

HST GRADUATE COMMITTEE
c/o Academic Office, E25-518

Dear HST Graduate Committee Chair,

Daryush Mehta presented his PhD thesis proposal on November 25, 2008 in the RLE Haus Room (36-428) at MIT to the following committee:

Chair: **Joseph S. Perkell, PhD**, MIT, whose areas of expertise are the sensory-motor control of speech production and voice/speech production and acoustics

Co-Supervisor: **Robert E. Hillman, PhD**, MGH, whose area of expertise is in the clinical evaluation of human voice production.

Co-Supervisor: **Thomas F. Quatieri, ScD**, MIT Lincoln Laboratory, whose area of expertise is in biologically-inspired speech signal processing.

Reader: **Dimitar D. Deliyski, PhD**, University of South Carolina, whose areas of expertise are in the acoustic analysis of voice and the development of laryngeal high-speed videoendoscopy.

The proposal, which is enclosed, was favorably received by the Committee, and we approved the scientific content and proposed work as being suitable for a PhD thesis.

All of the above members of the Committee have agreed to serve on the Thesis Committee.

Sincerely,

Joseph S. Perkell
Thesis Committee Chair

Robert E. Hillman, PhD
Thesis Co-Supervisor

Thomas F. Quatieri, ScD
Thesis Co-Supervisor

Enc: Thesis Proposal
Supervisor Agreements
Reader Agreement

PhD Thesis Committee Members

1. Joseph S. Perkell, PhD (Chair)
 - a. **Title:** Senior Research Scientist, Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology
 - b. **Major Discipline:** Sensory-motor control of speech production
 - c. **Justification:** Dr. Perkell fills the role of Chair as a non-supervisor and senior researcher at MIT. Dr. Perkell offers a wide knowledge range from voice and speech production to speech acoustics and motor involvement in pathological speakers.
2. Robert E. Hillman, PhD (Co-Supervisor)
 - a. **Title:** Co-Director/Research Director, Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital; Associate Professor, Harvard Medical School; Faculty of Harvard-MIT Program in Speech and Hearing Bioscience and Technology
 - b. **Major Discipline:** Voice function assessment
 - c. **Justification:** Dr. Hillman is co-adviser and supports the clinical aspects of the thesis project. The proposed research calls for data collection in the voice clinic and assessment of the voice production mechanisms and acoustic characteristics of human subjects. Subjects will be selected and evaluated under Dr. Hillman's supervision.
3. Thomas F. Quatieri, ScD (Co-Supervisor)
 - a. **Title:** Senior Member of Technical Staff, MIT Lincoln Laboratory; Faculty of Harvard-MIT Program in Speech and Hearing Bioscience and Technology
 - b. **Major Discipline:** Speech signal processing
 - c. **Justification:** Dr. Quatieri is co-adviser and supports the signal processing aspects of the proposed research. Dr. Quatieri's work includes the speech signal processing using multimodal analysis, and this work especially relates to the proposed research on characterizing vocal fold vibratory asymmetries from multimodal sensor measurements.
4. Dimitar D. Deliyski, PhD (Reader)
 - a. **Title:** Associate Professor, Department of Communication Sciences and Disorders, University of South Carolina; Director, USC Voice and Speech Laboratory
 - b. **Major Discipline:** Voice acoustics and laryngeal high-speed videoendoscopy
 - c. **Justification:** Dr. Deliyski's areas of expertise are in the acoustic analysis of voice and laryngeal high-speed videoendoscopy. Dr. Deliyski is a world-leader in the development of high-speed video camera technology for clinical voice assessment.

Massachusetts Institute of Technology
Harvard-MIT Division of Health Sciences and Technology
Speech and Hearing Bioscience and Technology Program

Proposal for Thesis Research in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Title: Investigating the impact of *in vivo* human vocal fold vibratory asymmetries: Co-variations among measures from laryngeal high-speed videoendoscopy, acoustic voice analysis, and auditory-perceptual voice assessment of sustained vowel phonation

Submitted by: Daryush Mehta
70 Pacific Street, Apt 516
Cambridge, MA 02139

Signature: _____

SHBT Track: Signal Processing

Date of Submission: November 25, 2008
Expected Date of Completion: July 2009

Thesis Co-Supervisors: Robert E. Hillman, PhD
Thomas F. Quatieri, ScD

Location of Research: Center for Laryngeal Surgery and Voice Rehabilitation,
Massachusetts General Hospital

Abstract:

Voice specialists make critical diagnostic, medical, therapeutic, and surgical decisions by coupling visual observations of vocal fold tissue motion with auditory-perceptual assessments of voice quality. The details of the relationship between vocal fold tissue motion and the acoustic voice signal are not fully understood, and there is recent evidence that the acoustic impact of visual judgments of vibratory asymmetry may be overestimated during clinical voice assessment. A series of three descriptive studies is proposed to systematically describe the co-variations among measures of vocal fold vibratory asymmetries and visual-perceptual judgments, acoustic voice properties, and auditory-perceptual ratings.

First, recent findings describing co-variations between subjective visual judgments and basic objective measures of vocal fold vibratory asymmetry in subjects with and without vocal pathologies will be validated with automated algorithms. After replicating these results, image-based measures will be further refined to consider additional dimensions in the left-right and anterior-posterior planes of the images.

Second, it is proposed to apply the developed objective measures of asymmetry to voice data from a new subject population with vocal pathologies that will be evaluated using a state-of-the-art system for laryngeal high-speed videoendoscopy. The new system will enable the validation of hypothesized relationships between vocal fold vibratory asymmetry measures and objective acoustic voice measures at unprecedented temporal resolution. Preliminary work has revealed mild co-variations between average values of vocal fold vibratory asymmetry and traditional acoustic perturbation measures, and new acoustic correlates of vocal fold vibratory asymmetries will be explored using knowledge of their effects on voice production.

Third, an initial study is proposed to characterize the influences of vocal fold vibratory asymmetry on the auditory perception of voice quality. This study more directly addresses the clinical reality that voices are assessed by relating vocal fold tissue vibratory patterns to the voice quality of a patient during a standard examination.

Table of Contents

1	Introduction.....	7
1.1	Motivation and goals	7
1.2	Thesis proposal structure.....	8
2	Specific Aims.....	8
2.1	Aim 1: Investigate co-variations between visual judgments of vocal fold vibratory asymmetry and objective measures of vocal fold vibratory asymmetry in subjects with and without vocal pathologies	8
2.2	Aim 2: Investigate hypothesized relationships between objective image-based measures of vocal fold vibratory asymmetry and characteristics of the acoustic voice signal in subjects with vocal pathologies.....	8
2.3	Aim 3: Characterize and map the space of vocal fold vibratory asymmetry measures onto auditory-perceptual dimensions of voice quality.....	10
2.4	Timeline.....	10
3	Background and Significance	10
3.1	Voice production	11
3.1.1	Theory.....	11
3.1.2	Role of symmetry.....	12
3.1.3	Endoscopic imaging	13
3.2	Image-based measures of vocal fold vibratory asymmetry.....	16
3.2.1	Visual judgments	16
3.2.2	Objective measures.....	17
3.3	Acoustic correlates of vocal fold vibratory asymmetry.....	21
3.4	Auditory-perceptual ratings of vocal fold vibratory asymmetry-affected acoustic waveforms.....	21
3.5	Summary of proposed thesis contributions	22
4	Preliminary Work.....	22
4.1	Developing a synchronous high-speed video and data acquisition system.....	23
4.1.1	Hardware setup	23
4.1.2	Graphical user interface for video and audio playback	24
4.2	Correlating acoustic perturbation measures to vocal fold vibratory asymmetry	25
4.2.1	Data collection	25
4.2.2	Measurement methods	25
4.2.3	Statistical analysis	28
4.2.4	Summary and conclusions	31
5	Research Design and Methods	32

5.1	Aim 1	32
5.1.1	Data collection	32
5.1.2	Previous measurement methods and statistical analysis.....	33
5.1.3	Proposed measurement methods	33
5.1.4	Statistical analysis	36
5.2	Aim 2	37
5.2.1	Data collection	37
5.2.2	Measurement methods.....	38
5.2.3	Statistical analysis	38
5.3	Aim 3	39
5.3.1	Data collection and rating methods	39
5.3.2	Statistical analysis	40
6	Use of Humans as Subjects.....	42
7	Literature Cited	43
8	Committee Agreements	47

1 Introduction

1.1 Motivation and goals

The goals of the proposed project are motivated by the clinical need for systematic studies that describe and develop acoustic correlates of vocal fold vibratory asymmetry, which could in turn help voice specialists manage voice disorders more effectively. Voice specialists make critical diagnostic, medical, therapeutic, and surgical decisions based on coupling visual observations of vocal fold tissue motion with auditory-perceptual assessments of voice quality (Zeitels *et al.*, 2007). While clinical experiences indicate that this approach is generally valid, it is inherently limited to case-by-case observations, and the details of the relationship between vocal fold tissue motion and the acoustic voice signal are not fully understood. Recent evidence indicates that visual judgments of vocal fold vibratory patterns may not adequately reflect changes in objective measures of the acoustic signal (Haben *et al.*, 2003). Furthermore, “[t]he anecdotal reports and stroboscopic findings of a prevalent typical amount of asymmetry cause a concern, in that it may indicate an increase in overdiagnoses of laryngeal pathology” (Shaw and Deliyski, 2008).

The overall goal of this project is to better understand the relationship between vocal fold tissue motion and the acoustic characteristics of the glottal voicing source so that clinical methods for assessing voice production can be improved. This work is made possible by recent advances in high-speed digital imaging, which provides adequate sampling for detailed intra- and inter-cycle comparisons between vocal fold tissue motion and the concomitant acoustic voice waveform. A series of three descriptive studies is proposed to systematically describe the co-variations among traditional and more advanced measures of vocal fold vibratory asymmetry and their impact on visual judgments, acoustic voice properties, and auditory-perceptual ratings.

First, it is proposed to replicate and improve upon recent findings describing co-variations between subjective visual judgments and basic objective measures of left-right vocal fold vibratory asymmetry in subjects with and without vocal pathologies (Bonilha *et al.*, 2008a; Bonilha *et al.*, 2008b). After validating the baseline co-variations with more automatic algorithms for computing left-right asymmetry, the image-based measures will be further developed and optimized based on the visual judgments of vocal fold vibratory asymmetry in both the left-right and anterior-posterior dimensions.

Second, the developed objective measures of asymmetry will be applied to voice data from a new subject population exhibiting vocal pathologies who will be evaluated using a state-of-the-art system for laryngeal high-speed videoendoscopy. The new system will allow for the validation of hypothesized relationships between vocal fold vibratory asymmetry measures and objective acoustic voice measures at unprecedented temporal resolution. Preliminary work has revealed mild co-variations between overall values of vocal fold vibratory asymmetry measures and traditional acoustic perturbation measures. Acoustic correlates of vocal fold vibratory asymmetries will be explored using knowledge of their effects on voice production.

Third, an initial study is proposed to characterize the influences of vocal fold vibratory asymmetry on the auditory perception of voice quality. This study more directly addresses the clinical reality that voices are assessed by relating vocal fold tissue vibratory patterns to the voice quality of a patient during a standard stroboscopic examination.

1.2 Thesis proposal structure

This thesis proposal is organized as follows. First, Section 2 outlines the three specific aims and associated hypotheses of the proposed investigation, along with a timeline of goals. Section 3 continues with background information on voice production mechanisms and reviews relevant research studies characterizing vocal fold vibratory asymmetries and the acoustic voice signal. Section 4 introduces work that investigated the co-variations between a preliminary measure of vocal fold vibratory asymmetry and traditional acoustic perturbation measures. Section 5 follows with the research design and methods for the three studies proposed. Finally, Section 6 concludes with information regarding the use of humans as subjects in these studies.

2 Specific Aims

A series of three studies is proposed to investigate the influence of vocal fold vibratory asymmetries on the acoustic voice signal. Specific aims and associated hypotheses of these studies are detailed below.

2.1 Aim 1: Investigate co-variations between visual judgments of vocal fold vibratory asymmetry and objective measures of vocal fold vibratory asymmetry in subjects with and without vocal pathologies

Aim 1 proposes to validate and improve upon recent findings describing co-variations between subjective visual judgments and objective image-based measures of left-right vocal fold vibratory asymmetry in a subject population without vocal pathologies. The recent findings have documented moderate correlations between visual-perceptual ratings and a basic objective measure of vibratory asymmetry of the left and right vocal folds (Bonilha *et al.*, 2008a). Completely automated image-based measures of asymmetry will be developed to replicate the published co-variations with visual ratings on the same data. After validating the automated algorithms for computing asymmetry measures, the image-based measures will be refined and optimized with respect to the visual-perceptual judgment data to improve upon the baseline co-variations in subject populations with and without vocal pathologies.

It is hypothesized that the new image-based measures of vocal fold vibratory asymmetry will co-vary with visual asymmetry judgments to a higher degree than previous image-based measure because of the ability to capture and integrate more temporal and spatial information from the image data.

2.2 Aim 2: Investigate hypothesized relationships between objective image-based measures of vocal fold vibratory asymmetry and characteristics of the acoustic voice signal in subjects with vocal pathologies

Aim 2 proposes to apply the developed objective measures of asymmetry to voice data collected from a new subject population with vocal pathologies that will be evaluated using a state-of-the-art system for laryngeal high-speed videoendoscopy. The system will allow the validation of hypothesized relationships between vocal fold vibratory measures and objective acoustic voice

measures, on an average and frame-by-frame basis. Preliminary work has revealed mild co-variations between average values of vocal fold vibratory asymmetry and traditional acoustic perturbation measures (jitter, shimmer, and harmonics-to-noise ratio). As a result, these measures will be applied to a larger subject population and explore new acoustic correlates of vocal fold vibratory asymmetries using knowledge of voice production mechanisms.

It is hypothesized that vocal fold vibratory asymmetries, especially time-varying (changing from period to period) asymmetries, will co-vary with acoustic measures of spectral slope, jitter, shimmer, and harmonics-to-noise ratio. The hypothesized effects of different types of vocal fold vibratory asymmetry on the acoustic voice signal are detailed in the following table:

Vocal fold vibratory asymmetry		Hypothesized effects on acoustic voice signal
Left-right amplitude asymmetry	Time-invariant	<ul style="list-style-type: none"> Strong negative correlation with spectral slope measures due to reduced vocal fold impact at closure. Mild positive correlation with shimmer and jitter due to vibratory instability. Strong positive correlation with harmonics-to-noise ratio due to increased air flow turbulence at the glottis.
	Time-varying	<ul style="list-style-type: none"> Strong positive correlation with shimmer due to cycle-to-cycle variations in vocal fold impact at closure. Mild positive correlation with jitter due to vibratory instability. Strong positive correlation with harmonics-to-noise ratio due to increased air flow turbulence at the glottis.
Left-right phase asymmetry	Time-invariant	<ul style="list-style-type: none"> Strong negative correlation with spectral slope measures due to reduced vocal fold impact at closure. Mild positive correlation with shimmer and jitter due to vibratory instability. Strong positive correlation with harmonics-to-noise ratio due to increased air flow turbulence at the glottis.
	Time-varying	<ul style="list-style-type: none"> Strong positive correlation with shimmer due to cycle-to-cycle variations in vocal fold impact at closure. Mild positive correlation with jitter due to vibratory instability. Strong positive correlation with harmonics-to-noise ratio due to increased air flow turbulence at the glottis.
Anterior-posterior phase difference	Time-invariant	<ul style="list-style-type: none"> Strong negative correlation with spectral slope measures due to asynchronous vocal fold impact at closure. Mild positive correlation with shimmer and jitter due to vibratory instability. Strong negative correlation with harmonics-to-noise ratio due to increased air flow turbulence at the glottis.
	Time-varying	<ul style="list-style-type: none"> Mild positive correlation with shimmer and jitter due to vibratory instability. Strong negative correlation with harmonics-to-noise ratio due to increased air flow turbulence at the glottis.

Period-to-period measurements of vocal fold vibratory asymmetry enable the ability to formulate short-time algorithms for analyzing shimmer- and jitter-type characteristics. The proposed signal processing techniques include harmonic/noise decomposition and perturbation-free algorithms for more sensitive calculations of the harmonics-to-noise ratio, spectral coherence functions for revealing underlying similarities among signals, and wavelet techniques for more accurate detection of glottal excitation in the acoustic signal.

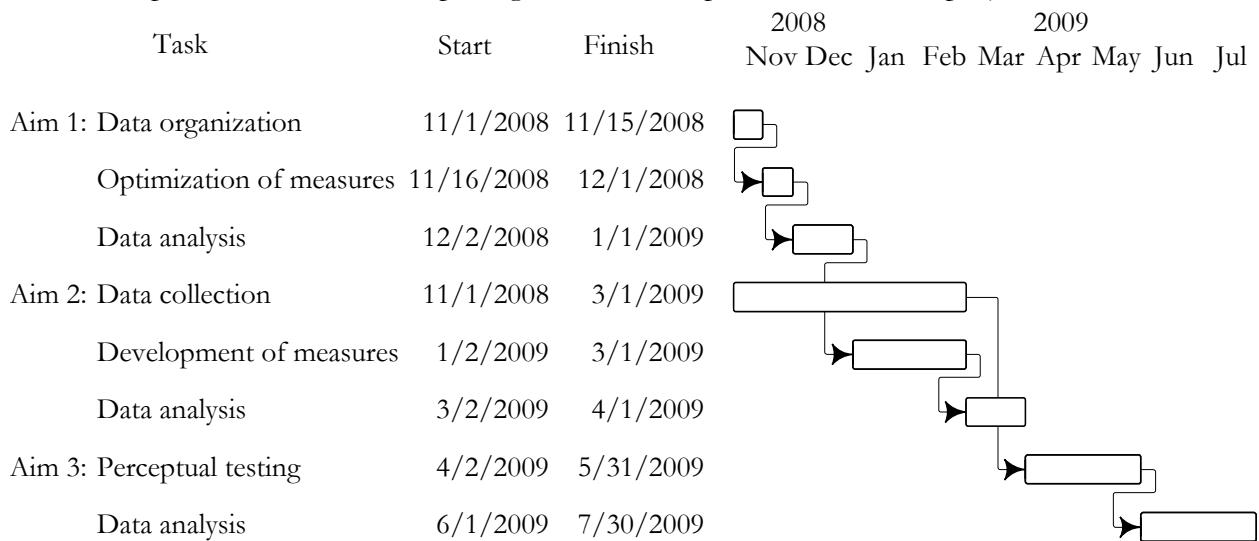
2.3 Aim 3: Characterize and map the space of vocal fold vibratory asymmetry measures onto auditory-perceptual dimensions of voice quality

Aim 3 proposes an initial study to characterize the influences of vocal fold vibratory asymmetry on the auditory perception of voice quality. This approach seeks to map vocal fold vibratory asymmetry measures to the dimension of perceptually-salient voice qualities. In this aim, the perceptual voice attributes of roughness, breathiness, and strain will be investigated in a select set of ten voice segments from the data set collected in Aim 2. The two image-based measures of left-right phase asymmetry and anterior-posterior phase differences will be investigated. Stimuli that represent a wide range values for these asymmetry measures will be selected to allow for the mapping of these characteristics onto auditory dimensions of breathiness, roughness, and strain voice qualities to reveal the ordering of the stimuli along these perceptual continua as well as the psychophysical distance separating the stimuli.

It is **hypothesized** that the vocal fold vibratory asymmetry measures that were found to covary to a large extent with acoustic measures will correlate, with respect to order, to the corresponding auditory perceptual rating of a voice quality. A linear correlation, however, is not expected. For example, a doubling of an image-based asymmetry measure will not be expected to result in a doubling of a perceptual voice attribute on its psychophysical scale.

2.4 Timeline

Proposed timeline for completing tasks for the specific aims of this project:



3 Background and Significance

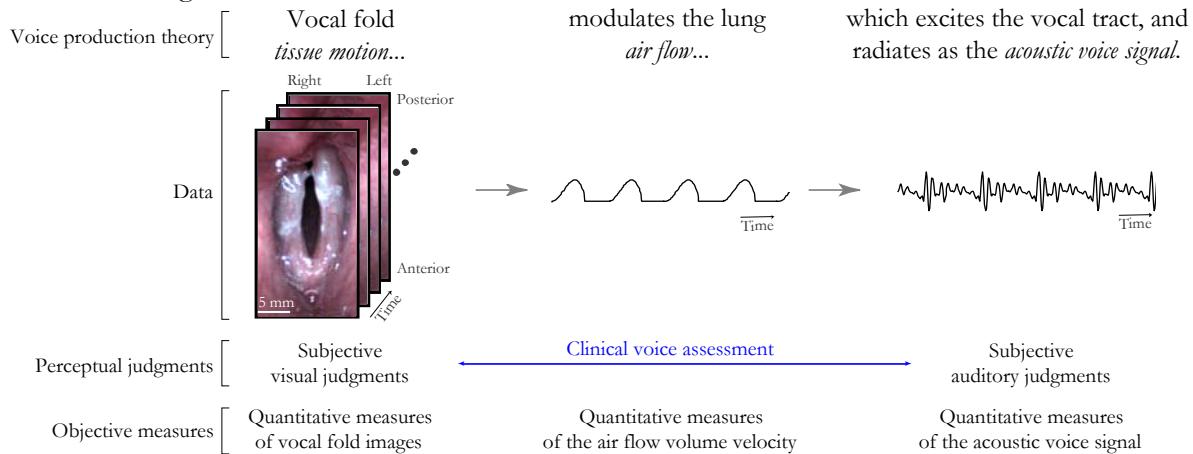
Section 3 provides important background information for understanding the goals of the thesis, which aim to elucidate co-variations among *in vivo* measures of the human voice in the physiological, acoustic, and perceptual dimensions. Section 3.1 provides an overview of voice production mechanisms, introduces the role of vocal fold asymmetry, and evaluates laryngeal imaging technologies available. Section 3.2 reviews methods for subjective and objective evaluation

of vocal fold asymmetry using laryngeal imaging data. Section 3.3, although meant to review acoustic correlates of vocal fold asymmetry, instead exposes the paucity of data in this area and provides the major motivation for the current studies. Section 3.4 gives an overview of auditory-perceptual rating methods used for the evaluation of voice quality produced by voices with vocal fold asymmetry. Finally, Section 3.5 summarizes the current state of knowledge based on the reviewed studies and the contributions of the proposed thesis to the field.

3.1 Voice production

3.1.1 Theory

The voice production system is often simplified to two independent mechanisms—the source and the filter. The source mechanism arises when the vocal folds of the larynx are set into periodic vibration by a combination of muscle tensions and aerodynamic forces—the myo-elastic aerodynamic theory (van den Berg, 1958). Vibratory patterns of the vocal folds provide an excitation source of quasi-periodic puffs of air that are subsequently input into the supraglottal system. Due to the relatively high acoustic impedance at the glottis, the supraglottal system effectively acts as a linear filter that shapes the spectral characteristics of the air flow source. Thus vocal fold tissue motion modulates the lung air flow, which excites the vocal tract, and radiates from the mouth as the acoustic voice signal:



Images of the vocal folds and the acoustic voice signal are available to clinicians for voice assessment, allowing for perceptual judgments of vocal fold tissue motion and voice quality.

It is acknowledged that vocal fold tissue motion plays only one role in the production of voice and that other considerations must be included in future work to be complete:

- aerodynamic factors of driving lung pressure and airflow velocities through the glottis, classically viewed as the source of acoustic energy (Granqvist *et al.*, 2003)
- coupling of the acoustic energy to tissue and to multiple cavities in the speech production system (Zhang *et al.*, 2006)
- aeroacoustic theories of voice production that take into account non-acoustic sources of energy and vortical flow during phonation (Zhang *et al.*, 2002; Zhao *et al.*, 2002; Krane, 2005; McGowan and Howe, 2007)

- hyperfunctional responses of the voice production mechanism that seek to compensate for voice quality disruptions and mask physiological deviations (Hillman *et al.*, 1990)

The proposed approach focuses on revealing correlations between objective and subjective characteristics of vocal fold vibrations and the acoustic waveform.

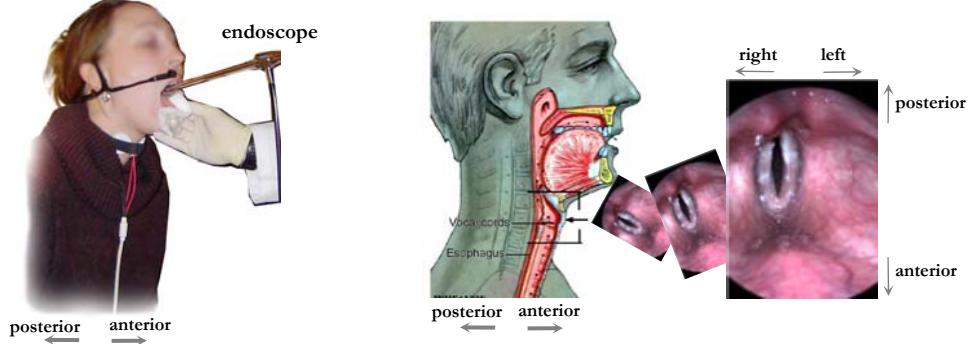
3.1.2 Role of symmetry

In the theory of voice production stated above, the vocal folds oppose each other in symmetric (in space) and regular (in time) manners. Asymmetries of the system have been observed in speakers with and without voice disorders (Bonilha *et al.*, 2008a; Shaw and Deliyski, 2008). Factors purported to influence vocal fold asymmetry within speakers include subglottal pressure (Berry *et al.*, 1996; Maunsell *et al.*, 2006), phonatory pitch, vocal fold mass and stiffness properties, vocal loading (Lohscheller *et al.*, 2008a), stress, and hydration. Across subjects with normal voices, possible factors affecting asymmetry include gender, age, profession, genetics, and language. The prevalence of vocal fold asymmetry in subjects with normal voices (Haben *et al.*, 2003; Bonilha *et al.*, 2008a; Shaw and Deliyski, 2008) indicates that the existence of asymmetry is not automatically associated with the presence of a vocal pathology (although the magnitude of asymmetry may reveal additional information). Pathologies linked to the presence of vocal fold asymmetry include vocal fold nodules, polyps, cysts (Qiu *et al.*, 2003; Gallivan *et al.*, 2008), unilateral recurrent laryngeal nerve paralysis (Švec *et al.*, 2007), Reinke's edema (Qiu *et al.*, 2003), vocal fold scarring (Haben *et al.*, 2003), hyperfunction (Eysholdt *et al.*, 2003; Gallivan *et al.*, 2008), laryngeal tuberculosis (Haben *et al.*, 2003), asymmetric cricothyroid muscle contraction (Maunsell *et al.*, 2006), and functional dysphonia (Neubauer *et al.*, 2001).

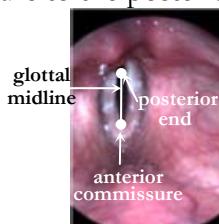
It is necessary here to clarify the types of vocal fold asymmetry that are of interest in this project. Vocal fold *vibratory symmetry* refers to spatial symmetry of the vocal folds over time during sustained phonation. Thus, if a mirror were placed at the glottal midline facing the left vocal fold, the right vocal fold would vibrate as if it were the reflected image of the left. In contrast, anatomical symmetry refers to symmetry of static structures on the left and right sides of the larynx, such as the static position of the arytenoid cartilages (Hirano *et al.*, 1989). The absence of anatomical symmetry does not automatically mean that vibratory asymmetry is present, and vice versa.

3.1.3 Endoscopic imaging

The mechanisms of normal and disordered human voice production have been difficult to investigate because the larynx is not naturally illuminated and the vocal folds vibrate too fast (100–1,000 cycles per second) to be seen with the naked eye. The larynx must be viewed using an endoscope that is passed through the nasal or oral cavity. The typical setup for transoral endoscopy is depicted here:

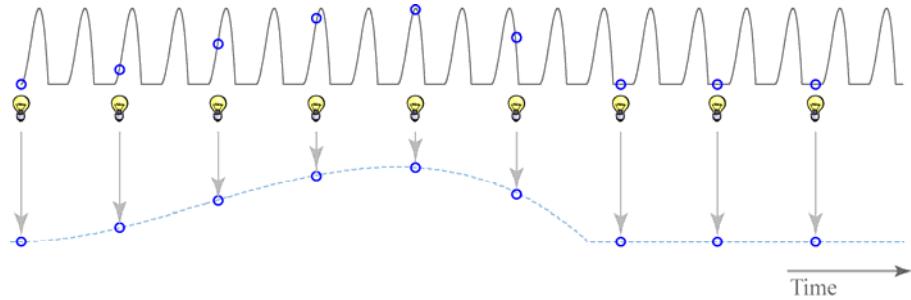


The leftmost image illustrates the position of a subject and the approach angle of a rigid endoscope, and the images on the right indicate the orientation of the vocal folds as viewed through the endoscope. The empty space between the folds is the glottis. Left and right directions are reversed, and the anterior-posterior plane is indicated. These definitions are crucial because measures of asymmetry based on these images rely on a precise spatial orientation. As mentioned before, although static anatomical symmetry might be important for assessing laryngeal irregularities, the current studies deal with *vibratory* symmetry, which refers to spatial symmetry of the vocal folds over time during phonation. The line of symmetry between the vocal folds is termed the glottal midline and is drawn between the anterior commissure to the posterior end of the vibrating vocal folds:



Videostroboscopy

The maximum frame rate of standard video cameras, approximately 30 frames per second (The Society of Motion Picture and Television Engineers, 2004), is too slow to adequately sample the vibrating vocal folds, which usually open and close over 100 times per second and approach velocities of one meter per second (Schuster *et al.*, 2005). To compensate for limitations on the unaided eye to observe the vibrating vocal folds, the video rate limitation, an imaging technique was developed to take advantage of the periodic nature of vocal fold vibrations [(Oertel, 1895), see (Zeitels, 1995)]. During a sustained vowel, the vocal folds often open and close in a regular pattern over time, and the temporal redundancy in information may be exploited when sampling the pattern. The following diagram illustrates this sampling principle by displaying a continuous periodic waveform that is too fast to sample adequately (the light bulbs cannot occur often enough):

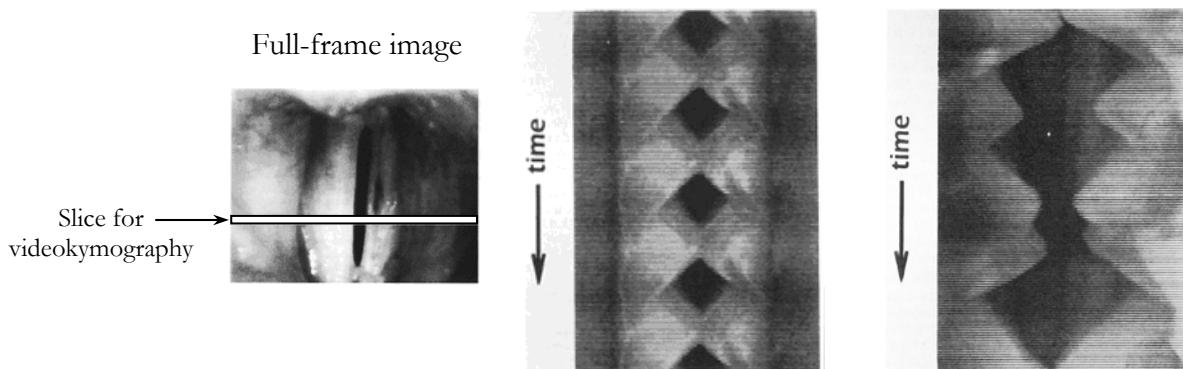


Videostroboscopy is the two-dimensional implementation of this sampling technique. Cycle-to-cycle regularity is assumed during sustained phonation, and a strobe light flashes at successive phases of the repeating vocal fold cycle to allow the video camera to record images at about 30 frames per second. A composite video sequence of the vocal fold vibratory cycle is reconstructed from images taken from several non-consecutive cycles. Since vocal fold vibratory asymmetries have been purported to occur in conjunction with period irregularities, the assumption of cycle-to-cycle regularity is too restrictive for the purposes of the proposed study. To adequately capture irregular vocal fold vibratory patterns, as well as enable direct correlations between vocal fold kinematics and acoustic voice properties, this project will use data obtained from high-speed imaging systems.

Videokymography

The drawbacks of videostroboscopy spurred the development of high-speed imaging technology. In a modification of existing video technology, Švec and Schutte (1996) developed a hardware solution termed videokymography that increased the video sampling rate by sacrificing spatial resolution. In videokymography, only one coronal *kyme* (Greek for *slice*) of the vocal folds was imaged, but this reduction in spatial resolution came with an increase in temporal sampling to about 8,000 kymes per second. Just as videostroboscopy took advantage of the temporal periodicity of vocal fold cycles, videokymography exploited the spatial redundancy exhibited by vocal folds during each cycle. Since the bulk movement of the vocal folds occurs in the medio-lateral direction to produce voice, one may assess an important aspect of vocal fold vibration using videokymography:

Videokymography of regular and irregular vocal fold vibratory patterns



The left image depicts a superior view of the vocal folds with the white horizontal line indicating the location imaged for the videokymographic images. Both regular and irregular vocal fold vibratory patterns can be imaged using videokymography.

For the purposes of the proposed studies, videokymography offers a visualization of vocal fold kinematics that is not affected by the assumption of strict periodicity necessary for using video stroboscopy. The gain in temporal resolution is, however, accompanied by a severe reduction in spatial resolution. Many unknown variables would be introduced, including variability associated with undesirable motions of the camera operator and the subject that would prevent the line scan to be taken from the exact same location over the duration of the phonatory gesture. Even more important for the analysis of vocal fold vibratory asymmetries is the simultaneous imaging of vocal fold tissue motion along the anterior-posterior aspect of the vocal folds. Thus, to answer research questions dealing with vocal fold vibratory asymmetries in both left-right and anterior-posterior dimensions and with possible irregularities in temporal vibratory patterns, a laryngeal imaging system is required that is not restricted by assumptions of the spatial or temporal aspects of vocal fold vibrations. One such solution is laryngeal high-speed videoendoscopy.

High-speed videoendoscopy

In the late 1930s and early 1940s, Farnsworth and colleagues at Bell Laboratories performed the first documented attempts at imaging vocal fold tissue motion using a high-speed motion picture camera (Farnsworth, 1940). Subsequently, high-speed filming of vocal fold vibration became more feasible and readily available with the development of digital technology for easier image acquisition and processing (Hammarberg, 1995; Maurer *et al.*, 1996). Characteristics of these early high-speed digital camera technologies included monochromatic visualization, sampling rates of up to 2,000 frames per second, low spatial resolutions, and light sensitivity that still did not approach video stroboscopic quality. To further improve the technology, Deliyski and colleagues worked with Vision Research, Inc. (Wayne, NJ), an industry leader in high-speed imaging hardware, to optimize the image sensor of high-speed cameras to provide much-improved light sensitivity that allowed the acquisition of high-resolution color images at rates up to 10,000 frames per second (Deliyski *et al.*, 2008). In addition, software algorithms were developed to help compensate for undesirable endoscopic motion artifacts, which enabled advanced image processing algorithms to extract features of vocal fold tissue motion from the high-speed image sequence (Deliyski, 2005).

High-quality color images, like those shown here—



—provide spatially-rich displays of the vocal folds during regular and irregular oscillation patterns. New higher-speed systems provided adequate imaging for examining higher-pitched phonation and facilitated direct correlations with recordings from other voice measurement devices that captured signals at comparable sampling rates. The potential for such systems to accurately synchronize and compare vocal fold vibrations captured at 10,000 images per second with the acoustic signal provides new insights into relationships between vocal fold tissue motion and sound production—specifically, relationships between asymmetries of tissue motion and characteristic deviations in the acoustic signal. High-speed videoendoscopy of the larynx using a rigid endoscope provides the image quality necessary for the analysis of vocal fold vibratory asymmetries and relations to acoustic properties proposed in the current research design.

3.2 Image-based measures of vocal fold vibratory asymmetry

The image data captured using the various videostroboscopy, videokymography, and high-speed videoendoscopy can yield valuable information regarding vocal fold vibratory asymmetry. Such information can be obtained through both subjective and objective methods. Section 3.2.1 reviews methods for obtaining subjective, visual-perceptual judgments of vocal fold vibratory asymmetry by human raters. Section 3.2.2 outlines methods for analyzing the image data for objective measures by computational algorithms that can estimate overall, as well as period-to-period, measures of vocal fold vibratory asymmetry.

3.2.1 Visual judgments

Visual judgments of vocal fold vibratory asymmetry have been made systematically in only a handful of research studies. Two studies documented binary judgments (ie, presence or absence) of vocal fold asymmetry. In the first study, raters documented the presence or absence of an asymmetric mucosal wave in subjects with no voice disorders by judging videostroboscopic data (Haben *et al.*, 2003). Although the videostroboscopic method could have only adequately imaged vocal fold tissue motion that was periodic, the investigators would have still been able to observe vocal fold asymmetries that occurred regularly in time. In the second study, the investigators utilized videokymographic data to document the presence or absence of left-right vocal fold asymmetries (videokymograms can only show left-to-right vocal fold tissue motion) (Švec *et al.*, 2007). However, an important step was taken to further categorize left-right asymmetries into four sub-categories: large amplitude differences, frequency differences, large phase differences, and axis shift during closure.

More formal perceptual rating methods have been introduced that provide information regarding the degree and extent of a particular property, improving on the binary judgments of existence or non-existence. In one study, clinicians were instructed to judge the magnitude of the mucosal wave of the left and right vocal folds, individually, on a 6-point ordinal scale (0 = absent, 1 = severely decreased, 2 = moderately decreased, 3 = typical, 4 = moderately increased, 5 = severely increased) (Shaw and Deliyski, 2008). A measure of mucosal wave asymmetry was obtained by taking the difference of the mucosal wave magnitude ratings. The clinicians' ratings were made on image data obtained using both videostroboscopy and high-speed videoendoscopy.

In subsequent studies, more direct ratings of vocal fold asymmetry were sought by asking raters to specifically characterize both left-right phase asymmetry and anterior-posterior asymmetry (Bonilha *et al.*, 2008a; Bonilha *et al.*, 2008b). Image data were obtained using videostroboscopy and high-speed videoendoscopy. To understand how raters would judge a videokymography-like view of vocal fold tissue motion, image processing algorithms generated an additional visualization termed digital kymography that digitally sliced a portion of the high-speed video images to yield an image related to that captured via videokymography. The various image data were visually rated for two types of asymmetry—left-right phase asymmetry and anterior-posterior asymmetry—on a 5-point ordinal scale (1 = completely asymmetrical, 2 = severely asymmetrical, 3 = moderately asymmetrical, 4 = mildly asymmetrical, 5 = symmetrical). Left-right phase asymmetry referred to the discrepancy between the times of maximum opening between the vocal folds. Anterior-posterior asymmetry was rated individually for each vocal fold and referred to the extent to which the vocal fold did not exhibit synchronous closure along the anterior-posterior axis.

Visual ratings of left-right and anterior-posterior asymmetry have been obtained on the 5-point ordinal scale mentioned above on 52 adult subjects with normal voices (Bonilha *et al.*, 2008a), as well as 54 adult subjects with vocal pathologies (Bonilha *et al.*, 2008b). The availability of these data allow for correlational analysis between perceptual and objective measures. Baseline covariations from the documented findings between subjective visual judgments of asymmetry and an objective measure of left-right asymmetry will be replicated with automatic algorithms. After this validation step, it is proposed to improve upon the baseline correlations obtained by refining image-based measures that consider asymmetry in both the left-right and anterior-posterior dimensions.

3.2.2 *Objective measures*

Several objective image-based measures of vocal fold vibratory asymmetry have been proposed in the voice literature. This section reviews two major types of measures: model-dependent and model-independent. Both types of measures have been applied to endoscopic image data obtained via videokymography and high-speed endoscopy.

Model-dependent measures

Model-dependent measures are derived from mathematical models of the vibrating vocal folds that are optimally fit to traces of vocal fold movements from high-speed video data. The most popular and modified vocal fold model in the speech field is the two-mass model developed by Ishizaka and Flanagan that synthesized voiced sounds using a self-oscillating mechanism with two coupled oscillators representing the superior and inferior portions of the vocal folds (Ishizaka and Flanagan, 1972). Later, a simplified adaptation of the two-mass model was developed to reduce the model's parameter space and increase the tractability of the equations (Steinecke and Herzel, 1995). Additionally, the later model did not constrain the vocal folds to symmetric vibratory patterns, allowing left and right vocal fold parameters to vary independently.

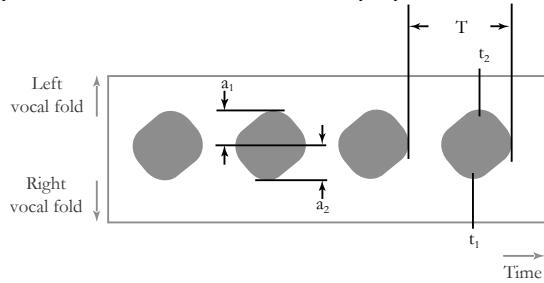
To characterize asymmetric vibratory patterns, the traces of individual vocal fold motion were fed into mathematical models based on the Steinecke and Herzel model. Döllinger and colleagues derived such vocal fold amplitude traces from three different anterior-posterior locations of the vocal folds in normal speakers (Döllinger *et al.*, 2003). Using a mathematical inversion procedure, the investigators arrived at a two-mass model of the vocal folds with unique parameters for (1) subglottal pressure, (2) mass/tension of the left vocal fold, and (3) mass/tension of the right vocal fold. Similar models have been optimized to describe laryngeal kinematics in subjects with unilateral recurrent nerve paralysis, vocal fold polyps, and functional voice disorders (Eysholdt *et al.*, 2003). These model-based methods have showed potential for characterizing vibrations in both the left-right and anterior-posterior dimensions, as well as classifying normal vibratory patterns from those exhibiting vocal fold pathologies (Schwarz *et al.*, 2006; Wurzbacher *et al.*, 2006; Wurzbacher *et al.*, 2008).

Although model-based approaches show promise of producing average asymmetry parameters that can characterize different voice types, the proposed project calls for more detailed measures of asymmetry that can be correlated with associated periods in the acoustic signal. Relying on an underlying model would create an undesirable averaging effect. It is hoped that future work would further refine existing vocal fold and vocal tract models and their interaction to better

understand the links between vocal fold tissue motion and overall measures of the acoustic voice signal.

Model-independent measures

Model-independent measures do not rely on an underlying vocal fold model. Image data are derived from algorithms that process videokymograms and full-frame high-speed videoendoscopic sequences. One proposed model-independent scheme extracted measures that tracked changes in voice-related measures using the videokymographic display (Qiu *et al.*, 2003). An outline of the relevant features of the kymogram (here displayed horizontally with time increasing to the right) shows the glottal area in gray for four vocal fold vibratory cycles:



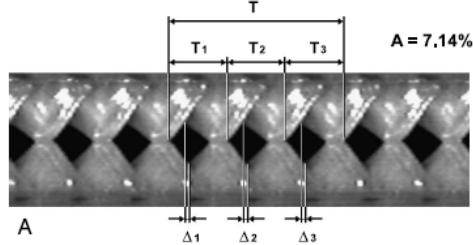
Analysis of the videokymogram yielded measures that described perturbation of the glottal period, open and opening phases, closed and closing phases, unilateral vibration amplitudes, and vocal fold asymmetries. Measures of glottal irregularity included the time periodicity index (variations in glottal period duration) and amplitude periodicity index (variations in glottal width). Of interest to the proposed study, vocal fold vibratory asymmetry measures included a phase symmetry index (PSI) and amplitude symmetry index (ASI) for each cycle:

$$\text{PSI} = \frac{t_1 - t_2}{T},$$

$$\text{ASI} = \frac{a_1 - a_2}{a_1 + a_2},$$

where the asymmetry indices PSI and ASI ranged from -1 to +1. Zero indicated perfect symmetry, the sign of the indices showed the directionality (toward the left or right vocal fold) of the asymmetry, and increasing magnitude corresponded to increasing degrees of asymmetry. In a group of 12 subjects with and without voice disorders, phase symmetry index magnitudes varied between 0.01 and 0.11 and amplitude symmetry index magnitudes varied between 0.01 and 0.59 (Qiu *et al.*, 2003).

Bonilha and colleagues, while investigating visual judgments of vocal fold vibratory asymmetry from various visualization methods, computed an objective measure of left-right phase asymmetry was very similar to the phase asymmetry index, PSI, above (Bonilha *et al.*, 2008a; Bonilha *et al.*, 2008b). The objective measure for left-right phase asymmetry was computed from a digital kymogram—

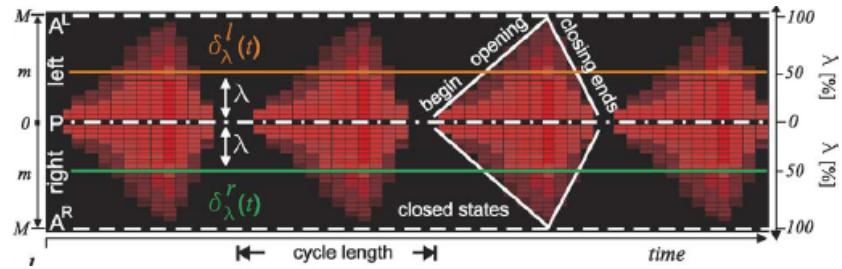


—derived from the full-frame high-speed video sequence. As illustrated, the left-right phase asymmetry measure was computed over three consecutive cycles by taking the average time delay (in pixels on the kymogram) between maximum intra-cycle amplitudes of the left and right vocal folds, normalized by the average period:

$$A = \frac{\Delta_1 + \Delta_2 + \Delta_3}{T_1 + T_2 + T_3} \times 100.$$

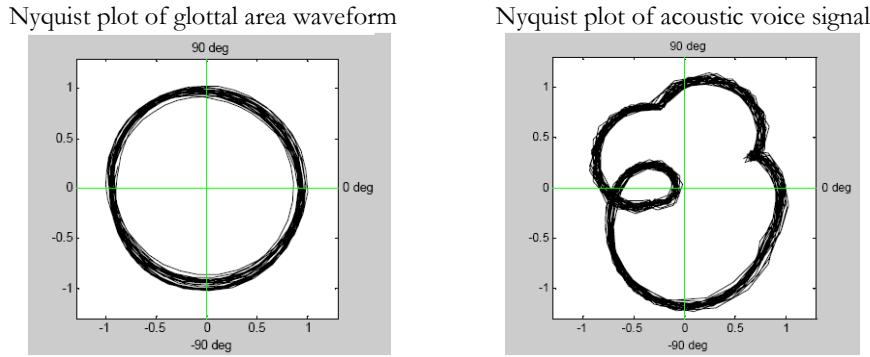
This measure of left-right phase asymmetry ranged from 0–36% and was shown to correlate moderately with visual judgments of phase asymmetry in subjects with and without voice disorders (Bonilha *et al.*, 2008a; Bonilha *et al.*, 2008b).

Most recently, the construction of a “phonovibrogram” has shown promise to quantify vocal fold vibratory asymmetries using a novel visualization of the glottal cycle (Lohscheller *et al.*, 2008b):

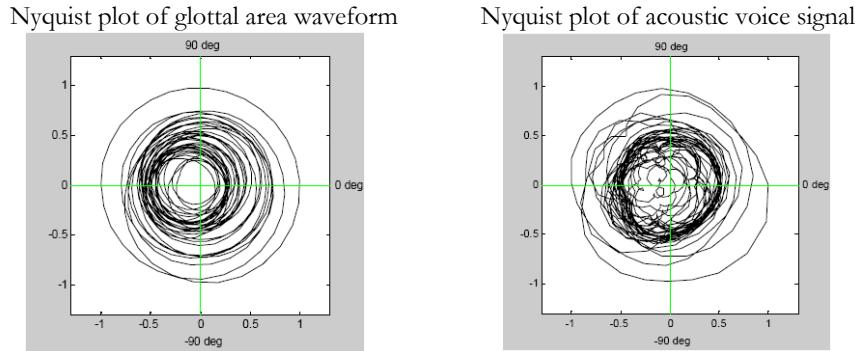


The phonovibrogram is a three-dimensional plot with the vertical axis representing the anterior-posterior lengths of each vocal fold, the horizontal axis representing time, and the brighter shades of red showing increasing amplitudes of the abduction of left and right vocal folds. Investigators parameterized the phonovibrogram using several angle measures that reflected the geometry and regularity of the shapes created in the phonovibrogram. Angle measures rely on linear regression lines that may smooth out details of vibratory characteristics. Although the technique served as an effective categorization of various glottal closure types, specific measures of left-right asymmetries and anterior-posterior differences of interest to the proposed studies are embedded within the phonovibrographic image.

Finally, since the proposed project involves analyzing the cycle-to-cycle regularity of vibratory measures to the acoustic voice signal, it is worth mentioning a series of papers that developed the ‘Nyquist plot’ method for visualizing glottal cycle irregularities (Yan *et al.*, 2005; Yan *et al.*, 2006; Yan *et al.*, 2007). In general, Nyquist plots operate on any waveform and generate a two-dimensional graph of the waveform compared with a Hilbert-transformed version of itself. The Nyquist plot of a periodic waveform would exhibit a particular shape that would fold back on itself due to its repetitive nature. Glottal area waveform and acoustic waveform analysis with this method produced the following Nyquist plots for a speaker with no vocal pathology (Yan *et al.*, 2007):



In contrast with this view of periodic phonation, analysis of an individual with recurrent respiratory papillomatosis revealed deviations in periodicity through erratic contours in the Nyquist plots of the glottal area waveform and the acoustic voice signal (Yan *et al.*, 2007):



Typing of voice signals from various pathologies using the Nyquist plot visualizations have not been fully validated but might be helpful in assessing certain irregularities that are not analyzable with traditional perturbation metrics.

The proposed study draws from the literature on model-independent measures of vocal fold vibratory asymmetries and aims to use information from the full-frame high-speed video sequence. A multi-slice kymographic approach will allow for the cycle-to-cycle computation of objective measures of anterior-posterior differences as well as left-right asymmetry. In addition, these measures will be meaningful and comparable to corresponding visual judgments of asymmetry that have been documented (see Section 3.4).

3.3 Acoustic correlates of vocal fold vibratory asymmetry

No published study has identified acoustic correlates of vocal fold vibratory asymmetry. This is the primary motivation of the proposed project. Traditional perturbation measures such as jitter, shimmer, and harmonics-to-noise ratio can be obtained from the acoustic voice signal. The validity and reliability of these acoustic measures, often used in during clinic voice assessment, are inherently limited by reliance on the accurate determination of fundamental frequency. Nonlinear dynamics analysis has been shown to be much more robust than traditional acoustic measures with respect to analyzing atypical signals such as aperiodic signals from pathological voices (Jiang *et al.*, 2006; Zhang *et al.*, 2008). Nevertheless, to enable correlational analysis over a wide range of measures, acoustic voice signals will be selected that would be categorized as Type 1 signals (Titze, 1995). In a study by Verdonck-de Leeuw *et al.* (2001), the harmonics-to-noise ratio was computed on sustained vowels elicited from four subjects with vocal fold asymmetry-producing vocal pathologies. The harmonics-to-noise ratio was compared to auditory-perceptual ratings of breathiness and roughness, and qualitative observations were made to link changes in the acoustic measure with vibratory characteristics. The proposed study is more ambitious in attempting a systematic study in a larger subject population that will allow for adequate statistical power to reveal significant co-variations among acoustic measures, vocal fold vibratory asymmetry measures, and auditory-perceptual judgments.

3.4 Auditory-perceptual ratings of vocal fold vibratory asymmetry-affected acoustic waveforms

Several voice quality descriptors have been implicated in reflecting vocal fold vibratory asymmetries, including hoarseness (Isshiki *et al.*, 1977; Niimi and Miyaji, 2000; Dresel *et al.*, 2006; Gallivan *et al.*, 2008; Schwarz *et al.*, 2008), roughness (Niimi and Miyaji, 2000; Verdonck-de Leeuw *et al.*, 2001), breathiness (Niimi and Miyaji, 2000; Verdonck-de Leeuw *et al.*, 2001), and vocal fry (Kiritani, 2000). Two studies have published data on quantitatively assigning auditory-perceptual judgments to voice segments where high-speed imaging has also been available to allow for evaluations of vocal fold asymmetry. In a study by Verdonck-de Leeuw *et al.* (2001), changes in the perceived quality of breathiness and roughness were qualitatively related to vocal fold characteristics from videokymographic data. Breathiness and roughness were rated on a 7-point ordinal scale (from 1 = severe breathiness or roughness to 7 = no breathiness or roughness) by one trained rater for voice segments from four subjects with vocal pathologies. Results indicated that physiological mechanisms occurring to create the rough voice quality included vibratory phase asymmetries between the left and right vocal folds and frequency and amplitude modulations of vocal fold amplitudes. Other results make similar qualitative relationships between physiology and voice quality.

A second study quantified the ability of auditory-perceptual ratings to reflect changes in vocal fold vibratory characteristics in 22 subjects with vocal pathologies that included vocal fold paralysis, vocal fold polyps/nodules, and vocal fold scarring (Niimi and Miyaji, 2000). On this larger data set, the investigators obtained auditory-perceptual ratings and vocal fold vibratory characteristics from a high-speed imaging system. Grade, roughness, and breathiness were judged by six raters on a 4-point ordinal scale (0 = none, 1 = slight, 2 = moderate, 3 = severe) drawn from a multi-dimensional scaling system developed in Japan. Vocal fold vibratory characteristics evaluated were presence of asymmetry, type of closure, amplitude differences, mucosal wave differences, and

frequency differences. Of particular interest to the current proposal were the asymmetry and amplitude difference characteristics. Results indicated that ratings of grade, roughness, and breathiness were significantly different for the symmetric and asymmetric voice segments. In addition, only breathiness was rated significantly differently when vocal fold amplitude differences were present (grade and roughness did not change significantly).

The proposed project builds on these results by using the consensus auditory-perceptual evaluation of voice (CAPE-V) to obtain continuous rating scales of breathiness, roughness, and strain (American Speech-Language-Hearing Association, 2002). In addition, a pair-wise comparison methodology that ranks voices on auditory-perceptual dimensions will be used to assess the actual perceptual distances of voices along psychophysical scales. The contributions of specific types of vocal fold vibratory asymmetries will be calculated, as well as an initial mapping of these asymmetries onto the voice quality dimensions.

3.5 Summary of proposed thesis contributions

Using a new system that accurately synchronizes and allows comparison of vocal fold vibrations captured at 10,000 images per second with the acoustic signal, relationships between vocal fold tissue motion and sound production can be revealed—especially relationships between asymmetries of vocal fold tissue motion and characteristics in the acoustic voice signal. New objective measures computed from high-speed video data will be developed to optimally correlate with visual judgments of left-right and anterior-posterior asymmetries. Measures will include different properties of asymmetry such as left-right asymmetry (including amplitude and phase asymmetries), anterior-posterior phase differences, and integrative asymmetry measures that combine all asymmetry properties. The influence of specific types of vocal fold vibratory asymmetries on the acoustic voice signal will be evaluated, and a statistical relationship will be outlined between asymmetry-related acoustic measures and voice quality judgments. Many vibratory measures have been developed to make binary distinctions between pathological from normal voices, as opposed to covering a range of acoustic characteristics. A key goal of this project is to appreciate the presence of a certain degree of vibratory asymmetry and of an acoustic property and begin to understand to what extent characteristics of the acoustic voice signal may reflect variations in the vocal fold vibratory asymmetries.

4 Preliminary Work

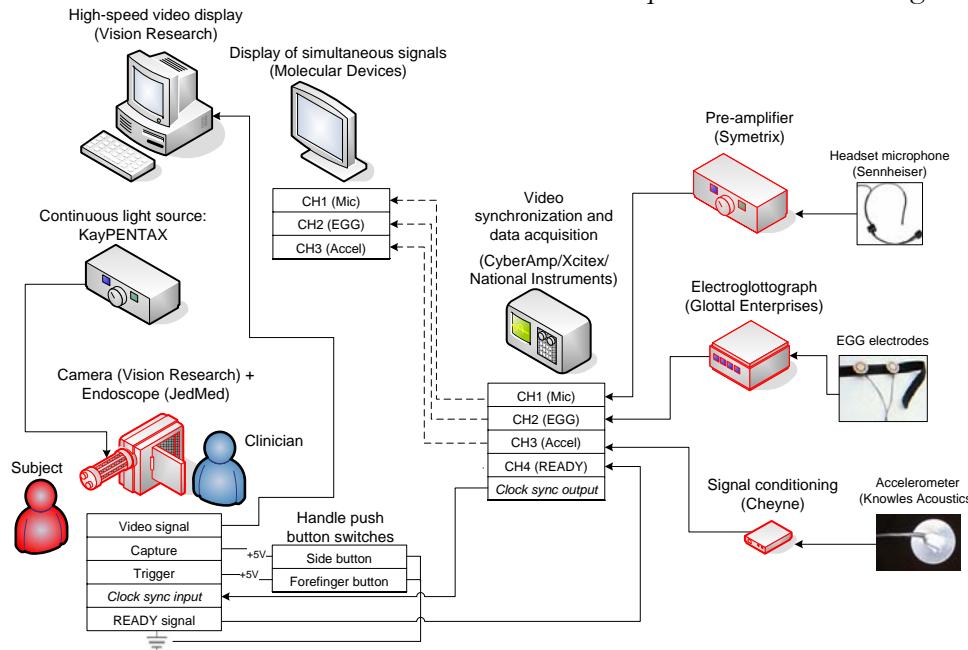
Several interim goals have been accomplished to establish the ability to analyze the details of vocal fold vibratory characteristics associated with corresponding changes in acoustic voice characteristics. Section 4.1 describes the implementation of an integrated system for synchronous acquisition of high-speed video images and voice-related signals. Section 4.2 documents an initial experiment to determine the degrees of co-variation between a preliminary measure of vocal fold asymmetry and traditional acoustic perturbation measures. Steps toward quantifying vocal fold tissue motion from high-speed video images are outlined.

4.1 Developing a synchronous high-speed video and data acquisition system

The components of the multi-sensor acquisition setup are described in Section 4.1.1. Section 4.1.2 illustrates the development of a graphical user interface that provides a valuable tool for interactive playback of the high-speed video images and synchronized acoustic data.

4.1.1 Hardware setup

The data acquisition system captures four separate voice-related signals: high-speed videoendoscopic images, acoustic signal, electroglottography signal, and neck skin acceleration. For the current investigation, the high-speed video and acoustic data are the focus, with the additional signals providing helpful data that might be used to better elucidate and explain selected results. The schematic below illustrates the electrical connections of the acquisition and recording hardware:



Light source: The light source consists of a short-arc Xenon lamp rated at 300 watts (KayPENTAX, Lincoln Park, NJ). The fan-cooled housing produces a collimated beam of light with a color temperature of over 6,000 K. Three glass infrared filters (two dichroic, one absorbing) block light from the infrared electromagnetic spectrum to reduce thermal energy of the light during endoscopy.

High-speed video acquisition: The Phantom v7.3 high-speed video camera (Vision Research, Inc., Wayne, NJ) enables high-quality color image capture at high image rates due to a highly-sensitive CMOS image sensor. A C-mount lens adapter with adjustable focal length (KayPENTAX, Lincoln Park, NJ) connects the camera to a 70° rigid endoscope (JEDMED, St. Louis, MO). Images are saved to partitions in a 4 GB on-board memory buffer and downloaded to the computer's hard drive via Ethernet after recordings are complete.

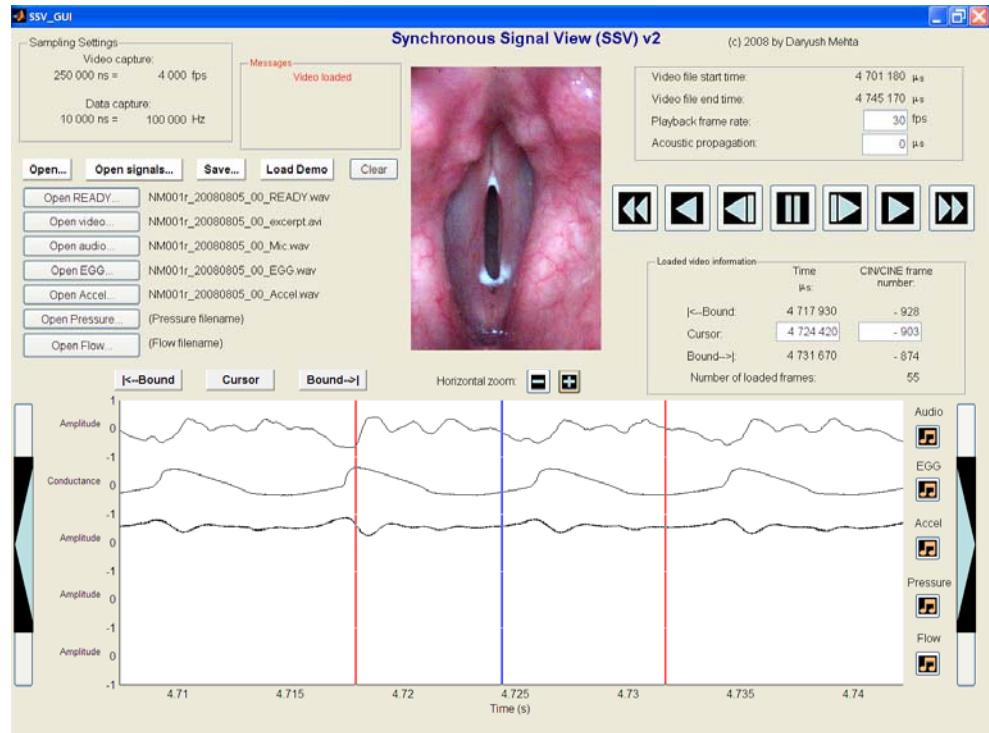
Acoustic signal acquisition: The acoustic data signal is captured by a head-mounted, high-quality condenser microphone (Model MKE104, Sennheiser electronic GmbH, Wennebostel, Germany) with a cardioid pattern, offering directional sensitivity and a wideband frequency response. The

microphone pre-amplifier (Model 302 Dual Microphone Preamplifier, Symetrix, Inc., Mountlake Terrace, WA) offers a low-noise, low-distortion gain input into the CyberAmp signal conditioner (Model 380, Axon Instruments, Inc., Union City, CA) for gain control to maximize the digitizer's dynamic range and analog anti-aliasing low-pass filtering at a 3 dB cutoff frequency of 30,000 Hz. The analog signals are digitized at a sampling rate of 100,000 Hz, 16-bit quantization, and a ± 10 V dynamic range by the NI 6259 M series PCI digital acquisition board (National Instruments, Austin, TX).

Data acquisition synchronization: The camera clock is supplied by an external clock source generated by the National Instruments board that is synchronized to the sampling of the data signals. The hardware clock division and data acquisition settings are controlled by MiDAS DA software (Xcitex Corporation, Cambridge, MA). Alignment of the high-speed video data and the data signals is accomplished by recording an analog signal from the camera that precisely indicates the time of the last recorded image.

4.1.2 Graphical user interface for video and audio playback

A graphical user interface has been developed for helping visualize the high-speed video and audio data in an integrated playback format. A screen shot of the tool displays the multi-faceted visualization:



In this figure, the acoustic waveform, electroglottographic signal, and neck acceleration signal are displayed, along with an endoscopic laryngeal image from a high-speed video sequence that corresponds to the time location of the blue cursor in the interface. Red cursors indicate bounds of the video images available for playback. The user is given control of video playback rate and the acoustic propagation time from larynx to microphone. With this integrated view, the user is able to navigate through the acoustic signal or the video images to investigate sources of acoustic

irregularities and speculate on possible physiological mechanisms observed in the corresponding vocal fold images.

4.2 Correlating acoustic perturbation measures to vocal fold vibratory asymmetry

The purpose of this study was to gain an initial first-hand impression of potential relationships between vocal fold vibratory mechanisms and acoustic perturbation measures.

4.2.1 Data collection

The preliminary study included high-speed video and acoustic voice data from ten subjects (8 male, 2 female) surgically managed for glottic cancer and other vocal pathologies disrupting vocal fold vibratory patterns:

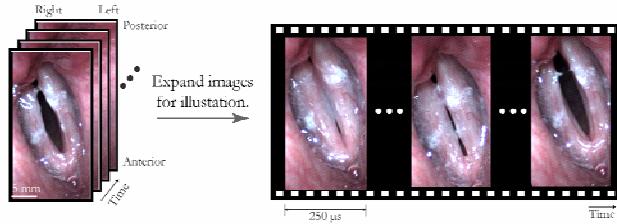
Subject	Gender	Voice diagnosis
P1	male	T2 glottic cancer
P2	male	T1 glottic cancer
P3	female	Bilateral mild loss of superficial lamina propria
P4	male	Dysplasia/keratosis
P5	male	Left vocal fold squamous cell carcinoma
P6	male	Cancer at anterior commissure
P7	male	Right arytenoid exophytic disease resembling squamous cell carcinoma
P8	male	Diffuse neoplasm
P9	female	Left vocal fold scarring
P10	male	Glottal insufficiency

The video images, collected via rigid endoscopy, were pre-screened for satisfactory image quality, a perpendicular endoscopic angle, and good exposure of the vocal folds. Subjects were instructed to sustain the vowel /i/ at a comfortable pitch and loudness for about four seconds (durations varied per subject). The sound produced approximated the vowel /ae/ due to the presence of the rigid endoscope in the mouth and protrusion of the tongue. Laryngeal high-speed videoendoscopic data were recorded at 4,000 images per second with maximum integration time and a spatial resolution of 320 horizontal x 480 vertical pixels for an approximately 2 cm² target area. Simultaneous recording of the acoustic signal was obtained using a directional microphone situated approximately 4 cm from the lips at a 45° offset in azimuth.

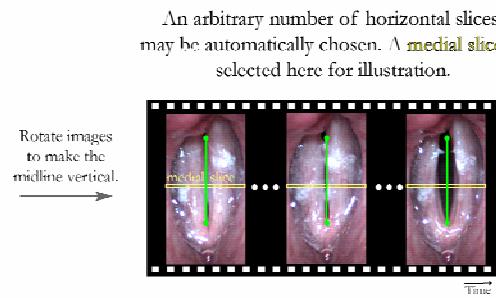
4.2.2 Measurement methods

The acoustic signal and the high-speed video images were divided into 500 ms segments for each subject. For each segment, vocal fold amplitude waveforms were obtained from the images using the following algorithm:

Over 4,000 high-speed video images are captured per second.

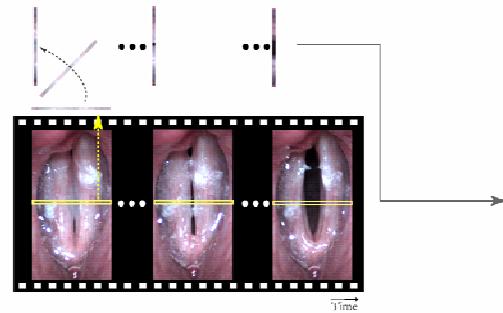


Three of the images are displayed here to illustrate the following analysis steps.



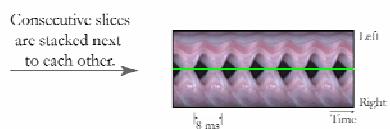
An arbitrary number of horizontal slices may be automatically chosen. A **medial slice** is selected here for illustration.

From one image with maximal glottal area, the user defines the **glottal midline** by indicating the anterior commissure and posterior vibrating end.



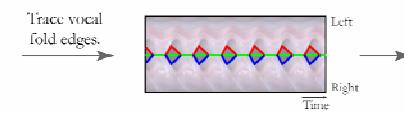
For each image, the selected slice is extracted and rotated.

Free edges of **left** and **right** vocal folds are traced.

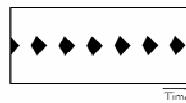


A digital kymogram is created from the medial position.

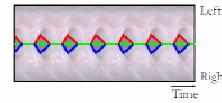
The glottis is segmented to determine the vocal fold edges.



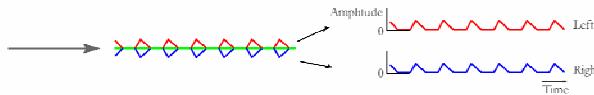
Binary segmentation using thresholding.



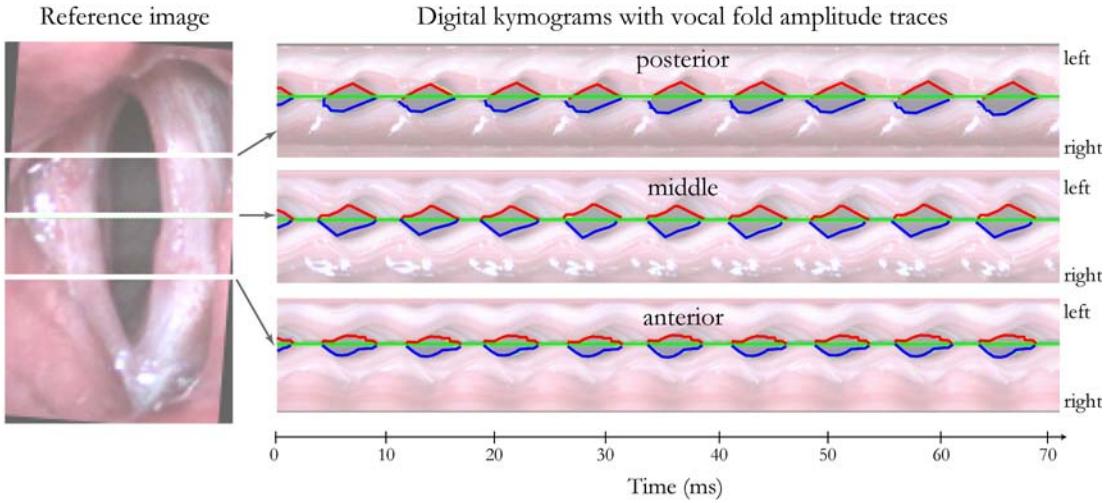
Trace vocal fold edges.



Amplitudes of vocal fold edges are calculated with reference to the glottal midline defined by the user.



Using the threshold-based edge detection method outlined above, vocal fold amplitude traces were derived from full-frame image sequences that were compensated for undesirable translational and rotational motion (Deliyski *et al.*, 2008). Left and right vocal fold amplitude waveforms— $x_l[n]$ and $x_r[n]$, respectively—were obtained from three distinct locations along the anterior-posterior vocal fold dimension. The locations were selected to be at an anterior, middle, and posterior position, equally spaced between the anterior commissure and the posterior end of the membranous glottis:



Thus, six amplitude waveforms $x_{l,\alpha}[n]$ and $x_{r,\alpha}[n]$ are computed, where the additional α subscript indicates the position of the associated kymogram in the reference image. To enable comparisons across subjects with varying sound pressure levels and fundamental frequencies, the amplitude waveforms were normalized by the maximum glottal width β_α at each location. The initial measure of asymmetry calculated the root-mean-square value for the arithmetic difference between the left and right vocal fold amplitude waveforms. The RMS values for the anterior, middle, and posterior locations were summed to arrive at an overall asymmetry measure that took into account left-right asymmetries and anterior-posterior differences—the integrative asymmetry (IA) measure:

$$IA = \sum_{\alpha=1}^M \frac{1}{\beta_\alpha} \left(\frac{1}{N} \sum_{n=0}^{N-1} (x_{l,\alpha}[n] - x_{r,\alpha}[n])^2 \right)^{\frac{1}{2}}$$

where

$n \triangleq$ image number

$\alpha \triangleq$ kymogram index {anterior, middle, posterior}

$x_{l,\alpha}[n] \triangleq$ left vocal fold excursion from midline

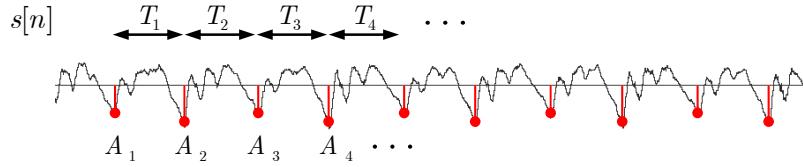
$x_{r,\alpha}[n] \triangleq$ right vocal fold excursion from midline

$N \triangleq$ total number of images

$M \triangleq$ total number of kymograms

$$\beta_\alpha = \max(x_{l,\alpha}[n] + x_{r,\alpha}[n])$$

In addition, shimmer, jitter, and harmonics-to-noise ratio values were computed from the synchronously-recorded acoustic voice signal to capture the time-varying variations in period duration and amplitude that may reflect variations in vocal fold vibratory asymmetry. These measures are calculated from the acoustic signal obtained from a phonatory segment $s[n]$:



Shimmer was defined as the absolute percent variation of the peaks around the mean amplitude over a given number of periods N in the acoustic waveform (Kay Elemetrics Corporation, 2006):

$$\text{shimmer } (\%) \triangleq 100 \cdot \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i}.$$

Similarly, jitter was defined as the absolute percent variation of period durations around the mean duration over a given number of periods N in the acoustic waveform (Kay Elemetrics Corporation, 2006):

$$\text{jitter}(\%) \triangleq 100 \cdot \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}|}{\frac{1}{N} \sum_{i=1}^N T_i}.$$

The harmonics-to-noise ratio (HNR) is an indication of the relative energy contributions of the harmonic or periodic component and the noise or aperiodic component of the acoustic voice signal. The HNR measure used time-domain analysis and calculation of autocorrelation sequences (Boersma and Weenink, 2007):

$$\text{HNR (dB)} \triangleq 10 \log_{10} \frac{r_s' [m_{\max}]}{1 - r_s' [m_{\max}]},$$

where

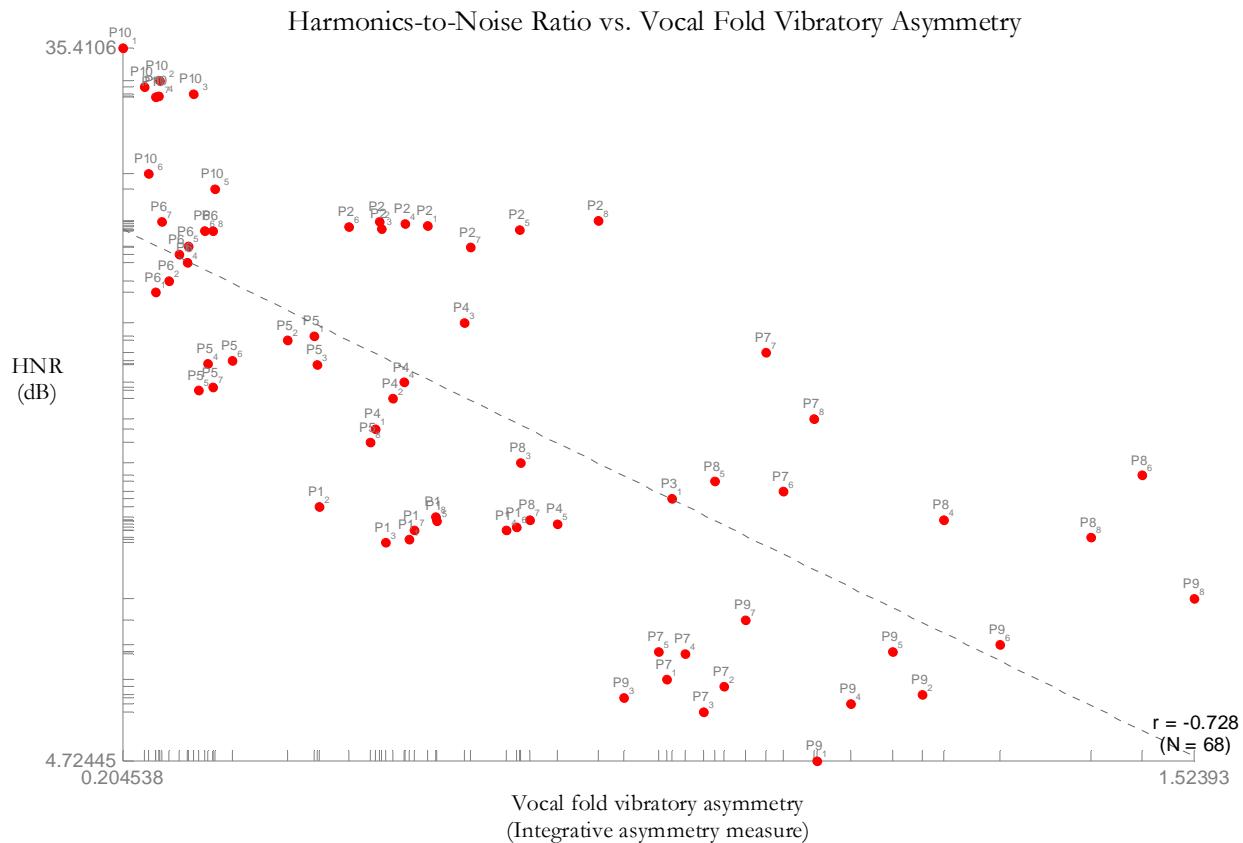
$$\begin{aligned} m_{\max} &= \arg \max_m r_s' [m] \\ r_s' [m] &\triangleq \frac{r_s [m]}{r_s [0]} \text{ (normalized autocorrelation)} \\ r_s [m] &\triangleq \sum_{m=0}^{N-1} s[n] s[n+m] \text{ (autocorrelation).} \end{aligned}$$

4.2.3 Statistical analysis

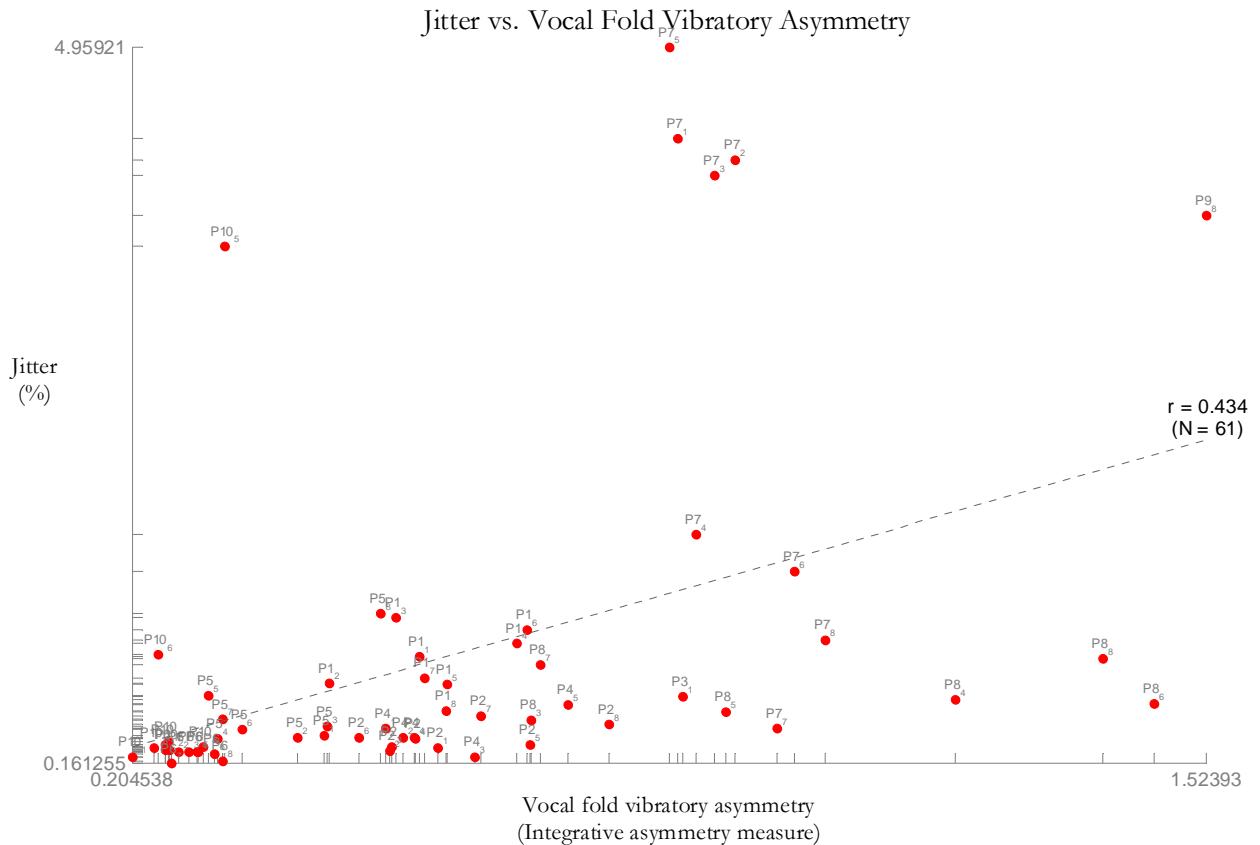
The integrative asymmetry measure and acoustic perturbation measures were calculated for each 500 ms phonatory segment within a subject's trial. Therefore, subjects have more than one set of measures associated with their voice signal ($P1_1, P1_2$, etc. in the scatter plots below). To enable

correlational analysis, independence of the data pairs and normality of the distributions were assumed. Prior to computing correlations for the jitter and shimmer measures, data with jitter and shimmer percentages greater than five percent were not included due to their unreliability (Titze, 1995). Pearson's correlation coefficient r analyzed the pair-wise relationships between the ratio-scaled asymmetry measure and ratio-scaled acoustic perturbation measures. Since jitter and shimmer measures of more than five percent are regarded as unreliable (Titze, 1995), data exhibiting extreme measures will not be included.

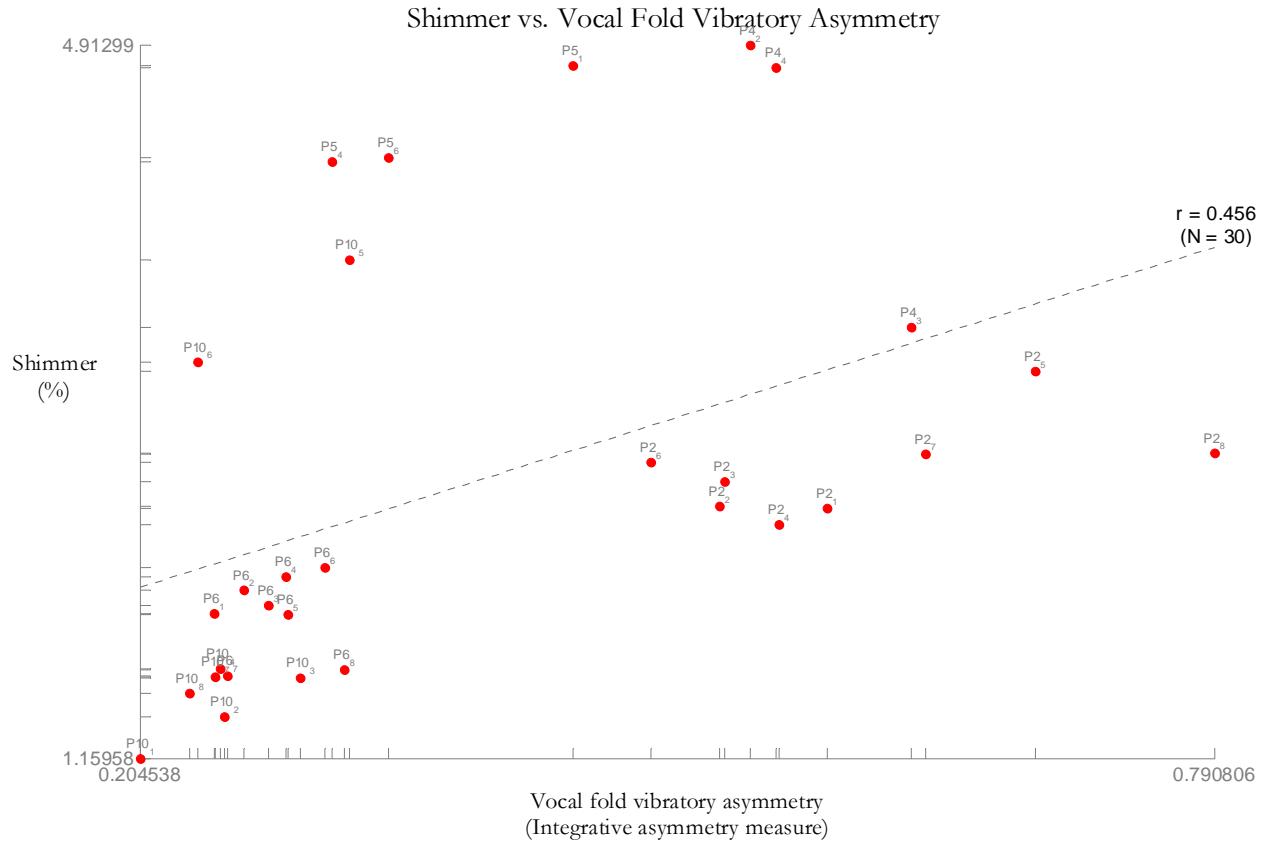
The correlation between harmonics-to-noise ratio and the integrative asymmetry measure was high ($r(68) = -0.728$, $p < 0.001$). The 95% confidence interval was also computed for Pearson's correlation coefficient ($-0.824 < r < -0.593$). Approximately 53% of the total variance in the harmonics-to-noise ratio was explained by taking into account the integrative asymmetry measure. A scatter plot displays the marginal distributions for harmonics-to-noise ratio and the integrative asymmetry measure (labels indicate subject numbers, with subscripts indicating the voice segment index):



When including only voice segments with jitter less than five percent, the correlation obtained between jitter percentage and the integrative asymmetry measure was moderate ($r(61) = 0.434$, $p < 0.001$). The 95% confidence interval for Pearson's correlation coefficient was computed ($0.204 < r < 0.618$). Thus, approximately 19% of the total variance in the jitter percentage is explained by the integrative asymmetry measure. A scatter plot displays the marginal distributions for jitter and the integrative asymmetry measure:



After pre-screening to include voice segments with shimmer percentages below five percent, the correlation between shimmer percentage and the integrative asymmetry measure was moderate ($r(30) = 0.456$, $p = 0.011$). The 95% confidence interval for Pearson's correlation coefficient was computed ($0.115 < r < 0.701$). Thus, approximately 21% of the total variance in the shimmer percentage is explained by the integrative asymmetry measure (although at a lower significance level compared with the HNR and jitter correlations). A scatter plot displays the marginal distributions for shimmer and the integrative asymmetry measure:



Due to the non-normal marginal distributions of jitter and shimmer (violating the assumptions for linear correlation analysis using Pearson's r), nonparametric statistical analysis using Spearman rank-order correlation methods was also performed. This verified the significance of the rank-based co-variations between the integrated asymmetry measure and shimmer and jitter. Spearman ρ was computed on the data after being transformed into rank-ordered measures. A summary of the correlation analysis of pair-wise measures follows:

Measures Correlated	Pearson's r	95% Confidence Interval for r	Spearman ρ
Jitter vs. Integrated Asymmetry	0.434	(0.204, 0.618)	0.638
Shimmer vs. Integrated Asymmetry	0.456	(0.115, 0.701)	0.661
HNR vs. Integrated Asymmetry	-0.728	(-0.824, -0.593)	-0.750

Note: All coefficients were significant at 95% confidence levels.

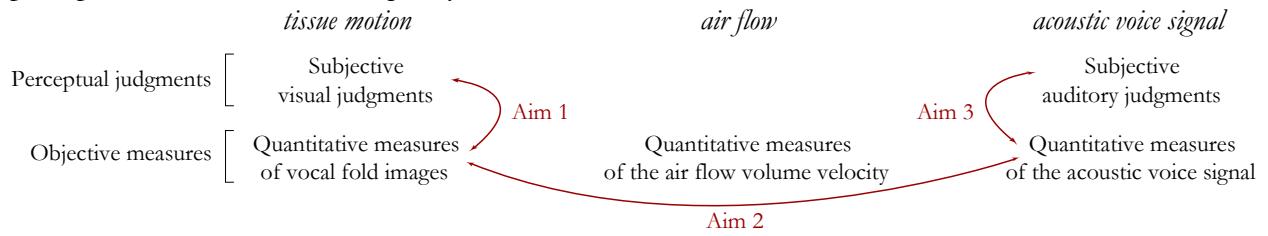
4.2.4 Summary and conclusions

General trends from the scatter plots indicate mild to moderate correlations between the integrated asymmetry measure and measures of shimmer and jitter. This result is expected since jitter and shimmer are only expected to increase with time-varying changes in asymmetry or glottal periods (Nardone, 2007). The higher absolute correlation between the integrative asymmetry measure and the harmonics-to-noise ratio supports the hypothesis that increases in asymmetry are coupled with a higher probability for noise turbulence at the glottis. This initial experiment shows promise for more in-depth analysis of the acoustic effects of different vocal fold vibratory asymmetries. Future experiments will not include voices that exhibit excessive jitter and shimmer

values for valid correlation analysis. The ability to quantify cycle-to-cycle changes in both the image and acoustic domains due to precise synchronization provides the opportunity to reveal the effects of physiological mechanisms at intra-cycle time scales.

5 Research Design and Methods

Details of research design and methodology for the three proposed aims are presented here. The proposed studies are designed to better understand the relationships between asymmetries of tissue motion during vocal fold vibration, properties of the acoustic voice signal, and auditory-perceptual attributes of voice quality:



Section 5.1 provides the study framework for Aim 1, which explores recent findings describing co-variations between subjective visual judgments and basic objective measures of vocal fold vibratory asymmetry in subject populations with and without vocal pathologies. Once baseline co-variations between visual ratings and image-based asymmetry measures are established and optimized, the focus shifts to Aim 2, as outlined in Section 5.2, which investigates the hypothesized relationships between the image-based asymmetry measures and acoustic voice measures. Finally, Section 5.3 outlines Aim 3, an initial approach to characterize the influences of vocal fold vibratory asymmetry on the auditory perception of voice quality.

5.1 Aim 1

The purpose of Aim 1 is to investigate co-variations observed between visual judgments of vocal fold vibratory asymmetry and objective measures of vocal fold vibratory asymmetry in subjects with and without vocal pathologies.

5.1.1 Data collection

Our collaborators at the University of South Carolina have collected high-speed video and acoustic data from 52 adult subjects (24 male, 28 female) with normal vocal function (Bonilha *et al.*, 2008a) and 54 adult subjects (11 male, 43 female) with vocal pathologies (Bonilha *et al.*, 2008b). Laryngeal high-speed videoendoscopic data were recorded at 2,000 images per second with maximum integration time and a spatial resolution of 120 horizontal x 256 vertical pixels for an approximately 2 cm² target area. Subjects were instructed to produce sustained vowels for about 2 seconds in habitual and pressed conditions at a normal pitch and loudness. Subjects from the pathological voice population only produced sustained vowels in a habitual manner at normal pitch and loudness. Although obtained with an earlier camera model, the high-speed video data available

with associated visual-perceptual ratings enables the development and optimization of the new image-based measures proposed in Aim 1.

5.1.2 Previous measurement methods and statistical analysis

To enable analysis of vocal fold vibratory asymmetries from the video data, image processing algorithms generated three types of visualizations (Deliyski *et al.*, 2008): high-speed video playback, digital kymography playback, and static kymography from the medial line. The pre-processed data were rated visually for two types of asymmetry—left-right and anterior-posterior vibratory asymmetry—on a five-point ordinal scale, with 1 representing “completely asymmetrical” and 5 indicating “symmetrical.” In addition, an objective measure of left-right asymmetry was computed from manual markings on the medial digital kymogram. The image-based asymmetry measure was the average time delay, normalized by the average period over three consecutive cycles, between maximum same-cycle amplitudes of the left and right vocal folds (see Section 3.2.2 for the exact formula).

Using the basic measure of left-right vocal fold vibratory asymmetry, mild to moderate co-variations were found with visual ratings of left-right asymmetry by the three voice specialists. Pearson’s correlation values between the objective measures and ratings are as follows (Bonilha *et al.*, 2008a):

Rater	Phonatory Condition	Digital Kymogram	Medial Digital Kymogram	High-Speed Video
1	Habitual	-0.76	-0.67	-0.47
	Pressed	-0.55	-0.70	-0.71
2	Habitual	-0.53	-0.64	-0.47
	Pressed	-0.42	-0.52	-0.71
3	Habitual	-0.26	-0.40	-0.64
	Pressed	-0.40	-0.61	-0.69

5.1.3 Proposed measurement methods

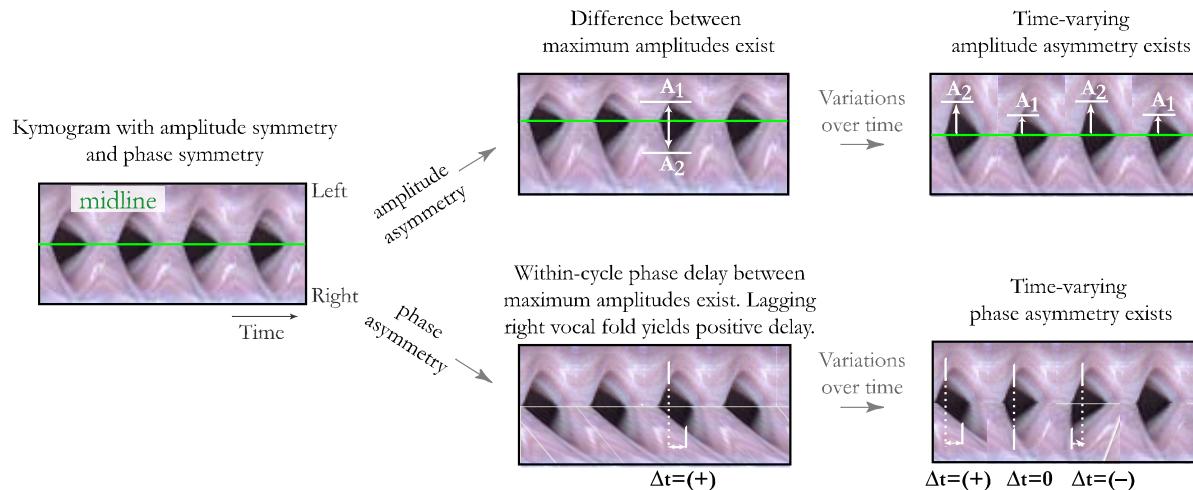
The first goal of Aim 1 is to develop and validate more automatic algorithms for computing left-right vocal fold vibratory asymmetry from the high-speed videoendoscopic data. The new image-based measures are extensions of the initial left-right asymmetry measure used in the correlation analysis above. However, the initial measure was computed only on the medial digital kymogram but correlated with the more information-rich high-speed video and digital kymographic video data. The second goal of Aim 1 is to improve upon the above baseline co-variations with visual left-right asymmetry judgments by refining image-based measures to consider left-right asymmetries from more than one kymographic slice. In addition, a multi-slice kymographic approach allows the computation of objective measures of anterior-posterior differences in vocal fold vibratory patterns that can be tested for co-variation with the visual ratings of anterior-posterior asymmetry.

Automatic algorithms improve upon manual markings of kymograms and include knowledge-based image processing of the kymograms and high-speed video images to extract vibratory asymmetry measures that can be tested against appropriate visual ratings. Vocal fold tissue

motion during phonation can be measured by tracking the left and right vocal fold edges closest to the glottal midline. The edges tracked collapse the three-dimensional rocking motion of the vocal folds that constitute the mucosal wave to a two-dimensional space that does not take into account inferior-superior phase differences. The tissue motion tracked, however, attempts to quantify the bulk of the medio-lateral tissue motion that modulates the airflow voice source. As detailed in the preliminary work section, the endoscopic vocal fold images can be segmented adequately because the glottal aperture typically forms a dark contrast to the illuminated foreground of the vocal folds. The image processing steps in Section 4.2.2 will be used to calculate trajectories of lateral vocal fold tissue motion at different locations along the anterior-posterior axis. Information from the full image can be integrated into overall measures of left-right and anterior-posterior asymmetry that could be shown to co-vary with data on visual judgments of left-right and anterior-posterior asymmetry.

Left-right asymmetries were further categorized into phase differences, amplitude differences, frequency differences, and axis shifts during closure. Švec and colleagues (2007) have validated this categorization by observing the different types of left-right asymmetries in videokymographic images recorded from subjects with various vocal pathologies. Motivated by these qualitative observations, quantitative measures will be derived to build on previous research that computed indices of left-right asymmetries from videokymographic data (Qiu *et al.*, 2003). Analysis will be restricted to the categories of phase differences and amplitude differences because of the overall interest in revealing co-variations of physiological properties with synchronous acoustic data. Although quantifiable, frequency differences between the left and right vocal folds are hypothesized to produce acoustic voice signals that exhibit nonlinear and chaotic properties that would not enable linear correlation analysis with acoustic and perceptual parameters. Axis shifts during closure refer to movement of the glottal midline (or axis) within one cycle. Having an unstable midline within and across periods creates an added factor that will be taken into account when calculating left-right asymmetry measures that do not rely on a stable midline.

With the full-frame high-speed video images available, digital kymograms will be generated and processed to extract left-right asymmetry measures related to phase and amplitude. In the following illustration, a prototypical digital kymogram—displaying no left-right differences in amplitude or phase—from a normal speaker is displayed:

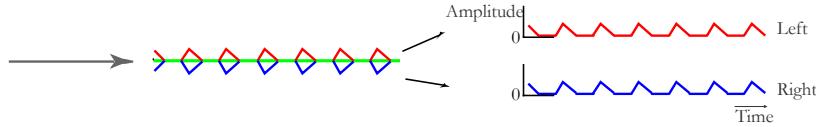


The green line defines zero amplitude at the glottal midline, i.e., the line of mirror symmetry between the left and right vibrating vocal folds. The diagram continues along two branches that

exemplify the differences between amplitude and phase asymmetry. The upper branch shows the effects of amplitude asymmetry on digitally-modified (ie, simulated) kymograms. Amplitude asymmetry refers to differences in the amplitudes (A_1 and A_2) of the left and right vocal folds. Amplitude differences potentially vary from period to period, and this time-varying phenomenon can be quantified as well. The lower branch displays the effects of phase asymmetry on the prototypical kymogram. Phase differences in each period can be quantified by measuring time delays between maxima of the left and right vocal fold amplitudes. This phase asymmetry measure (Δt) is similar to the objective measure of phase asymmetry, where the time delay was normalized by the period duration to obtain an asymmetry percent (Bonilha *et al.*, 2008a). Here phase asymmetries are further distinguished by including whether the left vocal fold vibration lags the right (negative delay) or the right lags the left (positive delay). As with amplitude asymmetry, the phase asymmetry may change from period to period.

Suitable normalization procedures are required to take into account variations in endoscope distance, anatomical differences across subjects, phonation sound pressure level, and the fundamental frequency. In the preliminary study described in Section 4.2, vocal fold amplitude traces were normalized by dividing by the maximum glottal width over the phonatory segment. A consistent definition of the glottal midline is also necessary; the midline is defined by manually marking its endpoints—the anterior commissure and the posterior extent of the membranous vocal folds. The image processing steps, detailed in Section 4.2.2, result in traces of lateral motion of the left and right vocal folds:

Amplitudes of vocal fold edges are calculated with
reference to the glottal midline defined by the user.



Anterior-posterior asymmetries must be defined differently. Above, the computation of left-right asymmetry was possible because a line of symmetry was able to be drawn between two anatomically similar structures, the left and right vocal folds. However, the description of anterior-posterior asymmetry is not well-defined due to anatomical differences of the anterior and posterior structures surrounding the glottis. The vocal folds come together to a point at the anterior commissure but connect posteriorly to arytenoid cartilages whose positions change depending on particular voice qualities and vocal pathologies. Due to this anatomical difference, the amplitude of the vocal fold trace from an anterior kymographic slice cannot be simply compared with the amplitude of a posterior amplitude trace. Therefore, amplitude asymmetries are ill-defined when comparing across the anterior-posterior length of the vocal folds. Phase differences, however, can be defined between waveforms of glottal area computed from the anterior and posterior sides of the mid-membranous glottis. The mid-membranous glottis is defined as half way along the glottal midline. The only method attempting to classify different phonatory segments according to differences in vibratory patterns in the anterior-posterior axis is that of phonovibrography (Lohscheller *et al.*, 2008b). Phase differences between the anterior and posterior glottal area waveforms will be denoted as a measure of anterior-posterior phase difference.

New objective measures computed from high-speed video data will be developed to better correlate with visual judgments of left-right and anterior-posterior asymmetries. Measures will include

- left-right asymmetry measure averaged over three cycles, as in (Bonilha *et al.*, 2008a; Bonilha *et al.*, 2008b)

- left-right amplitude asymmetry index, as in (Qiu *et al.*, 2003)
- left-right phase asymmetry index, as in (Qiu *et al.*, 2003)
- anterior-posterior phase difference measure
- integrative asymmetry measure from three kymographic slices, as in preliminary work and the symmetry waveforms proposed in (Deliyski and Petrushev, 2003)
- integrative asymmetry measure from all kymographic slices, as in the glottal area symmetry waveform in (Deliyski and Petrushev, 2003)

The main differences in the new measurement methods are the use of automatic algorithms and the development of measures that reflect additional types of vocal fold vibratory asymmetry. As a result, it is hypothesized that the integrative measures of asymmetry that take into account both left-right and anterior-posterior asymmetries from full-frame high-speed video images will correlate better with visual ratings of asymmetry on the video sequence. In addition, it is hypothesized that the new integrative asymmetry measures will correlate to a higher degree to visual asymmetry ratings than the manually derived left-right asymmetry measure in (Bonilha *et al.*, 2008a; Bonilha *et al.*, 2008b) because of the inclusion of phase asymmetry information.

5.1.4 Statistical analysis

Pearson's r will then be investigated as in (Bonilha *et al.*, 2008a; Bonilha *et al.*, 2008b), describing the pair-wise correlations between the new image-based asymmetry measures and the already collected visual ratings of left-right and anterior-posterior asymmetry. In addition, Spearman's ρ will be calculated for each pair-wise relationship due to the ordinal nature of the visual judgment scales (integers on a scale from 1 to 5). The visual data, as well as the acoustic measures, will be transformed to ranks to permit correlational analysis using Spearman's ρ . Significance levels and 95% confidence intervals will be computed for the correlation coefficients to enable statistical comparisons between the correlation coefficients obtained.

The following table shows which visual judgments of asymmetry (L-R: left-right, A-P: anterior-posterior) will be compared with the objective asymmetry measures for each visualization used for the judgments:

	High-Speed Video			Digital Kymogram Video			Medial Digital Kymogram	
	L-R <i>rating</i>	A-P <i>rating</i>	Objective <i>measures</i>	L-R <i>rating</i>	A-P <i>rating</i>	Objective <i>measures</i>	L-R <i>rating</i>	Objective <i>measures</i>
Subject 1
Subject 2
...
Subject 52

The non-integrative asymmetry measures that reflect certain types of asymmetry (left-right amplitude/phase asymmetry and anterior-posterior phase differences) allow for multiple linear regression analyses to determine optimal weights on these measures that, when combined, correlate optimally with the visual judgments asymmetry ratings.

5.2 Aim 2

The purpose of Aim 2 is to investigate hypothesized relationships between objective image-based measures of vocal fold vibratory asymmetry and characteristics of the acoustic voice signal in subjects with vocal pathologies. Given tools from Aim 1 for objectively and automatically extracting measures of left-right vocal fold vibratory asymmetries and anterior-posterior phase differences, the co-variations between image-based measures of vocal fold vibratory patterns and acoustic measures of the acoustic voice signal related to vocal fold vibratory asymmetry will be investigated. A new database is currently being compiled of subjects with vocal pathologies using high-quality high-speed video acquisition and synchronous acquisition of the acoustic voice signal (see Section 4.1.1 for system details). The new database aligns the acoustic signal and the video images on a sample-by-sample basis, enabling direct period-by-period comparisons of image-based asymmetry measures to acoustic voice measures.

Many vibratory measures have been developed to make binary distinctions between pathological from normal voices, as opposed to covering a range of acoustic characteristics. A key goal of this project is to appreciate the presence of a certain degree of vibratory asymmetry in normal speakers and begin to understand to what extent characteristics of the acoustic voice signal reflect low levels of asymmetries. Section 2.2 stated the hypothesized relationships between image-based asymmetry measures and proposed acoustic correlates.

5.2.1 *Data collection*

To rely on correlation coefficients as small as 0.2, the maximum number of subjects required to obtain a power of 0.85 at a 95% significance level ($\alpha = 0.05$, two-tailed) is 226. Conversely, to rely on correlation coefficients as large as 0.7, the minimum number of subjects required to obtain a power of 0.85 at a 95% significance level ($\alpha = 0.05$, two-tailed) is 20. To satisfy these statistical requirements, data from a minimum of 40 subjects with vocal pathologies will be collected at the MGH Voice Center. Vocal pathologies of interest include those that affect and perturb the vibratory mechanisms of the vocal folds, including organic lesions (nodules, polyps, and cysts), vocal fold scarring, vocal fold paresis, unilateral vocal fold paralysis, cancer, and glottic incompetence. High-speed videoendoscopy will be performed with a rigid endoscope at rates ranging from 6,000 to 10,000 images per second, depending on the subject's fundamental frequency, laryngeal anatomy, and the desired image quality. For select subjects, a higher frame rate will be required for adequate imaging of vocal fold dynamics (eg, female subjects at higher-than-normal pitch). Images will be acquired at a spatial resolution of 320 horizontal x 352 vertical pixels over an area of about 2.25 cm^2 . Simultaneous recording of the acoustic signal will be obtained using a directional microphone situated approximately 4 cm from the lips at a 45° offset in azimuth.

Subjects will be instructed to produce sustained vowels for about 2 seconds at normal pitch and loudness. For selected subjects who can tolerate it, a phonatory segment at higher pitch will be recorded. Since studies have shown that shimmer and jitter vary with sound pressure level (Orlikoff and Kahane, 1991) and aerodynamic glottal properties are affected by sound pressure level and fundamental frequency differences (Holmberg *et al.*, 1988, 1989; Holmberg *et al.*, 1994), efforts will be made to maintain acoustically-stable recordings of phonation at similar loudness and pitch levels.

5.2.2 Measurement methods

Time-varying asymmetries are expected to play a large role in affecting characteristics of the acoustic voice signal. In a recent study (Nardone, 2007), the physiological correlates of acoustic perturbation measures were investigated by simulating asymmetric vocal fold vibrations using a mathematical model. It was shown that time-varying changes in left-right asymmetry resulted in acoustic jitter and shimmer, whereas time-invariant left-right asymmetries yielded an acoustic output that did not exhibit increased values of jitter and shimmer. To capture the time-varying variations in period duration and amplitude that may reflect variations in vocal fold vibratory asymmetry, traditional acoustic perturbation measures—shimmer, jitter, and harmonics-to-noise ratio—will be employed. Section 4.2.2 stated the formulas for these acoustic measures. In addition, alternative algorithms for computing the harmonics-to-noise ratio will be explored. One method incorporates decomposition of the voice segment into two waveforms representing the additive harmonic and noise waveforms of the segment (Mehta, 2006). This method requires pitch-synchronous analysis of four periods. Another method showing promise computes a perturbation-free measure of HNR (Murphy, 1999), ie, a measure of HNR not influenced by jitter or shimmer.

In addition to quantifying acoustic measures that “perturb” the prototypical periodic waveform, it is hypothesized that variations in vocal fold vibratory asymmetry will affect measures of spectral tilt. Spectral tilt refers to the roll-off or tilt of the frequency spectrum of a sustained voice segment with a higher spectral tilt measure reflecting a steeper decline in the spectral magnitude toward higher frequency levels (Hanson, 1997). The purpose of the spectral tilt measure is to capture the degree of harmonic richness in the signal and thus is strongly related to the magnitude of the glottal excitation. Since the magnitude of glottal excitation is largely due to vocal fold impact forces during the closing phase, vocal fold vibratory asymmetries that affect these impact forces will have a corresponding effect on the spectral tilt measure. The formula for calculating spectral tilt on a voice segment is $H_1 - A_3$, the difference between amplitudes of the first harmonic and third formant in the magnitude spectrum.

Period-to-period measurements of vocal fold vibratory asymmetry enable the ability to formulate short-time algorithms for analyzing shimmer- and jitter-type characteristics. Algorithms are no longer tied to correlating average values over several glottal periods. For example, a measure of vocal fold vibratory asymmetry (eg, left-right amplitude asymmetry) can be calculated from a single period. This measure can then be directly related to an acoustic measure (eg, excitation amplitude) from the period of the acoustic waveform associated with the video-captured period. Thus an entire set of measures obtained from individual glottal cycles in both the image and acoustic domains allows for correlations at a new level.

5.2.3 Statistical analysis

Pearson’s r will be used for analyzing pair-wise relationships between ratio-scaled asymmetry-related measures from high-speed video images and the acoustic waveform. Scatter plots displaying two-dimensional trends and marginal distributions will be generated to visually validate the computed correlation coefficients. Significance levels and 95% confidence intervals will be computed for the correlation coefficients to enable statistical comparisons between the correlation coefficients obtained.

Once overall trends are revealed, acoustic measures will be identified that seem to explain large amounts of variance in the image-based asymmetry metrics. Optimal weights on each acoustic

measure will be calculated using multiple linear regression analysis techniques. Conversely, combinations of image-based vocal fold vibratory asymmetry measures will be explored that may predict acoustic properties of the corresponding voice waveform.

5.3 Aim 3

The purpose of Aim 3 is to determine the impact of different types and degrees of vocal fold vibratory asymmetry on the auditory perception of voice quality. This initial study more directly addresses the clinical reality that voices are assessed by relating vocal fold tissue vibratory patterns to the voice quality of a patient during a standard stroboscopic examination. The current clinical standard in perceptual voice assessment is the consensus auditory-perceptual evaluation of voice (CAPE-V), which provides a standardized procedure for perceptually evaluating abnormal voice quality using visual analog scaling of a closed set of perceptual voice attributes: overall severity of dysphonia, roughness, breathiness, strain, pitch, and loudness (American Speech-Language-Hearing Association, 2002). Each of these qualities constitutes a psychological scale or perceptual dimension of a listener, and the actual impact of a certain type of vocal fold vibratory asymmetry can be estimated by studying the perceptual salience of specific asymmetry measures that were developed in Aim 1.

The proposed approach develops an experimental framework to map the dimension of acoustic characteristics associated with changes in asymmetry to the auditory-perceptual dimension. The framework is based on Thurstone's law of comparative judgment (Thurstone, 1927) that enables the positioning of voices along a continuum on a psychological scale using data from a series of paired-comparison listening tasks. The auditory-perceptual dimensions hypothesized to be maximally influenced by vocal fold vibratory asymmetries are breathiness (strongly related to harmonics-to-noise ratio differences), roughness (linked to irregularities in the acoustic waveforms), and strain (affected by spectral shape characteristics). Perceptually, breathiness refers to "audible air escape in the voice," roughness refers to "perceived irregularity in the voice source," and strain refers to "perception of excessive vocal effort (hyperfunction)" (American Speech-Language-Hearing Association, 2002). Since the auditory-perceptual ratings can be transformed into measures on a ratio scale, correlational analysis can be performed between ratings from each perceptual dimension and the acoustic correlates of vocal fold vibratory asymmetry identified in Aim 2. For this initial study, the two image-based measures of left-right phase asymmetry and anterior-posterior phase differences will be investigated. Stimuli that represent a wide range values for these asymmetry measures will be selected to allow for the mapping of these characteristics onto auditory dimensions of voice quality to reveal the ordering of the stimuli along the perceptual continuum as well as the distance separating the stimuli.

5.3.1 *Data collection and rating methods*

The collection of laryngeal high-speed videoendoscopic data and associated synchronous-acquired acoustic recordings for 40 subjects with vocal pathologies was detailed in Section 5.2.1. Two sets of ten vowel stimuli will be selected to represent a range of values for the two physiological measures of left-right phase asymmetry and anterior-posterior phase difference. Five naïve, normal-hearing listeners will judge the vowel stimuli for breathiness, roughness, and strain by using both a paired-comparison paradigm and a visual analog scale rating method. The use of these two

perceptual rating methods has been shown to provide more information about the auditory-perceptual attributes and mappings than if only one paradigm were used (Meltzner and Hillman, 2005). A computer-based interface will be used to randomize the order of the auditory stimuli before presentation. All paired combinations (45) of the ten stimuli will be presented twice to each listener to assess intra-rater reliability. During each trial, the listener will be forced to select which of two voice samples has perceptually “more of” the dimension (breathiness, roughness, or strain) being assessed. In addition, the listener will be instructed to provide an absolute rating of this selected voice sample on the auditory dimension by using a mouse-controlled visual analog scale:

Mildly deviant	Moderately deviant	Severely deviant
----------------	--------------------	------------------

Thus, the raw data will include ordinal rankings from the five listeners of the 10 voice stimuli for each of the two vocal fold asymmetry types. Additionally, direct ratings of voice quality are available from the visual analog scale ratings.

5.3.2 Statistical analysis

Analysis of paired-comparison judgments

The statistical analysis of the paired-comparison data will follow techniques used in a past study that sought to place synthesized speech stimuli on the psychological scale of naturalness (Meltzner and Hillman, 2005). In that study, Thurstone’s law of comparative judgment (Thurstone, 1927) provided an effective model for analyzing paired-comparison data. In Thurstone’s model, the presentation of a stimulus generated a discriminative process in the listener. Due to the unreliability of human listeners, multiple presentations of the same stimulus yielded multiple discriminative processes, where the most common process elicited was the modal discriminative process and the variation of the elicited processes was termed the discriminative dispersion. To simplify the mathematical framework, Thurstone assumed that these processes followed a normal distribution. Thus the listener’s response on a psychological scale to a given stimulus followed a normal distribution, whose mean and standard deviation were equal to the modal discriminative process and the discriminative dispersion, respectively.

The application of this model to the current study comes from the utility of transforming paired-comparison judgments of several stimuli to absolute locations on a psychological scale (here, the auditory-perceptual scales of breathiness, roughness, and strain). Since listeners are forced to make comparisons between pairs of stimuli, the number of times that the listener selects a given stimulus to be “higher” than another indicate not only that the modal discriminative process (average response) to the given stimulus is higher on the psychological scale, but also that the psychological distance between the two stimuli can be calculated from the responses elicited. For example, if two voices are presented to the listeners and a first voice is more often judged to be “more breathy” than a second voice, it follows that the first voice lies higher on the breathiness scale than the second voice. Furthermore, the degree of breathiness separating the two voices in the listener’s auditory-perceptual dimension can be estimated by analyzing the proportion of times that the first voice was rated “more breathy” than the second voice.

Mathematically, two discriminative processes elicited by two stimuli 1 and 2 can be modeled by two normal distributions $N \sim (\mu_1, \sigma_1)$ and $N \sim (\mu_2, \sigma_2)$, respectively. The psychological distance between these two processes is modeled as a normal distribution with mean $\mu_{12} = \mu_1 - \mu_2$ and standard deviation $\sigma_{12} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2}$, where r_{12} is the correlation between the two stimuli. By defining the normal deviate z_{12} ,

$$z_{12} = \frac{\mu_1 - \mu_2}{\sigma_{12}},$$

the psychological distance between the responses to the stimuli is equal to μ_{12} :

$$\begin{aligned} \mu_{12} &= \mu_1 - \mu_2 \\ &= z_{12}\sigma_{12} \\ &= z_{12}\sqrt{\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2} \end{aligned}.$$

After assuming the equality of all standard deviations, this equation simplifies to

$$\mu_{12} = z_{12}$$

and so the main work for calculating the distance between two modal discriminative processes lies in estimating the theoretical normal deviate z_{12} from empirical data.

The details of calculating the normal deviates (see (Meltzner and Hillman, 2005)) will not be included here; however, in the example above, z_{12} would be proportional to the number of times that stimulus 1 is ranked higher than stimulus 2. In the current empirical analysis, the normal deviates will be calculated for all pair-wise sets of the ten voice stimuli. The entire scale will be shifted so that the psychological position of the lowest ranked stimulus is zero. Thus, the voice stimuli will be placed in a ratio-scaled manner on the auditory-perceptual dimensions of breathiness, roughness, and strain.

Analysis of visual analog scale ratings

The inclusion of visual analog scale ratings in addition to the paired-comparison judgments will provide for an internal reliability check. The distance from the left end of the scale will yield an estimate of breathiness, roughness, and strain in this study. Smaller estimates indicate that there is less of the particular perceptual in the voice stimulus. The differences between the overall positions of the voice stimuli along the visual analog scale will be tested for statistical significance using a one-way analysis of variance for each perceptual attribute.

Intra-rater and inter-rater reliability

The intra- and inter-listener reliability will be evaluated using Pearson's correlation coefficient (r), the percent of exact agreement among the listeners, and intra-class correlations. Ranges, confidence intervals, and statistical significance will be computed for each statistic. Reliability of the placement of the voice stimuli on psychological continua using the paired-comparison data will be assessed using the r statistic suggested in (Kaiser and Serlin, 1978) that models the data in a least squares sense.

6 Use of Humans as Subjects

Data collection from human subjects for this research is approved by institutional review boards at Massachusetts General Hospital (Assurance # FWA00003136) and MIT (Protocol # 0808002847).

Monitoring foreseeable risks and discomforts

Currently, some subjects require the administration of a topical anesthetic to inhibit gagging prior to oral endoscopic procedures. While most patients do not require topical anesthetization, it is available and has been used, as needed, on patients that request it. Participation in the study and the use of high-speed videoendoscopy may necessitate the re-application of topical anesthesia. As with all medications, side effects may include allergic reactions ranging from minor itching or rashes to major, life-threatening reactions.

There may be a slight risk associated with use of a high-intensity light source (300 W Xenon lamp) for recording high-speed videoendoscopy images. The Xenon light sources have been clinically used in laryngeal endoscopy for years, and no adverse effects have been reported to date. As a precaution, however, the exposure of the vocal folds to stationary light is limited to less than one minute, minimizing the potential for injury; further testing, however, is required.

Medical emergencies, should they arise, will be handled by calling the nearest within-hospital emergency care department under an existing agreement to provide emergency medical assistance. The instances will be reported promptly to the appropriate human studies review boards in accordance with established adverse event reporting guidelines. Per hospital regulations for outpatient clinics, 911 will also be called should a medical emergency arise.

Recruitment and selection of subjects

Enrollment includes all male and female adults who are 21 to 64 years old. Both males and females will be included in the subject population due to the historical demographics of patients at the Center. Patients who have been seen at the MGH Center for Laryngeal Surgery and Voice Rehabilitation will be asked by Dr. Hillman if they would be willing to consider participating in this project. Each candidate will be approached by Dr. Hillman during an appointment at the Center or via telephone.

Children will not be included in the data collection because the endoscopic images obtained in children are significantly different from that of the adult with respect to the size of the image, angle of approach of the endoscope relative to the plane of the vocal folds, and the physiology of the vocal folds. Also, voice disorders in children are limited and less serious than in adults (eg, laryngitis versus a more severe pathology).

All patients who meet the selection criteria will be approached about participation regardless of minority status. The equitable inclusion of minorities and women is expected to be satisfied based on historical demographics of the patient population at the Center. No remuneration will be given to patients as all data collection will be performed in the context of a routine voice assessment protocol. The individuals will be given time to read and understand the informed consent forms with opportunities to withdraw at any time during the experiment.

Data and safety monitoring

Video images derived from the high-speed videoendoscopy procedures, acoustic recordings, electroglottography and accelerometer signals, personal information attached to the electronic files, and all identifying information associated with the subject's patient file will be properly monitored and maintained electronically.

Safety monitoring will occur during the entire voice assessment procedure. Data monitoring will occur daily to ensure subject privacy and confidentiality. In addition, Dr. Hillman will determine whether the research should be altered or stopped due to data and safety monitoring concerns.

All of the data gathered will be coded and held in strict confidence, and any resulting reports will not contain information that identifies participants in any way. Data shared with research collaborators will always be de-identified.

7 Literature Cited

- American Speech-Language-Hearing Association (2002). "Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)."
- Berry, D. A., Herzel, H., Titze, I. R., and Story, B. H. (1996). "Bifurcations in excised larynx experiments," *J. Voice* 10(2), 129-138.
- Boersma, P., and Weenink, D. (2007). *Praat: Doing phonetics by computer* (Amsterdam, The Netherlands). Available at <http://www.praat.org>. Accessed on October 2, 2007.
- Bonilha, H. S., Deliyski, D. D., and Gerlach, T. T. (2008a). "Phase asymmetries in normophonic speakers: Visual judgments and objective findings," *Am. J. Speech Lang. Pathol.* **In press**.
- Bonilha, H. S., Gerlach, T. T., Whiteside, J. P., and Deliyski, D. D. (2008b). "Vocal fold phase asymmetries in patients: A study across visualization techniques," *Am. J. Speech Lang. Pathol.* **In review**.
- Deliyski, D. D. (2005). "Endoscope motion compensation for laryngeal high-speed videoendoscopy," *J. Voice* 19(3), 485-496.
- Deliyski, D. D., and Petrushev, P. (2003). "Methods for objective assessment of high-speed videoendoscopy," in *6th International Conference: Advances in Quantitative Laryngology, Voice and Speech Research AQL* (Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany), pp. 1-16.

- Deliyski, D. D., Petrushev, P. P., Bonilha, H. S., Gerlach, T. T., Martin-Harris, B., and Hillman, R. E. (2008). "Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution," *Folia Phoniatr. Logop.* **60**(1), 33-44.
- Döllinger, M., Braunschweig, T., Lohscheller, J., Eysholdt, U., and Hoppe, U. (2003). "Normal voice production: Computation of driving parameters from endoscopic digital high speed images," *Methods of Information in Medicine* **42**(3), 271-276.
- Dresel, C., Mergell, P., Hoppe, U., and Eysholdt, U. (2006). "An asymmetric smooth contour two-mass model for recurrent laryngeal nerve paralysis," *Log. Phon. Voc.* **31**(2), 61-75.
- Eysholdt, U., Rosanowski, F., and Hoppe, U. (2003). "Vocal fold vibration irregularities caused by different types of laryngeal asymmetry," *Eur. Arch. Otorhinolaryngol.* **260**(8), 412-417.
- Farnsworth, D. W. (1940). "High-speed motion pictures of the human vocal cords," *Bell Laboratories Record* **18**, 203-208.
- Gallivan, G. J., Gallivan, H. K., and Eitnier, C. M. (2008). "Dual intracordal unilateral vocal fold cysts: a perplexing diagnostic and therapeutic challenge," *J. Voice* **22**(1), 119-124.
- Granqvist, S., Hertegård, S., Larsson, H., and Sundberg, J. (2003). "Simultaneous analysis of vocal fold vibration and transglottal airflow: Exploring a new experimental setup," *J. Voice* **17**(3), 319-330.
- Haben, C. M., Kost, K., and Papagiannis, G. (2003). "Lateral phase mucosal wave asymmetries in the clinical voice laboratory," *J. Voice* **17**(1), 3-11.
- Hammarberg, B. (1995). "High-speed observations of diplophonic phonation," in *Vocal Fold Physiology: Voice Quality Control*, edited by O. Fujimura, and M. Hirano (Singular Publishing Group, Inc., San Diego), pp. 343-345.
- Hanson, H. M. (1997). "Glottal characteristics of female speakers: Acoustic correlates," *J. Acoust. Soc. Am.* **101**(1), 466-481.
- Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., and Vaughan, C. (1990). "Phonatory function associated with hyperfunctionally related vocal fold lesions," *J. Voice* **4**(1), 52-63.
- Hirano, M., Kurita, S., Yukizane, K., and Hibi, S. (1989). "Asymmetry of the laryngeal framework: A morphologic study of cadaver larynges," *Annals of Otology Rhinology and Laryngology* **98**(2), 135-140.
- Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (1988). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," *J. Acoust. Soc. Am.* **84**(2), 511-529.
- Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (1989). "Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch," *J. Voice* **3**(4), 294-305.
- Holmberg, E. B., Hillman, R. E., Perkell, J. S., and Gress, C. (1994). "Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings," *J. Speech Hear. Res.* **37**(3), 484-495.
- Ishizaka, K., and Flanagan, J. L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell System Technical Journal* **51**, 1233-1268.
- Isshiki, N., Tanabe, M., Ishizaka, K., and Broad, D. (1977). "Clinical significance of asymmetrical vocal cord tension," *Ann. Otol. Rhinol. Laryngol.* **86**(1 Pt 1), 58-66.
- Jiang, J. J., Zhang, Y., and McGilligan, C. (2006). "Chaos in voice, from modeling to measurement," *J. Voice* **20**(1), 2-17.
- Kaiser, H. F., and Serlin, R. H. (1978). "Contributions to the method of paired comparisons," *Applied Psychological Measurement* **2**, 421-430.

- Kay Elemetrics Corporation (2006). *Multi-Dimensional Voice Program (MDVP) Model 5105 Version 3.1.4: Software instruction manual* (Lincoln Park, NJ). Available at Accessed on
- Kiritani, S. (2000). "High-speed digital image recording for observing vocal fold vibration," in *Voice Quality Measurement*, edited by R. D. Kent, and M. J. Ball (Singular Publishing Group, San Diego, CA), pp. 269-283.
- Krane, M. H. (2005). "Aeroacoustic production of low-frequency unvoiced speech sounds," *J. Acoust. Soc. Am.* **118**(1), 410-427.
- Lohscheller, J., Döllinger, M., McWhorter, A. J., and Kunduk, M. (2008a). "Preliminary study on the quantitative analysis of vocal loading effects on vocal fold dynamics using phonovibrograms," *Ann. Otol. Rhinol. Laryngol.* **117**(7), 484-493.
- Lohscheller, J., Eysholdt, U., Toy, H., and Döllinger, M. (2008b). "Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics," *IEEE Trans. Med. Imaging* **27**(3), 300-309.
- Maunsell, R., Ouaknine, M., Giovanni, A., and Crespo, A. (2006). "Vibratory pattern of vocal folds under tension asymmetry," *Otolaryngol. Head Neck Surg.* **135**(3), 438-444.
- Maurer, D., Hess, M., and Gross, M. (1996). "High-speed imaging of vocal fold vibrations and larynx movements within vocalizations of different vowels," *Ann. Otol. Rhinol. Laryngol.* **105**(12), 975-981.
- McGowan, R. S., and Howe, M. S. (2007). "Compact Green's functions extend the acoustic theory of speech production," *Journal of Phonetics* **35**(2), 259-270.
- Mehta, D. (2006). "Aspiration noise during phonation: Synthesis, analysis, and pitch-scale modification," Master of Science thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 2006.
- Meltzner, G. S., and Hillman, R. E. (2005). "Impact of aberrant acoustic properties on the perception of sound quality in electrolarynx speech," *J. Speech. Lang. Hear. Res.* **48**(4), 766-779.
- Murphy, P. J. (1999). "Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis," *J. Acoust. Soc. Am.* **105**(5), 2866-2881.
- Nardone, M. (2007). "Analysis of voice perturbations using an asymmetric model of the vocal folds," Master of Science thesis, Physics, Bowling Green State University, 2007.
- Neubauer, J., Mergell, P., Eysholdt, U., and Herzl, H. (2001). "Spatio-temporal analysis of irregular vocal fold oscillations: Biphasication due to desynchronization of spatial modes," *J. Acoust. Soc. Am.* **110**(6), 3179-3192.
- Niimi, S., and Miyaji, M. (2000). "Vocal fold vibration and voice quality," *Folia Phoniatr. Logop.* **52**(1-3), 32-38.
- Oertel, M. (1895). "Das Laryngo-stroboskop und die laryngo-stroboskopische Untersuchung," *Arch. Laryng. Rhinol.* **3**, 1-16.
- Orlikoff, R. F., and Kahane, J. C. (1991). "Influence of mean sound pressure level on jitter and shimmer measures," *J. Voice* **5**(2), 113-119.
- Qiu, Q., Schutte, H. K., Gu, L., and Yu, Q. (2003). "An automatic method to quantify the vibration properties of human vocal folds via videokymography," *Folia Phoniatr. Logop.* **55**(3), 128-136.
- Schuster, M., Lohscheller, J., Kummer, P., Eysholdt, U., and Hoppe, U. (2005). "Laser projection in high-speed glottography for high-precision measurements of laryngeal dimensions and dynamics," *Eur. Arch. Otorhinolaryngol.* **262**(6), 477-481.

- Schwarz, R., Döllinger, M., Wurzbacher, T., Eysholdt, U., and Lohscheller, J. (2008). "Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model," *J. Acoust. Soc. Am.* **123**(5), 2717-2732.
- Schwarz, R., Hoppe, U., Schuster, M., Wurzbacher, T., Eysholdt, U., and Lohscheller, J. (2006). "Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model," *IEEE Trans. Biomed. Eng.* **53**(6), 1099-1108.
- Shaw, H. S., and Deliyski, D. D. (2008). "Mucosal wave: A normophonic study across visualization techniques," *J. Voice* **22**(1), 23-33.
- Steinecke, I., and Herzel, H. (1995). "Bifurcations in an asymmetric vocal-fold model," *J. Acoust. Soc. Am.* **97**(3), 1874-1884.
- Švec, J. G., and Schutte, H. K. (1996). "Videokymography: High-speed line scanning of vocal fold vibration," *J. Voice* **10**(2), 201-205.
- Švec, J. G., Šram, F., and Schutte, H. K. (2007). "Videokymography in voice disorders: What to look for?" *Ann. Otol. Rhinol. Laryngol.* **116**(3), 172-180.
- The Society of Motion Picture and Television Engineers (2004). "SMPTE 170M-2004. Television--Composite analog video signal--NTSC for studio applications (Revision of SMPTE 170M-1999)," SMPTE Standards.
- Thurstone, L. L. (1927). "A law of comparative judgment," *Psychology Review* **34**, 273-286.
- Titze, I. R. (1995). "Workshop on acoustic voice analysis: Summary statement," (National Center for Voice and Speech, Denver, CO), 1-36.
- van den Berg, J. (1958). "Myoelastic-aerodynamic theory of voice production," *J. Speech Hear. Res.* **1**(3), 227-244.
- Verdonck-de Leeuw, I. M., Festen, J. M., and Mahieu, H. F. (2001). "Deviant vocal fold vibration as observed during videokymography: The effect on voice quality," *J. Voice* **15**(3), 313-322.
- Wurzbacher, T., Dollinger, M., Schwarz, R., Hoppe, U., Eysholdt, U., and Lohscheller, J. (2008). "Spatiotemporal classification of vocal fold dynamics by a multimass model comprising time-dependent parameters," *J. Acoust. Soc. Am.* **123**(4), 2324-2334.
- Wurzbacher, T., Schwarz, R., Döllinger, M., Hoppe, U., Eysholdt, U., and Lohscheller, J. (2006). "Model-based classification of nonstationary vocal fold vibrations," *J. Acoust. Soc. Am.* **120**(2), 1012-1027.
- Yan, Y., Chen, X., and Bless, D. (2006). "Automatic tracing of vocal-fold motion from high-speed digital images," *IEEE Trans. Biomed. Eng.* **53**(7), 1394-1400.
- Yan, Y., Damrose, E., and Bless, D. (2007). "Functional analysis of voice using simultaneous high-speed imaging and acoustic recordings," *J. Voice* **21**(5), 604-616.
- Yan, Y. L., Ahmad, K., Kunduk, M., and Bless, D. (2005). "Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform-based methodology," *J. Voice* **19**(2), 161-175.
- Zeitels, S. M. (1995). "Premalignant epithelium and microinvasive cancer of the vocal fold: The evolution of phonometric surgical management," *Laryngoscope* **105**(3 Pt 2), 1-51.
- Zeitels, S. M., Blitzer, A., Hillman, R. E., and Anderson, R. R. (2007). "Foresight in laryngology and laryngeal surgery: A 2020 vision," *Annals of Otology, Rhinology, and Laryngology Supplement* **198**, 2-16.
- Zhang, C., Zhao, W., Frankel, S. H., and Mongeau, L. (2002). "Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds," *J. Acoust. Soc. Am.* **112**(5 Pt 1), 2147-2154.
- Zhang, Y., Bieging, E., Tsui, H., and Jiang, J. J. (2008). "Efficient and effective extraction of vocal fold vibratory patterns from high-speed digital imaging," *J. Voice* **proof**.

- Zhang, Z., Neubauer, J., and Berry, D. A. (2006). "The influence of subglottal acoustics on laboratory models of phonation," *J. Acoust. Soc. Am.* **120**(3), 1558-1569.
- Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (2002). "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms," *J. Acoust. Soc. Am.* **112**(5 Pt 1), 2134-2146.

8 Committee Agreements

The following committee agreements are included:

- Doctoral Thesis Supervisor Agreement
- Doctoral Thesis Reader Agreement

Harvard-MIT Division of Health Sciences and Technology
Speech and Hearing Bioscience and Technology Program

Doctoral Thesis Supervisor Agreement

To:
HST Graduate Committee Chair
MIT, E25-518

From:
Robert E. Hillman, PhD
Thomas F. Quatieri, ScD
Thesis Co-Supervisors

The program outlined in the proposal:

Title: Investigating the impact of *in vivo* human vocal fold vibratory asymmetries:
Co-variations among measures from laryngeal high-speed videoendoscopy,
acoustic voice analysis, and auditory-perceptual voice assessment of sustained
vowel phonation
Author: Daryush Mehta, SM
Date: November 25, 2008

is adequate for a Doctoral thesis. We believe that an appropriate reader for this thesis would be:

Reader: Dimitar D. Deliyski, PhD, University of South Carolina, whose areas of expertise are the acoustic analysis of voice and the development of laryngeal high-speed videoendoscopy.

Facilities and support for the research outlined in the proposal are available. We are willing to supervise the research and evaluate the thesis report. We further agree to hold a thesis committee meeting at least once per semester to review and guide the student's research.

Signed: _____
Title: Co-Director/Research Director, Center for Laryngeal Surgery
and Voice Rehabilitation, Massachusetts General Hospital;
Associate Professor, Harvard Medical School
Date:

Signed: _____
Title: Senior Member of Technical Staff
Lincoln Laboratory
Date:

Comments:

Harvard-MIT Division of Health Sciences and Technology
Speech and Hearing Bioscience and Technology Program

Doctoral Thesis Reader Agreement

To:
HST Graduate Committee Chair
MIT, E25-518

From:
Dimitar D. Deliyski, PhD
Thesis Reader

The program outlined in the proposal:

Title: Investigating the impact of *in vivo* human vocal fold vibratory asymmetries: Co-variations among measures from laryngeal high-speed videoendoscopy, acoustic voice analysis, and auditory-perceptual voice assessment of sustained vowel phonation
Author: Daryush Mehta, SM
Date: November 25, 2008
Co-Supervisors: Robert E. Hillman, PhD
Thomas F. Quatieri, ScD

is adequate for a Doctoral thesis. I am willing to aid in guiding the research and in evaluating the thesis report as a reader. Specifically, I agree to participate in a thesis committee meeting at least once each semester to review and guide the student's research.

Signed: _____
Title: Associate Professor
Department of Communication Sciences and Disorders
Director, Voice and Speech Laboratory
University of South Carolina
Date:

Comments: