FAST UPPER AIRWAY MRI OF SPEECH

by

Yoon-Chul Kim

A Dissertation Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(ELECTRICAL ENGINEERING)

December 2010

Copyright 2010 Yoon-Chul Kim
Acknowledgments

First of all I would like to express a sincere gratitude to my mentor and advisor Prof. Krishna S. Nayak for his guidance during my Ph.D. studies. He has introduced me to this exciting field of MRI. He has guided me to change in a positive and productive way. He has encouraged me and provided me with a broad range of research opportunities. Also, he was very clear about giving me insights into any research ideas and their impacts. All of these helped me to get motivated and work hard for many years of research.

I sincerely thank Prof. Shrikanth Narayanan for providing support in vocal tract MRI research and help with improving the manuscripts. In addition, I thank my dissertation and qualifying exam committee members Prof. Louis Goldstein, Prof. Antonio Ortega, and Prof. Richard Leahy for their valuable suggestions.

I appreciate all members of the Magnetic Resonance Engineering Laboratory (MREL) group. I thank Prof. Jon-Fredrik Nielsen who had worked with me on my first MRI project about real-time cardiac echo-planar imaging artifact correction. Jon also helped me to learn tools and software programming related to real-time MRI and conduct the challenging lab job (i.e., the MREL computer system and networking administrations). I thank Prof. Houchun Harry Hu who has kindly answered any questions on MR research, has helped me with data collection, and has proofread on my papers. I thank all of my former and current lab mates Kyunghyun Sung, Joao Carvalho, Hsu-Lei Lee, Taehoon
Shin, Zungho Zun, Mahender Makhijani, Samir Sharma, Travis Smith, Yinghua Zhu for discussions, help, and all favors of being subjects for the MRI scans at the USC hospitals.

I also express a sincere gratitude to Prof. Sungbok Lee who has been cheering and introduced background on speech research. I thank Erik Bresch, Stephen Tobin, Michael Proctor, Vikram Ramanarayanan from the Speech Production and Articulation Knowledge (SPAN) group for helping each other in our real-time speech MRI data collection on weekends.

I express a gratitude to outside faculty and senior researchers with whom I had a chance to get involved. I thank Prof. Cecil E. Hayes, Prof. Bradley Sutton, Prof. Jong-Chul Ye, Dr. Meng Law, Dr. John L. Go, Dr. Sara Banerjee, Dr. Holger Eggers, Dr. Stefanie Remmele, Prof. Michael Lustig for their comments/suggestions related to my research. I also thank MRI technicians and GE engineers at the hospitals for their support in MR scanner related issues, and thank anonymous journal referees who had reviewed and made valuable suggestions on my manuscripts.

Finally, I dedicate my thesis to my mom, dad, and sister who live in Seoul, Korea. Also I dedicate this thesis to my dad’s brother and his family. Moreover, I cannot forget to thank my uncle’s family who lives in Anaheim, CA. I am deeply indebted to my mom for her caring and praying for my safety and success.
# Table of Contents

Acknowledgments ii

List Of Tables vi

List Of Figures vii

Abstract x

Chapter 1: Introduction 1
  1.1 Outline ............................................. 3

Chapter 2: MRI Background 6
  2.1 MRI Physics .......................................... 6
    2.1.1 Basic Pulse Sequences ............................. 6
      2.1.1.1 Excitation .................................. 6
      2.1.1.2 Readout .................................... 8
    2.1.2 \( k \)-space Trajectories ......................... 10
      2.1.2.1 Practical Constraints ......................... 10
      2.1.2.2 2D Imaging .................................. 10
      2.1.2.3 3D Imaging .................................. 13
  2.2 Accelerated Imaging .................................. 15
    2.2.1 Parallel Imaging .................................. 15
    2.2.2 Compressed Sensing ............................... 18
  2.3 Upper Airway Imaging .................................. 21
    2.3.1 Air-tissue Magnetic Susceptibility ................. 21
    2.3.2 Motion of Articulators ............................ 22
    2.3.3 Imaging Tradeoffs ................................ 23
    2.3.4 Upper Airway MRI of Speech ....................... 26
      2.3.4.1 3D MRI of Sustained Speech ................... 26
      2.3.4.2 Real-time MRI of Fluent Speech ................ 28

Chapter 3: EPI Artifact Correction for Real-time MRI 31
  3.1 Introduction ......................................... 31
  3.2 Methods ............................................. 33
  3.3 Results .............................................. 37
Chapter 4: Accelerated 3D Imaging Using Compressed Sensing

4.1 Single-coil Imaging

4.2 Multi-coil Imaging

Chapter 5: Real-time Speech MRI Using Golden-ratio Spiral

5.1 Introduction

5.2 Materials and Methods

5.3 Results

5.4 Discussion

5.5 Conclusions

Chapter 6: Parallel Imaging with Novel 16-Channel Coil at 3 Tesla

6.1 Introduction

6.2 Materials and Methods

6.3 Results

6.4 Discussion

6.5 Conclusions

Chapter 7: Summary and Future Work

Bibliography
List Of Tables

3.1 Ghost-to-signal ratios from the phantom study (see Fig. 3.2 for 1D phase correction and the proposed method. Mean and standard deviation of g-factor values for the proposed method are also reported. . . . . . . . . . 37

6.1 Average SNR improvement. Average SNR was measured in a single subject (33 year old male) using the 16-channel UA coil, single-channel birdcage coil, and 8-channel NV coil. Eight regions of interest (see Fig. 6.3a) were identified in 2D midsagittal images with $1.88 \times 1.88 \times 2.50$ mm$^3$ spatial resolution, obtained without the use of parallel imaging. The 16-channel UA coil provided improved intrinsic SNR in all regions of interest. UA: upper airway, NV: neurovascular. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 102

6.2 Comparison of average SNR drop-off in three volunteers. Average SNR was measured in three volunteers using the 16-channel coil. Eight regions of interest were identified in 2D midsagittal images with $1.88 \times 1.88 \times 2.50$ mm$^3$ spatial resolution, obtained without the use of parallel imaging. Note the relatively less steep SNR drop-off from the small female subject. . . . 103
List Of Figures

2.1 A basic slice selective excitation. ......................... 8
2.2 Pulse sequence diagram and image formation process. ............ 11
2.3 Schematic descriptions of $k$-space trajectories for (a) 2DFT, (b) projection reconstruction (PR), (c) echo-planar imaging (EPI), and (d) spiral. .... 14
2.4 Schematic descriptions of $k$-space trajectories for (a) 3DFT and (b) 3D stack of spirals. ........................................ 15
2.5 Illustration of Cartesian SENSE reconstruction process. ............ 17
2.6 Spiral imaging using compressed sensing on a GE resolution phantom. .. 20
2.7 Effect of spiral readout duration on the upper airway images. ........ 22
2.8 An example of motion artifacts in midsagittal vocal tract MRI image ... 23
2.9 Spiral pulse sequence diagram and $k$-space trajectory ................ 24
2.10 Example MR image of a midsagittal slice of human upper airway. ...... 27
2.11 Conventional 3D vocal tract imaging approaches. ................... 28
2.12 Midsagittal real-time speech MRI for nasal speech production study. ... 30
3.1 Reconstruction flowchart for the proposed ghosting correction method for (a) non-accelerated and (b) two-fold accelerated EPI with automatic ghosting correction. ........................................ 43
3.2 Real-time cylindrical phantom images reconstructed with non-accelerated EPI data. ........................................ 44
3.3 *In vivo* real-time cardiac images reconstructed with non-accelerated EPI data. ........................................ 45
3.4 Automatic correction during continuous scan plane rotation. ............ 46
3.5 Automatic ghosting correction using no acceleration and two-fold acceleration (b,c,g,h) and corresponding g-factor maps (d,e,i,j) reconstructed using two coil calibration schemes.

3.6 Dynamics of vocal tract shaping during natural speech utterances of “all year”.

4.1 Illustration of scan plane prescription, which is used for the 3D upper airway imaging.

4.2 k-space sampling patterns used in the experimental studies.

4.3 L-curve for the selection of regularization parameter $\lambda$ for CS reconstruction of the 3D upper airway data with reduction factors of 3, 4, and 5.

4.4 Representative magnitude and phase images from axial slices.

4.5 Axial slice reconstructions from retrospective sub-sampling of fully sampled data.

4.6 Reformatted 2D midsagittal and coronal images after the PC-II CS reconstructions of the 5x undersampled 3DFT data set.

4.7 3D visualization of the tongue and lower jaw after the PC-II CS reconstructions from the data set prospectively acquired with 5x acceleration.

4.8 Flowchart of the proposed reconstruction scheme.

4.9 Midsagittal images for (a) /s/, (b) /ʃ/, and (c) /ɾ/. Their corresponding 3D tongue shapes for (d) /s/, (e) /ʃ/, and (f) /ɾ/.

4.10 The prospective use of accelerated 3D acquisition and multi-coil PC-CS reconstruction. (a) Reformatted midsagittal slices and their associated midlines drawn for cross-sectional slice prescription. (b) Area function plot. (c) 3D visualization of the tongue and lower jaw.

5.1 Schematic diagram of real-time continuous MRI data acquisition (DAQ) using a golden-ratio spiral view order.

5.2 k-space trajectories for conventional bit-reversed 13-interleaf UDS and golden-ratio spiral view order when samples from (a,b) 8, (c,d) 13, and (e,f) 21 consecutive TRs are combined.

5.3 Retrospective selection of temporal resolution: (a) Comparison of unaliased FOV between the golden-ratio view order and conventional bit-reversed 13-interleaf UDS sampling. (b) The enlargement of the region within the green rectangle in (a).
5.4 Midsagittal images with a large reconstruction field-of-view (FOV) of 38 \( \times \) 38 cm\(^2\) reconstructed from the data acquired in static posture.  90

5.5 Gridding reconstructed dynamic frames and time intensity profiles from (a,c) bit-reversed 13-interleaf uniform density spiral data and (b,d) golden-ratio spiral view order data.  91

5.6 Blockwise temporal resolution selection and synthesis of a single video from four temporal resolution videos.  92

5.7 Variable temporal resolution selection from real-time data acquired using the golden-ratio acquisition scheme.  93

6.1 16-channel upper airway receive coil array.  108

6.2 The 16-channel upper airway coil array on a volunteer. (Left) Side view. (Right) Top view.  109

6.3 Illustration of the eight upper airway regions of interest (ROIs) used in the evaluation of SNR and \( g \)-factor.  110

6.4 (a) Channel indices labeled on the coil layout. (b) Noise correlation matrix from one representative volunteer.  110

6.5 Midsagittal, axial, and coronal images at each coil element of the 16-channel upper airway coil.  111

6.6 2D midsagittal image reconstruction using 1D SENSE: Comparison of the two subjects with different head size.  112

6.7 Plots of \( g \)-factor values for 8 different ROIs as a function of reduction factor for 2DFT midsagittal parallel imaging.  113

6.8 Comparison of midsagittal, axial and coronal slice reconstruction using parallel imaging. (Top): 8-channel neurovascular coil. (Bottom): 16-channel upper airway coil.  114

6.9 3D image reconstruction using 2D SENSE.  115

6.10 2D image reconstruction of the dynamics of vocal tract shaping using spiral SENSE.  116
Abstract

Magnetic resonance imaging (MRI) is a powerful non-invasive imaging modality, but is relatively slow compared to alternatives such as X-ray and ultrasound. Accelerating MRI scans has been of the great interest over the past several years and the acceleration of upper airway MRI, in particular, is a primary focus of this thesis. Rapid and real-time MRI can be used to capture tissue dynamics (e.g., the tongue/velum during speech, or the beating heart) or to reduce scan time. Rapid MRI can be achieved through the use of novel acquisition and reconstruction methods. Acquisition technologies such as echo-planar imaging and spiral imaging are effective at improving temporal resolution, but introduce additional image artifacts. Image reconstruction techniques such as parallel imaging and compressed sensing accelerate acquisition speed by highly undersampling Fourier data and improve image quality by removing spatial aliasing artifacts.

In this dissertation, I present four methods that accelerate MRI and manage the associated artifacts. First, a new interleaved echo-planar imaging technique is presented and applied to real-time interactive cardiac MRI. When combined with parallel imaging reconstruction, this enables automatic correction of ghosting artifacts and maintains temporal resolution. Second, a method for accelerated three-dimensional (3D) upper airway imaging is presented. The method adopts compressed sensing for acceleration and enables extraction of a whole 3D vocal tract during a single trial of sustained sound production.
Hence, it is free from image mis-registration problem and dramatically improves throughput in data acquisition. Third, a novel acquisition scheme based on a golden-ratio spiral temporal view order is presented and applied to real-time speech MRI. This provides more flexible retrospective selection of temporal resolution in vocal tract imaging than conventional spiral acquisition. I demonstrate its effectiveness at capturing rapid motion of articulators (e.g., tongue and velum) in nasal sound production. Finally, a novel 16-channel receive coil is described that is highly sensitive to the upper airway regions of interest (e.g., lips, tongue, soft palate). Its higher signal-to-noise ratio (SNR) and acceleration in parallel imaging is demonstrated by comparing it with widely available commercial coils.
Chapter 1

Introduction

Magnetic resonance imaging (MRI) is a powerful medical imaging modality that is non-invasive, free from ionizing radiation, flexible in controlling soft tissue contrast, capable of imaging any arbitrary oblique planes of interest, and effective at producing high spatial resolution. However, it is inherently slow compared to other imaging technologies such as ultrasound and X-ray. Numerous methods that improve imaging speeds in either acquisition or reconstruction perspectives have been developed in the MRI research community.

Echo-planar imaging (EPI) is an ultrafast imaging technique that is based on time-efficient sampling of the acquisition space, referred to as $k$-space. However, reconstructed images from EPI data may suffer from geometric distortions and ghosting artifacts due to a variety of sources including echo-misalignment, off-resonance, flow, and motion. Ghosting artifacts due to echo-misalignment is a systemic problem and is hard to correct for in oblique scan planes. Prior to the actual scans, calibration scans are often necessary for subsequent artifact correction processes.

In the first part of the thesis, I present a novel and fully automatic correction method that eliminates ghosting artifacts due to echo-misalignment in any arbitrary oblique scan plane. Phantom and real-time cardiac in vivo experiments indicate that the proposed
method is superior to the conventional one-dimensional (1D) phase correction method in terms of ghost suppression in oblique or double-oblique scan planes.

In speech research, MR imaging of the human upper airway during sustained speech production has been used as a means to provide a full geometric information of the shaping of the vocal tract and data for its modeling. The acquisition of a whole 3D vocal tract requires a long scan time, typically exceeding normal breath-hold duration. 3D vocal tract constructed from many repetitions of the same articulation is likely to suffer from the image misregistration possibly due to different positioning of the jaw, lips, and tongue at each trial of sustaining the speech sound. Under the constraints in spatial resolution and scan time appropriate for a single sustained sound production, acceleration can be only achievable by highly undersampling 3D $k$-space. Conventional inverse Fourier transform reconstruction of the undersampled data set will produce severe spatial aliasing artifacts in reconstructed images, which will make it infeasible to quantitatively assess the geometry of vocal tract shaping with good accuracy.

In the second part of the thesis, I propose an application of compressed sensing to accelerated 3D imaging of human vocal tract shaping. A variable density pseudo random undersampling in $(k_y, k_z)$ space is exploited to promote incoherence of spatial aliasing artifacts. Image reconstruction from undersampled $k$-space data is performed based on a regularized iterative reconstruction, in which the $l_1$ norm of finite difference in the image is adopted for denoising and edge enhancement. Acquisition and reconstruction from retrospective and prospective studies demonstrates the effectiveness of the technique at improving the air-tissue boundary depiction at high acceleration factors.

In speech research using real-time MRI, the analysis of vocal tract dynamics is performed retrospectively after acquiring data in real-time. Conventional real-time speech
MRI is typically based on a constant temporal resolution. However, a flexible retrospective selection of temporal resolution is desirable because of natural variations in speaking rate and variations in the speed of different articulators.

In the third part of the thesis, a novel acquisition scheme based on a golden-ratio spiral temporal view order is proposed and applied to real-time speech MRI. Compared to a conventional spiral acquisition, the proposed method demonstrates improved aliasing artifact reduction for static postures and improved temporal resolution for capturing the dynamics of rapid articulator motion.

Novel pulse sequences and image reconstruction techniques have improved the SNR and spatio-temporal resolution in imaging the upper airway. The design and use of a receive coil that is highly sensitive to the upper airway regions of interest may be an additional source for improving image quality.

In the final part of the thesis, I describe a novel 16-channel 3 Tesla upper airway coil and present its performance in the SNR and parallel imaging by comparing it with other widely available commercial coils. With this coil and conventional parallel imaging, I demonstrate a 6-fold acceleration in 3D imaging of the vocal tract during sustained speech. I also demonstrate a 4.2-fold acceleration in 2D real-time imaging of the vocal tract during fluent speech.

1.1 Outline

This dissertation is outlined as follows.

**Chapter 2: MRI Background**

This chapter presents an overview of basic MR pulse sequences, $k$-space sampling, 2D/3D imaging, parallel imaging, compressed sensing, and a brief introduction to imaging issues in rapid MRI of the upper airway.
Chapter 3: EPI Artifact Correction for Real-time MRI

This chapter presents an EPI acquisition and reconstruction technique that eliminates ghosting artifacts due to mis-alignment of the odd and even echoes using parallel imaging. The effectiveness of the technique is demonstrated in oblique and double-oblique scan planes in real-time cardiac and upper airway MRI.

Chapter 4: Accelerated 3D Imaging Using Compressed Sensing

This chapter presents an accelerated MRI technology that enables to capture a high-resolution 3D vocal tract shape in a single trial of sustaining sound production. A proposed phase constrained compressed sensing method accelerates MR acquisition by sub-sampling the Fourier coefficients, and results in improved depiction of the air-tissue interface over conventional image reconstruction methods. Its extension to multi-coil imaging is also presented.

Chapter 5: Real-time Speech MRI Using Golden-ratio Spiral

This chapter presents a new acquisition technique that is based on golden-ratio spiral temporal view order. The golden-ratio view order provides more flexible retrospective selection of temporal resolution than conventional scheme. The method is applied to real-time imaging of the vocal tract shape during nasal sound production.

Chapter 6: Parallel Imaging with Novel 16-Channel Coil at 3 Tesla

This chapter describes a novel 3 Tesla 16-channel receive coil which is highly sensitive to the upper airway regions of interest. SNR and parallel imaging $g$-factor of the coil is evaluated by comparing with widely available single-channel birdcage coil and 8-channel neurovascular coil, and over several human subjects with different head size and upper airway geometry. Its parallel imaging performance is demonstrated using high resolution
3D imaging of the vocal tract shape during sustained speech production. High spatiotemporal real-time imaging is also demonstrated using 2D spiral parallel imaging of the vocal tract dynamics during natural speech production.

Chapter 7: Summary and Future Work

This chapter summarizes the thesis and presents future research topics that are worth of more investigation.
Chapter 2

MRI Background

I first describe the basics of MRI physics and the principle of MR image formation. I provide an overview of two fundamental MRI acceleration techniques: parallel imaging and compressed sensing. Finally, I present several considerations in imaging the upper airway and then review on conventional methods of imaging the vocal tract during speech.

2.1 MRI Physics

2.1.1 Basic Pulse Sequences

I introduce the basics and underlying physics of acquiring a two-dimensional MRI image from a certain imaging slice of the human body.

2.1.1.1 Excitation

The magnet inside the MRI scanner room produces a strong and permanent magnetic field. The subjects are placed inside the magnet bore. In microscopic level, a majority of the spins in human tissue are influenced by the strong magnet and are aligned along
the direction of the main magnetic field, i.e., longitudinal direction. They precess at a
Larmor frequency $\omega_0$ (rad/sec), which is proportional to the magnetic field strength $B_0$.

$$\omega_0 = \gamma B_0,$$

(2.1)

where $\gamma$ is a gyromagnetic ratio, whose value depends on the type of nuclei (e.g., $^1H, ^{13}C,$
$^{31}P$). From now on, I will assume imaging hydrogen $^1H$, which is most abundant in human
body. For example, $\gamma/2\pi$ for $^1H$ is 42.58 MHz/T. In $^1H$ imaging at 3 Tesla magnetic field
strength, the Larmor frequency is 128 MHz. The net magnetization $M_0$ of spins is in the
same longitudinal direction as the main magnetic field in thermal equilibrium condition.

In the presence of strong magnetic field, another magnetic field, called a $B_1$ transmit
field, is involved in MR imaging process. The $B_1$ field has its carrier frequency tuned
to the Larmor frequency, which is in the radio frequency (RF) range. The direction
of the $B_1$ field is perpendicular to the main magnetic field. The $B_1$ field is typically
applied only for a short duration (e.g., $1 \sim 3$ msec). The net effect results in excitation
of the spins, i.e., tipping the spin magnetization $M_0$ onto its transverse plane (i.e., the
plane perpendicular to the main magnetic field) by a certain angle. This phenomenon
can be described by the well-known Bloch equation [82]. The transversal component
of the excited spin magnetization is time varying in nature, and induces voltage in an
RF receive coil by Faraday’s induction law. After the pulse of the $B_1$ transmit field, the
transverse magnetization exponentially decays at certain rates characterized by $T_1$ and $T_2$
time constants and goes back to the thermal equilibrium state. This is called relaxation.
$T_1$ and $T_2$ represent longitudinal and transversal relaxation rates, respectively.

In one-dimensional (1D) slice selective excitation, a SINC-shaped $B_1$ transmit pulse is
typically generated along with a slice selective gradient $G_z$ (see Fig. 2.1). The combination
of the $B_1$ and $G_z$ results in a 1D slice profile, in which the spins within the passband of
Figure 2.1: A basic slice selective excitation. The duration and shape of the envelope of the $B_1$ waveform determines the bandwidth and transition width of the transverse magnetization. This is governed by the Fourier transform relationship between the envelope of the $B_1$ waveform and the spectral response of the transverse magnetization when the small tip approximation is assumed to hold [86]. Slice selection along the z direction is attained after applying the $G_z$ gradient waveform.

the slice profile have the transverse magnetization component and produce signals in the receive coil while the spins outside the slice profile have little or no transverse component and has no effect on the receive coil.

2.1.1.2 Readout

After the 1D slice selective excitation, time-varying gradient fields $G_x$ and $G_y$ are generated to spatially resolve the spin magnetization in 2D. The $G_x$ and $G_y$ gradient fields induce additional resonance offsets that linearly vary along the spatial $x$ and $y$ axes, respectively. The signal detected in the RF receive coil at an instantaneous time $t$ is given as a sum of the transverse components of all the spins, where each individual spin experiences their own instantaneous phase offsets depending on their spatial position.
The instantaneous phase offset $\triangle \phi$ can be described by:

$$\triangle \phi(t; x, y) = \int_0^t \triangle \omega(\tau) d\tau = \int_0^t \gamma \triangle B(\tau) d\tau = \gamma \int_0^t G_x(\tau) d\tau x + \gamma \int_0^t G_y(\tau) d\tau y.$$  \hspace{1cm} (2.2)

When deriving the third line of Equation 2.2, it is assumed that the spins at a position $(x, y)$ are stationary in time.

Here, I introduce the $k$-space representation [110]:

$$k_x(t) \triangleq \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau.$$ \hspace{1cm} (2.3)

$$k_y(t) \triangleq \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau.$$ \hspace{1cm} (2.4)

When ignoring effects such as spin relaxation and spatial resonance offset, the MR signal equation can be simply expressed as the following Fourier transform relationship:

$$s(t) = \int_x \int_y m(x, y) e^{-i \triangle \phi(t; x, y)} dx dy = \int_x \int_y m(x, y) e^{-i 2\pi (k_x(t)x + k_y(t)y)} dx dy = M(k_x(t), k_y(t)),$$ \hspace{1cm} (2.5)

where $s(t)$ is a signal from the receive coil, $M(k_x(t), k_y(t))$ is a function of 2D spatial frequencies $(k_x(t), k_y(t))$, both of which are parameterized by a time variable $t$. $m$ is the transverse component of the magnetized spin density, $(k_x, k_y)$ is the $k$-space location, and $(x, y)$ is the spatial position. Note that $k_x(t)$ and $k_y(t)$ are given as the integral of the gradient waveforms $G_x(t)$ and $G_y(t)$ respectively and thus are continuous with respect
to $t$. Figure 2.2(a-b) illustrates how the generation of $G_x(t)$ and $G_y(t)$ is related to the trajectory in $k$-space. The inverse Fourier transform of the samples $M(k_x(t), k_y(t))$ on the trajectories in $k$-space results in the MRI image (see Fig. 2.2(c)).

2.1.2 $k$-space Trajectories

2.1.2.1 Practical Constraints

The gradient amplifiers in MRI scanner are the sources that generate the $G_x$, $G_y$, and $G_z$ waveforms. They provide currents to the gradient coils, and their operations can be described by the L-R circuit model [59, 37]. The current in the gradient coils is proportional to the gradient amplitude. The voltage in the gradient coils is related to the gradient switching rate (i.e., gradient slew rate). For speed-up in the MR acquisition, gradient waveforms are designed to take full advantages of maximum limits of gradient amplitude and slew rate. Note that the use of maximum available slew-rate and amplitude in gradient waveform design often causes peripheral nerve stimulation and tissue heating in the subjects. Also, this is vulnerable to scanner heating, in particular for real-time MRI that involves several hours-long scans such as dynamic speech imaging experiments.

2.1.2.2 2D Imaging

Here, I briefly introduce four fundamental $k$-space sampling schemes (i.e., 2DFT, projection reconstruction (PR), echo-planar, and spiral), each of which have their own characteristics and exhibit advantages/disadvantages depending on the applications. Some representative $k$-space trajectories are illustrated in Fig. 2.3.
Figure 2.2: Pulse sequence diagram and image formation process. (a) Basic 2DFT fast gradient echo (GRE) pulse sequence. DAQ represents data acquisition period during which the signal is recorded by the receive coil. Note that the blip size in the $G_y$ gradient (i.e., phase-encode gradient) changes by a certain increment in the area at every TR (see the green arrows), enabling to acquire subsequent $k_y$ levels in the $k$-space shown in (b). (b) Data samples acquired during each DAQ period are mapped onto the $k$-space trajectory. (c) The inverse Fourier transform of the acquired $k$-space data yields the final image.
2DFT Imaging: Two-dimensional Fourier transform (2DFT) imaging acquires a single phase encode line after each excitation. It is popular and most widely used imaging method in clinical MRI studies. It is very robust to any system imperfections such as gradient/DAQ delay and magnetic field inhomogeneity. Since the acquired data are evenly distributed in $k$ space, image reconstruction procedure directly involves a simple and fast 2D inverse Fourier transform. However, 2DFT data acquisition is extremely slow and is very sensitive to motion artifacts.

PR Imaging: Projection reconstruction (PR) imaging covers $k$-space by acquiring radial lines at different azimuthal angles. Each radial line is designed to pass through the $k$-space origin. Since PR imaging acquires the $k$-space origin at every TR, it is relatively less sensitive to motion than 2DFT. The $k$-space samples on PR trajectories are distributed non-uniformly. The low spatial frequency region in $k$-space is sampled densely whereas the high spatial frequency region is sampled relatively sparsely. Spatial aliasing artifacts result from insufficient sampling of high spatial frequency. The “high-frequency” aliasing artifacts appear incoherent and are less visually disturbing than fold-over aliasing artifacts resulting from the regular undersampling in 2DFT imaging.

Echo-planar Imaging: Echo-Planar Imaging (EPI) samples $k$-space in a raster-like fashion, and its trajectory is somewhat similar to 2DFT sampling pattern. EPI is fast but is prone to artifacts including ghosting and geometric distortion in reconstructed images due to a variety of sources such as off-resonance, $T_2$ or $T_2^*$ relaxation, eddy currents, gradient/DAQ delay, etc. Geometric distortion is proportional to the amount of resonance offset and the bandwidth along the phase encode. Multi-shot (or interleaved) short echo-train-length EPI is often used for practical purposes. It needs a careful selection in the
number of shots (or interleaves) and readout duration. The proper choice compromises between 1) the improved image quality by reduction in image artifacts originating from either off-resonance or $T_2^*$ decay and 2) the improved imaging speed by the use of a fewer number of shots (or interleaves).

**Spiral Imaging:** Spiral imaging typically acquires data from the $k$-space origin, and its data acquisition ends at the $k$-space periphery after traversing the predetermined spiral trajectory. Spiral imaging is known to be the most time-efficient at sampling $k$-space. Moreover, it is insensitive to motion and flow effects. Spiral imaging is suitable for imaging rapidly moving objects (e.g., cardiac ventricular motion or flow imaging, vocal tract imaging during speech/swallowing). Imaging field-of-view (FOV) in spiral trajectories is inversely proportional to a distance along the radial line crossing two adjacent spiral lines. The design of spiral trajectory can be flexible and be parameterized as a function of the $k$-space radius (e.g., see Brian Hargreaves’ website http://www-mrsrl.stanford.edu/~brian/vdspiral). However, spiral imaging is prone to spatial blurring or geometric distortion in reconstructed images due to a variety of sources including off-resonance, concomitant gradient field, and gradient/DAQ timing delays.

### 2.1.2.3 3D Imaging

Three-dimensional (3D) imaging performs spatial encoding in 3D $k$-space (i.e., $k_x, k_y, k_z$). 3DFT (see Figure 2.4(a)) imaging is widely used and its pulse sequence can be described as follows: slab excitation pulses with a thick slice (e.g., 8-10 cm) are first applied and followed by a $G_z$ gradient encode blip, which performs $k_z$ encoding. $G_x$ and $G_y$ gradients are generated to perform $k_x$ and $k_y$ encoding, respectively. 3D imaging provides a full volumetric coverage of regions of interest, but it requires a prohibitively long scan time.
Figure 2.3: Schematic descriptions of $k$-space trajectories for (a) 2DFT, (b) projection reconstruction (PR), (c) echo-planar imaging (EPI), and (d) spiral.

However, 3D imaging has recently gained attention because its high dimensionality allows for substantial acceleration by highly undersampling $k$-space. Advanced reconstruction methods such as parallel imaging [90, 33, 113] and compressed sensing [66] or the combination of the two [73, 58] can be utilized to reconstruct an alias-free 3D volume from significantly undersampled 3D data set. Other non-Cartesian 3D trajectories such as 3D PR [8], 3D cones [34], and 3D stack of spirals [109, 62](see Figure 2.4(b)) have been developed to improve imaging speed.
2.2 Accelerated Imaging

2.2.1 Parallel Imaging

The use of multiple-channel phased array receive coil is widespread in MRI because it provides improved signal-to-noise ratio (SNR) over either single-channel coil or body coil. All coil elements in the array simultaneously detect the MR signals. The MR signal at one element differs from those at the others because each coil element has its own spatial coil sensitivity characterized by its direction and distance relative to tissue region of interest. Parallel imaging exploits this additional spatial sensitivity information for the purpose of accelerating MRI scans.

Numerous parallel imaging reconstruction techniques have been proposed in the literature. Sensitivity encoding (SENSE) [90] and generalized autocalibrating partially parallel acquisition (GRAPPA) [33] are two fundamental parallel imaging methods. SENSE performs parallel imaging reconstruction in the image domain while GRAPPA performs...
parallel imaging reconstruction in the $k$-space domain. In this subsection, I only introduce SENSE approaches in Cartesian and non-Cartesian trajectories.

After the incorporation of the spatial coil sensitivity, the MR signal equation is formulated as follows:

$$s_j(t) = \int \overrightarrow{r} c_j(\overrightarrow{r}) m(\overrightarrow{r}) e^{-i2\pi \overrightarrow{k}(t) \cdot \overrightarrow{r}} d\overrightarrow{r},$$  \hspace{1cm} (2.6)

where $s_j$ is the signal received from the $j^{th}$ coil element, $m$ is the transverse component of the spin magnetization, $c_j$ is the coil sensitivity from the $j^{th}$ coil element, $\overrightarrow{r}$ is the spatial position, and $\overrightarrow{k}(t)$ is the $k$-space location at time $t$. It is noted that coil sensitivity effect is modeled as a multiplicative factor in the MR signal equation.

In Cartesian $k$-space sampling, such as 2DFT and EPI, two-fold acceleration can be achieved by skipping every other phase encode lines in the acquisition. This introduces $\text{FOV}/2$ aliasing along the phase encode direction in reconstructed images. Two signals, which are originally $\text{FOV}/2$ apart along the phase encode direction, are superimposed onto a single pixel and this results in fold-over aliasing. For a given pixel location $(x, y)$, SENSE formulation can be described by the following linear system when four coil elements are considered and the acceleration rate is 2.

$$\begin{pmatrix}
    a_1(x, y) \\
    a_2(x, y) \\
    a_3(x, y) \\
    a_4(x, y)
\end{pmatrix} =
\begin{pmatrix}
    c_1(x, y) & c_1(x, y + \text{FOV}/2) \\
    c_2(x, y) & c_2(x, y + \text{FOV}/2) \\
    c_3(x, y) & c_3(x, y + \text{FOV}/2) \\
    c_4(x, y) & c_4(x, y + \text{FOV}/2)
\end{pmatrix} \begin{pmatrix}
    m(x, y) \\
    m(x, y + \text{FOV}/2)
\end{pmatrix}$$  \hspace{1cm} (2.7)

Here, $a_j$ is the aliased signal in the image domain for the $j^{th}$ coil, $c_j$ is the coil sensitivity for the $j^{th}$ coil, and $m$ is the unknown signal to be estimated.

Un-aliasing is performed by solving the least-squares problem described by Equation 2.7 at every pixel $(x, y)$. Figure 2.5 illustrates the Cartesian SENSE reconstruction.
process. The matrix inversion involved in solving the least-squares problem results in noise amplification in the final image. The degree of noise amplification is related to the condition number of its associated coil sensitivity matrix. Geometry factor (also called $g$-factor) is an important indicator of noise amplification [90] that results from the Cartesian SENSE reconstruction (see Fig. 2.5). $g$-factor values spatially vary and depend on the acceleration factor, coil geometry, and phase encode directions.
In non-Cartesian $k$-space sampling, such as spiral and radial trajectories, two-fold acceleration can be achieved by skipping every other interleaf/spoke. Unlike the Cartesian case, many pixels can contribute aliasing to a single pixel location. Therefore, when four coil elements are assumed to be considered, SENSE reconstruction is performed by inverting the following large scale over-determined linear MR system formulation.

\[
\begin{pmatrix}
    s_1(k) \\
    s_2(k) \\
    s_3(k) \\
    s_4(k)
\end{pmatrix} =
\begin{pmatrix}
    c_1(r_1)\phi(k, r_1) & c_1(r_2)\phi(k, r_2) & \cdots & c_1(r_N)\phi(k, r_N) \\
    c_2(r_1)\phi(k, r_1) & c_2(r_2)\phi(k, r_2) & \cdots & c_2(r_N)\phi(k, r_N) \\
    c_3(r_1)\phi(k, r_1) & c_3(r_2)\phi(k, r_2) & \cdots & c_3(r_N)\phi(k, r_N) \\
    c_4(r_1)\phi(k, r_1) & c_4(r_2)\phi(k, r_2) & \cdots & c_4(r_N)\phi(k, r_N)
\end{pmatrix}
\begin{pmatrix}
    m(r_1) \\
    m(r_2) \\
    \vdots \\
    m(r_N)
\end{pmatrix},
\tag{2.8}
\]

where $s_j(k)$ is the $M \times 1$ measured $k$-space data from the $j^{th}$ coil element, $\phi(k, r_p)$ is the $M \times 1$ vector $[e^{-i2\pi k_1 \cdot r_p}, ..., e^{-i2\pi k_M \cdot r_p}]^T$, and $m(r_p)$ is the unknown pixel value at the $p^{th}$ pixel location in the image domain. $M$ is the total number of $k$-space data samples per each coil, and $N$ is the total number of image pixels. Note that $M < N$ because of the undersampling for acceleration, but the fact that $4 \cdot M \geq N$ guarantees a unique solution in the image estimate $[m(r_1), m(r_2), ..., m(r_N)]^T$ in the least-squares sense.

### 2.2.2 Compressed Sensing

Compressed sensing [25, 20] (CS) has recently emerged as a promising theoretical framework in signal processing perspective. Compressed sensing states that a signal can be exactly recovered from incomplete (i.e., sub-Nyquist sampling rate) random measurement data with a very high probability via a minimum $l_p$ norm ($0 \leq p \leq 1$) reconstruction
provided that the signal is sparse (or compressible) in certain transform domain (e.g.,
wavelets, finite differences, curvelets). Compressed sensing is characterized by the follow-
ing three main components:

1. Sparsity of the solution in its sparsifying transform domain
2. Incoherence between the system matrix and the sparsifying basis
3. Minimum $l_p$ norm ($0 \leq p \leq 1$) reconstruction which promotes a sparse solution

Since Lustig et al. [66, 67] elaborated and proposed the applicability of compressed
sensing in MRI, many researchers have investigated its effectiveness and extended its use
to a variety of MR applications. Compressed sensing MRI (CS-MRI) can be used to
accelerate scan time or increase spatial resolution. CS-MRI can be combined with other
existing acceleration methods such as parallel imaging to promote further accelerations.
It is noted that CS-MRI has recently gained great attention among radiologists and
clinicians [112].

Transform sparsity can be exploited in wavelet coefficients of the images (e.g., $T_1$
or $T_2$ weighted brain imaging), finite differences of the images (e.g., contrast enhanced
angiography), or periodicity in the dynamic time series (e.g., gated cardiac cine). Pure
random undersampling is difficult to achieve in MRI because the data are sampled along a
smooth trajectory in $k$-space. However, random undersampling can be easily achieved in
special cases for example on $(k_y, k_z)$ encodes in 3DFT imaging [69], $(k_y, t)$ sampling [68,
31], and $(k_f, k_x)$ sampling for MR spectroscopic imaging [43]. Variable density pseudo-
random undersampling (i.e., low spatial frequencies are critically sampled and high spatial
frequencies are sparsely sampled) produces noise-like incoherent aliasing and are typically
preferred over uniformly random undersampling [66]. This is attributed to the fact that
the most energy in the $k$-space data is in the central part of the $k$-space. Figure 2.6 shows
an example of the use of compressed sensing. Variable density undersampling shown in Fig. 2.6(f) produces incoherent aliasing artifacts in gridding reconstructed image in Fig. 2.6(b), and aliasing is further suppressed using compressed sensing reconstruction as seen in Fig. 2.6(d). Sampling density can be optimized based on aliasing incoherence as a criterion and via the minimization of maximum sidelobe level in the sampling point spread function [56].

Compressed sensing reconstruction can be improved by taking into account a variety of system imperfections in MR acquisition and incorporating these into the MR signal equation. For example, factors that affect acquired MR signals may include: 1) spatial off-resonance, 2) relaxation, 3) gradient/DAQ delay, 4) eddy currents, and 5) coil sensitivity. Questions on how to determine appropriate sparsifying basis, find optimal
sampling schemes, and improve sparse reconstruction algorithms, still remain to be an-
swered. These will be answered in different context depending on the type of pulse
sequence, imaging parameters, the use of contrast agent, and the anatomy of interest.

2.3 Upper Airway Imaging

MRI is a powerful tool for the non-invasive assessment of upper airway anatomy including
the shaping of the tongue, lips, soft palate, and pharyngeal wall. Upper airway MRI has
been used to facilitate clinical assessments in patients with sleep apnea [98, 104], speech
disorders [9], swallowing disorders [4], and for surgical planning [65]. It has also been
used for research purposes within the speech research community [75, 79].

2.3.1 Air-tissue Magnetic Susceptibility

In upper airway imaging of speech and sleep apnea studies, air-tissue boundaries are
the regions of interest and are subject to substantial resonance offset due to a large
magnetic susceptibility between the air and tissue. Spatial off-resonance patterns are
related to the geometry of tongue, jaw, and velum positioning, and they are not spatially
smooth. These large resonance offsets result in severe artifacts in reconstructed images
when using a long readout duration in either EPI or spiral imaging. These artifacts can
be reduced by 1) measuring a field map (i.e., off-resonance map) and 2) reconstructing
the image after incorporating the field map into the reconstruction process. Field map
acquisition typically involves imaging with two different echo times and subtraction of
the phases of these two different echo images. Accurate estimation of the field map is
challenging in real-time MRI of the upper airway. This is due to the fact that the two
images from different echo-time may not be exactly the same due to motion. Figure 2.7
illustrates spatial blurring/distortion effects due to off-resonance from air-tissue magnetic susceptibility when using spiral imaging with long readout duration.

2.3.2 Motion of Articulators

In MR imaging of the upper airway for speech production or swallowing studies, resulting images are prone to artifacts due to motion. Conventional MRI acquisition is relatively much slower than the speed of articulators (e.g., jaw, tongue, lips, and velum) during fluent speech production. Several speech research groups have proposed gated imaging techniques which synchronize the data acquisitions from repetitions of the utterance as a means to improve temporal resolution [44]. The gated imaging schemes have limitations. The successful use of gated imaging relies on how consistently the speakers can keep the same positions of articulators and the same speech rate for each repetition trial. For practical purposes, speech tasks are often limited to many repetitions of a few words.

Non-gated real-time MRI may be more suitable for imaging the vocal tract dynamics during natural sound production [79]. To this end, real-time MRI of speech requires ultra-fast acquisition speed. Spiral imaging is time-efficient in sampling $k$-space, and is insensitive to motion. However, current real-time spiral MRI used for speech research still
Figure 2.8: An example of motion artifacts in midsagittal vocal tract MRI image. Shown are representative frames reconstructed from spiral data acquired in real-time when the subject pronounced /ipa/: (a) /i/, (b) transition from /i/ to /p/, and (c) /a/. Temporal resolution was 84 msec. Motion artifacts are prominent in (b) because the temporal resolution used for the imaging is not sufficient in capturing the rapidly moving articulators (e.g., tongue body and lips).

lacks in temporal resolution. Figure 2.8 shows that the temporal resolution of 84 msec is sufficient in capturing the air-tissue boundaries for the monophthongal vowel sounds such as /i/ and /a/, but it is not sufficient in capturing the shaping when there is a rapid transition from /i/ to /p/. Note the temporal blurring and swirling artifacts in Fig. 2.8(b) due to data inconsistency resulting from articulators’ motion.

2.3.3 Imaging Tradeoffs

Prior to imaging the upper airway, one should carefully determine pulse sequence design parameters: flip angle, field-of-view (FOV), spatial resolution, temporal resolution, the number of spiral interleaves, readout duration, slice thickness, etc. I focus on real-time fast gradient-echo spiral MRI, which is routinely being used for our speech production research at the University of Southern California.

Flip angle can be chosen based on the Ernst angle, which is the optimal flip angle that maximizes transverse magnetization in steady state given TR and T₁ values of the tissue.
Figure 2.9: Spiral pulse sequence diagram and $k$-space trajectory. (a) Pulse sequence diagram for spiral fast gradient echo sequence. Shown are 1D slice selective excitation (green box), spiral readout (red box), spiral rewinder (blue box), and gradient spoiler (magenta box). (b) Spiral $k$-space trajectory. This illustrates that spatial resolution is related to the spiral $k$-space radius, and imaging field-of-view (FOV) is inversely proportional to the distance between adjacent spiral lines.

FOV can be properly selected depending on the $k$-space trajectory. In Cartesian sampling, the frequency encode direction is typically chosen to be along the superior-inferior (S-I) direction. This suppresses signal from the brain and neck via analog low-pass filtering. In spiral imaging, FOV is rotationally symmetric (see Fig. 2.9(b)), and a smaller FOV imaging gives a faster acquisition speed. Reduced FOV imaging can be performed effectively when the receive coil sensitivity is localized to the upper airway regions of interest.

A proper selection of spatial and temporal resolution is important in real-time MRI. For example, in spiral trajectory design, when the readout duration and FOV is fixed, a selection of higher spatial resolution requires more spiral interleaves to cover the $k$-space and results in lower temporal resolution.
Readout duration (denoted by $T_{\text{read}}$ in Fig. 2.9(a)) offers a trade-off between the degree of blurring artifacts and the acquisition speed for spiral imaging. The longer is the readout duration, the less number of RF excitations is needed per image, implying faster imaging speed by reducing the overhead on the time spent for the slice selective excitation and spoiler gradients (see Fig. 2.9(a)). However, the longer readout duration incurs a larger phase accrual in the presence of off-resonance. This leads to more pronounced artifacts in reconstructed images.

Slice thickness needs a careful selection. The choice of thicker slice leads to higher SNR and also can lead to shorter duration of slice excitation pulse (i.e., gain in imaging speed), but it leads to poorer spatial resolution across the slice.

Signal-to-noise ratio (SNR) serves as one of the most important criteria in the assessment of MRI image quality. SNR is defined as the ratio of the signal intensity to standard deviation of the noise [82]. SNR is proportional to voxel size and the square root of the scan time. In addition, SNR is dependent on $T_1$ values of soft tissue of interest and pulse sequence parameters such as flip angle and repetition time ($TR$). In RF-spoiled gradient echo sequence with $TR \ll T_1$, the signal at the steady state has the following relationship:

$$s(TR, T_1, \theta_E) \propto M_0 \left(1 - e^{-TR/T_1}\right) e^{-TE/T_2^*} \sqrt{1 - e^{-2TR/T_1}},$$

where $s$ is the signal amplitude, $M_0$ is the magnetization at thermal equilibrium and is proportional to magnetic field strength $B_0$, $\theta_E$ is the Ernst flip angle that produces maximum transverse signal in steady state. $TE$ is the echo time (i.e., time interval between the center of RF excitation and the instant when the echo of the signal is formed), and $T_2^*$ is the effective transverse relaxation rate.
2.3.4 Upper Airway MRI of Speech

In speech research, understanding the shaping of the vocal tract and its temporal variation during speech has been of great interest in research communities from linguistics, speech signal processing, speech pathology, otolaryngology, etc. Imaging technologies such as X-ray, computed tomography (CT), ultrasound, and electromagnetic articulometer (EMA) have been adopted to examine the shaping of the vocal tract (or the tongue) [11]. However, each have limitations. For example, X-ray and CT involve exposure to ionizing radiation, which is harmful to the subjects. In ultrasound imaging, the probe contacts the jaw and obstructs natural speech production. In EMA, the sensors are difficult to attach in deep regions such as the pharyngeal wall and velum.

Magnetic resonance imaging (MRI) is a non-invasive imaging modality that involves no ionizing radiation and enables to provide excellent visualization of the soft tissue in 3D. A primary drawback is that MR image acquisition is notoriously slow. As seen in Figure 2.10, the midsagittal MR image shows clear depiction of articulators such as lips, tongue, jaw, and velum in open-mouthed position. However, the scan time (i.e., 31 seconds) is not adequate in capturing the dynamics of articulators during speech. Hence, technological development for improved spatiotemporal resolution is essential and should be targeted in the vocal tract regions of interest.

2.3.4.1 3D MRI of Sustained Speech

MRI has been adopted in speech research community as a means to extract full threedimensional (3D) anatomical information of the vocal tract shape during sustained speech production [103, 75, 6, 88]. This provides insights into the knowledge of a whole vocal tract shape and data for its modeling. High spatial resolution (i.e., 1 ~ 2 mm) 3D volume is acquired by prescribing multiple slices which are orthogonal to the midsagittal
slice (see Fig. 2.11(a)). Scan time in 3D imaging often exceeds normal breath-hold limit. The acquisition typically needs multiple repetitions of the same sound. This may result in image mis-registration problem, and thus accelerating MRI acquisition is desirable. In addition, 2D multi-slice imaging with a certain slice gap produces noncontiguous 3D shape of the soft-tissue as shown in Fig. 2.11(b). True 3D encoding in MR acquisition can eliminate this problem, but it may necessitate a longer scan time than 2D multi-slice imaging.
Figure 2.11: Conventional 3D vocal tract imaging approaches. (a) Slice prescription for 3D MRI of the vocal tract (adapted from O. Engwall et al. [26]. (b) 3D tongue shape extracted from data acquired in the oblique coronal slices (adapted from C. Shadle et al. [97]).

2.3.4.2 Real-time MRI of Fluent Speech

Real-time MRI continuously acquires data and reconstructs and displays image frames in real-time. It has been developed primarily for cardiovascular MRI [50, 80, 81]. Several research groups have applied real-time MRI to vocal tract imaging during speech [24, 79]. Speech Production and Articulation kNowledge group (SPAN, http://sail.usc.edu/span) at the University of Southern California (USC) is an interdisciplinary research group in which various research groups are in collaboration from linguistics, otolaryngology, computer science, and electrical and biomedical engineering. Currently, real-time speech MRI data collection is routinely conducted at the Los Angeles County hospital in the USC Health Science Campus. I have been working as an MRI operator in the SPAN group. MRI operator contacts the MRI chief technicians or administrators at the scanner sites for a schedule availability for our data collection. The operator performs the MRI safety screening on the subjects. After the consent and safety procedures are complete, the operator places the subject inside the magnet and runs a custom real-time imaging
software (RTHawk), which was originally developed by Juan M. Santos [96], for real-time MRI data collection. RTHawk graphic user interface provides a convenient way to operate vocal tract MRI scans: 1) The operator can be aware of when to stop the scan by monitoring the vocal tract movement in real-time, and 2) the operator can easily adjust scan parameters such as scan plane orientation, FOV, center frequency, linear shim values, DAQ/gradient delays, selection of coil elements, flip angle, slice thickness, etc.

In real-time speech MRI experiment, speech signal is recorded simultaneously with real-time MRI data acquisition that is based on fast spiral gradient echo sequence. Conventional gridding reconstruction [46] is used to generate the video of vocal tract dynamics. Noise cancellation based on adaptive signal processing [13] removes loud MRI gradient noise and recovers the speech signal. Then, the video reconstructed from MRI data is merged with the speech audio to produce synchronized audio/video of the vocal tract dynamics, typically midsagittal view. Real-time speech MRI experiments have been conducted in a variety of research and clinical settings. Figure 2.12 shows one example of real-time MRI nasal studies that investigate timing effects in the coordination of articulators under different syllable contexts [19].
Figure 2.12: Midsagittal real-time speech MRI for nasal speech production study. Syllable difference influences different timing of articulators. Articulation of /bono/ in different syllable contexts: (a) /n/ is initial in the second word. (b) /n/ is final in the first word. In (a), the velum starts lowering when the tongue tip touches the hard palate (see frame 6). In (b), the velum starts lowering (frame 4) before the tongue tip touches the hard palate (see frame 7). Temporal resolution was 78 msec. Frames are shown at every 42 msec.
Chapter 3

EPI Artifact Correction for Real-time MRI

This chapter introduces an echo-planar imaging (EPI) acquisition and reconstruction method which is effective at removing ghosting artifacts in real-time interactive cardiac MRI. In addition, I briefly present the application of the proposed acquisition scheme to real-time upper airway imaging of speech production.

3.1 Introduction

Echo-planar imaging (EPI) [71] is used in cardiac MRI because it accelerates image acquisition, while maintaining image quality comparable to 2DFT. Cardiac EPI imaging often introduces artifacts in reconstructed images which may include geometric distortion and ghosting due to a variety of sources including off-resonance, in-plane flow, cardiac motion, and echo-misalignment. Geometric distortions due to off-resonance can be mitigated by reducing the echo spacing in the readout gradient or acquiring the data using multiple RF excitations (i.e., shots), which increases the effective sampling rate (i.e., increases the acquisition bandwidth) along the phase-encode direction. Ghosting due to cardiac motion can be mitigated by making the acquisition faster or restricting data acquisition to a relatively stationary cardiac phase.
Ghosting artifacts caused by echo-misalignment are a systemic problem, and are a function of induced eddy currents and system timing errors, which are associated with the scanner hardware (e.g., eddy currents from the cryostat and relative delays of the physical x, y, and z gradients). These issues are further complicated in oblique scan planes, which are routinely used in real-time imaging (e.g., cardiac short-axis and long-axis views).

The conventional method for correcting echo-misalignment involves performing a calibration scan to determine the on-axis gradient/DAQ time delays, and using small gradient “blips” to align echoes. These blips are scan-plane dependent [92, 116], and should be redesigned upon each scan-plane change. Flyback readouts [29], which involve acquiring data only on the positive polarity of the readout gradient, can be used to avoid echo-misalignment artifacts without a calibration scan, but have the disadvantage of reduced scan time efficiency. Image-based correction [16], which takes into account 1D phase errors from reduced field of view (FOV) images reconstructed separately from odd and even echoes, is effective at reducing ghosting artifacts in on-axis scan planes, but its effectiveness may be degraded in oblique (and double-oblique) scan planes. Full 2D correction is possible using fully phase-encoded reference scans for each scan plane [21], but it may be impractical to acquire such reference scans in dynamic and/or real-time imaging applications. Furthermore, use of such 2D phase-error maps may be problematic when imaging rapidly moving structures such as the heart, in which the estimated phase-error between odd and even echoes will be biased by phase-accrual due to flow and motion.

Another alternative is to separate data from left-to-right and right-to-left traversals in k-space (which each have half the desired FOV), and to reconstruct ghost-free full-FOV images using parallel imaging [61, 42, 49, 32]. This approach is attractive for real-time imaging because it does not require additional calibration scans or any modification of
the pulse sequences at the time of scan plane change. The phased-array ghost elimination (PAGE) method [49] has provided a generalized framework for cancelling ghosting artifacts due to sources including off-resonance and echo-misalignment using the information of local coil sensitivity profiles from multiple coils. Herzka et al. [42] demonstrated gated cardiac imaging with a sequential non-interleaved EPI acquisition scheme, where the echo-train-length (ETL) was equal to the SENSE reduction factor (e.g., 2 or 3).

In this work, I propose an “interleaved” gradient-echo EPI acquisition scheme and associated reconstruction method. Compared to the original PAGE method, this acquisition uses a large ETL ranging from 15 to 50 and a small number of shots in order to achieve sufficient temporal resolution in free-breathing real-time cardiac EPI imaging. Using shot-to-shot interleaving of the phase-encode lines and a “double-alternating” $k$-space data acquisition scheme, the proposed method achieves ghost suppression using a SENSE reduction factor of 2, regardless of the ETL. The proposed technique is compared with the conventional 1D phase correction method [16] in oblique and double-oblique scan planes. Two-fold accelerated EPI imaging is also demonstrated in conjunction with the proposed automatic ghosting correction technique. Finally, the feasibility of real-time reconstruction is demonstrated using a custom real-time imaging platform [96].

3.2 Methods

When performing echo-planar imaging at oblique or double-oblique scan planes, two or all three physical gradients will oscillate during each readout. An important consideration is that physical gradients in $x$, $y$, and $z$ may have unequal delays [92]. For an oblique scan plane with unequal gradient delays, the $k$-space lines with different traversal directions in the logical coordinate frame will sample positions in $k$-space that are shifted in opposite directions along the physical coordinate axes. Combined data are not uniformly spaced,
which causes artifacts in reconstructed images. Within the set of lines having the same traversal direction, uniform spacing is maintained.

The proposed acquisition scheme involves acquiring $k$-space data with alternating polarity of the readout gradient. Figure 3.1(a) illustrates the proposed reconstruction method. Coil sensitivity maps are reconstructed separately from the left-to-right (L-R) and right-to-left (R-L) lines, which are acquired from the two most recent time frames, $n$ and $n-1$ [47]. The L-R and R-L lines each preserve uniform spacing between phase encode lines and prevent artifacts due to echo-misalignment [42]. In time frame $n$, SENSE [90] reconstruction with a reduction factor of 2 is applied separately to the L-R and R-L lines. The resulting two full-FOV images differ not in magnitude but in phase, and taking a root sum-of-squares combination eliminates distortions in phase and recovers signal-to-noise ratio (SNR). This non-accelerated double-alternating scheme can be applied to any number of interleaves as long as L-R and R-L lines are alternating along the phase encode. An even number of interleaves would require that the phase-encode blips alternate in size.

Figure 3.1(b) illustrates that automatic EPI ghosting correction and two-fold acceleration can be achieved simultaneously by controlling interleaf order as shown. In this case, full-FOV coil sensitivity maps are estimated on the fly based on four adjacent time frames. After separating L-R and R-L data, a reduction factor of 4 can be used to perform SENSE unaliasing operations. In general, to perform accelerated imaging with the proposed method and an acceleration factor of $A \geq 2$, the SENSE reconstruction will require a reduction factor of $2A$, which must not exceed the total number of coils, and the most recent $2A$ time frames will be used for forming coil sensitivity maps. The accelerated method can be applied to double-alternating EPI with $oA$ interleaves for any odd integer $o$ using phase-encode blips with constant size, and with $eA$ interleaves for any even integer $e$ using phase-encode blips with alternating size.
1D phase correction [16] is used for comparison in the phantom and in vivo studies because it also does not rely on any calibration, and is compatible with real-time imaging. For each coil, a 1D phase map representing constant and linear phase errors is computed using phase differences between the L-R and R-L images. 1D phase correction is performed using Eqn. (12) from Ref. [16]. A root-sum-of-squares operation is performed to produce the final corrected image, with the images from all coils considered for reconstruction.

Experiments were performed on a Signa Excite 3 T scanner (GE Healthcare, Waukesha, WI) with gradients capable of 40 mT/m amplitudes and 150 mT/m/ms slew rates. The receiver bandwidth was set to ±125 kHz, i.e., 4 μs sampling time. The body coil, capable of peak $B_1$ of 16 μT, was used for RF transmission and an 8-channel cardiac array coil was used for signal reception.

Circular EPI (CEPI) trajectories [85], which have a circular $k$-space footprint, were designed in MATLAB (The Mathworks, South Natick, MA). CEPI trajectories used in the experiments produced a 10 to 15% reduction in readout duration compared to conventional rectangular EPI trajectories. A bipolar pulse was added to the phase encode gradient waveform prior to data acquisition in order to null the first moment in the $y$ direction at $k_y = 0$. Alternating interleaved EPI readouts with an odd number of interleaves were used [17, 41], and this $k$-space traversal pattern was flipped in the readout direction at every time frame. Echo-time shifting [28] was used to mitigate off-resonance effects.

Real-time interactive cardiac scanning was performed using the RTHawk real-time imaging platform. Scan planes were changed interactively by the operator in order to test the performance of the proposed method at all angles. A spectral-spatial RF pulse was used to excite water spins, with a 5.2 mm slice thickness, and 440 Hz bandwidth [80].
Flip angles of 15 to 25 degrees were used. The proposed reconstruction was performed both off-line and on-line, and videos were produced off-line. Raw data from all four anterior receiver channels were used to perform SENSE reconstruction in the non-accelerated acquisition, and raw data from all eight elements were used to perform SENSE reconstruction in the accelerated case. The noise correlation matrix used in SENSE reconstruction was computed using the formula presented in the Appendix section of Ref. [90] from raw data obtained with the RF excitation turned off.

Phantom experiments were conducted to quantitatively evaluate the level of ghost suppression. A cylindrical phantom was imaged with axial, oblique, and double-oblique scan orientations. The proposed method was compared with conventional 1D phase correction [16]. The effectiveness of ghost suppression was evaluated by comparing ghost-to-signal ratios (GSR) within the same manually selected region of interest (ROI).

Cardiac in vivo experiments were performed on three healthy volunteers without gating or breath-holding, and were evaluated qualitatively. Each subject was screened and provided informed consent in accordance with institutional policy.

The non-accelerated automatic ghosting correction method was implemented in C++, within the RTHawk real-time reconstruction software [96]. The LAPACK linear algebra package was used for the matrix inversion operations required during SENSE reconstruction. A Linux personal computer (Compaq R3000) with a single 3.2 GHz Intel CPU and 896 MB RAM was used for real-time reconstruction. Data were sent from the host computer to the reconstruction computer via Ethernet after each TR [96]. Data from the four anterior elements of the eight element receiver array were used for reconstruction. Reconstruction time was measured separately using the built-in C++ “gettimeofday” function for the following four reconstruction steps: 1) coil sensitivity map estimation, 2) aliased image reconstruction, 3) SENSE matrix inversion, and 4) image display.
Table 3.1: Ghost-to-signal ratios from the phantom study (see Fig. 3.2 for 1D phase correction and the proposed method. Mean and standard deviation of $g$-factor values for the proposed method are also reported.

<table>
<thead>
<tr>
<th>Scan plane</th>
<th>1D phase correction</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ghost-to-signal ratio</td>
<td>Ghost-to-signal ratio</td>
</tr>
<tr>
<td>Axial</td>
<td>3.94 %</td>
<td>4.24 %</td>
</tr>
<tr>
<td>Oblique</td>
<td>8.76 %</td>
<td>3.40 %</td>
</tr>
<tr>
<td>Double oblique</td>
<td>6.18 %</td>
<td>2.03 %</td>
</tr>
</tbody>
</table>

### 3.3 Results

In Figures, unless otherwise noted, the readout and phase encode directions correspond to the horizontal and vertical axes, respectively. The uncorrected images in Figures 3.2 and 3.3 were reconstructed by performing an inverse Fourier transform of the raw data without separating the L-R and R-L acquisitions, and without applying any post-processing method for EPI ghosting correction.

Figure 3.2 contains phantom images reconstructed from data acquired in axial, oblique, and double oblique scan planes. Both 1D phase correction and the proposed method produce visually comparable ghost-free images in the axial scan plane. However, in oblique and double-oblique scan planes, ghosting artifacts are prominent when using conventional 1D phase correction but are significantly reduced when using the proposed method. Ghost-to-signal ratios (GSR) are listed in Table 3.1 and demonstrate the effectiveness of the proposed method. The GSRs for oblique and double oblique scan planes for the proposed method were 3.40% and 2.03% respectively, whereas those for 1D phase correction were 8.76% and 6.18%. The mean and standard deviation of $g$-factor values for the proposed method are also listed in Table 3.1. The average $g$-factor values ranged from 1.2 to 1.3 in all three scan planes considered.
Figure 3.3 contains *in vivo* cardiac images for four standard cardiac views in one representative volunteer. Uncorrected images are shown alongside images reconstructed using 1D phase correction and the proposed method. Uncorrected images exhibit ghosting artifacts in all four cardiac views, and are most severe in the two chamber and four chamber views. Both 1D phase correction and the proposed method produce appreciable suppression of ghost artifacts in all views. The proposed method produces better image quality and improved ghost suppression compared to 1D phase correction (indicated by white arrows in Fig. 3.3). However, the proposed method experiences noise amplification due to the use of parallel imaging with a reduction factor of 2. The mean and standard deviation of g-factor values were $1.45 \pm 0.50$, $1.75 \pm 0.94$, $1.34 \pm 0.33$, and $1.34 \pm 0.28$, for the axial, two-chamber, four-chamber, and short axis views, respectively.

Figure 3.4 illustrates a double-oblique short-axis view during rapid scan-plane rotation. The scan plane was rotated in increments of 10 degrees continuously throughout the acquisition. Ghosting artifacts are suppressed even during rapid changes in scan orientation. Note that images at 0.9, 1.2, and 3.3 seconds appear blurred because they occur during scan plane changes.

Figure 3.5 compares two-fold acceleration with no acceleration and also illustrates the effect of motion on low temporal resolution coil sensitivity maps. Coil sensitivity maps were reconstructed using either I) data temporally adjacent to the target data, or II) data obtained from a stable diastolic phase. A FOV of 33 cm (rather than 25 cm) was used to mitigate the g-factor increase when using a reduction factor of 4 with our 8-element cardiac array. Two-fold accelerated images (Fig. 3.5c,h) exhibit much lower SNR than non-accelerated images (Fig. 3.5b,g) because of the reduced acquisition time and the highly elevated g-factor. In some areas the g-factor reached 10 (Fig. 3.5e). Coil sensitivity maps obtained using calibration scheme I exhibit motion induced ghosting
artifacts (Fig. 3.5a, solid arrow), while those obtained using calibration scheme II exhibit substantially reduced motion artifacts (Fig. 3.5f, solid arrow). Ghosting artifacts in the coil sensitivity maps lead to residual ghosting in the final images (Fig. 3.5b,c compared to Fig. 3.5g,h). The mean and standard deviation of $g$-factor values were $1.25 \pm 0.37$ and $2.05 \pm 0.90$ for the non-accelerated and two-fold accelerated cases using calibration scheme I, and were $1.16 \pm 0.25$ and $1.79 \pm 0.58$ for the non-accelerated and two-fold accelerated cases using calibration scheme II. Two-fold accelerated images exhibit less temporal blurring of structures because of the reduced acquisition time (e.g., descending aorta indicated by dashed arrows).

Reconstruction time for the non-accelerated method was measured for the real-time reconstruction algorithm. Average run time measurements for the different reconstruction steps were: 45.04 ms for the computation of coil sensitivity maps, 41.45 ms for the computation of aliased images, 57.44 ms for SENSE matrix inversion operations, and 0.84 ms for image display. The total reconstruction time was approximately 144 ms per frame, while the acquisition time per image was 60 ms. This indicates that pipelining this computation across three processors will be sufficient for real-time reconstruction using commercially available personal computer hardware.

### 3.4 Discussion

The proposed method effectively eliminated ghosting artifacts due to EPI echo-misalignment in arbitrary double-oblique scan planes. When applied to real-time interactive imaging, it automatically corrected ghosting artifacts without a calibration scan whenever a scan plane change occurred. This automatic capability is attributed to the fact that ghost-free coil sensitivity maps are updated with a few recent time frames, e.g., two time frames for
the non-accelerated acquisition and four time frames for the two-fold accelerated acquisition. In this method, a factor of two in SENSE reduction is used solely for correcting EPI ghosting artifacts, and not for accelerating data acquisition. In the non-accelerated acquisition, where a SENSE reduction factor of two is used, noise amplification due to the SENSE matrix inversion was relatively insignificant. However, in the two-fold accelerated acquisition, where a SENSE reduction factor of four is used, the SENSE noise amplification was severe. This can be mitigated by using 16-channel, 32-channel, or larger receiver coil arrays for which rate-4 parallel imaging has been demonstrated with reasonable $g$-factors [36].

A drawback of the proposed method is that lower temporal resolution coil sensitivity maps suffer from motion induced ghosting artifacts when cardiac motion occurs during coil calibration. Coil sensitivity maps corrupted with motion artifacts produce residual artifacts in final corrected images. A simple way of alleviating this effect is to use the data from a relatively stationary diastolic cardiac phase to construct coil sensitivity maps that are free from motion induced ghosting. In the experiments, the use of stationary frames for coil calibration substantially reduced ghosting artifacts in final corrected images. Alternatively, the use of temporal low-pass filtering in coil calibration may also mitigate ghosting artifacts in coil sensitivity maps [47].

When the non-accelerated correction method was applied to real-time cardiac imaging at 3 T, overall reconstructed images demonstrated high temporal resolution, excellent suppression of ghosting artifacts, and high blood-myocardium contrast. In systolic cardiac phases, subtle but noticeable FOV/2 ghosting was observed due to rapid cardiac motion. In end-diastolic cardiac phases, almost no ghosting was observed. While three-interleaf gradient-echo EPI was primarily used, I also experimented with other odd numbers of interleaves. Single shot EPI was considered as a way to further accelerate acquisition, but
signal loss and blurring due to $T_2^*$ relaxation proved to be limiting, given the same spatial resolution as the imaging protocol used in Fig. 3.3. The use of five or more interleaves increased the prevalence of artifacts due to cardiac motion and the lowered temporal resolution.

In conclusion, an interleaved gradient-echo EPI acquisition strategy, and PAGE-based reconstruction technique have been presented as a means for automatically correcting EPI ghosting artifacts due to echo-misalignment. The method was applied successfully to real-time interactive cardiac imaging at 3 T, with superior performance compared to conventional 1D correction. Ghosting artifacts were automatically corrected at arbitrary oblique scan planes, and high-quality ghost-free images were obtained with 3.1 mm spatial resolution and 60 ms temporal resolution. The automatic EPI ghosting correction method utilizes parallel imaging with a reduction factor of 2, and may be also compatible with further acceleration using higher reduction factors when 16, 32, or higher channel receiver coil arrays are used. The feasibility of real-time reconstruction using commercially available workstations was demonstrated.

### 3.5 Application to Upper Airway Imaging

The proposed “double-alternating” EPI acquisition method was tested on a midsagittal slice of the human upper airway. To reduce the readout duration and shorten the echo time, partial $k$-space sampling was used where only 66.7 % $k$-space lines were acquired. Pixel resolution was 60 × 60. ETL was 8. The $k$-space data from every 10 interleaves (i.e., 10 TRs) were combined and used for image reconstruction. This resulted in a frame with 77 ms temporal resolution. Frames were updated at every 5 TRs, in which TR was 7.704 ms. Designed FOV was 18 × 18 cm$^2$ and 110% FOV scale factor was used so that the final in-plane spatial resolution was 3.3 × 3.3 mm$^2$. 
Experiments were performed on a Signa Excite 1.5 T scanner (GE Healthcare, Waukesha, WI) with gradients capable of 40 mT/m amplitudes and 150 mT/m/ms slew rates. The receiver bandwidth was set to ±125 kHz, i.e., 4 µs sampling time. The body coil, capable of peak $B_1$ of 16 µT, was used for RF transmission and a custom 4-channel upper airway coil was used for signal reception.

One subject was screened and provided informed consent in accordance with institutional policy. Audio signals were recorded inside the magnet. The subject was scanned in supine position and repeated the following TIMIT [30, 118] sentences: “she had your dark suit in greasy wash water all year” and “Don’t ask me to carry an oily rag like that”. MRI gradient acoustic noise was eliminated based on Ref. [13]. The frames corresponding to “all year” were able to be identified with the aid of the synchronized audio and video. Figure 3.6 shows 20 frames representing the utterance of “all year”. Images are free from ghosting artifacts even though the scan plane is slightly double-oblique from the on-axis sagittal plane. Unlike spiral imaging, air-tissue boundaries do not suffer from a large degree of spatial blurring, but may suffer from geometric distortion along the phase encode direction.
Figure 3.1: Reconstruction flowchart for the proposed ghosting correction method for (a) non-accelerated and (b) two-fold accelerated EPI with automatic ghosting correction. (a) The acquisition and reconstruction as a function of time is shown for three-interleaves, and can be applied to any odd number of interleaves with constant phase encode blip size, or to any even number of interleaves with an alternating phase encode blip size. At time frame n-1, a full $k$-space data set is acquired with the interleaf ordering: $i_1^-$, $i_2^+$, $i_3^-$. At time frame n, another full data set is acquired with ordering: $i_1^+$, $i_2^-$, $i_3^+$. These two sets are repeated continuously. Note that '-' indicates that the readout gradient is flipped. To reconstruct a ghost-free image for time frame n, L-R data lines and R-L data lines are separated. The most recent two temporal frames are used to form coil sensitivity maps separately from L-R and R-L full-FOV data (middle row). SENSE reconstruction is used to form full-FOV L-R and R-L images for the current time frame, which are then combined using root-sum-of-squares, to produce a final image representing time frame n. (b) When accelerating data acquisition time by a factor of two, a SENSE reduction factor of four is needed. The most recent four temporal frames are used to form coil sensitivity maps. In general, if a reduction factor of R can be achieved for a particular coil geometry and scan planes, it can be combined with the proposed EPI strategy with an acceleration factor of R/2 since one half of the reduction factor is used for separating L-R and R-L lines during reconstruction and the remainder can still be used for acceleration.
Figure 3.2: Real-time cylindrical phantom images reconstructed with non-accelerated EPI data. Data from four receiver coils (two from anterior and the other two from posterior receiver coils) in 8-channel cardiac array coil were used for reconstruction. Reconstruction results for axial (top row), oblique (middle row), and double-oblique (bottom row) scan planes are shown using no correction, 1D phase correction, and the proposed correction method. The blue and red contours, superimposed on uncorrected images, represent regions of interest (ROIs) for signal and ghost, respectively. The pixel values in the ROIs were used to compute ghost-to-signal ratios. Images reconstructed with the proposed method are ghost-free even in oblique and double-oblique scan planes. Scan parameters: Double-alternating CEPI with five interleaves, ETL = 19, FOV = 21 × 21 cm², spatial resolution = 2.2 × 2.2 mm², TR = 39 ms, and time-per-image = 195 ms.
Figure 3.3: *In vivo* real-time cardiac images reconstructed with non-accelerated EPI data. Data from four anterior receiver coils in 8-channel cardiac array coil were used for reconstruction. Uncorrected (left column), 1D phase corrected (middle column), and corrected images with the proposed correction scheme (right column) are shown for four standard views: axial, two chamber, four chamber, and short axis. Ghosting artifacts are substantially reduced in all four corrected images with the proposed method. The white arrows indicate that residual ghosting artifacts are clearly visible in images reconstructed with the 1D phase correction method, but they are not observed in images reconstructed with the proposed method. Scan parameters: Double-alternating CEPI with three interleaves, ETL = 27, FOV = 25 × 25 cm², spatial resolution = 3.1 × 3.1 mm², TR = 20 ms, and time-per-image = 60 ms.
Figure 3.4: Automatic correction during continuous scan plane rotation. Scan parameters were identical to those described in Fig. 3.3. Reconstructed images using the proposed method are shown with respect to time. Numbers in the lower right corners in the images denote time in second. The interval between images is 300 ms, which corresponds to a spacing of five time frames. Note the decrease in contrast between myocardium and blood during the systolic phase in the cardiac cycle (see images at 3.6 and 4.5 second).
Figure 3.5: Automatic ghosting correction using no acceleration and two-fold acceleration (b,c,g,h) and corresponding $g$-factor maps (d,e,i,j) reconstructed using two coil calibration schemes. The target data are from a systolic cardiac phase where cardiac motion is substantial. $g$-factor maps are shown using a scale from 1 to 10. Note that $g$-factor values from calibration scheme II (i,j) are significantly lower than those from calibration scheme I (d,e), especially for the two-fold accelerated case. Motion artifacts are noticeably reduced for the two-fold accelerated case when using calibration scheme II (solid arrows). Two-fold accelerated images have lower SNR because of reduced acquisition time and the elevated $g$-factor, but show less temporal blurring in the descending aorta (dashed arrows) compared to non-accelerated images. Scan parameters: Double-alternating CEPI with two interleaves, ETL = 50, FOV = 33 × 33 cm$^2$, spatial resolution = 3.3 × 3.3 mm$^2$, TR = 34 ms, and time-per-image = 136 ms (full FOV low temporal resolution for coil calibration), 68 ms (non-accelerated), 34 ms (two-fold accelerated).
Figure 3.6: Dynamics of vocal tract shaping during natural speech utterances of “all year”. The phase encode is along the horizontal axis and the frequency encode is along the vertical axis. SENSE reconstruction was not used primarily because images from the two posterior coil elements exhibited aliasing artifacts in the vocal tract ROIs. Temporal resolution was 77 ms. Frame update rate was 38.52 ms.
4.1 Single-coil Imaging

4.1.1 Introduction

Three-dimensional (3D) imaging of the upper airway during sustained sound production has recently emerged as a promising tool in speech production research as a means to capture the full geometry of the vocal tract. The diversity of tongue shapes and dynamics are made possible, at least in part, through different lingua-palatal bracing mechanisms \([101, 102, 75, 2]\) leading to complex airway geometries, the understanding of which is critical for investigations into the production of both normal and disordered speech. In addition to helping shed light on the intricate airway shaping mechanisms underlying the production of various linguistically-meaningful speech sounds, 3D imaging also lends itself to providing quantitative volumetric information of the airway regions. The shaping of the tongue and other articulators, and the temporal characteristics of their shaping, give rise to characteristic patterns of acoustic resonance behavior of the vocal tract that define the properties of human speech that can be modeled with such quantitative information.
Recent work has shown that three-dimensional tongue shape and the dynamics underlying shape formation are critical to understanding natural linguistic classes and issues of phonological representation as evidenced in speech motor control. Previous models of speech production often assumed that the position of maximum constriction, defined in the midsagittal plane, was the main “place of articulation” parameter. Imaging studies such as those by Narayanan et al. [78] have suggested that articulation cannot be characterized solely by identifying a constriction position and that speech production targets go beyond the midsagittal plane. Initial speech studies using MRI focused on vowel sounds [6, 103]. The models of the vocal tract constructed from the MR images of different vowels yielded good estimations of vowel formant frequencies and formant patterns, which agreed with the general acoustic implication of the notion of the tongue height and backness on vowel articulation. For example, the study by Narayanan et al. [77] that focused on tongue shaping and 3D vocal tract data and models for the American English vowels /a/, /i/, /u/ showed distinct differences in tongue shaping: the anterior tongue was raised and convex for /i/ compared to the lowered concave shape for /a/ while the tongue back showed an opposite trend in the degree of concavity. These data were used in a finite element based simulation of the vocal tract models to study the acoustic properties of the vowel sounds. Other studies have investigated a variety of continuant consonant sounds such as fricatives and liquids. Narayanan et al. [75] examined vocal tract shaping of consonants using MRI and other articulatory measurements, and have presented data and results on three dimensional vocal tract and tongue shapes for fricative sounds produced by talkers of American English. These data showed key differences in tongue shaping between the sibilants /s/ (concave, grooved) and /ʃ/ (convex, cupped) and were helpful in deriving meaningful acoustic source models for these sounds [74]. Using insights gained in imaging work, in conjunction with the quantitative data of vocal tract
area functions and sublingual cavity of Alwan et al. [2], Espy-Wilson et al. [27] created acoustic models for the American-English /r/ delineating clearly the role of the oral and pharyngeal constrictions and the sublingual volume. Similar advances have been made toward understanding the acoustics of lateral sounds [7, 115]. While these studies represent significant progress in speech research, they can be further improved by addressing certain technological limitations.

These previous MRI studies were based on 2D multi-slice acquisitions, requiring multiple repetitions of the same sound and scan-time on the order of several minutes [75, 2, 78, 6, 103, 7, 76, 117]. These procedures are prone to data inconsistency, resulting from slightly different positions of the jaw, head, and tongue during each repetition. Compared to 2D multi-slice, it is well known that 3D encoding provides contiguous coverage with the potential for thinner slices and improved signal-to-noise ratio (SNR) efficiency. However, 3D encoding with high spatial resolution currently requires prohibitively long scan time and easily exceeds the normal duration of sustained sound production with minimal subject motion.

3D MRI scans may be accelerated using time-efficient $k$-space sampling [45], parallel imaging [33, 90], or with the recently developed approach of compressed sensing [20, 25, 66, 10]. Many of the efficient $k$-space sampling schemes (based on spiral and echo-planar trajectories) are prone to severe blurring artifacts and geometric distortions due to off-resonance at the air-tissue boundaries. Parallel imaging requires the design and use of receiver coil arrays where the coil elements have differing sensitivity over the anatomic region of interest [39]. Compressed sensing MRI (CS-MRI) relies only on sparsity of the final reconstructed image in a transform domain [20, 25, 66, 10].

In this manuscript, I investigate the use of CS-MRI for accelerated 3D upper airway imaging, and investigate the potential benefit of phase constraints. MR images often have
spatially varying phases whose sources may include receiver coil phase, gradient/DAQ delays, off-resonance, flow and motion. Phase constrained (PC) CS, originally proposed by Lustig et al. [66], applies a low spatial resolution phase estimate as part of the encoding function. This is expected to increase sparsity of the solution in certain transform domains (e.g., finite difference). I explore the use of PC-CS in this application, because air-tissue boundaries are the primary features of interest and are expected to experience substantial phase variation due to air-tissue susceptibility. I compare phase estimation from a low-resolution fully sampled regime with a two-stage approach that estimates the object phase map from a non-PC CS reconstruction. In retrospective sub-sampling experiments with no sound production, CS reconstructed images with and without phase constraints were compared qualitatively. Undersampled 3DFT acquisition and PC-CS reconstruction was then prospectively applied with acceleration factors of 3, 4, and 5, to high-resolution 3D vocal tract scanning during sustained sound production of English consonants /s/, /ʃ/, /l/, /r/, sounds characterized by complex tongue and airway shaping.

4.1.2 Theory

Consider 3DFT imaging, where \( k_y \) and \( k_z \) are the phase encoding directions, and therefore the axes of undersampling. After 1D Fourier transformation along \( k_x \), the signal for each \( x \) position can be expressed as:

\[
s(k_j) = \sum_{l=1}^{L} e^{-i2\pi k_j \cdot r_l} e^{i\phi(r_l)} m(r_l) + n(k_j).
\] (4.1)

Here, \( k_j \) is the \( j^{th} \) sampled \( k \)-space sample location in the \( (k_y, k_z) \) domain and \( 1 \leq j \leq J \), where \( J \) is the total number of phase encodes. \( r_l \) is the \( l^{th} \) spatial position in the \( (y, z) \) image domain, and \( L \) is the total number of pixels. \( \phi \) is the phase in the \( (y, z) \) image domain and \( m \) is the desired magnitude image (representing amplitude of
transverse magnetization) in the \((y, z)\) image domain, and \(n\) is the i.i.d. (independent and identically-distributed) additive white Gaussian noise. Because Eq. 4.1 holds for \(1 \leq j \leq J\), there exist \(J\) linear equations that can be expressed as one matrix equation:

\[
s = \Phi Pm + n. \tag{4.2}
\]

Here, the signal vector \(s\) is \([s(k_1) \ s(k_2) \ ... \ s(k_J)]^T\), \(\Phi\) is the \(J \times L\) Fourier encoding matrix, where \(\Phi(j, l) = e^{-i2\pi k_j \cdot r_l}\), \(P\) is an \(L \times L\) diagonal matrix, where the \(l\)th diagonal element is \(e^{i\phi(r_l)}\). \(m = [m(r_1) \ m(r_2) \ ... \ m(r_L)]^T\) is the unknown image estimate and \(n = [n(k_1) \ n(k_2) \ ... \ n(k_J)]^T\). When \(J << L\), Eq. 4.2 becomes a highly underdetermined linear system, and infinitely many solutions for \(m\) exist. Compressed sensing theory states that \(m\) can be exactly recovered with a very high probability when \(m\) is sparse in a transform domain, by minimizing the \(l_1\)-norm of the sparsifying transform of the solution under the constraint that \(\|s - \Phi Pm\|_2\) is close to zero. Unconstrained optimization is more practical for large-scale reconstruction problems such as MRI image reconstruction, therefore, the unknown image estimate \(m\) is obtained by minimizing the following convex function:

\[
f(m) = \|s - \Phi Pm\|_2^2 + \lambda \|\Psi m\|_1. \tag{4.3}
\]

Here, \(\lambda\) is a regularization parameter that controls the relative weight of sparsity and data fitting, and \(\Psi\) is a sparsifying transform (e.g., wavelets, curvelets, or finite difference). In this work, I adopted the finite difference sparsifier that contains the horizontal and vertical gradients of the image. In the absence of \(P\) (i.e., \(P\) is the identity matrix), the optimization problem is referred to as non-phase-constrained CS reconstruction. In phase constrained CS reconstruction, \(P\) contains a predetermined estimate of the object phase,
which may originate from system delays, receiver coil phase, and phase accrual due to off-resonance.

4.1.3 Materials and Methods

4.1.3.1 Data Acquisition

Experiments were performed on a Signa Excite HD 3.0 T scanner (GE Healthcare, Waukesha, WI) with gradients capable of 40 mT/m amplitude and 150 mT/m/ms slew rate. The receiver bandwidth was set to ±125 kHz (i.e., 4 µs sampling rate). A birdcage head coil was used for RF transmission and signal reception. Each subject was screened and provided informed consent in accordance with institutional policy.

The vocal tract region of interest was imaged using a single midsagittal slab with 8-cm thickness in the right-left (R-L) direction. The readout direction was superior-inferior (S-I) and the phase encode directions were anterior-posterior (A-P) and right-left (R-L) (see Fig 4.1). A gradient echo sequence was used with TE = 2.2 msec, TR = 4.6 msec, flip angle = 5°, NEX = 1, spatial resolution = 1.5 × 1.5 × 2.0 mm³, and FOV = 24 × 24 × 10 cm³.

Pseudo-random undersampling was implemented as follows. First, two independent and uniformly distributed random numbers corresponding to k-space radius and azimuthal angle were generated to create pseudo random \((k_y, k_z)\) location in polar form. From the randomly chosen samples, the nearest \((k_y, k_z)\) Cartesian phase encodes were selected for sampling. This scheme achieves a sampling density that is inversely proportional to k-space radius. Second, a low spatial frequency, whose outermost k-space radius was 30% of the full k-space radius, was fully sampled. The final sampling patterns and corresponding reduction factors are shown in Fig 4.2.
Figure 4.1: Illustration of scan plane prescription, which is used for the 3D upper airway imaging. The dashed lines indicate the orthogonal slice orientation of each image. The largest-width, medium-width, and smallest-width dashed lines are for the prescription of the midsagittal, coronal, and axial slices, respectively. An 8 cm sagittal slab excitation is applied to cover the vocal tract volume of interest. The readout direction is along S-I so that the analog low-pass filter suppresses uninteresting regions (e.g., the brain and neck). The features of interest include: [LL] lower lip, [UL] upper lip, [P] palate, [T] tongue surface, [V] velum, [PW] pharyngeal wall, [E] epiglottis.

4.1.3.2 Image Reconstruction

Since all data sets were fully sampled along the readout ($k_x$) direction, data were first inverse-Fourier transformed along the readout direction, and image reconstruction was performed separately for each $y - z$ planar section. For each $x$ position, fully sampled data sets were reconstructed using 2D inverse Fourier transform (IFT). For the simulated and real undersampled acquisitions, un-acquired $k$-space locations were filled with zeros prior to inverse Fourier transformation.

For PC-CS, the phase map was calculated in two ways: (PC-I) Taking a 2D inverse Fourier transform of fully sampled low spatial frequency data. In order to remove Gibbs
Figure 4.2: $k$-space sampling patterns used in the experimental studies. Relative reduction factors are (a) 1, (b) 1.3, (c) 3, (d) 4, and (e) 5. Note that the region inside the ellipse with a radii 30% of the overall $k$-space was fully sampled in all cases for the estimation of low-resolution image phase.

ringing artifacts due to $k$-space truncation, the low spatial frequency data set was multiplied by a 2D Hanning window. (PC-II) Taking the phase of the complex-valued image estimate obtained from a non-PC CS iterative reconstruction. To avoid noise contamination, the PC-II phase map was masked to contain only spatial locations where the magnitude image was greater than 20% of its maximum value.

CS reconstructions from undersampled data sets were based on an iterative non-linear conjugate gradient algorithm [66] which sought to find a global minimum for the cost function in Eq. 4.3. The $l_1$-norm of the finite difference of the solution (also known as Total Variation [95]) was used as a regularizer. The regularization parameter $\lambda$ was chosen based on the L-curve method [35]. I examined the tradeoff between data consistency and total variation for a broad range of $\lambda$ values (see Fig. 4.3) prior to selecting a $\lambda$ value for image reconstruction from prospectively acquired data. To speedup reconstructions over a broad range of $\lambda$ values, the final image from a particular $\lambda$ value was used as the initial image estimate for the CS reconstruction with the next higher $\lambda$ value.
4.1.3.3  **In Vivo Experiments**

Subjects were in supine position and their heads were immobilized by inserting foam pads between their ears and the receiver coil. A fully sampled data set without sound production was acquired in one trained subject. Their mouth was held open for 36 seconds without swallowing. A total of 8000 \((k_y, k_z)\) encodes, where the number of \(k_y\) and \(k_z\) encodes was 160 and 50, respectively, was used to fully cover 3D \(k\)-space at the Nyquist rate. This data set was retrospectively sub-sampled to simulate the sampling patterns shown in Fig. 4.2. The CS reconstructions were performed both without and with phase constraints.

Prospective accelerated acquisitions were performed by imaging the vocal tract shaping during each sustained sound production of English consonants /s/, /ʃ/, /l/, and /r/. Scan time for the 3, 4, and 5-fold accelerated acquisitions took 12, 9, and 7 seconds, respectively. 2D CS reconstruction was performed for each axial slice. The initial estimate for the CS reconstruction of a slice was taken from the final image estimate obtained from the CS reconstruction of its adjacent slice. PC-CS reconstruction was applied with $$\lambda = 0.005$$ and 100 iterations for 65 contiguous slices of interest along \(x\) (i.e., S-I direction). 3D visualization of tongue shape was realized by manually segmenting the tongue in each reconstructed coronal image, stacking the segmented slices, and finally performing 3D volume rendering using the vol3d.m Matlab routine (publicly available at http://www.mathworks.com). The final volume rendered tongue surfaces were able to be displayed at any view angle, providing efficient visualization of tongue shaping.

4.1.4  **Results**

Figure 4.3 shows an L-curve obtained from the non-PC CS reconstruction. The corner of the L-curve was not sharp and $$\lambda = 0.005$$, which lies on highest curvature, was chosen as
an optimal regularization parameter for both the non-PC and PC CS reconstructions. For large values of $\lambda$ (i.e., $\lambda > 0.01$ in Fig. 4.3), reconstruction strongly favored minimization of total variation so that reconstructed images were observed to be overly smooth.

Figure 4.4 shows some representative axial slice images from 3D data set when the subject was in open-mouthed position during the 36-seconds scan. The phase variation pattern resulting from large off-resonance due to air-tissue susceptibility is not smooth, and is related to the geometry of air-tissue interface. This is observed especially in the orofacial tissue and the lateral sides of the tongue (see the yellow arrows).

Figure 4.5 shows images from one axial slice extracted from 3D volume in the retrospective sub-sampling experiment. Figure 4.5a contains images obtained from IFT, non-PC CS, PC-I CS, and PC-II CS reconstructions of the data sets sub-sampled with different reduction factors. The image from the elliptic $k$-space full sampling (1.3x) was comparable in image quality to that from the rectangular $k$-space full sampling (1x). The IFT reconstructed images from the undersampled data exhibited incoherent aliasing artifacts and the image quality was degraded with higher reduction factors. The non-PC CS reconstruction improved image quality over the IFT reconstruction in terms of de-noising and enhancement of the air-tissue boundaries. The PC-I and PC-II CS reconstructions further improved the air-tissue boundary depiction quality for reduction factors 3, 4, and 5. Figure 4.5b contains phase difference images after the low and high resolution phase maps were subtracted from the full resolution fully sampled reference phase map. Notice the larger phase errors in the low resolution phase map (see Fig. 4.5b(iii)) particularly in the ROIs with rapid phase variations (indicated by the white arrows in Fig. 4.5b(i,iii,v)). Figure 4.5c compares the boundary depiction in ROIs with rapid phase variation for
different reconstruction schemes. The PC-II CS reconstruction clearly improved the depiction of the air-tissue boundaries compared to PC-I CS reconstruction (see white arrows in Fig. 4.5c).

Figure 4.6 shows a midsagittal slice and eight equally spaced coronal slices reformatted from a 3D vocal tract volume obtained after the PC-II CS reconstructions. 3D imaging provided many useful vocal tract shaping features that cannot be captured by 2D midsagittal imaging alone. The groove of the tongue surface could be clearly observed in coronal sections in /s/ (see the white arrow in the /s/ row of Fig. 4.6). /ʃ/ and /l/ sounds exhibited very similar vocal tract shaping patterns in the midsagittal scan plane, but when comparing the coronal slices, the vocal tract cross sectional areas were significantly different (see the white arrows in the /ʃ/ and /l/ rows in Fig. 4.6). Figure 4.7 shows a 3D visualization of the tongue surface for each sound production of /s/, /ʃ/, /l/, and /r/. The groove of the tongue was clearly seen for the fricative /s/ and /ʃ/ sounds, but it was not observed for the /l/ sound. The cupping of the tongue was observed in the /r/ sound.

4.1.5 Discussion

The major sources of phase include: 1) receiver coil phase, 2) spatial frequency offset due to field inhomogeneity and air-tissue magnetic susceptibility difference, and 3) gradient/DAQ timing delay. These may be estimated from separate calibration scans or via self-calibration, which was chosen in this study. Self-calibration avoids possible errors caused by the vocal tract geometry changing between calibration scans and accelerated scans. It is noted that the features of interest are air-tissue boundaries such as the tongue surface, lips, hard palate, velum, and epiglottis which are coordinated for the generation
of unique gestures depending on different articulation tasks. Even two separate productions of the same sound/articulation task could result in slightly different vocal tract shaping, and has been a source of difficulty widely reported in the literature.

The PC-II CS reconstruction utilized a relatively high spatial resolution phase map obtained from the non-PC CS reconstruction and improved the depiction of the air-tissue boundaries with large degree of phase variation, particularly at high acceleration factors. The phase map estimate may be prone to artifacts due to imperfect CS reconstruction, but it does tend to contain the rapidly varying phase information, while the low spatial resolution phase map does not. A drawback of the PC-II CS reconstruction is an increased reconstruction time because of the need for an additional iterative CS reconstruction just for the phase estimate.

The use of Total Variation (TV) regularization was effective at improving the depiction of air-tissue boundaries and suppressing noise-like aliasing artifacts, and was more effective when combined with the phase-constrained reconstruction technique. The de-noising and edge-preserving characteristics can improve the performance of the subsequent image processing tasks (e.g., Canny edge detection, image segmentation) for the quantification process such as the measurement of the vocal tract area function. The degree of the influence of TV regularizer was controlled by the choice of the regularization parameter $\lambda$. The L-curve analysis provided the insight of choosing an appropriate $\lambda$. Moreover, the wavelet or curvelet transform can be used as another sparsifying basis and the reconstruction may be improved by incorporating an additional regularizer into the optimization function.

A drawback of the method is that reconstruction is computationally intensive and requires a considerable reconstruction time. The convergence speed of the algorithm was observed to decrease as either a higher acceleration factor or a large value of the
regularization parameter is used. For the generation of a 3D volume of the upper airway, 100 iterations were used to reconstruct a single image and this iterative reconstruction was processed for 65 contiguous slices of interest along $x$. The generation of a 3D volume when using PC CS took approximately 4 hours on a 3.4 GHz of CPU with 3.0 GB of RAM.

In this work, the CS reconstructions were performed in two dimensions ($y, z$) after 1D IFT along $k_x$. If computation time and memory size were not issues, there would be potential benefits to solving the CS optimization in 3D directly. Sparsity along $x$ would allow for some additional de-noising, and there would be an opportunity to correct shifts in $x$-position due to off-resonance if the different sources of image phase could be separated.

Although not shown here, the use of coil arrays (e.g., 8-channel neurovascular array in our work) can improve the SNR in 3D upper airway imaging. If the combined use of parallel imaging and compressed sensing were adopted, significantly higher accelerations would be achievable [66, 10, 58, 73]. Linguistically relevant high resolution features such as tongue tip constrictions and epiglottis would be easily resolved. Moreover, it may be possible to measure the vocal tract area function with greater precision, therefore improving the accuracy of the quantitative analysis of vocal tract shaping in both normal and disordered speech production.

### 4.1.6 Conclusions

I have demonstrated the application of compressed sensing (CS) MRI to high-resolution 3D imaging of the vocal tract during a single sustained sound production task (no repetitions needed). Phase constrained CS outperformed conventional CS in spatial locations with large phase variations (lateral edges of the tongue). I have demonstrated that 5x
acceleration is achievable with PC CS, with negligible loss of tissue boundary information that is relevant to speech production research. I have demonstrated a 3D upper airway imaging using an undersampled 3DFT gradient echo acquisition with a $1.5 \times 1.5 \times 2.0$ mm$^3$ spatial resolution in 7 seconds, which is a duration practical for sustained sound production.

4.2 Multi-coil Imaging

In this section, I extend the accelerated 3D single-coil imaging and phase-constrained CS (PC-CS) reconstruction to imaging with a multiple channel receive coil array (parallel imaging). This combined use of compressed sensing and parallel imaging has been recently proposed by several groups [58, 73], but the notion of incorporating high-resolution phase information is unique to this work. I present a two-stage reconstruction approach that first estimates phase maps for each coil element via conventional CS reconstructions, and then reconstructs final image iteratively after incorporating the high-resolution phase maps and low-resolution magnitude coil sensitivity maps into a multi-coil CS reconstruction.

4.2.1 Methods

4.2.1.1 Data Acquisition

Experiments were performed on a 3.0 T Signa Excite HD MRI scanner (GE Healthcare, Waukesha, WI). The receiver bandwidth was set to $\pm 125$ kHz (4$\mu$s sampling rate). The body coil was used for RF transmission, and 8-channel neurovascular array coil was used for signal reception (only 4 superior elements were used for reconstruction). The vocal tract region of interest (ROI) was imaged using a single thick midsagittal slab with 8 cm thickness in the right-left (R-L) direction. The readout direction was superior-inferior (S-I) and the phase encode directions were anterior-posterior (A-P) and right-left (R-L).
A gradient echo (GRE) sequence was used with TE = 2.3 msec, TR = 5.0 msec, flip angle = 10°, NEX = 1, spatial resolution = 1.33 × 1.33 × 1.33 mm³, and FOV = 20 × 24 × 8 cm³.

4.2.1.2 In Vivo Experiments

A fully sampled data set, without sound production, was acquired when one trained subject held the mouth open for 54 seconds, without swallowing. A total of 10800 (k_y, k_z) encodes, where the number of k_y and k_z encodes was 180 and 60, respectively, fully covered 3D k-space at the Nyquist rate. The undersampling of (k_y, k_z) was based on 1) full sampling of low-spatial frequencies and 2) random undersampling of the remaining high-spatial frequencies. The outermost k-space radius of the fully sampled region was chosen to be 20% of the full k-space radius.

Prospective accelerated acquisitions were performed by imaging the vocal tract shaping during sustained sound production of American English /s/, /ʃ/, /l/, /ɹ/. The scan time for 6x, 8x, and 10x acquisitions were 9.0, 6.8, and 5.4 seconds, respectively.

4.2.1.3 Image Reconstruction

Data were first inverse-Fourier transformed (IFT) along k_x. At each x position, reconstruction was performed in 2D planar section. For comparison, two different conventional reconstruction schemes were used: 1) root-sum-of-squared (RSS) reconstruction, where images from all four coil elements were root-sum-of-squared (RSS) to produce final image, and 2) iterative conjugate-gradient-based un-regularized SENSE reconstruction based on the work of Pruessmann et al. [89].

The multi-coil PC-CS reconstruction is illustrated in Fig. 4.8. In the first stage, high-resolution phase map was estimated using CS reconstruction for each coil element. This is effective at capturing rapidly varying phases in the air-tissue boundaries, where
rapid phase variation is expected due to large susceptibility difference between the air and tissue. Its incorporation into a PC-CS optimization leads to increased sparsity of the transform coefficients of the final solution [66]. In the second stage, multi-coil PC-CS reconstruction was performed by minimizing the convex function in Eq 4.4. Here, $s_l$ is the data vector for the $l^{th}$ coil element, $\Phi$ is the Fourier encoding matrix, $P_l$ is a diagonal matrix containing the phase estimate, $C_l$ is a diagonal matrix containing coil intensity map, and $m$ is the unknown image estimate. $\beta_{TV}$ and $\beta_W$ are regularization parameters for total variation and $l_1$-norm of wavelet transform, respectively. Daubechies-8 wavelet transform was adopted in this study using WaveLab850 software (see the website http://www-stat.stanford.edu/~wavelab/). Their values were chosen after visual inspection of reconstructed images representing a broad range of the values from retrospective studies.

$$f(m) = \sum_{l=1}^{L} ||s_l - \Phi P_l C_l m||_2^2 + \beta_{TV}||\Psi_F m||_1 + \beta_W||\Psi_W m||_1.$$ \hspace{1cm} (4.4)

4.2.1.4 Data Processing and Analysis

Vocal tract area functions were measured by 1) manually drawing the vocal tract midline on a midsagittal slice, 2) prescribing several cross-sectional slices orthogonal to the midline, from the lips to the glottis, and 3) calculating the vocal tract areas from each cross-sectional slice. All analysis was done using OsiriX software [94].

4.2.2 Results and Discussion

Reconstruction results from retrospectively undersampled data (not shown) indicated that 6x and 8x produced little or no air-tissue boundary errors but 10x produced significant boundary errors in the airway and lateral sides of the tongue.
Figure 4.9 contains midsagittal images and their corresponding 3D visualization of tongue shapes for /ʃ/ and /ɾ/ sounds from prospectively acquired 8x data. The use of 1.33 mm isotropic resolution allows for sufficiently resolving the narrowing of the vocal tract between the tongue blade and alveolar ridge (the yellow arrow in (b)). The degree of the tongue grooving is clearly seen for the English fricative /ʃ/ (the black arrow in (e)). The /ɾ/ sound characterizes a complex geometry of the tongue shape (e.g., large volume of the sublingual cavity (the white arrow in (f)) and cupping of the frontal tongue (the red arrow in (f))).

Figure 4.10 contains a midsagittal slice, vocal tract area function, and 3D visualization of tongue surface for /s/, /ʃ/, /i/, and /ɾ/. Midlines (thin lines marked on each midsagittal image) that were manually drawn have different shapes depending on the sound being produced. The midline tongue contour for /ɾ/ was highly tortuous because of the large space from the sublingual cavity and the upward position of the tongue tip. Figure 4.10(b) shows that the measured area functions are different for the different articulations. The fricative sounds /s/ and /ʃ/ have increased areas near the glottis region unlike the vowel sound /i/ because of the backward movement of the tongue root (compare the shaping of the epiglottis in /s/, /ʃ/, /i/ from the midsagittal in Fig. 4.10(a)). Figure 4.10(c) suggests that 3D vocal tract geometry provides additional information such as the degree of the tongue grooving (compare the arrows in /s/, /ʃ/, /i/), cupping of the tongue (see the thick arrow in /ɾ/), and the volume of the sublingual cavity (see the thin arrow in /ɾ/).

4.2.3 Summary

The proposed reconstruction can produce a clear depiction of 3D tongue shaping with 1.33 × 1.33 × 1.33 mm³ resolution, which is higher than the single-coil imaging data set discussed in the previous section, from data acquired during 7 seconds scan and sustained
sound production. It adopts a phase-constrained CS combined with multi-coil data. It demonstrates clear depiction of air-tissue boundaries, which are the features of interest. However, it is computationally intensive because it requires L+1 iterative reconstructions as shown in Fig. 4.8, where L is the number of coil elements.
Figure 4.3: L-curve for the selection of regularization parameter $\lambda$ for CS reconstruction of the 3D upper airway data with reduction factors of 3, 4, and 5. The CS reconstruction was terminated at the 1000$^{th}$ iterate. The plotted points (x) and their corresponding regularization parameter values ($\lambda$) are shown for reduction factor 3. Virtually identical patterns were observed for reduction factors 4 and 5. The corners of the L-curve are not sharp, but provide a clear trade-off between total variation (sparsity) and data consistency.
Figure 4.4: Representative magnitude and phase images from axial slices. The adjacent images shown were 3 mm apart along the S-I direction. The data were acquired in the open-mouthed position during 36 seconds. Large phase variations are observed particularly in the orofacial tissue and the lateral sides of the tongue (see the yellow arrows).
Figure 4.5: Axial slice reconstructions from retrospective sub-sampling of fully sampled data. (a) Magnitude images reconstructed by use of inverse Fourier transform (iFT), non-phase-constrained compressed sensing (CS), PC-I CS, and PC-II CS reconstructions of 1x, 1.3x, 3x, 4x, 5x sub-sampled data. (b) (i) Full-resolution phase map from fully sampled 1x data. (ii) Low-resolution phase map from fully sampled low-frequency data. (iii) Phase difference between phase maps (i) and (ii). (iv) Phase map from non-PC CS reconstruction of 5x sub-sampled data. (v) Phase difference between phase maps (i) and (iv). (c) Magnified ROIs inside the red rectangle in (a). Notice the sharp depiction of the air tissue boundaries in 5x PC-II CS reconstructed image (see the white arrows in (c)).
Figure 4.6: Reformatted 2D midsagittal and coronal images after the PC-II CS reconstructions of the 5x undersampled 3DFT data set. The prospective use of accelerated 3DFT scanning required just 7 seconds of scan time during which one trained subject produced each sustained English consonant /s/, /f/, /l/, and /r/. This achieved 1.5 × 1.5 × 2.0 mm³ resolution over a 24 × 24 × 10 cm³ FOV. Representative 2D midsagittal images are shown in the leftmost column. Eight representative coronal slices of interest are shown that are ordered from lips to pharyngeal wall. Important articulatory features provided by the 3D vocal tract dataset include: (1) groove of the tongue surface for fricative sound /s/ (see the arrow in the /s/ row) and (2) wider shaping of the vocal tract between the hard palate and the tongue front for /l/ indicating the curving of the tongue sides to allow airflow along the sides (for the comparison, see the arrows in the /f/ and /l/ rows) although their 2D midsagittal slices exhibit similar shaping patterns.
Figure 4.7: 3D visualization of the tongue and lower jaw after the PC-II CS reconstructions from the data set prospectively acquired with 5x acceleration. Tongue grooves are seen for /s/ and /ʃ/, further forward in /s/ than /ʃ/, but not for /l/ (see the arrows in /s/, /ʃ/, and /l/). Cupping of the tongue (i.e., cavity behind the tongue front) is seen for /r/ (see the arrow in /r/).
Figure 4.8: Flowchart of the proposed reconstruction scheme.
Figure 4.9: Midsagittal images for (a) /s/, (b) /ʃ/, and (c) /r/. Their corresponding 3D tongue shapes for (d) /s/, (e) /ʃ/, and (f) /r/.
Figure 4.10: The prospective use of accelerated 3D acquisition and multi-coil PC-CS reconstruction. (a) Reformatted midsagittal slices and their associated midlines drawn for cross-sectional slice prescription. (b) Area function plot. (c) 3D visualization of the tongue and lower jaw.
Chapter 5

Real-time Speech MRI Using Golden-ratio Spiral

5.1 Introduction

Real-time MRI has provided new insight into the dynamics of vocal tract shaping during natural speech production [79, 11, 24, 107]. In real-time speech MRI experiments, image data and speech signals are simultaneously acquired. Real-time movies, typically of a 2D midsagittal slice, are reconstructed and displayed in real-time. The shape of the vocal tract, from the lips to the glottis, is identified using air-tissue boundary detection performed at each frame [12]. Adaptive noise cancellation is used to produce speech signals free from the MRI gradient noise [13]. Articulatory and acoustic analysis is then performed using synchronized audio and video information [87]. Although MRI data is acquired and reconstructed in real-time, the processes of segmentation and analysis are performed retrospectively.

Speech rate is highly dependent on the subject’s speaking style and the speech task, and it affects speed of articulatory movement [1, 108]. Variations in the velocity of articulators such as tongue dorsum, lips, and jaw result from the nature of the sequences of the vowels and consonants being produced [84]. The motion of articulators (e.g., tongue, velum, lips) is relatively slow during production of monophthongal vowel sounds.
or during/vicinity of pauses. Vocal tract variables such as tongue tip constriction, lip aperture, and velum aperture are dynamically controlled and coordinated to produce target words [15]. The speeds among articulators can also differ during the coordination of different articulators, for example, the movement of the velum and the tongue tip during the production of the nasal /n/.

Current speech MRI protocols do not provide a mechanism for flexible selection of temporal resolution. This is of potential value, because higher temporal resolution is necessary for frames that reflect rapid articulator motion while lower temporal resolution is sufficient for capturing the frames that correspond to static postures. As recently shown by Winkelmann et al. [114], golden-ratio sampling enables flexible retrospective selection of temporal resolution. It may be suited for speech imaging, in which the motion patterning of articulators varies significantly in time, and in which it is difficult to determine an appropriate temporal resolution \textit{a priori}.

In this manuscript, I present a first application of spiral golden-ratio sampling scheme (see Fig. 5.1) to real-time speech MRI and investigate its performance by comparison with conventional bit-reversed temporal view order sampling scheme. Simulation studies are performed to compare unaliased field-of-view (FOV) from spiral golden-ratio sampling with that from conventional bit-reversed sampling at different levels of temporal resolution after a retrospective selection. \textit{In vivo} experiments are performed to qualitatively compare image signal-to-noise ratio (SNR), level of spatial aliasing, and degree of temporal fidelity. Finally, I present an automated technique in which a composite movie can be produced using data reconstructed at several different temporal resolutions. I demonstrate its effectiveness at improving articulator visualization during production of nasal consonant /n/.
Figure 5.1: Schematic diagram of real-time continuous MRI data acquisition (DAQ) using a golden-ratio spiral view order. A sequence of only first five TRs is shown. (Top) Pulse sequence diagram. (Bottom) Accumulation of spiral interleaves in the sampling of $k$-space as time elapses. Every spiral interleaf acquired during current DAQ period is indicated in blue color. It is noted that temporal resolution can be controlled by the number of adjacent TRs chosen when reconstructing a frame. In the golden-ratio sampling scheme, next spiral interleaf never overlaps with previously acquired interleaves. Hence, sampling density (i.e., imaging field-of-view) increases as the number of adjacent TRs used for a frame reconstruction increases.

5.2 Materials and Methods

5.2.1 Simulation

A simulation study was performed to compare unaliased FOVs from conventional bit-reversed view order sampling [19] (see Fig. 5.2a,c,e) and spiral golden-ratio view order sampling (see Fig. 5.2b,d,f) for a variety of temporal resolutions selected retrospectively. The spiral trajectory design was based on the imaging protocol routinely used in our laboratory at the University of Southern California [79, 19]. The design parameters were: 13-interleaf uniform density spiral (UDS), $20 \times 20$ cm$^2$ FOV, $3.0 \times 3.0$ mm$^2$ in-plane spatial resolution, maximum gradient amplitude = 22 mT/m, maximum slew rate = 77 T/m/s, and conventional bit-reversed view order. Bit-reversed temporal view order is often adopted in real-time MRI because it shortens the spiral interleaf angle gaps in a few adjacent interleaves at any time point and reduces motion artifacts [100, 81]. Spiral golden-ratio view order was performed by sequentially incrementing the spiral interleaf
angle by the golden-ratio angle $360^\circ \cdot 2/(\sqrt{5} + 1) \approx 222.4969^\circ$ at every repetition time (TR) (see Fig. 5.1 and Fig. 5.2d). The unaliased FOV was defined as the reciprocal of the maximum sample spacing in $k$-space.

### 5.2.2 In Vivo Experiments

MRI experiments were performed on a commercial 1.5 Tesla scanner (Signa Excite HD, GE Healthcare, Waukesha, WI). A body coil was used for radio frequency (RF) transmission, and a custom 4-channel upper airway receive coil array was used for RF signal reception. The receiver bandwidth was set to ±125 kHz (i.e., 4 µs sampling rate). One subject was scanned in supine position after providing informed consent in accordance with institutional policy.

A midsagittal scan plane of the upper airway was imaged using custom real-time imaging software [96]. The spiral trajectory design followed those described in the Simulation section. The imaging protocol was: slice thickness = 5 mm, TR = 6.164 ms, temporal resolution = 80.1 ms. The golden-ratio view order scheme was compared with the conventional bit-reversed 13-interleaf UDS scheme with all other imaging and scan parameters fixed (e.g., scan plane, shim and other calibrations, etc.). The volunteer was instructed to repeat “go pee shop okay bow know” for both the conventional bit-reversed UDS and golden-ratio acquisitions. The speech rate was maintained using a 160 bpm metronome sound that was communicated to the subject using the scanner intercom.

In conventional bit-reversed 13-interleaf UDS data, gridding reconstructions were performed using temporal windows of 8-TR and 13-TR. In golden-ratio spiral data, gridding reconstructions were performed using temporal windows of 8-TR, 13-TR, 21-TR, and 34-TR. Gridding reconstructions were based on interpolating the convolution of density
compensated spiral \( k \)-space data with a \( 6 \times 6 \) Kaiser-Bessel kernel onto a two-fold over-sampled grids followed by taking 2D inverse fast Fourier transform (FFT) and deapodization [46]. Root sum-of-squares (SOS) reconstruction from the 2 anterior elements of the coil was performed to obtain the final images. For comparison of images reconstructed from different temporal windows, image frames were reconstructed from the data in which the centers of each temporal window were aligned.

5.2.3 Blockwise Temporal Resolution Selection

In golden-ratio data sets, multiple temporal resolution videos can be produced retrospectively. I sought a procedure for automatic selection of the temporal window that was appropriate for each image region in each time frame, and the ability to use this to synthesize a single video.

I used time difference energy (\( TDE \)), as described in Eq. 5.1, as an indicator of motion. This was calculated for each block \( B_j \) and each time \( t \):

\[
TDE(B_j, t) = \sum_{t' = -T}^{T} \sum_{(x,y) \in B_j} |I(x,y,t) - I(x,y,t - t')|^2,
\]

where \( I(x,y,t) \) is image intensity at pixel location \( (x, y) \) and time \( t \), and \( 2T \) is the number of adjacent time frames that are considered.

Temporal resolution selection was performed based on the alias-free high temporal resolution frames. I used sensitivity encoding (SENSE) reconstructed frames from 8-TR temporal resolution data (i.e., 49.3 ms temporal resolution) for the calculation of \( TDE \). In addition, SENSE reconstructions were performed at each frame from 13-TR, 21-TR, and 34-TR temporal windows, whose corresponding temporal resolutions were 80.1 ms, 129.4 ms, and 209.6 ms. Data from all 4 elements of the coil were considered for the reconstruction. Alias-free coil sensitivity maps were obtained from 34-TR data.
Sensitivity maps for each coil element were obtained by dividing the image at each element by the root SOS image of all elements. Frames were updated at every 4-TR = 24.7 ms (i.e., 40.6 frames per second). The spiral SENSE reconstruction was based on a non-linear iterative conjugate gradient algorithm with a total variation regularizer [58, 66]. Total variation regularization was effective at removing image noise while preserving high contrast signals such as the air-tissue boundaries. Iterations were terminated at the 20\textsuperscript{th} iterate after visual inspection.

Intensity correction was performed at each frame using a thin plate spline fitting method [64]. The SENSE reconstructed frames were first cropped to a 64 × 64 size that only contained the vocal tract regions of interest and then were interpolated to a 128 × 128 size in order to avoid the blockiness of the images. An 8 × 8 block and T = 2 was used to calculate $TDE$. Two spatially adjacent blocks were overlapped by 4 pixels in either the vertical or horizontal direction. The calculated $TDE$ at each block was assigned to the central 4 × 4 block. $TDE$ was normalized through the entire time frames. Temporal resolution selection at each 4 × 4 block was performed after a simple thresholding of the normalized $TDE$ ($TDE_{\text{norm}}$). For $TDE_{\text{norm}} \geq 0.6$, 8-TR SENSE was assigned. For $0.4 \leq TDE_{\text{norm}} < 0.6$, 13-TR SENSE was assigned. For $0.2 \leq TDE_{\text{norm}} < 0.4$, 21-TR SENSE was assigned. For $TDE_{\text{norm}} < 0.2$, 34-TR SENSE was assigned. These settings were chosen empirically based on quality of the final synthesized video.

5.2.4 Oral-Velar Coordination

Another set of experiment was performed to investigate the effectiveness of variable temporal resolution selection in golden-ratio spiral acquisition. The imaging protocol was: slice thickness = 5 mm, repetition time (TR) = 6.004 ms, in-plane spatial resolution = 2.4 × 2.4 mm\textsuperscript{2}, receiver bandwidth = ±125 kHz. The golden-ratio view order scheme was
compared with the conventional bit-reversed 13-interleaf UDS method on the same mid-
sagittal scan plane. The experiment was performed under a standard mirror-projector
setup. In one set of RT-MRI scan, 6 slides were sequentially presented as “type bow know
five”, “type bone oh five”, “type toe node five”, “type bone know five”, “type tone oh
five”, and “don’t carry an oily rag like that”. The presentation of the stimuli was con-
trolled by Microsoft Powerpoint software, in which there was a 1 second pause between
the adjacent slides. The subject was instructed to lie in supine position and read/speak
the words at a normal speech rate.

5.3 Results

Figure 5.3 contains a plot of the unaliased FOV as a function of temporal resolution
(i.e., the number of TRs) when retrospectively selecting a temporal window. For 13-TR,
the 13-interleaf UDS supports a larger unaliased FOV than the golden-ratio view order.
When the number of TRs becomes a Fibonacci number (e.g. 2, 3, 5, 8, 13, 21, 34), there
is a change in the unaliased FOV for the golden-ratio method. Note the sudden increase
in the unaliased FOV from 17.1 cm to 27.6 cm when the number of TRs changes from
20 to 21. For 8-TR, the 13-interleaf UDS provides inconsistent unaliased FOVs, which
are 6.7 cm and 10 cm. The unaliased FOVs for the 13-interleaf UDS are smaller than
those for the golden-ratio when the number of TRs is between 8 and 12. The golden-ratio
view order provided a consistent unaliased FOV at any time point and at any temporal
window chosen.

Figure 5.4 contains the images reconstructed from the data acquired when the subject
was stationary. Note that the midsagittal slice of interest has a regional support of
roughly 38 cm and has significant intensity shading due to coil sensitivity. The images
reconstructed from the 13-interleaf UDS data have spatial aliasing artifacts in the regions
posterior to the pharyngeal wall from coil 1 and in the regions superior to the hard palate and velum from coil 2. The root SOS image in Fig. 5.4(c) contains little or no aliasing artifacts within the vocal tract region of interest (denoted by the dashed box). Although the unaliased FOV (i.e., 20 cm) from the 13-interleaf UDS is larger than that from the golden-ratio sampling for the choice of 13-TR, aliasing artifacts indicated by white arrows in Fig. 5.4(c) are more prominent than in Fig. 5.4(d). The spatial aliasing pattern can be understood by examining the point spread functions (PSF) for each sampling pattern. The PSF from the 13-TR UDS had a ratio of maximum sidelobe to mainlobe peak ($PSF_{\text{max-sl}}$) of 0.061 and exhibited single sidelobe ring with a radius of 20 cm. The PSF from the 13-TR golden-ratio sampling had a $PSF_{\text{max-sl}}$ of 0.044 and showed less coherent pattern with multiple sidelobe rings and lower sidelobe amplitude than the 13-TR UDS. Unlike the 13-interleaf UDS acquisition, a selection of long temporal window provides a large unaliased FOV in the golden-ratio acquisition. Note that aliasing artifacts are removed in the entire image from a selection of 34-TR window in Fig. 5.4(f).

Figure 5.5 contains image frames and time-varying intensity profiles from the conventional bit-reversed 13-interleaf UDS and golden-ratio methods when the subject produced the speech utterance “bow know”. Frames were updated at every TR. As seen in the undersampled 8-TR case of Fig. 5.5(c) and Fig. 5.5(d), the 13-interleaf UDS method produces aliasing artifacts that are periodic in time while the golden-ratio method produces less coherent aliasing in time. This periodicity in the aliasing is attributed to the inconsistent FOVs from the conventional bit-reversed 13-interleaf UDS as shown in Fig. 5.3. As seen from the 13-TR case of Fig. 5.5(c) and (d), the level of aliasing is higher for the golden-ratio result. Figure 5.5(d) shows that the intensity profile from the 8-TR result exhibits the sharpest transition of tongue tip motion (compare the yellow arrows in 8-TR, 13-TR, and 21-TR results).
The region-based temporal resolution selection method required the use of multiple temporal resolution videos reconstructed from iterative SENSE reconstructions, which substantially increased computation time. The generation of 160 dynamic frames of SENSE reconstructions from 8-TR, 13-TR, 21-TR, and 34-TR took approximately 45 min, 49 min, 53 min, and 60 min, respectively, with a 3.06 GHz CPU and 3.48 GB RAM. The blockwise temporal resolution selection algorithm took approximately 2 min.

Figure 5.6 shows a result of the blockwise temporal resolution selection from the SENSE reconstructed golden-ratio framesets. The synthesized frames in Fig. 5.6(c) exhibit good assignment of four distinct temporal resolution videos. Less tongue tip blurring is seen as indicated by the yellow hollow arrows in Fig. 5.6(c) and (d) than the 34-TR result (see the red hollow arrow in Fig. 5.6(e)). A better visualization of the velum opening is seen as indicated by the yellow solid arrows in Fig. 5.6(c) and (e) than the 8-TR result (see the red solid arrow in Fig. 5.6(d)).

Nasal speech imaging studies were performed using the golden-ratio acquisition and the results are shown in Fig. 5.7. Time intensity profiles are shown from the tongue tip and velum in the production of nasal consonants in three different syllable conditions at onset, coda, and juncture geminate. Note that time intensity profiles available from multiple temporal resolution videos facilitate a proper selection of temporal resolution on each articulator. Among the four temporal resolutions considered, tongue tip dynamics is depicted clearly with least temporal blurring from a selection of 48 ms temporal resolution. The depiction of the velum lowering is clearly defined with sufficient SNR from a selection of 126 ms temporal resolution.
5.4 Discussion

A new acquisition scheme that adopts a spiral golden-ratio view order has been demonstrated as a means to provide flexibility in retrospective selection of temporal resolution. The golden-ratio scheme has been compared with conventional bit-reversed 13-interleaf UDS acquisition, which is routinely used in our real-time speech MRI data collection at the University of Southern California. The spiral golden-ratio view order provides larger and consistent unaliased FOV when undersampling real-time data for higher temporal resolution. In addition, spiral interleaves are evenly distributed for any choices of the number of spiral interleaves at any time point, and hence a parallel imaging reduction factor can be flexibly chosen and applied to dynamic golden-ratio data. Auto-calibration with high resolution full FOV coil sensitivity maps is possible at any time point by utilizing fully-sampled temporal window data centered on that time point.

The proposed region-based temporal resolution selection method has limitations. The 8-TR (i.e., 49.3 ms temporal window) SENSE reconstructed frames served as a guide to select proper temporal resolution. However, they inherently lacked in temporal resolution and contained low image SNR. The velum and pharyngeal wall suffered from much lower SNR due to low coil sensitivity and potentially due to high parallel imaging $g$-factor. This can result in higher $TDE$ regardless of motion. In addition, I performed SENSE reconstruction from the data whose temporal window is smaller than 8-TR, but resulting SENSE images produced inadequate image quality with significantly low SNR or blurred air-tissue boundaries with the use of a large regularization parameter. Higher acceleration may be possible by the use of a highly sensitive upper airway receive coil with higher channel counts [39].

The motivation for using blockwise processing was based on the following. First, air-tissue boundaries within a block typically move with similar speed. Second, blockwise
processing helps to stabilize the calculation of $TDE$ in the presence of noise. There is a trade-off in selection of the block size. For example, the choice of larger block size can result in improper assignment of temporal resolution for a block within which motion is not at uniform speed. The choice of smaller block size causes $TDE$ to be more sensitive to noise.

The temporal resolution assignment procedure does not provide a strong link between the needed temporal resolution for an event and choice of retrospective temporal resolution. An additional navigator sequence may help to obtain the required temporal resolution information although it reduces scan efficiency. Ref. [108] by Tasko and McClean reports that the speed of the tongue tip during fluent speech was measured using electromagnetic articulometer (EMA) and was up to 200 mm/s. With the 3 mm spatial and 49 ms temporal resolution from 8-TR SENSE, it is anticipated that more than 3 pixels will experience temporal blurring around the tongue tip region of interest if the speed is 200 mm/s. Hence, higher spatio-temporal resolution frames will be necessary for estimating temporal bandwidth more reliably.

Recent real-time spiral speech MRI has been demonstrated with 80 $\sim$ 100 ms temporal resolution [79, 12, 19], but it lacks in temporal resolution compared to other speech imaging technologies such as EMA and ultrasound. Higher temporal resolution can be achieved by lowering spatial resolution or designing longer spiral readout with a fewer number of interleaves. Lower spatial resolution imaging may lose details of fine structures such as the epiglottis or lead to more difficulties in resolving the narrowing in the airway, e.g., the constriction between the alveolar ridge and tongue tip in certain sound productions such as the fricative /s/. Lengthening the spiral readout would cause images to be more susceptible to blurring or distortion in the air-tissue boundaries due to a large amount of resonance offset from air-tissue magnetic susceptibility. Correction of blurring
or distortion artifacts is challenging in real-time upper airway MRI because of difficulty in estimating accurate field map. Effective off-resonance correction from real-time golden-ratio view order data is an interesting area for investigation. Alternatively, real-time radial speech MRI with 55 ms temporal resolution has been demonstrated using a short-TR radial fast gradient echo sequence and parallel imaging reconstruction in combination with temporal filtering [111]. Improved temporal resolution imaging may be achieved by adopting an additional navigator sequence in the acquisition and a spatiotemporal model in the reconstruction [107, 63].

I have focused on an example of nasal sound production study in which knowledge of the timing of oral and velar coordination is important for modeling temporal changes in the constriction degrees of articulators under different syllable contexts [19]. This particular articulation involves a rapid tongue tip motion and relatively slow velar motion when producing nasal consonant /n/, and is well suited for investigating the importance of flexibility in temporal resolution selection. It is noted that velar movement highly varies depending on the subjects and involved speech tasks. Kuehn [60] reported that the velocity of the velar flesh point was measured using the cineradiographic equipment and reached up to 120 mm/s. EMA studies on dynamics of French nasal vowels such as /ã/ and /œ/ have been recently proposed [3]. Improved real-time MRI of dynamics of nasal vowels using the golden-ratio scheme will be of interest to linguistic community. The golden-ratio method can be applied to other articulatory timing studies which are investigated in the literature [72].

The acoustic noise generated by the MRI gradients during conventional bit-reversed 13-interleaf UDS imaging is temporally periodic. This can be exploited for high quality adaptive noise cancellation [13] and audio recordings during speech production in the magnet. One difficulty with golden-ratio imaging is that the MRI gradient noise is no
longer periodic. Advanced models are therefore required for acoustic noise cancellation during golden-ratio spiral imaging, and remain as future work.

5.5 Conclusions

I have demonstrated the application of a spiral golden-ratio temporal view order to imaging a midsagittal slice of the vocal tract during fluent speech. Simulation studies showed that the golden-ratio method provided larger and consistent unaliased FOV when retrospectively undersampling real-time data than the conventional bit-reversed 13-interleaf uniform density spiral. In nasal speech imaging studies, the proposed method provided an improved depiction of rapid tongue tip movement with less temporal blurring and velar lowering with higher SNR and potentially reduced aliasing artifacts. The region-based temporal resolution selection method synthesizes a single video from multiple temporal resolution videos available in the golden-ratio real-time data and potentially facilitates subsequent vocal tract shape analysis.
Figure 5.2: $k$-space trajectories for conventional bit-reversed 13-interleaf UDS and golden-ratio spiral view order when samples from (a,b) 8, (c,d) 13, and (e,f) 21 consecutive TRs are combined. Temporal view orders are marked with the numbers on each end of the spiral interleaves. Note that for the 8-TR case the spiral interleaves in (b) are distributed more uniformly than in (a). In (c) and (e), the angle spacing between spatially adjacent spiral interleaves is uniform with an angle of $360^{\circ}/13$. In (b,d,f), the angle spacing between spatially adjacent two spiral interleaves is not uniform but the angle increment between successive view numbers is constant with an angle of $360^{\circ} \cdot 2/(\sqrt{5}+1) \approx 222.4969^{\circ}$. 
Figure 5.3: Retrospective selection of temporal resolution: (a) Comparison of unaliased FOV between the golden-ratio view order and conventional bit-reversed 13-interleaf UDS sampling. (b) The enlargement of the region within the green rectangle in (a). The blue shaded region in (a, b) indicates that unaliased FOV varies in conventional bit-reversed 13-interleaf UDS when the number of TRs is less than 10. The black solid line illustrates a linear relationship between unaliased FOV and temporal resolution when UDS trajectories are designed under the constraints of the same spatial resolution and readout duration. Note that the unaliased FOV is fixed as 20 cm for a temporal window length (≥ 13-TR) for the conventional bit-reversed 13-interleaf UDS, but it increases with temporal window length for the golden-ratio view order.
Figure 5.4: Midsagittal images with a large reconstruction field-of-view (FOV) of 38 × 38 cm$^2$ reconstructed from the data acquired in static posture. (a-c) Conventional bit-reversed 13-interleaf UDS. (a) Image from coil 1, (b) image from coil 2, (c) root sum-of-squares (SOS) of the coil 1 and coil 2 images. The region within the dashed box in (c) is the vocal tract regions of interest (ROIs). For speed-up of the spiral acquisition, the FOV of the 13-interleaf UDS is typically chosen to be small such that aliasing artifacts are not observed in the vocal tract ROIs. (d-f) Root SOS of the coil 1 and coil 2 images reconstructed from data acquired using the spiral golden-ratio acquisition: reconstruction from (d) 13-TR, (e) 21-TR, and (f) 34-TR data. Spatial aliasing artifacts are completely removed in (f) because of larger FOV available from the golden-ratio method. The image SNR in (f) is higher than that in (d) and (e).
Figure 5.5: Gridding reconstructed dynamic frames and time intensity profiles from (a,c) bit-reversed 13-interleaf uniform density spiral data and (b,d) golden-ratio spiral view order data. A $40 \times 40$ matrix containing only the vocal tract region of interest is shown. Retrospective selection of temporal resolution is performed using 8-TR and 13-TR in (a), and 8-TR, 13-TR, and 21-TR in (b). Frame update rate was 1-TR = 6.164 ms. Two example frames (frame 1, 6) are shown (frame 1 is relatively stationary, frame 6 is captured during rapid tongue tip motion). Time intensity profiles from the image column indicated by the dashed lines are shown for (c) bit-reversed 13-interleaf UDS and (d) golden-ratio view order data.
Figure 5.6: Blockwise temporal resolution selection and synthesis of a single video from four temporal resolution videos. Five representative frames are shown which are captured when the subject produced the speech sound /bono/. (a) Normalized time difference energy ($TDE$) map. (b) Temporal resolution selection map [white: 48 ms (= 8-TR) temporal resolution, bright gray: 78 ms (= 13-TR) temporal resolution, dark gray: 126 ms (= 21-TR) temporal resolution, black: 204 ms (= 34-TR) temporal resolution]. (c) Synthesized temporal resolution frames based on the temporal resolution selection map in (b). (d) 48 ms (= 8-TR) temporal resolution frames. (e) 204 ms (= 34-TR) temporal resolution frames.
Figure 5.7: Variable temporal resolution selection from real-time data acquired using the golden-ratio acquisition scheme. (a) A midsagittal MR image from the subject. (b) Time intensity profiles for the tongue tip aperture (from the solid prescribed line in (a)) and the velum aperture (from the dashed prescribed line in (a)) for onset, coda, and juncture geminate in the articulations of “bow know”, “bone oh”, and “bone know”. In (b), the hollow arrows in the tongue tip and velum aperture profiles each indicate tongue tip closure onto the alveolar ridge and velar lowering, respectively. Temporal resolutions were (i) 204 ms, (ii) 126 ms, (iii) 78 ms, and (iv) 48 ms.
Chapter 6

Parallel Imaging with Novel 16-Channel Coil at 3 Tesla

6.1 Introduction

MRI is a powerful tool for the non-invasive assessment of upper airway anatomy and function. Upper airway MRI has been used to investigate the feasibility of clinical assessments in patients with obstructive sleep apnea [99, 5] and swallowing disorders [38]. In sleep apnea studies, rapid volumetric imaging is desirable for identifying the sites of the narrowing or occlusion of the airway and measuring its volume. In swallowing studies, high temporal resolution real-time MRI is desirable for the evaluation of swallowing disorders. This enables improved assessment of the dynamics of a bolus of food without introducing motion artifacts.

Rapid upper airway MRI is also valuable for basic research into human speech production. In such studies volumetric coverage is typically obtained during sustained sound production, using a two-dimensional (2D) multi-slice [75] or native 3D acquisition [53]. Tract dynamics (e.g., movement of the tongue and velum) is typically assessed using real-time acquisition of a single 2D slice (e.g., midsagittal) [79]. An objective and quantitative knowledge of the orchestration of articulatory activity that creates speech is a necessary element in understanding of human communication. A perennial challenge in
speech production research is the ability to examine 3D real-time changes in the shaping of the vocal tract. Spatio-temporal information about speech movements is critical not only to understanding and modeling the speech production process but also to a thorough understanding of speech acoustics; this in turn has significant implications for developing technological applications of machine synthesis and recognition of speech.

From a translational application perspective, understanding speech production deficits due to neurological damage or other disease etiology directly from articulation details (versus what is offered by speech acoustic recordings) is critical to assess and plan treatment. For example, people with certain neurological disease (e.g., apraxia) are known to exhibit speech timing patterns differing from neurologically unimpaired speakers (see articulatory data in Ref. [18]). The fact that both irregularities in the implementation of linguistic prosody and irregularities in articulatory timing patterns occur in neurologically impaired populations implies that investigation of the influence of prosodic structure on articulatory timing may illuminate the general question of how language-specific knowledge is related to motor control. Rapid upper airway MRI offers tools to look at clinical disorders in a new way.

A variety of pulse sequences and image reconstruction techniques have been applied to improve the spatio-temporal resolution of upper airway MRI. Spiral gradient echo (GRE) imaging techniques have been shown to be effective at capturing vocal tract dynamics during natural speech production [79, 107, 23]. Improved spatio-temporal resolution imaging has been demonstrated in the assessment of swallowing disorder using parallel imaging on a standard 12-channel head and neck array coil [14]. However, a design and use of dedicated multiple channel receive coil which is highly sensitive to the upper airway anatomy has not been reported in the literature. In this manuscript, a novel 3 Tesla 16-channel upper airway coil is described and its parallel imaging performance is
demonstrated. In volunteers, the signal-to-noise ratio (SNR) and parallel imaging $g$-factor values are compared over the upper airway regions of interest in the midsagittal slice. 3D parallel imaging of the upper airway is demonstrated to investigate the feasibility of high resolution accelerated upper airway MRI during sustained sound production. Real-time spiral acquisition and parallel imaging reconstruction is applied to capture vocal tract dynamics during natural sound production and is demonstrated with $2.0 \times 2.0 \text{ mm}^2$ spatial and 84 ms temporal resolution.

6.2 Materials and Methods

6.2.1 Coil Design and Construction

The desired upper airway field-of-view (FOV) extends in the superior-inferior (S-I) direction from above the hard palate down the throat to the level of the third cervical vertebra (C3) and in the anterior-posterior (A-P) direction, from several centimeters in front of the lips to the anterior surface of the cervical spine.

Our coil configuration consists of seven subgroups of elements: five longitudinal pairs and two longitudinal triplets (see Fig. 6.1a). The coils in each subgroup share a common conducting element along one edge (uniform color in Fig. 6.1a) but form two layers along the opposite side (dashed line). The array has an overall dimension of $43 \times 18 \text{ cm}^2$. Fourteen elements have a dimension of $8.2 \times 7.5 \text{ cm}^2$ whereas the third coil of each triplet has a dimension of $8.5 \times 7.5 \text{ cm}^2$. The top two bands of seven coil elements each are mounted on a thin plastic substrate that wraps around the lower face. The third coil of each triplet extends down the side of the neck and is encased in a flexible foam cover that allows it to curve under the chin for a tighter coupling to the neck region (Fig. 6.1c). The overlaps between neighboring coil elements form three different areas: A, B, and C
in Fig. 6.1a. The contour of the coil elements was adjusted so that area B cancels the mutual inductive coupling between diagonal elements. The sum of the areas A and B cancel the coupling between the circumferentially adjacent coils. The overlap C between longitudinally adjacent coil elements is larger than that needed to cancel mutual inductive coupling. For coils spaced along the z-direction with the conventional overlap [93], the SNR profile shows two peaks with a valley where the coils intersect. The overlapping area C was enlarged to eliminate the valley and give a flat SNR response along the z-direction. The resulting excess mutual inductance can be canceled with the decoupling capacitor Cd placed in the common conducting element. In all cases, the coupling between adjacent coil elements could be compensated far below the inherent inductive coupling between non-adjacent coil elements. The individual coil elements were tuned with the trimmer capacitors to give a minimal reactive component to the input impedance when the coils were loaded with a volunteer. The input match capacitor of each element (marked with an asterisk in Fig. 6.1a) is connected to an input match circuit (Fig. 6.1b). The pin diode MA4P1250NM provides active blocking during the transmit pulse. The signal diode 1N6640 allows passive blocking if the coil is not plugged into the scanner. The ratio of the unloaded $Q$ to loaded $Q$ for each coil element was deduced from measurements of the coil input impedances. The ratio was taken from the peak values of the resistive component of the input impedance of the unloaded and loaded condition.

Each coil element is connected to a 27 dB-gain, low-input-impedance preamplifier (Rich Spring Technologies, Arcadia, CA) located in a box on either side of the head. Each coaxial cable connecting a coil element and its preamp is fitted with an RF trap at the input of the preamp to eliminate multiple ground connections to different points on the subgroup’s common conductor and to prevent spurious eddy currents in the shields of the coax. The RF trap consists of three turns of 0.040” semi-rigid coax (Fig. 6.1d) inside
a cylindrical shell (Fig. 6.1e) made from double-sided Teflon circuit board. The outer surface of the Teflon is etched to form two capacitors between two outer conducting pads and the common inner pad. The outer pads are segmented into many smaller, binary weighted pads connected with small bridges. Each trap was tuned to 127.72 MHz by cutting a sufficient number of the bridges with a file while monitoring its insertion loss with a network analyzer. This construction is an economical way to produce numerous traps that can be fine-tuned without stocking a large number of different values of high voltage capacitors or using expensive variable capacitors.

The preamp boxes are mounted on a base plate that serves as the platform for an adjustable headrest padded with memory foam. The two preamp boxes serve as the base of a pivoting cross bar that holds an adjustable cantilever connected to the facemask coil array. The coil array can be held close to the face (Fig. 6.2) or folded up and back to permit patient entry and exit. The outputs of the two preamp boxes were connected to the scanner’s input sockets by way of two cables that were each fitted with a single RF trap. The latter trap consists of two concentric conducting cylinders that are connected at each end by several capacitors in parallel.

6.2.2 Experimental Methods

Experiments were performed on a Signa EXCITE HDx 3.0T scanner (GE Healthcare, Waukesha, WI) with gradients capable of 40 mT/m amplitude and 150 mT/m/ms slew rate. The receiver bandwidth was set to $\pm 125$ kHz (i.e., 4 $\mu$s sampling rate) for all studies. The noise correlation matrix as a measure of coil coupling was computed by 1) acquiring data with no RF excitation and 2) calculating noise correlation from every pair of the coil elements as described in Eq. (6) of Ref. [83].
For the SNR and g-factor evaluation, 3D volume of the upper airway was acquired in supine position using a 3DFT gradient echo sequence. Imaging parameters were: 3DFT gradient echo; \( k_y \) and \( k_z \) phase encoding along anterior-posterior and right-left directions, respectively; TE = 2.1 ms; TR = 3.9 ms; 1.88 \( \times \) 1.88 \( \times \) 2.50 mm\(^3\) spatial resolution; 24 \( \times \) 24 \( \times \) 18 cm\(^3\) FOV; pixel dimension = 128 \( \times \) 128 \( \times \) 72; scan time = 30 seconds. The subjects held their mouth closed without swallowing during each 30 seconds scan. I considered a retrospective 1D regular undersampling on a midsagittal slice and 2D regular undersampling on axial and coronal slices. The undersampling was along A-P direction. I performed the SNR evaluation of the 16-channel coil by comparing it with widely available single-channel birdcage transmit/receive coil and 8-channel neurovascular (NV) receive coil and also by comparing the performance of the 16-channel coil among three subjects. The single-channel birdcage coil is cylindrical-shaped with the inner diameter of 28 cm and the length of 38 cm, and usually covers the brain. It is high-pass and quadrature driven. The 8-channel NV coil consists of a 4-element head array and 4-element neck array: the head array has an inner diameter of 22 cm and length of 33 cm. The neck array has a dimension of 34 \( \times \) 17 cm\(^2\) in the anterior neck region and a dimension of 44 \( \times \) 22 cm\(^2\) in the posterior neck region. Eight ROIs (see Fig. 6.3a) were selected manually on a midsagittal scan plane, and the average SNRs were evaluated on each ROI separately. For the data acquired from the 8-channel NV and 16-channel UA coils, the average SNRs were calculated from images in SNR units reconstructed from the \( B_1 \)-weighted combining method described in Eq. (6) of Ref. [48]. Low resolution coil sensitivity maps were calculated from 32 central \( k \)-space lines by dividing the image at each coil element by the root sum-of-squares (RSS) image. I compared image quality between one large male and one small female on a midsagittal scan plane after performing parallel imaging.
reconstruction [90] for reduction factors of $R = 2, 3, 4, \text{ and } 5$. Parallel imaging $g$-factors were calculated on a pixel-by-pixel basis as described in Ref. [90].

High resolution 3D imaging of the vocal tract was demonstrated using a 3DFT gradient echo sequence. Imaging parameters were: 8 cm midsagittal slab excitation, flip angle = $5^\circ$; $k_y$ and $k_z$ phase encoding along A-P and R-L directions, respectively; TE = 2.1 ms; TR = 4.2 ms; $1.25 \times 1.25 \times 1.25$ mm$^3$ spatial resolution; $20 \times 20 \times 10$ cm$^3$ FOV; pixel dimension = $160 \times 160 \times 80$; scan time = 54 seconds. During each 54 seconds scan, the subjects held their mouth open without swallowing. For the evaluation of parallel imaging performance, I considered a retrospective 2D regular undersampling on the $k_y$ and $k_z$ encodes after taking the inverse Fourier transform along the readout direction (i.e., S-I). SENSE reconstructions were performed at each axial slice, in which coil calibration was performed using the central $32 \times 32$ $(k_y, k_z)$ encodes.

Real-time upper airway MRI during natural speech sound production was demonstrated on the 16-channel UA coil using a 2D spiral gradient echo sequence. Imaging parameters were: TE = 1.4 ms, TR = 4.0 ms, readout duration = 1.2 ms, slice thickness = 5 mm, flip angle = $10^\circ$, FOV = $30 \times 30$ cm$^2$, spatial resolution = $2.0 \times 2.0$ mm$^2$, image matrix size = $150 \times 150$. The subject was instructed to repeat “go pee shop okay bow know” during the scan. The angular increment of spiral interleaves in $k$-space was based on the golden-ratio temporal view order [114], which supports a retrospective selection of temporal resolution at any arbitrary time point.

From the 2D spiral data set, five different acceleration factors of 1.0, 1.6, 2.6, 4.2, and 6.8 were considered in frame reconstruction. Their corresponding numbers of spiral interleaves were 89, 55, 34, 21, and 13 which led to temporal resolution of 356 ms, 220 ms, 136 ms, 84 ms, and 52 ms, respectively. Coil sensitivity calibration was performed by utilizing multi-coil data corresponding to fully-sampled temporal window and
by dividing each individual coil image by the RSS image. The coil sensitivity maps and noise covariance matrix were applied to image reconstruction. Image reconstruction from the undersampled multi-coil spiral data was based on an iterative conjugate gradient algorithm after taking the noise decorrelation steps as described in Ref. [89]. Iterations were stopped at the 15th iterate for the acceleration factors of 1.0, 1.6, and 2.6 and at the 20th iterate for the acceleration factors of 4.2, and 6.8 based on Ref. [91]. After the reconstruction of the frames, temporal median filtering with a filter length of 5 was applied pixel-by-pixel to successive frames in order to eliminate residual aliasing artifacts in reconstructed frames [111].

6.3 Results

The majority of the $Q_{\text{unloaded}}/Q_{\text{loaded}}$ ratios ranged from 4 to 6 with a few as high as 8 depending on how closely the array was placed on the subject. Figure 6.4 contains the locations of the coil indices and the magnitude plot of a noise correlation matrix from a healthy volunteer. The off-diagonal elements in the noise correlation matrix ranged from 0.0029 to 0.7039 with a mean of 0.1738. The maximum noise correlation occurred in a pair of longitudinally adjacent coil elements which has a relatively larger degree of overlap than a pair of horizontally adjacent coil elements.

Figure 6.5 shows midsagittal, axial, and coronal slices of the 3D upper airway from a healthy subject. From the three orthogonal slice images, it is observed that each coil element has its unique coil sensitivity pattern. For example, coil 8 shows highly localized sensitivities in midsagittal slice but shows relatively uniform sensitivity in coronal slice. Coil 12 shows uniform sensitivity in midsagittal slice but shows localized sensitivities in axial and coronal slices.
Table 6.1: Average SNR improvement. Average SNR was measured in a single subject (33 year old male) using the 16-channel UA coil, single-channel birdcage coil, and 8-channel NV coil. Eight regions of interest (see Fig. 6.3a) were identified in 2D midsagittal images with $1.88 \times 1.88 \times 2.50$ mm$^3$ spatial resolution, obtained without the use of parallel imaging. The 16-channel UA coil provided improved intrinsic SNR in all regions of interest. UA: upper airway, NV: neurovascular.

<table>
<thead>
<tr>
<th>ROI</th>
<th>SNR$<em>{UA}$/SNR$</em>{Birdcage}$</th>
<th>SNR$<em>{UA}$/SNR$</em>{NV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Upper lip</td>
<td>11.6</td>
<td>5.2</td>
</tr>
<tr>
<td>2 Lower lip</td>
<td>19.0</td>
<td>8.8</td>
</tr>
<tr>
<td>3 Front tongue</td>
<td>10.0</td>
<td>5.5</td>
</tr>
<tr>
<td>4 Mid tongue</td>
<td>5.4</td>
<td>3.5</td>
</tr>
<tr>
<td>5 Back tongue</td>
<td>4.1</td>
<td>2.6</td>
</tr>
<tr>
<td>6 Palate</td>
<td>3.2</td>
<td>1.9</td>
</tr>
<tr>
<td>7 Velum</td>
<td>2.9</td>
<td>2.0</td>
</tr>
<tr>
<td>8 Pharyngeal wall</td>
<td>2.0</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Table 6.1 compares SNR for 8 different ROIs with the three different coils on one subject. The 16-channel coil produced highest SNR in every region. Particularly, the average SNR in the lower lip provided a 8.8-fold and 19-fold increase compared to the 8-channel NV coil and birdcage coil, respectively. The 16-channel coil produced the smallest SNR improvement in the pharyngeal wall, 1.5 over the 8-channel coil and 2.0 over the birdcage coil.

Table 6.2 compares SNR ratio for 8 different ROIs on three subjects with the 16-channel UA coil. The maximum SNR ratio from Subject 2 was 3.1, which was far lower than 7.1 and 8.0 from Subject 1 and Subject 3, respectively. This results from the fact that Subject 2 has the relatively smaller head size and the coil array is placed close to the orofacial part of the head.

Figure 6.6 contains SENSE reconstructed images and g-factor maps for R = 2 to 5 from one large male (Subject 1) and one small female (Subject 2) using the 16-channel coil. The pharyngeal wall of Subject 1 in Fig. 6.6a is seen relatively darker than that
Table 6.2: Comparison of average SNR drop-off in three volunteers. Average SNR was measured in three volunteers using the 16-channel coil. Eight regions of interest were identified in 2D midsagittal images with $1.88 \times 1.88 \times 2.50 \text{mm}^3$ spatial resolution, obtained without the use of parallel imaging. Note the relatively less steep SNR drop-off from the small female subject.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age/Sex</th>
<th>Size (A-P/S-I)*</th>
<th>ROI SNR</th>
<th>ROI $\text{SNR}/\text{SNR}_{pw}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Upper lip</td>
<td>7.1</td>
<td>3.1</td>
<td>6.4</td>
</tr>
<tr>
<td>2</td>
<td>Lower lip</td>
<td>7.0</td>
<td>3.0</td>
<td>8.0</td>
</tr>
<tr>
<td>3</td>
<td>Front tongue</td>
<td>3.9</td>
<td>1.9</td>
<td>3.2</td>
</tr>
<tr>
<td>4</td>
<td>Mid tongue</td>
<td>2.6</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>5</td>
<td>Back tongue</td>
<td>2.2</td>
<td>1.4</td>
<td>1.6</td>
</tr>
<tr>
<td>6</td>
<td>Palate</td>
<td>3.8</td>
<td>1.4</td>
<td>3.9</td>
</tr>
<tr>
<td>7</td>
<td>Velum</td>
<td>2.0</td>
<td>1.1</td>
<td>1.4</td>
</tr>
<tr>
<td>8</td>
<td>Pharyngeal wall</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

*: A-P size was measured as the distance from the lower lip to the pharyngeal wall, and S-I size was measured as the distance from the hard palate to the glottis.

of Subject 2 in Fig. 6.6c because of the larger head size of Subject 1 and the geometry of the close-fitting coil. In both subjects the SENSE images up to a reduction factor 4 exhibited good depiction of air-tissue boundaries in the upper airway ROIs as indicated by the red contours in Fig. 6.6a and c). Note that the spiky artifacts resulting from noise amplification due to sensitivity matrix inversion are observed in the $R = 3$, 4, and 5 SENSE images in both subjects. The mean $g$-factors in the upper airway ROIs were comparable in both subjects for all reduction factors considered.

Figures 6.7a and b show $g$-factor plots for the eight ROIs and entire image as a function of reduction factor for 2DFT midsagittal imaging when using the 8-channel neurovascular coil and 16-channel upper airway coil, respectively. The direction of the phase encode was
along A-P. The \( g \)-factor values were substantially lower for the 16-channel coil compared to the 8-channel coil. For reduction factor 4, the average \( g \)-factor was 40% lower and the difference in \( g \)-factor values among the ROIs were 30% lower for the 16-channel coil. As indicated by the shaded region, \( g \)-factor values are in the range of 3-12 among the ROIs for the 8-channel coil while they are in the range of 2-4 for the 16-channel coil. The dramatic improvement achieved from the 16-channel coil may be due to more coil elements having distinct coil sensitivities in the mid-sagittal ROIs (see Figure 6.5a). Since only 4-element of the 8-channel NV coil is close to the ROIs, it is expected that the \( g \)-factor values dramatically increase for reduction factors greater than 4. As seen from Fig. 6.7a, \( g \)-factor values for the pharyngeal wall ROI doubled when reduction factor changed from 5 to 6.

Figures 6.8a and b contain midsagittal images reconstructed via SENSE (R=4) from the Cartesian sampled data. The image from the 8-channel neurovascular coil shows severe noise amplification in most upper airway ROIs (see the arrow in Fig. 6.8a). The image from the 16-channel coil exhibits improved image quality with much higher SNR (see Fig. 6.8b). Figures 6.8c-f show axial and coronal slice reconstructions using SENSE with a rate 6 (\( R_y = 3 \) along the vertical axis, \( R_z = 2 \) along the horizontal axis). The images from the 8-channel coil exhibit significant noise amplification (arrows in Figures 6.8c and e). The images from the 16-channel coil exhibit substantially reduced noise amplification because of lower \( g \)-factor values and higher intrinsic SNR available from the 16-channel upper airway coil. The average \( g \)-factors for the axial slices in Fig. 6.8c and d were 4.64 and 2.67, respectively. The average \( g \)-factors for the coronal slices in Fig. 6.8e and f were 2.23 and 1.52, respectively. Note that the coronal slices exhibit lower \( g \)-factors than the axial slices for both the 8-channel and 16-channel coils.
Figure 6.9 compares R = 1 and R = 6 SENSE images on one reformatted midsagittal, four reformatted coronal, and four axial slices. Slice prescriptions for the coronal and axial slices are indicated by the solid lines in Fig. 6.9a. These slices were chosen because they contain information on the air-tissue boundaries necessary for the extraction of the 3D vocal tract shape [75]. Although the R = 6 SENSE images are noisier than the R = 1 images, they preserve the air-tissue boundary features that are relevant to the measurement of the vocal tract area function.

Figure 6.10 contains spiral SENSE reconstructed images of the midsagittal slice for five different acceleration factors of 1.0, 1.6, 2.6, and 4.2, and 6.8. Parallel imaging with reduction factors of up to 4.2 produced the images that retained sufficient image quality for speech imaging applications. In other words, the boundaries of the tongue, velum, and other articulators were clearly depicted. The R = 6.8 image (see Fig. 6.10a) exhibited poor depiction of these boundaries because of lower image SNR. The constriction between the tongue blade and the alveolar ridge is seen during the utterance of /i/ (the solid arrows in Fig. 6.10b). Also, the constriction between the tongue tip and the alveolar ridge is seen during the utterance of /ʃ/ (the hollow arrows in Fig. 6.10b).

6.4 Discussion

The 16-channel coil produced acceptable image quality in static upper airway imaging with conventional Cartesian SENSE from a rate-4 in 1D undersampling and a rate-6 in 2D undersampling. This opens up new opportunities for upper airway MRI with improved spatial resolution. For example, 1.25 mm isotropic resolution 3D imaging of the vocal tract is achievable with a rate-6 conventional SENSE during 9 seconds of the scan time. This is already a substantial improvement over the recently reported accelerated 3D sustained vocal tract imaging that utilizes the single-channel birdcage coil and achieves
a 1.5 × 1.5 × 2.0 mm³ resolution and 7 seconds scan time using compressed sensing [53]. Higher acceleration than rate-6 may be achievable by using advanced image reconstruction methods such as the combined use of compressed sensing and parallel imaging [70].

The localized sensitivity with the 16-channel coil enables spiral imaging with a smaller FOV, and leading to better spatio-temporal resolution compared to what is possible with the 8-channel NV coil. As reduction factor increases to 6.8 in the spiral parallel imaging, severe SNR degradation occurred from the iterative SENSE reconstruction. Iterative SENSE with an explicit regularization (e.g., total variation regularization) may be beneficial since it mitigates noise amplification and preserves the edges of the air-tissue interface. It is worth noting that the use of SENSE-based non-Cartesian parallel imaging requires: 1) Full FOV image acquisition (including the brain that is not our region-of-interest) for estimating coil sensitivity. 2) Extrapolation to estimate coil sensitivity in image regions occupied by air. Data inconsistency due to motion of articulators can also lead to inaccuracy in data fitting term and degrade image quality. As an alternative, \(k\)-space based parallel imaging approaches, such as non-Cartesian GRAPPA [40], do not require the explicit estimation of coil sensitivity maps and may be more robust in practice.

Compared to the 8-channel neurovascular coil array which has large-sized coil elements, coil calibration may be more difficult with the 16-channel coil which has a larger number of smaller diameter elements. As such, coil sensitivities are more highly localized and rapidly moving anatomy (e.g., lips, tongue tip) that is close to the coil elements has a strong signal contribution.

In coil design there is also much room for improvement. For example, future designs can benefit from: rearrangement of the coil elements to reduce the \(g\)-factor in its worst places, better coverage of the lower airway with additional coil elements, better coverage of the pharyngeal wall with a few large coils around the back of the head, access for
a respirator or mask covering the face and nose for experiments related to airway collapse [22], and rearrangement of the coils to allow for easier opening of the lower jaw. It is also worth of investigating the benefit of using higher channel counts (32 or more) for even higher acceleration.

6.5 Conclusions

This novel 16-channel coil array provides rate-4 acceleration to 2DFT midsagittal imaging, rate-6 acceleration to 3DFT imaging, and rate-4.2 acceleration to 2D spiral midsagittal imaging of the upper airway. This will lead to either higher spatial resolution imaging or reduced scan time in capturing a 3D vocal tract shape during sustained sound production. This coil has the potential to allow for improved spatio-temporal resolution in dynamic upper airway MRI studies of swallowing and normal/disordered speech.
Figure 6.1: 16-channel upper airway receive coil array. (a): Coil layout. Individual coil elements are overlapped with their neighboring coil elements to minimize mutual inductive coupling. The two lower coil elements, which correspond to the arrows in (c), extend down the side of the neck region to improve SNR of the lower airway. (b): Input circuit diagram. (c): Photograph of the 16-channel coil. Each coil element is connected to a preamplifier. A pivoting cross bar is supported by the two preamplifier boxes and holds an adjustable cantilever connected to the coil array. The coil array can be held close to the face or folded up and back to permit patient entry and exit. The outputs of the two preamp boxes are connected to the scanner’s input sockets by way of two cables that are each fitted with a single RF trap. (d): RF trap coaxial inductor. (e): RF trap with etched capacitors [39].
Figure 6.2: The 16-channel upper airway coil array on a volunteer. (Left) Side view. (Right) Top view.
Figure 6.3: Illustration of the eight upper airway regions of interest (ROIs) used in the evaluation of SNR and $g$-factor. (a) The ROIs are identified by red contours and are superimposed onto the midsagittal anatomic image acquired using the 16-channel upper airway coil: 1-upper lip, 2-lower lip, 3-front tongue, 4-mid tongue, 5-back tongue, 6-palate, 7-velum, 8-pharyngeal wall, and 9-epiglottis. (b) and (c) shows images acquired using the birdcage coil and 8-channel neurovascular coil, respectively. Note that the image acquired from the 16-channel coil has the localized sensitivity on the upper airway regions of interest.

Figure 6.4: (a) Channel indices labeled on the coil layout. (b) Noise correlation matrix from one representative volunteer.
Figure 6.5: Midsagittal, axial, and coronal images at each coil element of the 16-channel upper airway coil. 3D data set was acquired using 3DFT GRE sequence. The images shown are ordered based on the coil layout in Fig. 6.4a. (a) Midsagittal magnitude image for each coil element. In the upper left image, the dotted and solid lines indicate axial and coronal slice prescriptions, respectively. From the prescribed lines in (a), shown are (b) axial and (c) coronal images from each coil element. The numbers shown on each image indicate channel indices.
Figure 6.6: 2D midsagittal image reconstruction using 1D SENSE. The direction of undersampling was A-P. SENSE reconstructed images with a spatial resolution of $1.88 \times 1.88 \times 2.50$ mm$^3$ are compared with respect to reduction factors from 2 to 5 for the (a) Subject 1 and (c) Subject 2. Corresponding $g$-factor maps are shown for the (b) Subject 1 and (d) Subject 2. The mean and maximum $g$-factors shown were computed from the ROIs indicated by the red contours of the $R = 1$ images in (a) and (c).
Figure 6.7: Plots of $g$-factor values for 8 different ROIs as a function of reduction factor for 2DFT midsagittal parallel imaging. The axis of undersampling was along A-P. (a) 8-channel neurovascular coil. (b) 16-channel upper airway coil. Note that the scales in the vertical axes are different for (a) and (b) and substantial reduction in $g$-factors is achieved with the 16-channel coil.
Figure 6.8: Comparison of midsagittal, axial and coronal slice reconstruction using parallel imaging. (Top): 8-channel neurovascular coil. (Bottom): 16-channel upper airway coil. (a, b): Midsagittal SENSE reconstructed images (spatial resolution: $1.88 \times 1.88 \times 2.5 \text{ mm}^3$) with a rate 4. The direction of phase encode was horizontal. (c, d): Axial SENSE reconstructed images (spatial resolution: $1.88 \times 1.88 \times 2.5 \text{ mm}^3$) with a rate 6 ($R = 3$ along the vertical axis and $R = 2$ along the horizontal axis). (e, f): Coronal SENSE reconstructed images (spatial resolution: $1.88 \times 1.88 \times 2.5 \text{ mm}^3$) with a rate 6 ($R = 3$ along the vertical axis and $R = 2$ along the horizontal axis). The 8-channel coil produced unacceptable image quality with low image SNR and substantial noise amplification (see the arrows in (a, c, e)). The 16-channel coil produced significantly improved image SNR with smaller noise amplification than the 8-channel coil.
Figure 6.9: 3D image reconstruction using 2D SENSE. The directions of undersampling were A-P and R-L. Fully sampled 1.25 mm isotropic resolution 3D data set was acquired in open-mouthed position during 54 seconds scan. Retrospective undersampling was performed with a regular undersampling factor of 6, where two-fold acceleration was along R-L direction and three-fold acceleration was along A-P direction. Reformatted midsagittal images are shown for (a) SENSE $R = 1$, and (c) SENSE $R = 6$. Eight representative coronal and axial images are shown for (b) SENSE $R = 1$ and (d) SENSE $R = 6$. Air-tissue boundaries in the vocal tract are well defined in all the SENSE $R=6$ images.
Figure 6.10: 2D image reconstruction of the dynamics of vocal tract shaping using spiral SENSE. (a) Comparison of spiral SENSE images from a single representative frame. For the evaluation of SENSE reconstruction performance under different reduction factors, the frame was chosen from the data acquired when the subject (Subject 1) was stationary. Spatial aliasing was substantially reduced for all the SENSE reconstructed images. The R = 6.8 image exhibited relatively lower image SNR in the posterior regions of the upper airway (e.g., velum and pharyngeal wall). (b) Dynamic frames that represent the subject’s utterance of “ee shop”. The R = 4.2 SENSE reconstruction was used to generate the image frames with its frame update rate of 44 ms. The frame sequence is ordered from the top left to the bottom right.
Chapter 7

Summary and Future Work

In summary, I have presented four imaging methodologies that improve image quality over conventional techniques in rapid MRI of the beating heart and upper airway:

- **Automatic correction of echo-planar imaging (EPI) ghosting artifacts in real-time MRI**

  I proposed “double-alternating” echo-planar imaging and sensitivity encoding reconstruction to effectively remove ghosting artifacts and maintain its temporal resolution [57]. I demonstrated its effectiveness at reducing ghosting artifacts resulting from echo-misalignment in oblique/double-oblique scan planes by comparing it with traditional 1D phase correction method.

- **Accelerated 3D upper airway MRI using compressed sensing: Application to sustained speech imaging**

  I demonstrated a first application of compressed sensing to sustained speech imaging [53]. Compared to traditional 3D MRI of the vocal tract, the proposed method acquires data in true 3D Fourier space and its acquisition does not involve any repetition of the target utterance. Hence, it is not prone to the image mis-registration
effect and also results in a more time-efficient data collection procedure. Pseudo-random Fourier encoding and compressed sensing reconstruction is exploited to further improve spatial resolution within sustained sound duration. Its combination with parallel imaging significantly improves the image quality at high acceleration rate [52].

- **Improved real-time speech MRI using a golden-ratio spiral view order**

  I proposed a golden-ratio spiral view order and demonstrated its first application to real-time speech MRI [55, 54]. The golden-ratio spiral view order provides more flexible retrospective selection of temporal resolution. This is well suited to real-time speech MRI in which speech rate and motion of articulators highly vary depending on speech task or speaking style. I demonstrated its effectiveness at capturing rapid movement of the tongue tip and relatively slow movement of the velum by comparing it to traditional bit-reversed spiral view order.

- **Improved upper airway MRI using a 16-channel coil at 3 Tesla**

  I described a novel 16-channel upper airway 3 Tesla receive coil that was originally designed by Hayes et al. [39]. I demonstrated its SNR and parallel imaging $g$-factor performance which is far superior to widely available single channel birdcage and 8-channel neurovascular array coils [51]. I demonstrated a 1.25 mm isotropic resolution 3D imaging of the vocal tract with a 6-fold parallel imaging acceleration. This coil also has the potential to improve SNR and spatio-temporal resolution in real-time speech MRI at 3 Tesla.

  Finally, I wish to present some of the topics that I have been working on but have not completed. These are worth of more investigation as future work.
• **High resolution vocal tract imaging during sustained speech production**

I have performed a preliminary study on collecting 3D MRI data acquired during sustained sound production of English vowels and fricatives. Improvement in imaging work was attributed to a state-of-the-art technology equipped with the 16-channel upper airway receive coil, pseudo-random Poisson disk undersampling, and phase-constrained multi-coil compressed sensing reconstruction. I obtained high quality 3D vocal tract shape data sets with 1.25 mm isotropic resolution and 6 seconds scan time (i.e., 8x acceleration). Future work includes: 1) Recruit 3 American English male and 3 American English female who have mid-west or Californian accents, 2) collect multiple sets of sustained sound production of American English vowels and fricatives from the subjects, and 3) extract vocal tract shape and report any differences in the area functions and 3D geometries of the vocal tract shapes across the subjects.

• **Multi-coil phase constrained compressed sensing for accelerated 3D upper airway MRI**

I have proposed a multi-coil phase constrained compressed sensing reconstruction algorithm that involves multiple iterations to improve the estimation of the phase map and thus ultimately improves image quality in the depiction of air-tissue boundaries. I will address the issue on convergence of the iterative algorithm and seek to build up a theoretical framework on the method.

• **Real-time speech MRI at 3 Tesla**

SPAN data collection has been performed dominantly at 1.5 Tesla scanner sites. The reconstructed frames have low SNR, which leads to a poor depiction of the air-tissue boundaries. Speech imaging at 3 Tesla is an interesting research area,
and some imaging work has been reported in the literature [111, 105]. I have developed a real-time continuous spiral imaging sequence at 3 Tesla, but this does not provide interactive monitoring of the vocal tract like what is normally done on a custom real-time imaging software (i.e., RTHawk). Adaptation of RTHawk into this new 3 Tesla scanner site is considered as a long-term future work. Optimizing spiral imaging pulse sequences suited for speech imaging at 3 Tesla may require substantial amounts of experimental work.

- **Obstructive sleep apnea imaging under a less noisy MRI scan**

So far, technological advances have been made towards the improvement in speech imaging. For a side project related to obstructive sleep apnea imaging, I have developed a pulse sequence that is based on the design of the excitation and readout gradients with a far lower slew rate and thus significantly reduces acoustic noise from the MRI gradients. The pulse sequence had an increased TR such that the idle time exceeded the duration of the readout gradients. Scanning was able to continuously proceed without any automatic pause for more than 30 minutes. The amplitudes for the excitation and readout gradients were able to be adjusted from 0 to 100% in real-time. This setting may help the subjects to naturally fall asleep without any use of sedatives and provide a better experimental setup for investigating the airway collapse in sleep apnea MRI.

- **Time interleaved imaging of arbitrary scan planes**

Conventional real-time speech MRI typically acquires the dynamics of vocal tract shape from a single slice (typically midsagittal slice). This provides insights into the dynamics of all articulators, but does not allow for visualizing important features in vocal tract shaping such as grooving/doming of the tongue, asymmetries in tongue
shaping, and lateral shaping of the pharyngeal airway. I have developed a real-time speech imaging technique that provides acquisition of multiple arbitrary scan planes in time-interleaved fashion. Additional benefit is that partial saturation indicates where the slices are located in relation to each other. I have applied this technique to speech imaging on Mandarin Chinese fricatives and English voiced/voiceless fricatives. Moreover, time-interleaved imaging of multiple axial slices of the brain and one midsagittal slice of the upper airway is another application area. This enables monitoring speech or swallowing during fMRI study [106].
Bibliography


