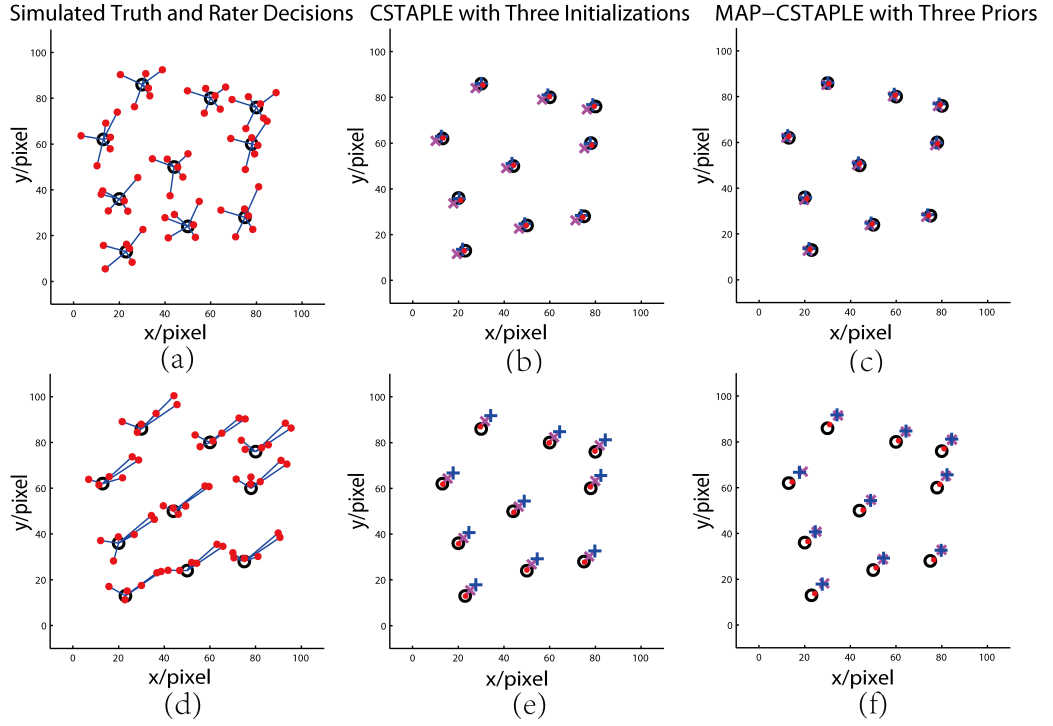


Figure A.3: CSTAPLE simulation of 2D point identification. In (a) (d) circles are generated truth and dots are rater decisions. In (b) (e) “x” markers are the fusion of zero initialization, crosses are average initialization, and dots are informed initialization. In (c) (f) “x” markers are fusion of weak prior, crosses are data-adaptive prior, and dots are informed prior.

APPENDIX A. APPENDIX

Table A.1: RMSE (in pixels) of Estimated Truth from Generated Truth with Six Fusion Techniques in 500 Monte Carlos of 2D Simulation.

	CSTAPLE Initialization			MAP-CSTAPLE Priors		
	Zero	Average	Informed	Weak	Data-adaptive	Informed
Model 1	3.3±1.2	2.0±0.7	0.9±0.3	3.4±1.3	2.0±0.7	1.0±0.3
Model 2	3.4±1.5	5.9±1.8	0.8±0.3	3.6±2.6	5.9±1.8	1.0±0.4

in units of pixels of the estimated truth from the generated truth for all six approaches. It is observed that the informed versions of both initialized CSTAPLE and MAP-CSTAPLE perform best. The average initialization and data-adaptive MAP-CSTAPLE are excellent when the raters are uniformly biased (Model 1) but these approaches are quite bad when a fraction of the raters are biased (Model 2). The zero initialization in CSTAPLE and the weak prior in MAP-STAPLE have intermediate and approximately equal performance for both rater models; thus, they represent “safe” choices when there is no available rater information (and the possibility of large rater bias exists).

It is also worth to mention that the estimation of covariance matrices Σ_j is regarded accurate with an average absolute error of $\begin{bmatrix} 1.1 & 0.8 \\ 0.8 & 0.5 \end{bmatrix}$ (pixel²), regardless of its initialization. In our experiments, the variance estimation does not tend to cause any major problem to the algorithm.

A.7 Empirical Fusion: RV Insertion Points Identification in Cardiac Images

A high-resolution Cine MR short axis image set of the heart of a pig was obtained in a steady-state free suppression acquisition with breath holds on a commercial Philips 3T Achieva whole body system (Philips, Andover, MA). The scan acquisition parameters are FOV: $280 \times 72 \times 280$ mm³, Size: 176×215 , Scan Duration: 124 s and Repetition Time: 3.333 ms. Six human raters with no previous experience on labeling cardiac data were given a 15-minute training session and were asked to identify 82 RV insertion points in 41 designated image slices. An expert on cardiac anatomy labeled the same data using the same in-house software.

Rater performance and true RV locations were estimated with the six fusion techniques using the point-wise data from the six inexperienced raters. The fused RV locations were compared with the location specified by the expert rater as “truth”. To implement the informed CSTAPLE methods we used half of the dataset (20 images) as training data and compared the rater decisions in the training data directly with the expert’s decision, obtaining the rater’s average deviation from the truth and its covariance as the prior mean and prior covariance. The experiment was repeated in 100 Monte Carlo trials, each with 20 random selected training images and 21 remaining test images. In each Monte Carlo, fusions of the test image RV points with six methods and their RMSE from expert decision were computed. Finally, the av-

APPENDIX A. APPENDIX

Table A.2: RMSE (in pixels) of Estimated Truth from Expert Truth with Six Fusion Techniques in 100 Monte Carlos of Real RV Insertion Points Data.

CSTAPLE Initialization			MAP-CSTAPLE Priors		
Zero	Average	Informed	Weak	Data-adaptive	Informed
5.5 ± 0.6	4.7 ± 0.4	4.2 ± 0.4	5.2 ± 0.6	4.7 ± 0.4	4.2 ± 0.4

erage and standard deviation of the RMSE through all Monte Carlos were evaluated (Table A.2).

The results of all six methods on one slice are shown in Figs. A.4(a) and A.4(b). To better visualize the differences between MAP-CSTAPLE with the data-adaptive prior and MAP-CSTAPLE with the informed prior we plotted their two distances from the truth as an ordered pair on the x - y axis. Five hundred of these points, one from each Monte Carlo trial, are shown in Fig. A.4(c). The fact that more points fall above the $y = x$ line reveals that the informed prior is generally better. This confirms the RMSE results shown in Table A.2.

A.8 Empirical Fusion: Contour Identification in Cardiac Images

The cardiac MRI dataset, as described in the previous section, was used for endocardium contour identification. The same six inexperienced raters manually labeled the endocardium on all slices after a 15-minute training session. The expert performed the same task. The labeling was achieved by direct delineation (painting the

APPENDIX A. APPENDIX

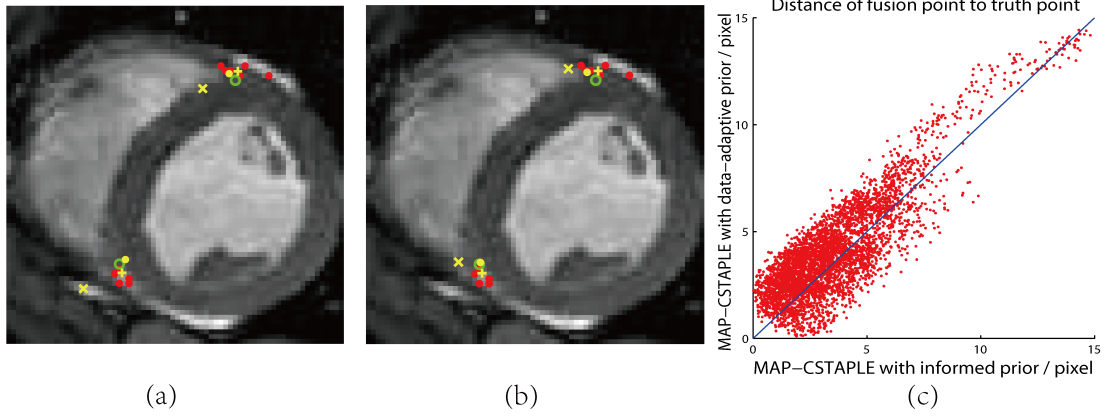


Figure A.4: Identification of RV insertion points in cardiac MRI. In (a) (b) red dots are rater decisions and green circles are an expert's decision as ground truth. Fusions are shown in yellow, where "x" markers, crosses and dots are respectively zero, average, informed initialization in (a) and weak, data-adaptive, informed prior in (b). The error distance of all fusion points from corresponding truth points shown in (c) which compares data-adaptive prior and informed prior methods.

endocardium area) using the same in-house software as in the previous section. In order to compare the continuous contour fusion result with discrete label fusion result, we performed classic STAPLE on rater decisions in discrete domain by assigning Label 1 as endocardium and Label 0 as background.

We considered one image slice for detailed evaluation. The pixel size of the region of interest was 80×80 so that the total pixel count was 6400. Before performing continuous fusion, we computed the signed distance function from the contour of the manually delineated endocardium, which resulted in six decision sets, each of 6400 one-dimensional vectors (scalars). They were then fused by the six continuous fusion techniques respectively. Finally the fusion's zero level set was regarded as the estimated contour. As in the previous section, to implement the informed CSTAPLE

APPENDIX A. APPENDIX

Table A.3: Dice Coefficients (in percentage) of Estimated Truth from Expert Truth with Six Fusion Techniques in 50 Monte Carlos and Discrete STAPLE for Endocardium Identification.

CSTAPLE Initialization			MAP-CSTAPLE Priors			STAPLE
Zero	Average	Informed	Weak	Data-adaptive	Informed	
81.9±3.0	93.3±0.0	92.6±0.1	93.2±0.0	93.3±0.0	93.3±0.0	92.3

methods we used part of the dataset (1000 pixels) as training data and compared the rater distance functions in the training data directly with the expert’s distance function, obtaining the rater’s average deviation from the truth and its covariance as the prior mean and prior covariance. The experiment was repeated in 50 Monte Carlo trials, each with 1000 random selected training pixels. An example of the expert decision and fusion results of all methods are shown in Fig. A.5.

The Dice coefficients of the endocardium fusions in comparison to the expert decision were computed, and the average and standard deviation of all Monte Carlo trials are shown in Table A.3. The Dice coefficients show that well initialized CSTAPLE (average, informed) and all MAP-CSTAPLE methods perform better than classic STAPLE. Also, poorly initialized CSTAPLE (zero in this case) can lead to very poor results (which can also be seen in Fig. A.5(b)).

We then changed the number of training pixels to alter the prior mean and variance for informed CSTAPLE methods. The results in Table A.4. show that both informed methods are quite stable with respect to numbers of training samples.

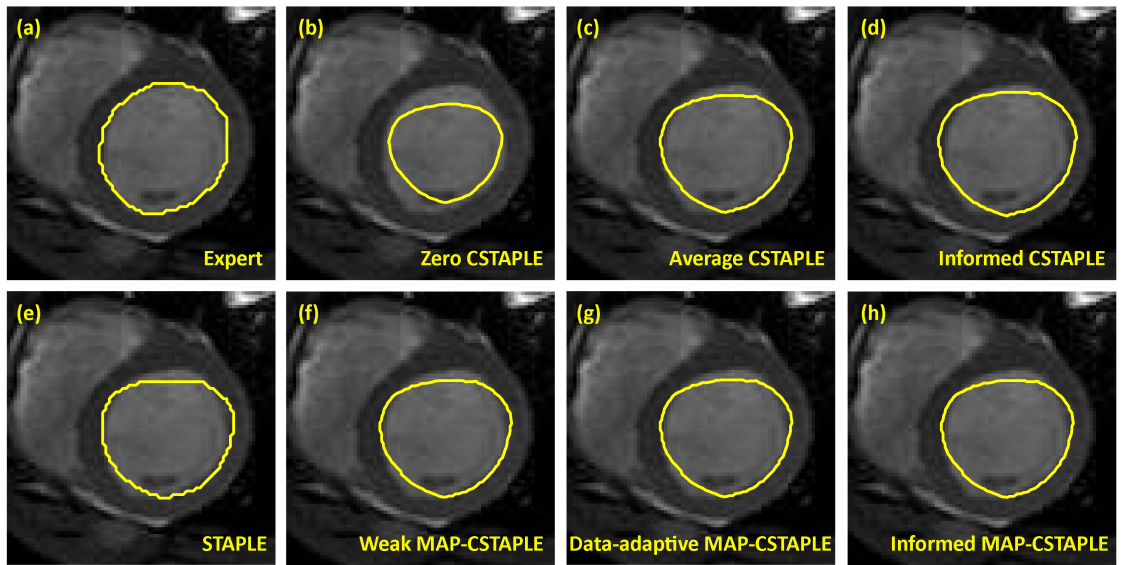


Figure A.5: Different fusion methods for endocardium contour identification. (a) Expert decision from manual delineation regarded as truth. (b) CSTAPLE with zero initialization. (c) CSTAPLE with average initialization. (d) CSTAPLE with informed initialization. (e) Classic STAPLE fusion of discrete labels. (f) MAP-CSTAPLE with weak prior. (g) MAP-CSTAPLE with data-adaptive prior. (h) MAP-CSTAPLE with informed prior.

APPENDIX A. APPENDIX

Table A.4: Dice Coefficients (in percentage) of Estimated Truth from Expert Truth with Two Informed CSTAPLE Methods Subject to Training Dataset Size Change.

Number of Training Pixels	10	50	100	500	1000	2000	3000	4000	5000	6000
Informed Initialization Dice	92.5	92.6	92.5	92.5	92.6	92.6	92.5	92.6	92.5	92.5
Informed Prior Dice	93.3	93.3	93.3	93.3	93.3	93.3	93.3	93.3	93.3	93.3

A.9 Discussion

We observed that zero initialization CSTAPLE and weak prior MAP-CSTAPLE led to mediocre performance in all experiments. Thus these approaches are not recommended for use. Average initialization CSTAPLE and data-adaptive MAP-CSTAPLE lead to superior fusion results except in the pathological case of Model 2 in Table A.1. The limitation of these two methods is that they require most raters to perform well. This problem can be addressed by using either the informed initialization CSTAPLE or informed prior MAP-CSTAPLE, but only when appropriate information about the rater biases are known beforehand. In the human rater experiments, the methods using training data to estimate an informed approach were the most successful.

In the contour identification task, we saw that as long as the bias was handled appropriately (i.e., did not use a zero initialization), the continuous fusion result was similar to that of discrete STAPLE. Average initialization CSTAPLE and data-adaptive MAP-CSTAPLE provided an excellent fusion result and, as in the RV insertion points example, informed CSTAPLE methods did not show apparent advantages. Except for the zero initialization CSTAPLE case, all other proposed methods are slightly better than discrete STAPLE, and they have the potential advantage of providing

APPENDIX A. APPENDIX

subvoxel delineations.

The comparison of different methods demonstrates that informed approaches are better on both simulated and real data. Although incorporating prior knowledge of human raters performance can be particularly challenging, recent developments have shown that learning the performance of automated methods is possible [108]. Otherwise, average initialization and data-adaptive prior methods can be regarded as proper continuous fusion techniques in general without the presence of prior information. The major contribution of this work is to provide a theoretical correction to the CSTAPLE algorithm. In practice, since various reasonable parameter-tweaking methods (e.g., tuning the covariance matrix parameter) can lead to reasonable solutions, this work may be perceived as subtle. However, severe pathological failures may arise if the user is not aware of the fundamental shortcomings. For example, in the cardiac RV insertion points picking task, one rater is seen to consistently make the same mistake for every image slice by identifying the top right RV insertion point to the right of its correct position (Figs. A.4(a) and A.4(b)). Although this rater has a small variance, the rater’s bias to the right of the truth is not estimated by the algorithm, and also cannot be compensated by the estimation of the variance. It can be argued that a more straightforward solution is not to consider the misinformed rater’s decision. But in practice, it is not always possible to manually examine each rater’s decision prior to fusing the data facing a great number of dataset.

Finally, the focus of this discussion is on the general CSTAPLE algorithm, where

APPENDIX A. APPENDIX

the assumptions are inherited from those of the discrete STAPLE. However, in certain special cases where prior knowledge of the truth is known, assumptions can be changed to include a non-uniform prior ($f(T)$ in Eqn. A.6) or to use a non-Gaussian framework, which will result in a change of derivation of equations and is likely to eliminate the bias invariance problem. Details of these methods are not discussed in this paper.

A.10 Summary

In this chapter, we first proved that rater bias as a performance parameter is not updated after the first step in the CSTAPLE algorithm. We then presented two families of solutions, each with three variations, to address this problem. Although informed methods—known biases or their statistics—are best, the original CSTAPLE algorithm initialized with biases computed from the group average or MAP-CSTAPLE using a data-adaptive prior provide essentially equivalent results in realistic scenarios. We note that in some cases the differences between these approaches could be considered clinically nominal (e.g., DSC differences of 0.01%); the major contribution confirmed by experimental results is the finding that poor bias initialization may lead to very poor results when using a naive fusion approach. Hence, it is important to evaluate these considerations in practice.

Bibliography

- [1] S. Abd-El-Malek, “The part played by the tongue in mastication and deglutition,” *Journal of anatomy*, vol. 89, no. Pt 2, p. 250, 1955.
- [2] H. Takemoto, “Morphological analyses of the human tongue musculature for three-dimensional modeling,” *Journal of Speech, Language, and Hearing Research*, vol. 44, no. 1, pp. 95–107, 2001.
- [3] A. Maton, *Human biology and health*. Prentice Hall, 1997.
- [4] 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,” *Plastic and reconstructive surgery*, vol. 113, no. 1, pp. 114–125, 2004.
- [5] “American Cancer Society webpage,” <http://www.cancer.org/statistics>, accessed: 2014.
- [6] K. S. Heller, J. Levy, and J. J. Sciubba, “Speech patterns following partial

BIBLIOGRAPHY

- glossectomy for small tumors of the tongue,” *Head & neck*, vol. 13, no. 4, pp. 340–343, 1991.
- [7] B. R. Pauloski, J. A. Logemann, L. A. Colangelo, A. W. Rademaker, F. McConnel, M. A. Heiser, S. Cardinale, D. Shedd, D. Stein, Q. Beery *et al.*, “Surgical variables affecting speech in treated patients with oral and oropharyngeal cancer,” *The Laryngoscope*, vol. 108, no. 6, pp. 908–916, 1998.
- [8] R. Mayeux, R. Ottman, M.-X. Tang, L. Noboa-Bauza, K. Marder, B. Gurland, and Y. Stern, “Genetic susceptibility and head injury as risk factors for alzheimer’s disease among community-dwelling elderly persons and their first-degree relatives,” *Annals of neurology*, vol. 33, no. 5, pp. 494–501, 1993.
- [9] J. A. Mortimer, C. M. Van Duijn, V. Chandra, L. Fratiglioni, A. B. Graves, A. Heyman, A. F. Jorm, E. Kokmen, K. Kondo, W. A. Rocca *et al.*, “Head trauma as a risk factor for Alzheimer’s disease: a collaborative re-analysis of case-control studies,” *International journal of epidemiology*, vol. 20, no. Supplement 2, pp. S28–S35, 1991.
- [10] C. M. Baugh, J. M. Stamm, D. O. Riley, B. E. Gavett, M. E. Shenton, A. Lin, C. J. Nowinski, R. C. Cantu, A. C. McKee, and R. A. Stern, “Chronic traumatic encephalopathy: neurodegeneration following repetitive concussive and subconcussive brain trauma,” *Brain imaging and behavior*, vol. 6, no. 2, pp. 244–254, 2012.

BIBLIOGRAPHY

- [11] J. A. Langlois, W. Rutland-Brown, and M. M. Wald, “The epidemiology and impact of traumatic brain injury: a brief overview,” *The Journal of head trauma rehabilitation*, vol. 21, no. 5, pp. 375–378, 2006.
- [12] R. Drake, A. W. Vogl, and A. W. M. Mitchell, *Gray’s anatomy for students*. Elsevier Health Sciences, 2014.
- [13] W. Kier and K. Smith, “Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats,” *Zoological Journal of the Linnean Society*, vol. 83, no. 4, pp. 307–324, 1985.
- [14] E. Zerhouni, D. Parish, W. Rogers, A. Yang, and E. Shapiro, “Human heart: tagging with MR imaging—a method for noninvasive assessment of myocardial motion,” *Radiology*, vol. 169, no. 1, pp. 59–63, 1988.
- [15] V. Parthasarathy, J. L. Prince, M. Stone, E. Murano, and M. NessAiver, “Measuring tongue motion from tagged Cine-MR using harmonic phase (HARP) processing,” *J. Acoustic Society of America*, vol. 121, no. 1, pp. 491–504, 2007.
- [16] A. Knutsen, E. Magrath, J. McEntee, F. Xing, J. Prince, P. Bayly, J. Butman, and D. Pham, “Improved measurement of brain deformation during mild head acceleration using a novel tagged MRI sequence,” *Journal of Biomechanics*, vol. 47, no. 14, pp. 3475–3481, 2014.
- [17] M. Guttman, J. Prince, and E. McVeigh, “Tag and contour detection in tagged

BIBLIOGRAPHY

- MR images of the left ventricle,” *Medical Imaging, IEEE Transactions on*, vol. 13, no. 1, pp. 74–88, 1994.
- [18] M. Guttman, E. Zerhouni, E. McVeigh *et al.*, “Analysis of cardiac function from MR images,” *Computer Graphics and Applications, IEEE*, vol. 17, no. 1, pp. 30–38, 1997.
- [19] A. Young, D. Kraitchman, L. Dougherty, L. Axel *et al.*, “Tracking and finite element analysis of stripe deformation in magnetic resonance tagging,” *Medical Imaging, IEEE Transactions on*, vol. 14, no. 3, pp. 413–421, 1995.
- [20] A. Amini, Y. Chen, R. Curwen, V. Mani, J. Sun *et al.*, “Coupled B-snake grids and constrained thin-plate splines for analysis of 2-D tissue deformations from tagged MRI,” *Medical Imaging, IEEE Transactions on*, vol. 17, no. 3, pp. 344–356, 1998.
- [21] W. Kerwin and J. Prince, “Cardiac material markers from tagged MR images,” *Medical Image Analysis*, vol. 2, no. 4, pp. 339–353, 1998.
- [22] N. Osman, W. Kerwin, E. McVeigh, and J. Prince, “Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging,” *Magnetic Resonance Medicine*, vol. 42, pp. 1048–1060, 1999.
- [23] N. Osman, E. McVeigh, and J. Prince, “Imaging heart motion using harmonic

BIBLIOGRAPHY

- phase MRI,” *Medical Imaging, IEEE Transactions on*, vol. 19, no. 3, pp. 186–202, 2000.
- [24] N. Osman and J. Prince, “Visualizing myocardial function using HARP MRI,” *Physics in medicine and biology*, vol. 45, no. 6, p. 1665, 2000.
- [25] T. Chen, X. Wang, S. Chung, D. Metaxas, and L. Axel, “Automated 3d motion tracking using gabor filter bank, robust point matching, and deformable models,” *Medical Imaging, IEEE Transactions on*, vol. 29, no. 1, pp. 1–11, 2010.
- [26] S. Ryf, M. Spiegel, M. Gerber, and P. Boesiger, “Myocardial tagging with 3D-CSPAMM,” *J. Magnetic Resonance Imaging*, vol. 16, pp. 320–325, 2002.
- [27] T. Reese, D. Feinberg, J. Dou, and V. Wedeen, “Phase contrast MRI of myocardial 3D strain by encoding contiguous slices in a single shot,” *Magnetic resonance in medicine*, vol. 47, no. 4, pp. 665–676, 2002.
- [28] K. Abd-Elmoniem, N. Osman, J. Prince, and M. Stuber, “Three-dimensional magnetic resonance myocardial motion tracking from a single image plane,” *Magnetic Resonance in Medicine*, vol. 58, no. 1, pp. 92–102, 2007.
- [29] B. Spottiswoode, X. Zhong, C. Lorenz, B. Mayosi, E. Meintjes, and F. Epstein, “3D myocardial tissue tracking with slice followed cine DENSE MRI,” *Journal of Magnetic Resonance Imaging*, vol. 27, no. 5, pp. 1019–1027, 2008.

BIBLIOGRAPHY

- [30] F. Yin, C. Chan, and R. Judd, “Compressibility of perfused passive myocardium,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 271, no. 5, pp. H1864–H1870, 1996.
- [31] I. Rodriguez, D. Ennis, and H. Wen, “Noninvasive measurement of myocardial tissue volume change during systolic contraction and diastolic relaxation in the canine left ventricle,” *Magnetic resonance in medicine*, vol. 55, no. 3, pp. 484–490, 2006.
- [32] R. Gilbert, V. Napadow, T. Gaige, and V. Wedeen, “Anatomical basis of lingual hydrostatic deformation,” *Journal of Experimental Biology*, vol. 210, no. 23, pp. 4069–4082, 2007.
- [33] G. E. Christensen and H. J. Johnson, “Consistent image registration,” *Medical Imaging, IEEE Transactions on*, vol. 20, no. 7, pp. 568–582, 2001.
- [34] A. Leow, S.-C. Huang, A. Geng, J. Becker, S. Davis, A. Toga, and P. Thompson, “Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties,” in *Information Processing in Medical Imaging*. Springer, 2005, pp. 493–503.
- [35] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, “Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain,” *Medical image analysis*, vol. 12, no. 1, pp. 26–41, 2008.

BIBLIOGRAPHY

- [36] M. Chen, W. Lu, Q. Chen, K. Ruchala, and G. Olivera, “A simple fixed-point approach to invert a deformation field,” *Medical physics*, vol. 35, no. 1, pp. 81–88, 2007.
- [37] F. Xing, J. Lee, E. Murano, J. Woo, M. Stone, and J. Prince, “Estimating 3D tongue motion with MR images,” in *Asilomar Conference on Signals, Systems, and Computers 2012*. Springer, 2012.
- [38] W. O’Dell, C. Moore, W. Hunter, E. Zerhouni, and E. McVeigh, “Three-dimensional myocardial deformations: calculation with displacement field fitting to tagged MR images,” *Radiology*, vol. 195, no. 3, pp. 829–835, Jun 1995.
- [39] T. Denney and J. Prince, “Reconstruction of 3-D left ventricular motion from planar tagged cardiac MR images: an estimation theoretic approach,” *Medical Imaging, IEEE Transactions on*, vol. 14, no. 4, pp. 625–635, 1995.
- [40] I. Haber, D. Metaxas, and L. Axel, “Three-dimensional motion reconstruction and analysis of the right ventricle using tagged MRI,” *Medical Image Analysis*, vol. 4, no. 4, pp. 335–355, 2000.
- [41] J. Huang, D. Abendschein, V. Davila-Roman, and A. Amini, “Spatio-temporal tracking of myocardial deformations with a 4-D B-spline model from tagged MRI,” *Medical Imaging, IEEE Transactions on*, vol. 18, no. 10, pp. 957–972, 1999.

BIBLIOGRAPHY

- [42] C. Ozturk and E. McVeigh, “Four-dimensional B-spline based motion analysis of tagged MR images: introduction and in vivo validation,” *Physics in medicine and biology*, vol. 45, no. 6, p. 1683, 2000.
- [43] X. Liu, K. Abd-Elmoniem, M. Stone, E. Murano, J. Zhuo, R. Gullapalli, and J. Prince, “Incompressible deformation estimation algorithm (IDEA) from tagged MR images,” *Medical Imaging, IEEE Transactions on*, vol. 31, no. 2, pp. 326–340, 2012.
- [44] P. Shi, A. Sinusas, R. Constable, and J. Duncan, “Volumetric deformation analysis using mechanics-based data fusion: Applications in cardiac motion recovery,” *International Journal of Computer Vision*, vol. 35, no. 1, pp. 87–107, 1999.
- [45] X. Papademetris, A. Sinusas, D. Dione, R. Constable, and J. Duncan, “Estimation of 3-D left ventricular deformation from medical images using biomechanical models,” *Medical Imaging, IEEE Transactions on*, vol. 21, no. 7, pp. 786–800, 2002.
- [46] A. Amini and J. Prince, *Measurement of cardiac deformations from MRI: Physical and mathematical models*. Springer Science & Business Media, 2013, vol. 23.
- [47] J. Lee, J. Woo, F. Xing, E. Murano, M. Stone, and J. Prince, “Semi-automatic segmentation for 3D motion analysis of the tongue with dynamic MRI,” *Computerized Medical Imaging and Graphics*, vol. 38, no. 8, pp. 714–724, 2014.

BIBLIOGRAPHY

- [48] D. Pham, C. Xu, and J. Prince, “Current methods in medical image segmentation 1,” *Annual review of biomedical engineering*, vol. 2, no. 1, pp. 315–337, 2000.
- [49] A. Bistoquet, J. Oshinski, and O. Skrinjar, “Left ventricular deformation recovery from cine MRI using an incompressible model,” *Medical Imaging, IEEE Transactions on*, vol. 26, no. 9, pp. 1136–1153, 2007.
- [50] A. Bistoquet, J. Oshinski, and O. Skrinja, “Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model,” *Medical image analysis*, vol. 12, no. 1, pp. 69–85, 2008.
- [51] F. Xing, J. Woo, E. Murano, J. Lee, M. Stone, and J. Prince, “3D tongue motion from tagged and cine MR images,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 41–48.
- [52] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, “Diffeomorphic demons: efficient non-parametric image registration,” *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.
- [53] T. Mansi, J.-M. Peyrat, M. Sermesant, H. Delingette, J. Blanc, Y. Boudjemline, and N. Ayache, “Physically-constrained diffeomorphic demons for the estimation of 3D myocardium strain from cine-MRI,” in *Functional Imaging and Modeling of the Heart*. Springer, 2009, pp. 201–210.

BIBLIOGRAPHY

- [54] T. Mansi, X. Pennec, M. Sermesant, H. Delingette, and N. Ayache, “ilogdemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues,” *International journal of computer vision*, vol. 92, no. 1, pp. 92–111, 2011.
- [55] J. Prince and E. McVeigh, “Motion estimation from tagged MR image sequences,” *Medical Imaging, IEEE Transactions on*, vol. 11, no. 2, pp. 238–249, 1992.
- [56] M. NessAiver and J. L. Prince, “Magnitude image CSPAMM reconstruction (MICSR),” *Magnetic Resonance in Medicine*, vol. 50, no. 2, pp. 331–342, 2003.
- [57] L. Mannelli, G. J. Wilson, T. J. Dubinsky, C. A. Potter, P. Bhargava, C. Cuevas, K. F. Linnau, O. Kolokythas, M. L. Gunn, and J. H. Maki, “Assessment of the liver strain among cirrhotic and normal livers using tagged MRI,” *Journal of Magnetic Resonance Imaging*, vol. 36, no. 6, pp. 1490–1495, 2012.
- [58] M. A. Green, S. C. Gandevia, and L. E. Bilston, “A strain analysis of the lower leg during ankle rotation using HARP at 3 Tesla,” in *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 16, 2008, p. 2594.
- [59] Y. B. Fu, C.-K. Chui, C. L. Teo, and E. Kobayashi, “Motion tracking and strain map computation for quasi-static magnetic resonance elastography,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*. Springer, 2011, pp. 428–435.

BIBLIOGRAPHY

- [60] T. J. Ypma, “Historical development of the newton-raphson method,” *SIAM review*, vol. 37, no. 4, pp. 531–551, 1995.
- [61] X. Liu and J. L. Prince, “Shortest path refinement for motion estimation from tagged MR images,” *Medical Imaging, IEEE Transactions on*, vol. 29, no. 8, pp. 1560–1572, 2010.
- [62] V. Delgado, C. Ypenburg, R. J. van Bommel, L. F. Tops, S. A. Mollema, N. A. Marsan, G. B. Bleeker, M. J. Schalij, and J. J. Bax, “Assessment of left ventricular dyssynchrony by speckle tracking strain imaging: Comparison between longitudinal, circumferential, and radial strain in cardiac resynchronization therapy,” *Journal of the American College of Cardiology*, vol. 51, no. 20, pp. 1944–1952, 2008.
- [63] J. Woo, E. Murano, M. Stone, and J. Prince, “Reconstruction of high-resolution tongue volumes from MRI,” *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 12, pp. 3511–3524, 2012.
- [64] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.
- [65] L. Grady, “Random walks for image segmentation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 11, pp. 1768–1783, 2006.
- [66] J. Serra and P. Soille, *Mathematical morphology and its applications to image processing*. Springer Science & Business Media, 2012, vol. 2.

BIBLIOGRAPHY

- [67] R. D. Kent, *Acoustic analysis of speech*. Singular Publishing Group, 2002.
- [68] C. De Boor, “A practical guide to splines,” *Mathematics of Computation*, 1978.
- [69] M. Stone, J. M. Langguth, J. Woo, H. Chen, and J. L. Prince, “Tongue motion patterns in post-glossectomy and typical speakers: A principal components analysis,” *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 3, pp. 707–717, 2014.
- [70] K. Nakayama and G. Silverman, “The aperture problem—I. perception of non-rigidity and motion direction in translating sinusoidal lines,” *Vision research*, vol. 28, no. 6, pp. 739–746, 1988.
- [71] S. Shimojo, G. H. Silverman, and K. Nakayama, “Occlusion and the solution to the aperture problem for motion,” *Vision research*, vol. 29, no. 5, pp. 619–626, 1989.
- [72] A. Leow, I. Yanovsky, M. Chiang, A. Lee, A. Klunder, A. Lu, J. Becker, S. Davis, A. Toga, and P. Thompson, “Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration,” *Medical Imaging, IEEE Transactions on*, vol. 26, no. 6, pp. 822–832, 2007.
- [73] T. Lehmann, C. Gönner, and K. Spitzer, “Survey: Interpolation methods in medical image processing,” *Medical Imaging, IEEE Transactions on*, vol. 18, no. 11, pp. 1049–1075, 1999.

BIBLIOGRAPHY

- [74] M. Unser, A. Aldroubi, and M. Eden, “B-spline signal processing. I. Theory,” *Signal Processing, IEEE Transactions on*, vol. 41, no. 2, pp. 821–833, 1993.
- [75] S. Helgason, *Differential geometry and symmetric spaces*. American Mathematical Soc., 2001, vol. 341.
- [76] S. Nithiananthan, K. Brock, M. Daly, H. Chan, J. Irish, and J. Siewerdsen, “Demons deformable registration for CBCT-guided procedures in the head and neck: convergence and accuracy,” *Medical Physics*, vol. 36, no. 10, pp. 4755–4764, 2009.
- [77] B. K. Horn and B. G. Schunck, “Determining optical flow,” in *1981 Technical symposium east*. International Society for Optics and Photonics, 1981, pp. 319–331.
- [78] H. Takajo and T. Takahashi, “Least-squares phase estimation from the phase difference,” *JOSA A*, vol. 5, no. 3, pp. 416–425, 1988.
- [79] D. Ghiglia and L. Romero, “Direct phase estimation from phase differences using fast elliptic partial differential equation solvers,” *Optics letters*, vol. 14, no. 20, pp. 1107–1109, 1989.
- [80] D. Ghiglia and M. Pritt, *Two-dimensional phase unwrapping: theory, algorithms, and software*, vol. 4.

BIBLIOGRAPHY

- [81] S. Song, S. Napel, N. Pelc, and G. Glover, “Phase unwrapping of MR phase images using Poisson equation,” *Image Processing, IEEE Transactions on*, vol. 4, no. 5, pp. 667–676, 1995.
- [82] J.-M. Peyrat, H. Delingette, M. Sermesant, C. Xu, and N. Ayache, “Registration of 4D cardiac CT sequences under trajectory constraints with multichannel diffeomorphic demons,” *Medical Imaging, IEEE Transactions on*, vol. 29, no. 7, pp. 1351–1368, 2010.
- [83] J. Lloyd, I. Stavness, and S. Fels, “ArtiSynth: a fast interactive biomechanical modeling toolkit combining multibody and finite element simulation,” in *Soft tissue biomechanical modeling for computer assisted surgery*. Springer, 2012, pp. 355–394.
- [84] S. Maas, B. Ellis, G. Ateshian, and J. Weiss, “FEBio: finite elements for biomechanics,” *Journal of biomechanical engineering*, vol. 134, no. 1, p. 011005, 2012.
- [85] M. E. Gurtin, *An introduction to continuum mechanics*. Academic press, 1982.
- [86] M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, M. S. Verani *et al.*, “Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart a statement for healthcare professionals from the cardiac imaging committee of the Council on Clinical Cardiology of the American Heart Association,” *Circulation*, vol. 105, no. 4, pp. 539–542, 2002.

BIBLIOGRAPHY

- [87] J. K. Udupa, V. R. Leblanc, Y. Zhuge, C. Imielinska, H. Schmidt, L. M. Currie, B. E. Hirsch, and J. Woodburn, “A framework for evaluating image segmentation algorithms,” *Computerized Medical Imaging and Graphics*, vol. 30, no. 2, pp. 75–87, 2006.
- [88] S. K. Warfield, K. H. Zou, and W. M. Wells, “Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation,” *Medical Imaging, IEEE Transactions on*, vol. 23, no. 7, pp. 903–921, 2004.
- [89] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*. John Wiley & Sons, 2007, vol. 382.
- [90] T. Rohlfing, D. B. Russakoff, and C. R. Maurer, “Expectation maximization strategies for multi-atlas multi-label segmentation,” in *Information Processing in Medical Imaging*. Springer, 2003, pp. 210–221.
- [91] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman, “Linear time euclidean distance transform algorithms,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 17, no. 5, pp. 529–533, 1995.
- [92] O. Commowick and S. K. Warfield, “A continuous STAPLE for scalar, vector, and tensor images: an application to DTI analysis,” *Medical Imaging, IEEE Transactions on*, vol. 28, no. 6, pp. 838–846, 2009.

BIBLIOGRAPHY

- [93] S. K. Warfield, K. H. Zou, and W. M. Wells, “Validation of image segmentation by estimating rater bias and variance,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*. Springer, 2006, pp. 839–847.
- [94] F. Xing, S. Soleimanifard, J. L. Prince, and B. A. Landman, “Statistical fusion of continuous labels: identification of cardiac landmarks,” in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2011, pp. 796 206–796 206.
- [95] B. Landman, A. J. Asman, A. G. Scoggins, J. Bogovic, F. Xing, J. L. Prince *et al.*, “Robust statistical fusion of image labels,” *Medical Imaging, IEEE Transactions on*, vol. 31, no. 2, pp. 512–522, 2012.
- [96] A. J. Asman, B. Landman *et al.*, “Robust statistical label fusion through consensus level, labeler accuracy, and truth estimation (COLLATE),” *Medical Imaging, IEEE Transactions on*, vol. 30, no. 10, pp. 1779–1794, 2011.
- [97] F. Xing, A. J. Asman, J. L. Prince, and B. A. Landman, “Finding seeds for segmentation using statistical fusion,” in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2012, pp. 831 430–831 430.
- [98] T. A. Lampert, A. Stumpf, and P. Gañarski, “An empirical study into annotator agreement, ground truth estimation, and algorithm evaluation,” *arXiv preprint arXiv:1307.0426*, 2013.

BIBLIOGRAPHY

- [99] H. Wang, J. W. Suh, S. Das, J. Pluta, M. Altinay, and P. Yushkevich, “Regression-based label fusion for multi-atlas segmentation,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1113–1120.
- [100] M. R. Sabuncu, B. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland, “A generative model for image segmentation based on label fusion,” *Medical Imaging, IEEE Transactions on*, vol. 29, no. 10, pp. 1714–1729, 2010.
- [101] T. Rohlfing and C. R. Maurer Jr, “Shape-based averaging,” *Image Processing, IEEE Transactions on*, vol. 16, no. 1, pp. 153–161, 2007.
- [102] A. Suinesiaputra, B. R. Cowan, A. O. Al-Agamy, M. A. Elattar, N. Ayache, A. S. Fahmy, A. M. Khalifa, P. Medrano-Gracia, M.-P. Jolly, A. H. Kadish *et al.*, “A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images,” *Medical image analysis*, vol. 18, no. 1, pp. 50–62, 2014.
- [103] C. J. Wu, “On the convergence properties of the EM algorithm,” *The Annals of statistics*, pp. 95–103, 1983.
- [104] J. A. Bilmes *et al.*, “A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models,” *International Computer Science Institute*, vol. 4, no. 510, p. 126, 1998.

BIBLIOGRAPHY

- [105] J.-L. Gauvain and C.-H. Lee, “Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains,” *Speech and audio processing, IEEE transactions on*, vol. 2, no. 2, pp. 291–298, 1994.
- [106] J. V. Graça, K. Ganchev, and B. Taskar, “Expectation maximization and posterior constraints,” 2007.
- [107] D. L. Snyder and M. I. Miller, *Random point processes in time and space*. Springer Science & Business Media, 2012.
- [108] H. Wang, S. R. Das, J. W. Suh, M. Altinay, J. Pluta, C. Craige, B. Avants, P. A. Yushkevich, A. D. N. Initiative *et al.*, “A learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex and brain segmentation,” *NeuroImage*, vol. 55, no. 3, pp. 968–985, 2011.

Vita

Fangxu Xing was born in Qingdao, China on March 3rd, 1987. He received his B.S. degree in Mathematics and Physics from Tsinghua University in 2009. Starting from 2009, he was a graduate student at Johns Hopkins University and joined the Image Analysis and Communications Laboratory under the supervision of Dr. Jerry L. Prince. In 2011 and 2015, he completed M.S.E. and Ph.D. degrees in Electrical and Computer Engineering, respectively. His research interests include medical image analysis, computer vision, tissue motion estimation, tagged magnetic resonance imaging, speech analysis, and traumatic brain injury studies.