Impact of Human Vocal Fold Vibratory Asymmetries
on Acoustic Characteristics of Sustained Vowel Phonation

by

Daryush Dinyar Mehta

SM, Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2006
BS, Electrical Engineering, University of Florida, 2003

SUBMITTED TO THE HARVARD-MIT DIVISION OF HEALTH SCIENCES AND
TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY IN
SPEECH AND HEARING BIOSCIENCE AND TECHNOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2010

Author ....
Harvard-MIT Division of Health Sciences & Technology
January 29, 2010

Certified by...
Robert E. Hillman, PhD
Associate Professor of Surgery and Health Sciences & Technology, Harvard Medical School
Co-Director/Research Director, Center for Laryngeal Surgery & Voice Rehabilitation
Massachusetts General Hospital
Thesis Co-Supervisor

Certified by...
Thomas F. Quatieri, ScD
Faculty Affiliate of Health Sciences & Technology
Senior Member of Technical Staff, Lincoln Laboratory
Massachusetts Institute of Technology
Thesis Co-Supervisor

Accepted by ....
Ram Sasisekharan, PhD
Director, Harvard-MIT Division of Health Sciences & Technology
Edward Hood Taplin Professor of Health Sciences & Technology and Biological Engineering

Impact of Human Vocal Fold Vibratory Asymmetries on Acoustic Characteristics of Sustained Vowel Phonation

by

Daryush Dinyar Mehta

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on January 29, 2010, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Speech and Hearing Bioscience and Technology

ABSTRACT

Clinical 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 voice produced are not fully understood, and there is recent evidence that the diagnostic significance of asymmetries during vocal fold vibration may be over-interpreted during clinical voice assessment.

An automated system based on high-speed videoendoscopy recordings was developed to objectively quantify vocal fold vibratory asymmetry with initial validation from manual markings and visual-perceptual judgments. Efficient estimation of these measures was possible due to recent technological advances in high-speed imaging of the larynx that enabled the capture and processing of high-resolution video (up to 10,000 images per second) of rapid vocal fold vibrations (100–1000 times per second). Synchronized recordings of the acoustic voice signal were made to explore physiological-acoustic relationships that were not possible using clinical stroboscopic imaging systems.

In an initial study of asymmetric vibration in 14 patients treated for laryngeal cancer, perturbations in the voice signal were most associated with asymmetry that changed across vibratory cycles, while the overall level of asymmetry did not contribute to degradations in voice quality measures. Thus, since stroboscopic imaging is only able to capture vibratory asymmetry that occurs periodically, voice clinicians are not able to observe the time-varying nature of asymmetry that presumably affects acoustic perturbations to a higher degree.

The impact of asymmetric vibration on spectral characteristics was explored in a computational voice production model and an expanded group of 47 human subjects. Surprisingly, in both model and subject data, measures of vocal fold vibratory asymmetry did not correlate with spectral tilt measures. In the subject data, left-right phase asymmetry and closing quotient exhibited a mild inverse correlation. This result conflicted with model simulations in which the glottal area waveform exhibited higher closing quotients (less abrupt glottal closure) with increasing levels of phase asymmetry. Results call for further studies into the applicability of traditional spectral tilt measures and the role of asymmetric vocal fold vibration in efficient voice production.

Thesis Co-Supervisor: Robert E. Hillman, PhD
Title: Associate Professor
Surgery and Health Sciences & Technology

Thesis Co-Supervisor: Thomas F. Quatieri, ScD
Title: Senior Member of Technical Staff
MIT Lincoln Laboratory

Biographical Note

Daryush Dinyar Mehta was born and raised in Kissimmee, Florida. In August 2003, Daryush received a Bachelor of Science degree in electrical engineering with a Minor in music performance (clarinet) from the University of Florida. That Fall, Daryush pursued doctoral studies at MIT in the Speech and Hearing Bioscience and Technology Program within the Harvard-MIT Division of Health Sciences and Technology. In February 2006, he received a Master of Science degree in electrical engineering and computer science from the Massachusetts Institute of Technology while performing research in the speech group at MIT Lincoln Laboratory.

Daryush subsequently became interested in clinical research and initiated a joint collaboration between Lincoln Laboratory and the Center for Laryngeal Surgery & Voice Rehabilitation at Massachusetts General Hospital. As a research assistant on a newly funded NIH research project, he helped in building a comprehensive data acquisition system to aid in determining the efficacy of high-speed video technology for clinical voice assessment.

During his time at MIT, Daryush was a member of the SHBT Admissions Committee, headed the SHBT Distinguished Lecture Series committee (bringing in Christy Ludlow, PhD, and Ray Kent, PhD), and was a Teaching Assistant (Acoustics of Speech and Hearing). He was involved with creating the Voice Quality Study Group and the Voice Center Research Forum to spur discussion and host eminent researchers in the voice and speech area. As side gigs, Daryush performed with the MIT Symphony Orchestra, Longwood Symphony Orchestra, MIT woodwind quintets, and jazz ensembles.

Daryush is a member of the Acoustical Society of America and the Institute of Electrical and Electronics Engineers.

Selected Publications

Hillman, R. E., and Mehta, D. D. (**2010: in press**). "The science of stroboscopic imaging," in *Laryngeal examination: Indirect laryngoscopy to high-speed digital imaging*, edited by K. A. Kendall, and R. J. Leonard (Thieme Medical Publishers, Inc., New York, NY).

Mehta, D., and Quatieri, T. F. (**2005**). "Synthesis, analysis, and pitch modification of the breathy vowel," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY).

Mehta, D. D., Deliyski, D. D., Zeitels, S. M., Quatieri, T. F., and Hillman, R. E. (**2010**). "Voice production mechanisms following phonosurgical treatment of early glottic cancer," Ann. Otol. Rhinol. Laryngol. **119**(1), 1–9.

Mehta, D. D., and Hillman, R. E. (**2008**). "Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods," Curr. Opin. Otolaryngol. Head Neck Surg. **16**(3), 211–215.

Acknowledgments

My primary thanks goes to my co-advisers Bob and Tom, who patiently guided me and allowed me to figure things out for myself during the PhD experience. And also to my colleagues at the MGH Voice Center and Lincoln Laboratory, who created an enjoyable atmosphere in which to perform good research.

Big thanks to Dimitar and Heather at the University of South Carolina, for sharing their high-speed videoendoscopy databases that aided in validating the measures developed in Chapter 2. To Dimitar, thank you for motivating this line of research—to determine the benefits of high-speed imaging that can be translated to the voice clinic. Ultimately, the goal of all this work is to help people. To Heather, special thanks for the many hours of conversation during the sharing of hundreds of video, audio, and data files used for your dissertation.

To Matías, thank you for the countless hours of discussion we had about phonation that have really made me a better scientist and engineer and, especially, for implementing and helping me understand the computational voice production model that was a significant portion of Chapter 4.

On April 7, 2005, the Voice Quality Study Group (VQSG) met for the first time at the Speech Communication Group's conference room at MIT. The group grew out of a desire to form collaborations and maintain the flow of ideas among faculty and students at MGH (Bob, Asa) and Lincoln Laboratory (Tom, Nick and me, as I had just completed my Master's with LL). The first VQSG meeting spawned a series of 90 biweekly discussions (as of November 13, 2009) on various topics related to voice quality research.

Thanks to the SHBT Class of 2003 for waiting patiently for me and for being a Boston-based family for me throughout the graduate years.

Last, but never least, thanks to my family for unmeasurable love and support—to my Mom and Dad and my sisters, Nazneen and Parendi. And…to my wife, Andrea, for offering me her unyielding compassion and motivation.

Contents

Figures

Tables

Chapter 1

Introduction

This thesis project grew out of a critical need to assess the true vibratory characteristics of synthetic and autologous bio-implants that are currently under development by the Center for Laryngeal Surgery & Voice Rehabilitation at the Massachusetts General Hospital. These bio-implants potentially aid in the reconstruction of the phonatory region of the vocal folds known as the superficial lamina propria (SLP). One of the major issues facing voice patients is the formation of a sulcus, or scar, on the vocal folds that often develops as a result of vocal abuse or a necessary side effect of surgical procedures. Vocal fold scars essentially act as a stiff layer in place of SLP tissue, reducing the pliability, or deformability, of the vocal fold during voice production. Healthy vocal fold tissue vibrates freely during sustained phonation because of its ability to deform as air flows from the lungs through the larynx.

The current gold standard in laryngeal imaging involves a system that couples standard video capture technology to endoscopes that allow adequate exposure of the vocal folds through the oral or nasal cavities. Because the vocal folds oscillate on the order of hundreds of times each second at velocities approaching one meter per second (Schuster *et al.*, 2005), standard video frame rates are too slow to capture any meaningful tissue motion and would simply record blurred images. To compensate, laryngeal videostroboscopy systems apply stroboscopic principles to reconstruct illusory vocal fold cycles during vowel production so that clinicians can make visual judgments of vocal fold function (Hillman and Mehta, 2010: in press). Detailed analysis, however, of vocal fold tissue motion is not possible using stroboscopic imaging. See Appendix A for further discussion about laryngeal videostroboscopy.

Laryngeal high-speed videoendoscopy (HSV) has vastly improved over the past decade to enable high-resolution imaging up to 16 000 images per second (Mehta and Hillman, 2008). The MGH Voice Center has been collaborating for several years with Dimitar Deliyski, PhD, at the University of South Carolina, on a long-term project to assess the validity and efficacy of HSV technology in the voice clinic. HSV-derived data provide novel visualizations, such as slow-motion video, digital kymography, mucosal wave kymography, and dynamic three-dimensional reconstruction, that must be better understood before they can prove useful for clinical voice assessment (Deliyski *et al.*, 2008). To perform a large portion of the data collection, we acquired a color high-speed video camera and assembled a comprehensive system for capturing HSV data along with time-synchronized, multi-channel signal acquisition. A myriad of objective measures can be obtained with access to thousands of digital images that provide cycle-to-cycle information of vocal fold vibratory characteristics. Since post-processing and data management is particular challenging with so much data, an extensive MATLAB code base was developed to enable data conversion, video-signal time alignment, and database organization.

The analysis of asymmetries during vocal fold vibration became the central focus of this thesis research due to their inclusion as a presumed indicator of vocal pathology (Bless *et al.*, 1987). In addition, recent studies have shown that even normal voice production can exhibit certain levels of vibratory asymmetry (Haben *et al.*, 2003; Bonilha *et al.*, 2008), thus prompting the question as to the impact of vocal fold vibratory asymmetry on acoustic voice production. These lines of inquiry directly relate to the development of vocal fold bio-implants stated at the outset. It is becoming necessary to better understand the details of how the phonatory mechanism works, with special focus here on how humans exhibit different types and levels of vocal fold vibratory asymmetry so that bio-compatible materials can be formulated with appropriate properties.
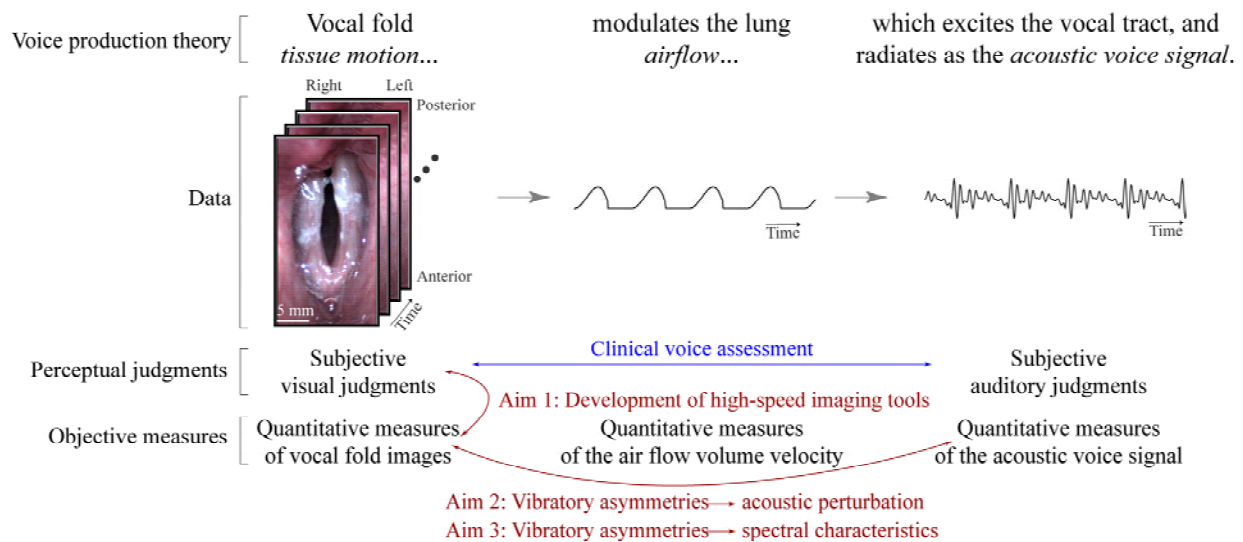
Figure 1.1 **The three study aims of this thesis project placed in the context of voice production theory and clinical voice assessment.**

1.1 Thesis Contributions

This project centered on three studies, which are described in three separate chapters, respectively. Figure 1.1 illustrates the aims of the three studies in the context of voice production theory and clinical voice assessment. Vocal fold tissue motion modulates the lung airflow, which excites the vocal tract, and radiates from the mouth as the acoustic voice signal. Access to vocal fold images and the acoustic voice signal enable visual-perceptual judgments of vocal fold tissue motion and auditory-perceptual assessments of voice quality by voice clinicians. The three aims of this project (with corresponding chapters):

1. To develop an automated signal processing framework to obtain cycle-to-cycle measures of vocal fold vibratory asymmetry from HSV recordings (Chapter 2)

2. To investigate the effects of asymmetric vocal fold vibration on acoustic perturbation measures (Chapter 3)

3. To investigate the effects of asymmetric vocal fold vibration on spectral characteristics of the acoustic voice signal (Chapter 4)

Both hypothesis-driven and data-driven approaches were applied in complementary manners to investigate these three aims. Along the way, an integrative data acquisition system was built to allow for time-synchronized recordings of laryngeal high-speed videoendoscopy data and multi-channel sensor data. This hardware and software framework is used extensively to compute voice function measures in subjects with normal voices and in patients with voice disorders. The

framework shows promise for aiding voice surgeons and clinicians in better integrating visual observations of vocal fold vibratory characteristics with additional objective measures to enhance clinical voice assessment (Hillman *et al.*, 1997).

1.2 Outline

The organization of this thesis is as follows. Each chapter is self contained with its own abstract, background, methods, results, conclusions, and references. Chapter 3 has been published, and Chapters 2 and 4 are organized in the form of manuscripts to be submitted to peer-reviewed journals. A list of all references cited in this document is presented at the end.

Chapter 2 describes the first study, which began with the custom development of a signal processing framework to efficiently compute vocal fold vibratory asymmetry measures derived from HSV data. We were fortunate to have access to databases containing acoustic and HSV data from 52 subjects with normal voices. In particular, the availability of manually derived asymmetry measures and visual judgments of vibratory asymmetry enabled initial cross validations of the automated measures.

Chapter 3 describes the second study that applied the automated measures of vocal fold vibratory asymmetry to a group of 14 patients being managed for glottic cancer at the MGH Voice Center. This initial study explored correlations between the HSV-based measures and acoustic perturbation. Results of this study were recently published (Mehta *et al.*, 2010). The paper is reproduced in this chapter with permission.

Chapter 4 describes the study that addressed the third aim of the thesis project: to investigate the impact of asymmetric vocal fold vibration on the acoustic spectrum. A computational voice production model was implemented to simulate asymmetries during vocal fold vibration. Results of model simulations provided model-based hypotheses that were explored in a group of 47 subjects with and without voice disorders.

Chapter 5 summarizes the major conclusions of the three studies and concludes with directions for future research.

Appendix A provides a brief overview of laryngeal videostroboscopy and the results of a direct comparison with high-speed videoendoscopy.

Appendix B presents supplementary information to Chapter 2. More details are presented regarding the algorithms implemented to measure vocal fold vibratory asymmetry from HSV recordings.

Appendix C provides supplementary information to Chapter 3. In particular, a schematic of the data acquisition system shows how the necessary time synchronization between HSV recordings and multi-channel data was accomplished and validated. Also, a MATLAB-based graphical user interface is shown that provided an integrated visual playback of the video and time-aligned signals.

Appendix D is a supplement to Chapter 4. Results are presented regarding the effects of asymmetric vocal fold vibration on acoustic perturbation in the larger subject sample (cf. results in Chapter 3).

Appendix E includes the experimental protocol followed to collect data from the subjects participating in the studies of Chapters 3 and 4.

Appendix F contains the informed consent form provided to the subjects participating in the studies described in Chapters 3 and 4.

Chapter 2

Automated Measurement of Vocal Fold Vibratory Asymmetry from High-Speed Videoendoscopy Recordings

In prior work, a manually derived measure of vocal fold vibratory phase asymmetry correlated to varying degrees with visual judgments made from laryngeal high-speed videoendoscopy (HSV) recordings. The current investigation extended upon this work by establishing an automated HSV-based framework to quantify three categories of vocal fold vibratory asymmetry. HSV-based analysis provided for cycle-to-cycle estimates of left-right phase asymmetry, left-right amplitude asymmetry, and axis shift during glottal closure for 52 normal speakers producing comfortable and pressed phonation. An initial cross validation of the automated left-right phase asymmetry measure was performed by correlating the measure with other objective and subjective assessments of phase asymmetry. Vocal fold vibratory asymmetry is exhibited to a similar extent in both comfortable and pressed phonations. The automated measure of left-right phase asymmetry strongly correlated with manually derived measures and moderately correlated with visual-perceptual ratings. Correlations with the visual-perceptual ratings remained relatively consistent as the automated measure was derived from kymograms taken at different glottal locations. An automated HSV-based framework for the quantification of vocal fold vibratory asymmetry was developed and initially validated. This framework serves as a platform for investigating relationships between vocal fold tissue motion and acoustic measures of voice function.

2.1 Introduction

The purpose of this chapter is to describe an automated framework that has been developed to obtain measures of vocal fold vibratory asymmetry from laryngeal high-speed videoendoscopy (HSV) recordings. Measurements were designed to provide objective quantification of clinically based visual-perceptual judgments and to also serve as a platform for investigating the effects of asymmetric vocal fold vibration on voice quality.

Normal voice production is typically described as the periodic vibration of left and right vocal folds that mirror each other as they oscillate. Vocal fold vibratory asymmetry generally refers to deviations from this mirroring between the left and right vocal folds during phonation. Recent work has demonstrated that HSV-based methods are required to obtain a comprehensive assessment of asymmetric vocal fold vibration because videostroboscopy is not capable of revealing cycle-to-cycle variations between the left and right vocal fold vibratory patterns (Mehta *et al.*, 2010).

Visual observations of deviations in left-right symmetry have been treated as clinical indicators of vocal fold pathology (Bless *et al.*, 1987) and have been linked to the presence of vocal lesions (Qiu *et al.*, 2003; Gallivan *et al.*, 2008), paralysis (Švec *et al.*, 2007), and scarring (Haben *et al.*, 2003). In addition, certain levels of vocal fold vibratory asymmetry have been exhibited by speakers without voice disorders (Haben *et al.*, 2003; Bonilha *et al.*, 2008; Shaw and Deliyski, 2008). Auditory-perceptual descriptions of voice quality have been implicated as reflections of asymmetric vocal fold vibration, including the perceptions of roughness and breathiness (Niimi and Miyaji, 2000; Verdonck-de Leeuw *et al.*, 2001).

Categories of vocal fold vibratory asymmetry observable with high-speed imaging technology include left-right amplitude asymmetry, left-right phase asymmetry, and axis shift during glottal closure (Švec *et al.*, 2007). The cycle-to-cycle quantification of these measures has already proven critical in an initial assessment of the impact of asymmetric vocal fold vibration on the acoustic voice signal (Mehta *et al.*, 2010).

Alternative HSV-based approaches have been previously described to obtain vibratory measures along the anterior-posterior length of the glottis, including the comprehensive phonovibrogram (Lohscheller and Eysholdt, 2008; Lohscheller *et al.*, 2008b). Phonovibrography is an attempt to represent, in one summary image, the results of detailed measures of the relative opening and closing phases of the left and right vocal folds for all of the individual vibratory cycles in a sample of sustained phonation. A display based on this type of data reduction requires more formal evaluation to assess its clinical validity and utility. In contrast, the current approach aims to

obtain HSV-based measures that are directly related to the way in which voice clinicians visually assess videoendoscopic images.

This chapter describes our current approach for extracting measures of vocal fold vibratory asymmetry from HSV recordings and the results from initial attempts to validate these measures.

2.2   Methods

2.2.1   Subject sample

HSV recordings from a previous study of 52 adult subjects (24 male, 28 female) with normal voice function were available for analysis (Bonilha *et al.*, 2008). HSV data were recorded in monochrome using the High-Speed Video System Model 9700 (KayPENTAX, Lincoln Park, NJ). The frame rate was set to 2000 images per second with maximum integration time, and the spatial resolution was 120 horizontal pixels by 256 vertical pixels for an approximately 2 cm^2 target area. Subjects were instructed to produce sustained vowels for 5 s in both comfortable and pressed manners at normal pitch and loudness. Phonatory segments of 2.2 s in duration were recorded and saved. From each saved token, a stable region of 160 ms (320 images) in duration was extracted for processing.

2.2.2   Objective measures of vocal fold vibratory asymmetry

Figure 2.1 illustrates how lateral displacement waveforms were obtained from HSV recordings. Translational motion artifacts were compensated for so that edge detection was equivalent to motion tracking of the vocal fold edge closest to the glottal midline (Deliyski, 2005). The glottal midline was defined interactively by endpoints at the anterior commissure and the posterior end of the musculo-membranous glottis on the first HSV image capturing maximum vocal fold abduction. All images in the HSV recording were rotated and cropped such that the glottal midline was oriented vertically. A digital kymogram (DKG) was extracted from a horizontal cross section of the HSV images, and cubic interpolation increased the effective DKG sampling rate by a factor of four to enhance edge visualization (Deliyski et al., 2008). A user-defined intensity threshold was selected to segment the glottis on the DKG. Lateral displacement waveforms of the left and the right vocal folds were obtained by tracking the edges defined by the boundary between the glottis and vocal fold tissue regions. See Appendix B for details of this algorithm.

Glottal midline
endpoints defined

Cropping box defined

posterior ↑

Rotate →

Cross section selected

Crop →

anterior ↓

← right   left →

Create digital
kymogram →

Middle digital kymogram with edges detected

left

right

| | | | | | | | |
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 |

Time (ms)

Extract
edges →

Lateral displacement waveforms obtained

—— Left edge
—— Right edge
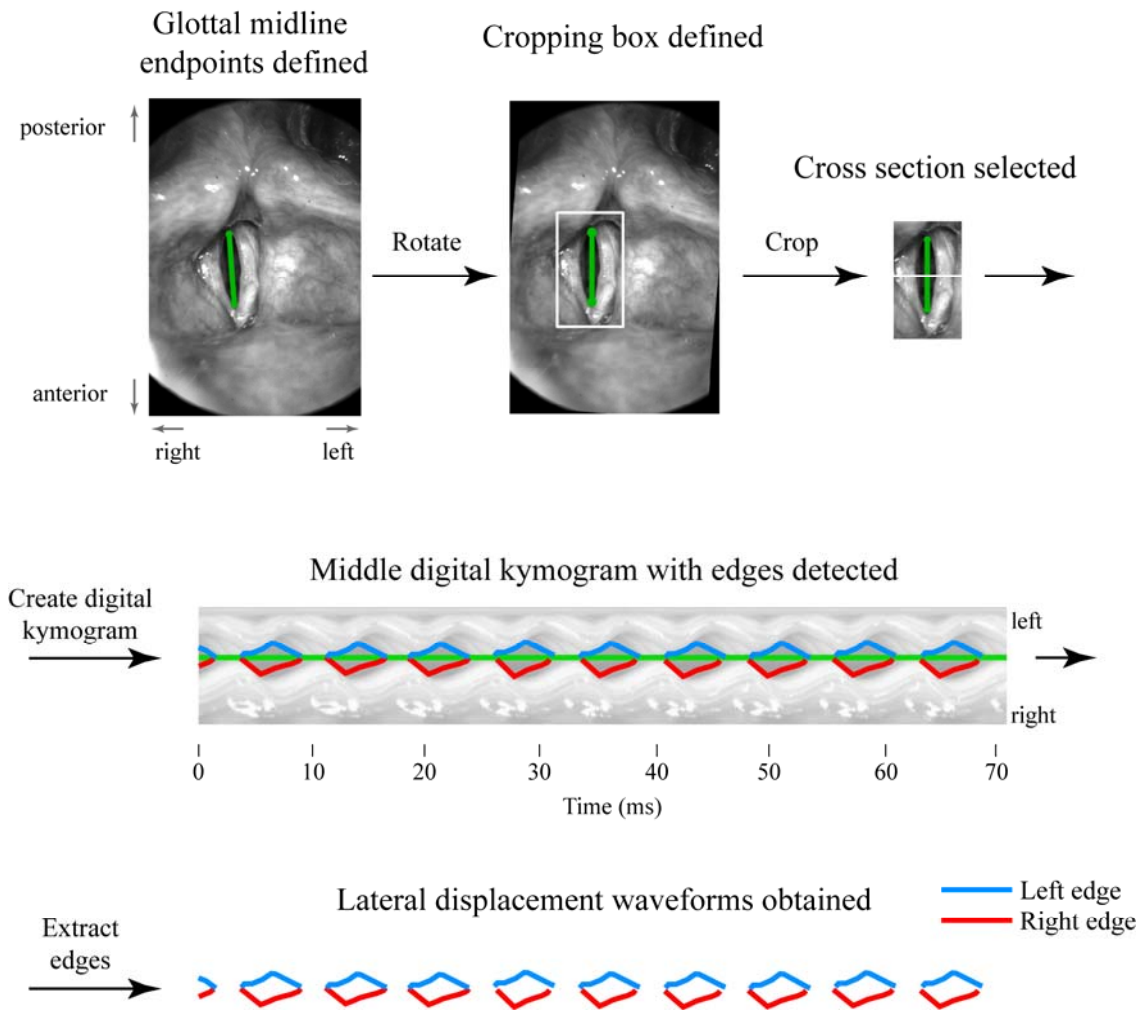
Figure 2.1   Steps for obtaining lateral displacement waveforms from the digital kymogram at a specified horizontal cross section of the glottis.
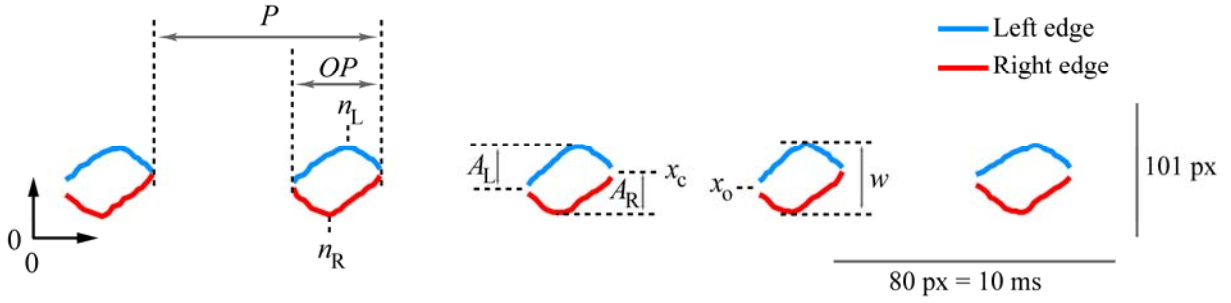
Figure 2.2 **Parameterization of lateral displacement waveforms to obtain measures of vocal fold vibratory asymmetry. See text for definitions of variables.**

Figure 2.2 defines variables used for computing measures of left-right phase asymmetry, left-right amplitude asymmetry, and axis shift during closure. Left-right phase asymmetry PA quantified the phase delay between the left and right vocal folds (Bonilha *et al.*, 2008; Lohscheller *et al.*, 2008b):

$$PA = \frac{n_R - n_L}{P},\tag{2.1}$$

where zero indicated phase symmetry; positive values indicated that the left vocal fold led the right vocal fold in phase, and vice versa. Left-right amplitude asymmetry PA quantified the relative peak-to-peak displacements of the left and right vocal folds (Qiu et al., 2003):

$$AA = \frac{A_L - A_R}{A_L + A_R},\tag{2.2}$$

where zero indicated amplitude symmetry; positive values reflected a relative reduction of the right vocal fold displacement, and vice versa. Finally, the axis shift AS quantified the distance traveled by the vocal folds during glottal closure (Švec et al., 2007):

$$AS = \frac{x_O - x_C}{w},\tag{2.3}$$

where zero indicated no axis shift; positive values indicated that the axis shifted toward the left vocal fold during closure, and vice versa.

Histogram-based analysis and first-order descriptive statistics allowed for comparisons of the prevalence of the different types of vocal fold vibratory asymmetry in the two phonatory conditions.

27

2.2.3 Cross validation

The automated measures were initially validated by determining how well they correlated with other measures based on computing Pearson's *r* statistic at a 95 % confidence level.

PA measures at mid-glottis were compared to results in previous work that obtained a manually derived measure of left-right phase asymmetry from three cycles in the DKG (Bonilha et al., 2008). Pearson's *r* was computed between absolute values of the automated and manually derived measures of left-right phase asymmetry to determine the agreement of these two approaches.

The correlation between *AS* and *PA* was computed to determine the extent of a purported relationship between axis shift and left-right phase asymmetry (Isshiki et al., 1977; Švec et al., 2007).

Finally, data were available from three voice specialists who were instructed to judge the degree of left-right phase asymmetry in each HSV recording using a 5-point rating scale (1 = *completely asymmetrical*, 2 = *severely asymmetrical*, 3 = *moderately asymmetrical*, 4 = *mildly asymmetrical*, 5 = *symmetrical*). Details of the rating procedure have been previously documented (Bonilha et al., 2008). Pearson's *r* measured the relationship between the visual ratings of left-right phase asymmetry and the average magnitudes of *PA*. *PA* measures were computed from each horizontal cross section along the anterior-posterior length of the glottis to determine if particular cross sections received perceptual preference by the raters.

2.3 Results

Figure 2.3 displays the prevalence of the average measures of left-right phase asymmetry *PA*, left-right amplitude asymmetry *PA*, and axis shift during closure *AS* for the comfortable and pressed phonatory conditions.
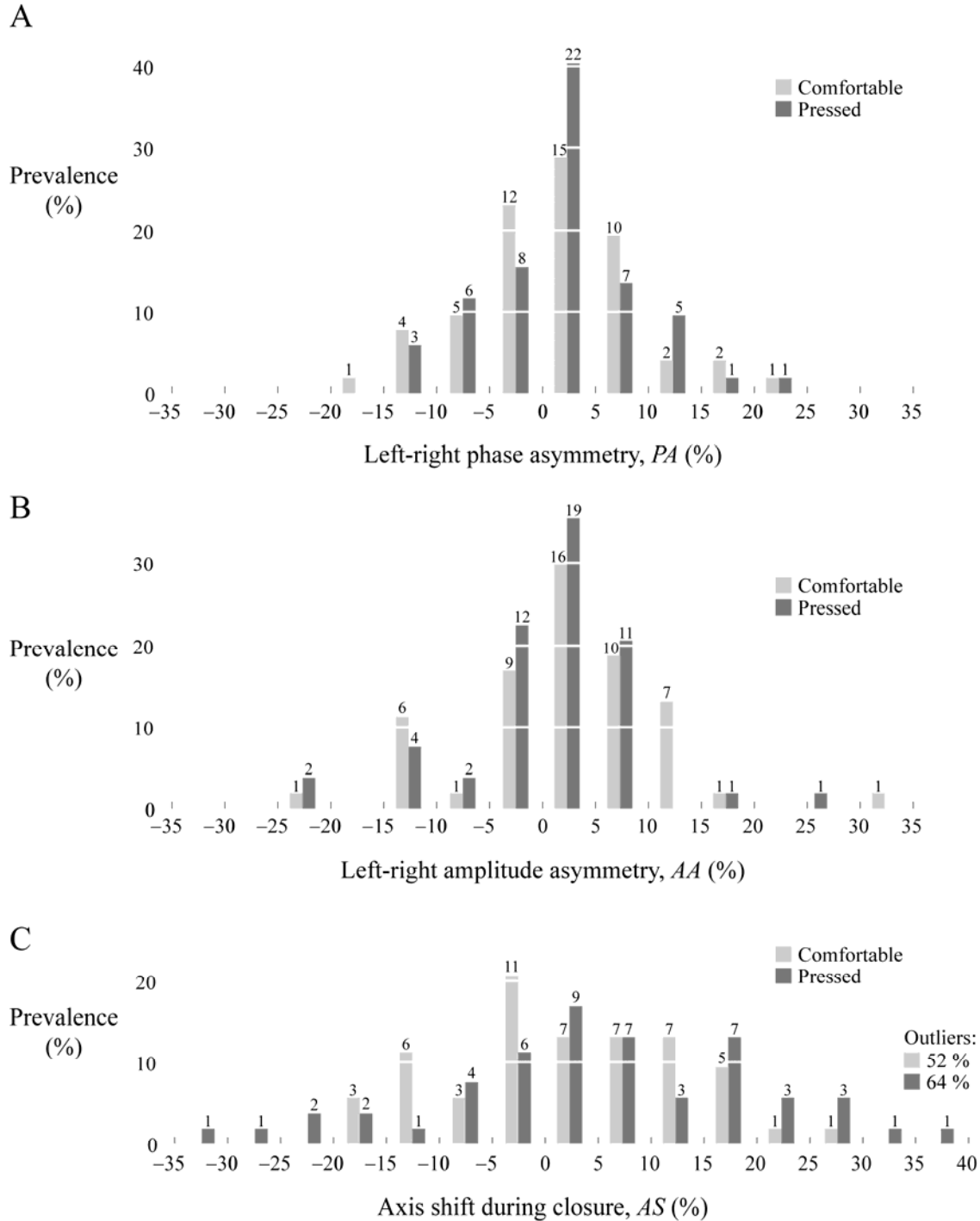
Figure 2.3 Distributions of average measures of HSV-based vocal fold vibratory asymmetry for the two phonatory conditions ($N = 52$): (A) left-right phase asymmetry PA, (B) left-right amplitude asymmetry PA, and (C) axis shift during closure AS. Values above each bar refer to number of subjects. Cf. Figure 5 in (Bonilha *et al.*, 2008).

Table 2.1 shows summary statistics of the magnitudes of the average asymmetry measures. At mid-glottis, there was a strong correlation between the automated measure *PA* and the manually derived measure of left-right phase asymmetry ($r = 0.85$, $p < 0.001$). A strong correlation was also found between average measures of *PA* and *AS* across both subject groups ($r = 0.86$, $p < 0.001$).

Table 2.1 **Summary statistics for HSV-based measures of vocal fold vibratory asymmetry across trials of the comfortable and pressed phonatory conditions. Measures were averaged over all cycles in the middle DKG, and directionality was removed by computing the absolute values prior to finding the mean, standard deviation (SD), and range.**

Measure	Subject group	Mean	SD	Range
Left-right phase asymmetry, *PA* (%)	Comfortable	6.27	4.33	[0.13, 22.2]
	Pressed	5.40	4.63	[0.00, 20.2]
Left-right amplitude asymmetry, *PA* (%)	Comfortable	6.45	4.66	[0.25, 31.5]
	Pressed	5.97	5.92	[0.12, 29.8]
Axis shift during closure, *AS* (%)	Comfortable	10.4	8.30	[0.36, 52.1]
	Pressed	13.5	12.1	[0.16, 64.0]

Table 2.2 lists the results of the cross validation of the average measure of *PA* with visual ratings of left-right phase asymmetry. Correlations obtained were comparable or higher than reported correlations using the manually derived measure of left-right phase asymmetry (Bonilha et al., 2008).

Table 2.2 **Pearson's correlation coefficient *r* ($p < 0.05$) between objective measures of left-right phase asymmetry and visual judgments of left-right phase asymmetry for the two phonatory conditions. The objective measures were the absolute value of average *PA* and the manually derived measure of left-right phase asymmetry, both from the DKG taken halfway between the endpoints of the glottal midline. Cf. Table 5 in (Bonilha *et al.*, 2008).**

Rater	Subject group	*PA*	Manual
1	Comfortable	−.66	−.47
	Pressed	−.84	−.71
2	Comfortable	−.41	−.38
	Pressed	−.51	−.55
3	Comfortable	−.67	−.63
	Pressed	−.67	−.70
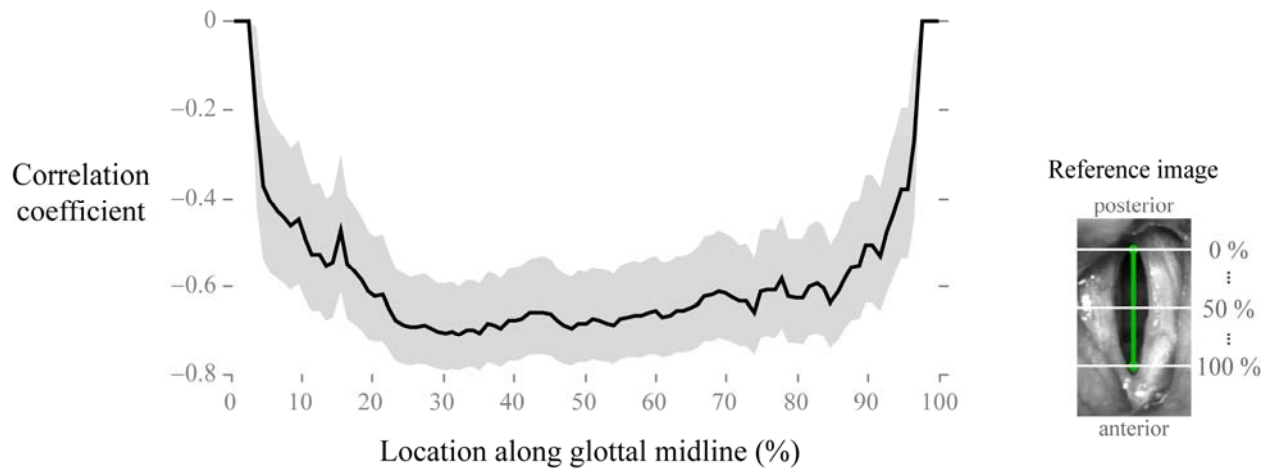
Figure 2.4   Correlation (Pearson's coefficient *r* and 95 % confidence interval) between visual judgments of left-right phase asymmetry and magnitude of average *PA* measured from different horizontal cross sections along the glottal midline (see Reference image). Visual ratings were taken from Rater 1. Data were aggregated across phonatory conditions (N = 101) with 3 HSV recordings considered unsuitable for rating.

Figure 2.4 displays the effect of position along the glottis on the correlation between average *PA* and visual judgments of left-right asymmetry by Rater 1. The strength of the correlation is similar across over half of the glottal area around the mid-glottis. Correlations weaken toward the endpoints of the glottal midline.

2.4 Discussion

The diagnosis of voice disorders depends upon a number of evaluation procedures that describe abnormal voice production to aid in establishing appropriate intervention strategies. Although the relationship between asymmetric vocal fold vibration and voice quality is not well understood, the presence of vocal fold vibratory asymmetry is often considered a clinical indicator of abnormal vocal function that would be linked to any signs or complaints of degraded voice quality and/or inefficient voice production. In the current study, three types of vocal fold vibratory asymmetry were measured from each glottal cycle in an automated HSV-based system. Average values of left-right phase asymmetry, left-right amplitude asymmetry, and axis shift during closure were obtained in a group of normal speakers phonating in comfortable and pressed manners.

The initial attempt at validation produced a strong, but not perfect, correlation between automated (*PA*) and manually derived measures of left-right phase asymmetry ($r = 0.85$, $p < 0.001$).

There were clearly some discrepancies between these two estimates. In some cases, manual markings labeled regions of mucosal wave deformation that were not tracked using the automated system.

Although an improvement over the manual method, the automated system required user interaction to define the glottal midline and set glottal segmentation thresholds. Efficient glottal segmentation has been the subject of intense study (Lohscheller et al., 2007; Yan et al., 2007; Moukalled et al., 2009; Zhang et al., 2010). It was found that a simple threshold-based approach resulted in computational efficiency and robustness. Errant endoscopic angles potentially generated artifacts in measures of vocal fold vibratory asymmetry. Partial exposure of the glottis can affect even effective segmentation algorithms of HSV segments due to arytenoid hooding or laryngeal compression. In addition, significant amounts of mucus on the vocal fold epithelium can prevent accurate estimation of underlying tissue motion.

The three types of vocal fold vibratory asymmetry were observed to similar degrees across phonatory conditions. Results yielded quantitative evidence for a strong correlation between axis shift and left-right phase asymmetry, suggesting a dynamic spatiotemporal relationship during phonation. Visual judgments of left-right phase asymmetry served as another validation of the automated measure of left-right phase asymmetry to reflect the clinical situation where objective measures must be complemented by auditory and visual assessments of voice production (Hillman et al., 1997).

The automated measures of left-right asymmetry (*PA*) maximally explained about 70 % of the variance in the visual judgments. Substantial differences were evident among raters, underscoring the need to incorporate objective measurements of vibratory asymmetry into clinical voice assessment. Intra-rater and inter-rater reliability of auditory-perceptual judgments are known to vary depending on stimulus context and experience. Listener disagreement can stem from drifts in internal representations of a particular quality, scale resolution, the absolute magnitude of the quality, and the presence of anchor stimuli (Kreiman et al., 2007). The visual-perceptual rating task potentially suffered from similar issues. HSV-based measurement could potentially contribute to improving visual rating protocols because of a better understanding of the nature and extent of asymmetric vocal fold vibration.

The correlations between visual judgments and the automated measures were relatively consistent over a wide range of locations along the glottis from which the automated measures were computed. Thus, at least in terms of visual perception, objective correlates of left-right phase asymmetry were similarly estimated from any position around the mid-glottis. Objective measures

estimated near the anterior and posterior endpoints of the glottal midline correlated weakly with the visual judgments due to limited resolution in the lateral displacement waveforms.

The ability to quantify measures of cycle-to-cycle variation in asymmetry using HSV-based approaches has received recent attention (Lohscheller *et al.*, 2008b; Mehta *et al.*, 2010). In particular, average levels of left-right vibratory asymmetry did not necessarily correlate with elevated acoustic perturbation measures in a study on the acoustic effects of asymmetric vocal fold vibration (Mehta et al., 2010). Instead, standard deviations of the measures were shown to correlate moderately with measures of acoustic jitter and shimmer. This sort of detailed study directly relating vocal fold vibratory properties to acoustic measures of voice quality is becoming possible with current imaging technology. The digital kymography–based approach provides a processing framework that has the same temporal and spatial dimensions as the original recorded data, such that the objective measures are always traceable to vocal fold tissue motion. The specific categories of asymmetry measured in this study complement the asymmetry measures obtainable from the phonovibrogram (Lohscheller *et al.*, 2008b).

2.5   Conclusions

An HSV-based automated system for quantifying vocal fold vibratory asymmetries found asymmetry to be prevalent in speakers without voice disorders. Statistically-significant correlations between objective measures of left-right phase asymmetry and visual-perceptual judgments substantiate the use of these types of measures in future investigations into the impact of asymmetric vocal fold vibration on voice quality measures. As high-speed imaging becomes less novel and its properties better understood, the additional temporal information gained by this technology has the potential to complement current diagnostic tools in the voice clinic.

References

Bless, D. M., Hirano, M., and Feder, R. J. (**1987**). "Videostroboscopic evaluation of the larynx," Ear. Nose. Throat J. **66**(7), 289–296.

Bonilha, H. S., Deliyski, D. D., and Gerlach, T. T. (**2008**). "Phase asymmetries in normophonic speakers: Visual judgments and objective findings," Am. J. Speech Lang. Pathol. **17**(4), 367–376.

Deliyski, D. D. (**2005**). "Endoscope motion compensation for laryngeal high-speed videoendoscopy," J. Voice **19**(3), 485–496.

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.

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.

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.

Hillman, R. E., Montgomery, W. W., and Zeitels, S. M. (**1997**). "Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders," Curr. Opin. Otolaryngol. Head Neck Surg. **5**, 172–175.

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.

Kreiman, J., Gerratt, B. R., and Ito, M. (**2007**). "When and why listeners disagree in voice quality assessment tasks," J. Acoust. Soc. Am. **122**(4), 2354–2364.

Lohscheller, J., and Eysholdt, U. (**2008**). "Phonovibrogram visualization of entire vocal fold dynamics," Laryngoscope **118**(4), 753–758.

Lohscheller, J., Eysholdt, U., Toy, H., and Döllinger, M. (**2008**). "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.

Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U., and Döllinger, M. (**2007**). "Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos," Med. Image Anal. **11**(4), 400–413.

Mehta, D. D., Deliyski, D. D., Zeitels, S. M., Quatieri, T. F., and Hillman, R. E. (**2010**). "Voice production mechanisms following phonosurgical treatment of early glottic cancer," Ann. Otol. Rhinol. Laryngol. **119**(1), 1–9.

Moukalled, H. J., Deliyski, D. D., Schwarz, R. R., and Wang, S. (**2009**). "Segmentation of laryngeal high-speed videoendoscopy in temporal domain using paired active contours," in *Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications MAVEBA*, edited by C. Manfredi (Firenze University Press, Firenze, Italy), p. 137–140.

Niimi, S., and Miyaji, M. (**2000**). "Vocal fold vibration and voice quality," Folia Phoniatr. Logop. **52**(1–3), 32–38.

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.

Shaw, H. S., and Deliyski, D. D. (**2008**). "Mucosal wave: A normophonic study across visualization techniques," J. Voice **22**(1), 23–33.

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.

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.

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.

Zhang, Y., Bieging, E., Tsui, H., and Jiang, J. J. (**2010**). "Efficient and effective extraction of vocal fold vibratory patterns from high-speed digital imaging," J. Voice **24**(1), 21–29.

Chapter 3

Voice Production Mechanisms Following Phonosurgical Treatment of Early Glottic Cancer

The following paper was recently published and is reproduced here with permission from the Annals of Otology, Rhinology, and Laryngology. Full citation:

Mehta, D. D., Deliyski, D. D., Zeitels, S. M., Quatieri, T. F., and Hillman, R. E. (**2010**). "Voice production mechanisms following phonosurgical treatment of early glottic cancer," Ann. Otol. Rhinol. Laryngol. **119**(1), 1–9.

Voice Production Mechanisms Following Phonosurgical Treatment of Early Glottic Cancer

Daryush D. Mehta, SM; Dimitar D. Deliyski, PhD; Steven M. Zeitels, MD;
Thomas F. Quatieri, ScD; Robert E. Hillman, PhD

Objectives: Although near-normal conversational voices can be achieved with the phonosurgical management of early glottic cancer, there are still acoustic and aerodynamic deficits in vocal function that must be better understood to help further optimize phonosurgical interventions. Stroboscopic assessment is inadequate for this purpose.

Methods: A newly developed color high-speed videoendoscopy (HSV) system that included time-synchronized recordings of the acoustic signal was used to perform a detailed examination of voice production mechanisms in 14 subjects. Digital image processing techniques were used to quantify glottal phonatory function and to delineate relationships between vocal fold vibratory properties and acoustic perturbation measures.

Results: The results for multiple measurements of vibratory asymmetry showed that 31% to 62% of subjects displayed higher-than-normal average values, whereas the mean values for glottal closure duration (open quotient) and periodicity of vibration fell within normal limits. The average HSV-based measures did not correlate significantly with the acoustic perturbation measures, but moderate correlations were exhibited between the acoustic measures and the SDs of the HSV-based parameters.

Conclusions: The use of simultaneous, time-synchronized HSV and acoustic recordings can provide new insights into postoperative voice production mechanisms that cannot be obtained with stroboscopic assessment.

Key Words: acoustic perturbation, glottic cancer, high-speed videoendoscopy, vocal fold asymmetry, voice production.

INTRODUCTION

Early glottic cancer has been associated with high cure rates since direct endolaryngeal treatment techniques were established almost a century ago.[1] Because it is relatively easy to cure these neoplasms, voice outcome has always been a key metric of success of the treatment. In the mid-20th century, this concern about voice outcome led many surgeons to resect the thyroid lamina lateral to the tumor when a laryngofissure and cordectomy would have provided adequate oncological treatment. The phonatory function of most cordectomy patients reflected aerodynamic glottal incompetence. After removal of the thyroid cartilage,[2] the reconstructed soft tissue collapsed to form a competent neoglottic valve. How-

ever, at present, open laryngeal treatment approaches are seldom done as initial primary treatment.

As radiotherapy became the mainstay of treatment for early glottic cancer, those who espoused surgical management relied primarily on microlaryngeal treatment strategies. Most surgeons, however, lack confidence about obtaining consistent voice outcomes from endoscopic treatment, so radiotherapy remains as the dominant treatment strategy.[3] To enhance the voice outcome while treating early glottic cancer, we performed narrow-margin resections for the more superficial lesions[4,5] and reconstructed the paraglottic space with the laryngeal framework by means of reconstructive methods[6-8] transferred from skills acquired in managing paralytic dyspho-

nia. These later techniques became the key strategy in reestablishing phonatory aerodynamic competency. Most recently, we have been able to further diminish loss of soft tissue from the paraglottic region by employing 532 nm KTP (potassium titanyl phosphate) angiolytic laser treatments.[3] Although we have previously reported that near-normal conversational voices can be achieved with the phonosurgical management of early glottic cancer, there are still acoustic and aerodynamic deficits in vocal function that must be better understood to help further optimize phonosurgical interventions.[3,9]

The purpose of this study was to gain better insight into the voice production mechanisms of patients who have undergone endoscopic phonosurgical treatment of early glottic cancer. The investigation was accomplished with a recently developed system that could acquire digital color images of vocal fold vibration by use of high-speed videoendoscopy (HSV) with time-synchronized recordings of the acoustic voice signal during phonation. This system enabled direct correlations between acoustic parameters and measures of glottal closure and vibratory symmetry extracted from the high-speed digital recordings by means of custom-designed image processing algorithms. The focus of the study was to determine the impact of vocal fold vibratory asymmetries and glottal closure on acoustic perturbation measures used in the clinic. Normative ranges for the HSV-based measures and acoustic parameters were derived from previous studies and provide a descriptive context for the current subject sample.

METHODS

This study used HSV and acoustic voice data from 14 subjects (12 male, 2 female) who had undergone endoscopic phonosurgical management of early glottic cancer. This comprised primarily angiolytic laser treatment with ultranarrow margins that allowed for maximal preservation of paraglottic soft tissue. The average age of the 14 subjects in the current study was 66 years (range, 50 to 84 years). Ten of the subjects were treated with the KTP angiolytic laser, 2 were treated with the pulsed dye angiolytic laser, and 2 were treated with both types of lasers. An HSV assessment was performed within a year of the most recent procedure for all the subjects, except for 2 subjects who were treated more than 2 years before the data acquisition. The subjects were evaluated in a sound-treated room with transoral rigid endoscopy and were instructed to sustain the vowel /i/ at a comfortable and steady pitch and loudness for 2 to 4 seconds.

HSV Data Acquisition. The HSV data were acquired with the Phantom v7.3 high-speed video camera (Vision Research, Inc, Wayne, New Jersey), which enabled high-quality color image capture at fast rates because of a sensitive CMOS (complementary metal-oxide semiconductor) image sensor. A C-mount lens adapter with an adjustable focal length (KayPENTAX, Lincoln Park, New Jersey) was placed between the image sensor and a 70° transoral rigid endoscope (Jedmed, St Louis, Missouri). Images were saved to partitions in a 4 GB onboard memory buffer and downloaded to a personal computer's hard drive via Ethernet after the recordings were complete. The HSV data were recorded at 4,000 or 6,250 images per second with a maximum integration time and a spatial resolution of 320 horizontal × 480 vertical pixels to capture an approximately 2 cm^2 target area. The light source contained a short-arc xenon lamp rated at 300 W (KayPENTAX). The fan-cooled housing produced a collimated beam of light with a color temperature of more than 6,000 K. Three glass infrared (2 dichroic, 1 absorbing) filters blocked infrared light to reduce the thermal energy buildup during endoscopy.

Acoustic Voice Signal Acquisition. The acoustic signal was recorded with a head-mounted, high-quality condenser microphone (model MKE104, Sennheiser Electronic GmbH, Wedemark-Wennebostel, Germany) with a cardioid pattern that offers directional sensitivity and a wideband frequency response. The microphone was situated approximately 4 cm from the lips at a 45° azimuthal offset. The microphone's gain circuitry (model 302 Dual Microphone Preamplifier, Symetrix, Inc, Mountlake Terrace, Washington) offered a low-noise, low-distortion input signal to additional preconditioning electronics (CyberAmp model 380, Danaher Corp, Washington, DC) for gain control and anti-aliasing with low-pass filtering at a 3-dB cutoff frequency of 30 kHz. The analog signals were digitized at a 100 kHz sampling rate, 16-bit quantization, and a ±10 V dynamic range by a PCI digital acquisition board (6259 M series, National Instruments, Austin, Texas).

Time Synchronization of HSV Data and Microphone Signal. Time synchronization of the HSV data and the microphone signal was critical for enabling correlations between voice production measures and acoustic perturbation from the same phonatory segment. The camera clock was supplied by the National Instruments board's clock source, which was synchronized to the sampling of the acoustic signal. The hardware clock division and data acquisition settings were controlled by MiDAS DA software (Xcitex Corporation, Cambridge, Massachusetts). Alignment of the HSV data and the microphone signals was accomplished by recording an analog sig-

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

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

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

Expand images for illustration.

Compensate for translational motion artifacts.

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

Rotate images to make the midline vertical.

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

A digital kymogram is created from the medial position.

The glottis is segmented to determine the vocal fold edges.

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

Consecutive slices are stacked next to each other.

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.

Fig 1. Steps for generating vocal fold amplitude waveforms from which all high-speed videoendoscopy (HSV)–based measures were derived.

nal from the camera that precisely indicated the time of the last recorded image. To compensate for the larynx-to-microphone acoustic propagation time, we shifted the microphone signal into the past by 600 μs relative to the HSV data. This shift duration was approximated from a propagation length of 21 cm (17 cm vocal tract plus 4 cm lip-to-microphone distance) and a speed of sound of 35,400 cm/s. Precise compensation for differences in the vocal tract lengths of each subject was not necessary for the goals of this study, because data were based on averages derived from period-to-period measures across entire phonatory segments. Future investigations into acoustic-physiological correlations on a cycle-by-cycle basis would benefit from a more rigorous compensation scheme.

Signal Processing of HSV Data. Vocal fold tissue motion during phonation was measured by tracking the mediolateral motion of the left and right vocal fold edges closest to the glottal midline. The tracked edges collapsed the 3-dimensional motion of the vocal folds that constituted the mucosal wave to a 2-dimensional space that ignored inferior-superior phase differences.

In this study, the acoustic signal and the HSV images were segmented into approximately 400 ms segments for each subject who exhibited stable phonatory characteristics. Semiautomatic algorithms improved upon manual markings of digital kymograms[10] and included knowledge-based image processing of the kymograms and high-speed video images to extract vibratory asymmetry measures that were compared with properties of the acoustic voice signal. For each segment, vocal fold amplitude waveforms were obtained from the images by the steps illustrated in Fig 1.

The HSV-based measures depended on accurate extraction of vocal fold tissue motion extracted by demarcating the edges of the digital kymograms. Motion artifacts were compensated for[11,12] so that edge detection was equivalent to motion tracking of the medial vocal fold edge. The anterior-posterior axis (glottal midline) was defined interactively on the first image that captured maximum vocal fold abduction. The midline defined the line of symmetry by connecting the anterior commissure to the posterior end of the musculomembranous glottis. All HSV frames were rotated such that the glottal midline was oriented vertically, and a cropping stage framed the glottis to reduce processing time and constrain the image to the most relevant features for accurate edge detection. Using threshold-based edge detection, we derived vocal fold amplitude traces from digital kymograms taken at the midpoint of the musculomembranous glottis, ie, midway between the anterior commissure and the posterior end of the musculomembranous glottis.

Objective Measures of Vocal Fold Vibratory Asymmetry. The HSV-based measures included left-right phase asymmetry, left-right amplitude asymmetry, frequency differences, and axis shifts during closure. Švec et al[13] have validated this categorization by observing these different types of vibratory asymmetries in videokymographic images recorded from subjects with various vocal disorders. Quantitative measures were derived to build on these findings and other studies that computed indices of left-right asymmetries from kymographic data.[10,14] Analyses were restricted to phase differences, amplitude differences, and axis shifts during closure because of the overall interest in revealing covariations of physiological properties with synchronous acoustic data.

HSV-Based Measures. Five HSV-based measures were automatically extracted from the digital kymograms. *Left-right amplitude asymmetry* quantified the difference in the maximum excursions of the left and right vocal fold amplitude traces within a period. The zero amplitude reference was defined as the vocal fold position at the onset of its opening phase. *Left-right phase asymmetry* quantified temporal delays between maximum excursion of the vocal folds. The time delay was normalized by the period duration to obtain a percentage.[10] The degree of *axis shift during closure* calculated the spatial distance, in the mediolateral axis, between onsets of closure and subsequent opening phases, and yielded a percentage after normalization by the maximum glottal width. The HSV-based *open quotient* was the ratio, in percent, between the open phase duration and the

entire period. For each of the above cycle-to-cycle measures, mean and SD values were obtained over the number of cycles in the phonatory segment. Finally, *period irregularity* was computed from the digital kymogram to reflect a jitter-like phenomenon. Period durations were defined as the difference between successive onsets of glottal closure. The absolute difference between successive period durations was calculated, and the period irregularity percentage yielded the average of these absolute differences divided by the mean period duration.

Acoustic Perturbation Measures. Standard acoustic measures of *shimmer*, *jitter*, and *noise-to-harmonics ratio* (NHR) were computed from the synchronously recorded acoustic voice signal with the Multi-Dimensional Voice Program (MDVP, model 5105, KayPENTAX). The acoustic shimmer was defined as the average absolute difference between peak amplitudes of consecutive periods divided by the average peak amplitude during the phonatory segment. Similarly, the acoustic jitter was defined as the average absolute difference between successive periods divided by the average period duration. The NHR reflected the relative spectral energy contributions of the noise component and the harmonic component of the acoustic voice signal.

Comparison With Normative Data. Limited data sets exist for making extensive comparisons among subjects with and without vocal disorders. Normative values for HSV-based measures have only been published for the left-right phase asymmetry parameter based on manual measurements of digital kymograms from 52 subjects.[10] The average age of these subjects was 41 years (range, 18 to 65 years). The subjects were age- and sex-matched, with 24 male and 28 female subjects divided among 3 age ranges: 18 to 33 years, 34 to 49 years, and 50 to 65 years. Normative values for the other HSV-based measures were obtained by reanalyzing the HSV recordings from that normative investigation with the signal processing methods described above. Averaged values (means, SDs, ranges) for the entire normal cohort were used to provide a descriptive context for the data obtained from the postsurgical subjects, because statistical testing showed no age- or sex-based differences in measures within the normal group.

The normative range for the open quotient was defined as the mean ± 2 SD. For all other HSV-based measures, the normative range was defined as values less than or equal to 2 SD above the mean. Normal ranges of measures derived from the normative study were included to provide a descriptive context to which HSV-based measures of the subjects in the current study could be compared. A more rigor-

TABLE 1. SUMMARY OF HSV-BASED MEASURES AND ACOUSTIC MEASURES FOR PATIENT GROUP (N = 13)

Measure	Normals			Patients	
	Mean	*SD*	*Thresholds*	*Mean*	*SD*
HSV-based left-right phase asymmetry (%)	6.27	4.33	≤14.9	14.1	9.10
HSV-based left-right amplitude asymmetry (%)	6.45	4.66	≤15.8	34.1	26.5
HSV-based axis shift during closure (%)	10.4	8.30	≤27.0	21.4	17.4
HSV-based open quotient (%)	49.8	13.4	23.0 to 76.6	54.7	11.4
HSV-based period irregularity (%)	1.84	0.880	≤3.60	1.91	0.882
Acoustic jitter (%)	0.616	0.434	≤1.04	2.06	1.10
Acoustic shimmer (%)	2.21	0.915	≤3.81	4.21	1.67
Acoustic noise-to-harmonics ratio	0.116	0.0123	≤0.190	0.143	0.0246

Normative values for high-speed videoendoscopy (HSV)–based measures were obtained by analyzing recordings from 52 normal subjects.[10] Normative mean, SD, and threshold for each acoustic measure were from Multi-Dimensional Voice Program (MDVP; model 5105, KayPENTAX, Lincoln Park, New Jersey). Male and female data were combined.

ous analysis inferring statistical differences between subjects with and without vocal disorders is a subject for future work in this area. Normative ranges for the acoustic perturbation measures were taken from standard clinical assessment thresholds reported by MDVP.

RESULTS

Data from 1 of the 14 subjects (a 64-year-old man) were eliminated before further statistical assessment of the group results, because the acoustic signal that was recorded during endoscopy was too unstable for standard perturbation analyses (exhibiting type II signal characteristics: sudden qualitative changes in the form of subharmonics or modulation frequencies that interfere with the fundamental frequency[15]). Auditory-perceptual ratings were documented for a majority of the subjects by a certified speech-language pathologist on staff using the CAPE-V protocol.[16] Voices fell within the mild-to-moderate range for overall severity of dysphonia, agreeing with our previous observations that although near-normal conversational voices can be

achieved with phonosurgical treatment, persistent perceptual acoustic deficits require additional investigation.

Table 1[10] shows a summary of HSV-based and acoustic measures for 13 patients relative to normative values. As described above, the normative values for the HSV data were obtained by analyzing recordings from a previous investigation of 52 normal subjects.[10] Normative mean, SD, and threshold for each acoustic perturbation measure were taken from standard clinical assessment values reported by MDVP.

The distributions of all HSV-based measures for the patient group are displayed in Fig 2, with indications of the normal ranges that were defined above. Five of the 13 subjects (38%) exhibited left-right phase asymmetry values that were above the normative threshold. Eight of the 13 subjects (62%) exhibited abnormal amplitude asymmetry values, whereas 4 of the 13 subjects (31%) exhibited axis shift values above the normative threshold. None of the subjects displayed average open quotient or period
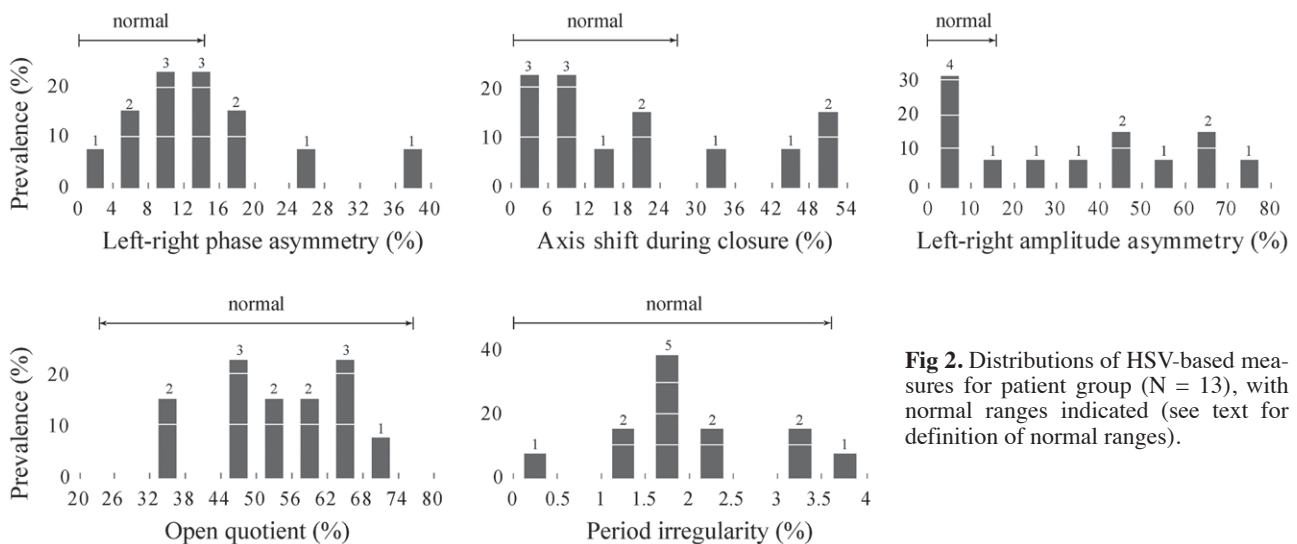
Fig 2. Distributions of HSV-based measures for patient group (N = 13), with normal ranges indicated (see text for definition of normal ranges).

TABLE 2. SIGNIFICANT PAIRWISE CORRELATIONS BETWEEN HSV-BASED MEASURES AND
ACOUSTIC MEASURES FOR PATIENT GROUP (N = 13)

Pairwise Comparison	*Spearman rho*	*p*	*95% Confidence Interval for rho*
Acoustic jitter vs SD of left-right phase asymmetry	0.604	0.03	0.080 to 0.867
Acoustic jitter vs SD of left-right amplitude asymmetry	0.621	0.02	0.106 to 0.873
Acoustic shimmer vs SD of open quotient	0.659	0.01	0.170 to 0.888

irregularity values that were outside the normative range. The average values for the acoustic measures (Table 1) were comparable to those from previous reports on vocal function in this patient group,[3,9] with a postoperative persistence of elevations in jitter and shimmer, and essentially normal NHR values.

Because of the nonnormal marginal distributions of all of the measures and the limited number of subjects in the sample (violating the assumptions for linear correlation analysis using Pearson's *r*), nonparametric statistical analysis using the Spearman rank-order correlation method was performed. Spearman rho values were computed from transformed, rank-ordered measures. Pairwise correlation analysis was performed between HSV-based measures and acoustic voice parameters. Although none of the correlations between average values of the measures were significant, statistically significant correlations ($p \leq 0.05$) were found between the SDs of the HSV-based measures and acoustic perturbation parameters in 3 instances (Table 2). Scatterplots with corresponding regression lines for each of these statistically significant correlations are depicted in Fig 3.

As Table 2 shows, jitter in the acoustic voice signal significantly correlated with the HSV-based measure reflecting the SD of left-right phase asymmetry. The Spearman rho for this correlation was 0.604, indicating that approximately 36% of the variance (ie, rho squared) in jitter was explained by

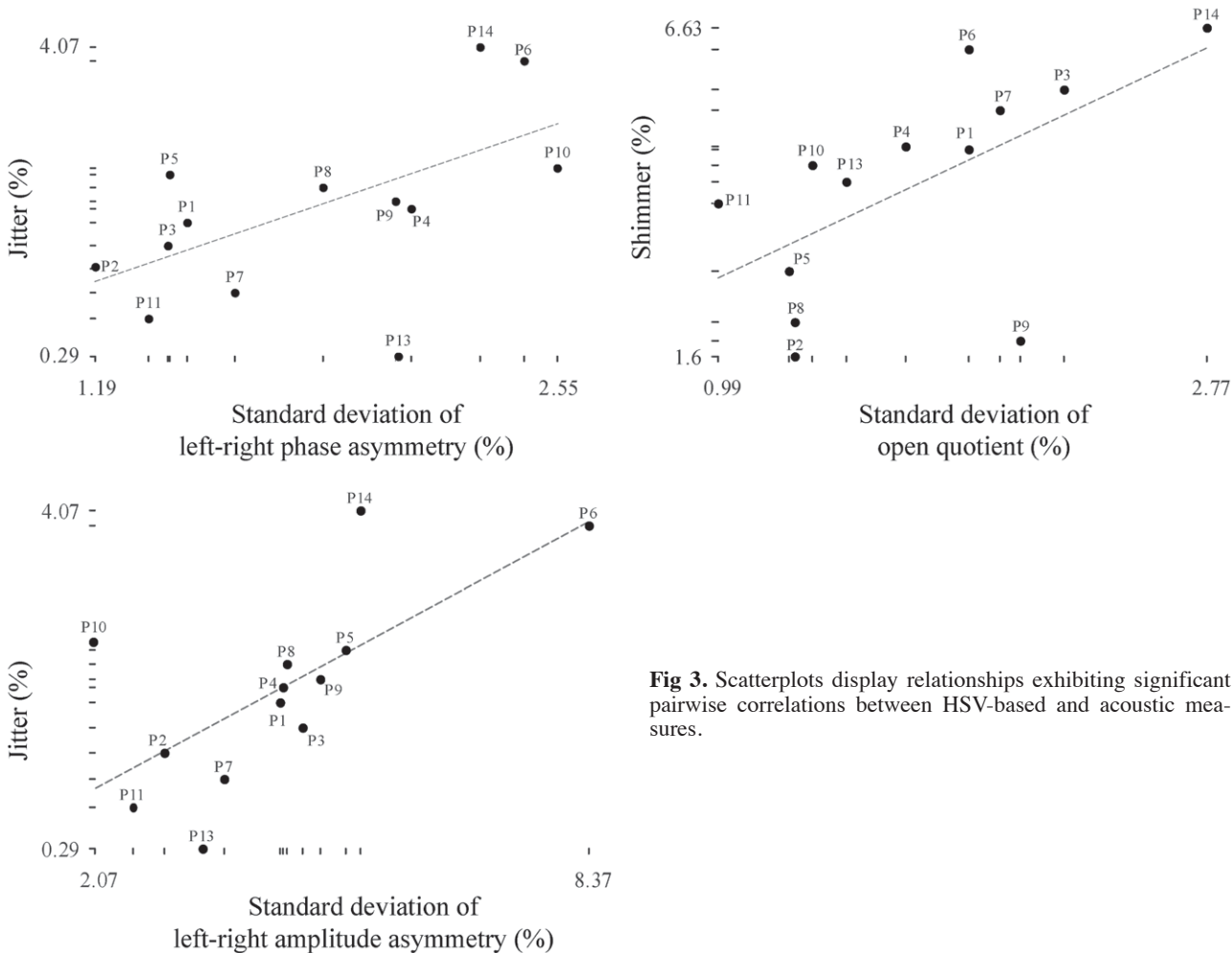**Fig 3.** Scatterplots display relationships exhibiting significant pairwise correlations between HSV-based and acoustic measures.

the variation in left-right phase asymmetry. Acoustic jitter also correlated significantly with the SD of left-right amplitude asymmetry (rho = 0.621). Thus, approximately 39% of the variance in jitter was accounted for by variation in the left-right amplitude asymmetry measure. Finally, shimmer in the acoustic voice signal was most significantly correlated (rho = 0.659) with the HSV-based SD of the open quotient, corresponding to an explained variance of 43%. The NHR did not correlate significantly with any of the HSV-based measures computed.

DISCUSSION

This study used a newly developed HSV-based system to examine voice production mechanisms in patients who had undergone phonosurgical management of glottic cancer. The analyses focused on quantitative measures of glottal closure and vibratory symmetry (ie, vocal fold kinematics) extracted via custom-designed digital image processing routines and on the correlations of these measures with acoustic perturbation parameters.

Taken together, the HSV and acoustic measurement results for this patient group described postoperative voice production mechanisms in which the glottal closure (open quotient) and the overall periodicity of tissue vibration during phonation were within the normal ranges that were defined for this study. The restoration and/or maintenance of normal-like glottal closure durations and overall vibratory periodicity probably contributed to the normal acoustic NHR values that have also been observed previously in similar groups of patients.[3,9] On the other hand, the voice production mechanisms of these patients also displayed higher-than-normal within-cycle asymmetries in vocal fold tissue vibration (phase, amplitude, and closure axis) and, as previously reported,[3,9] postoperative persistence of elevations in acoustic jitter and shimmer.

The results of pairwise correlations between average values for HSV and acoustic perturbation measures revealed that the variations in levels of acoustic jitter and shimmer were not related to within-cycle asymmetries in tissue vibration. Instead, a significant amount of the variation in acoustic jitter was accounted for by the changes (SD) in the symmetry of the phase and amplitude of tissue motion across vibratory cycles, and a significant amount of the shimmer variation was correlated with changes (SD) in glottal closure durations (open quotients) across cycles of vibration. Thus, it appeared that even though the open quotient values for patients fell within the normal range that was defined for this study (mean ± 2 SD), this range was quite broad, with the larger

variations within this range actually being associated with elevated levels of acoustic shimmer.

Time-varying asymmetries were expected to play a large role in affecting characteristics of the acoustic voice signal. In a recent study, the physiological correlates of acoustic perturbation measures were investigated by simulating asymmetric vocal fold vibrations with a mathematical model.[17] 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. This conclusion was supported by the results of the current study.

The apparently critical cycle-to-cycle variations in tissue vibratory behavior that were shown to be significantly correlated with the degradation of the acoustic signal in this investigation would not be reliably revealed with stroboscopy. In fact, because stroboscopy would only provide stable imaging of periodic (repeating) vocal fold motion, it would be capable of only capturing the kind of highly repetitive asymmetries that do not appear to make a major contribution to disrupting acoustic sound generation, assuming that glottal closure is adequate.

The results of this study indicated that the postoperative vocal function of this patient group could be further improved by reducing cycle-to-cycle variations in the phase relationships of vibratory tissue. This might be achievable if it were possible to restore lost pliability to specific regions of the vocal folds, depending on the nature and extent of the resection that was needed to ameliorate the cancer. The need to restore pliability in such cases is one of the major motivations behind recent efforts to develop vocal fold phonatory mucosal implants.[18]

The postresection healed glottal epithelium assumes the biomechanical properties of the underlying scar and/or residual soft tissue of the neocord.[9] Therefore, the overall goal of a superficial lamina propria replacement implant would be to more closely equilibrate the biomechanical properties within and between the vocal folds and thereby reduce sources of cycle-to-cycle asymmetric vibration. Given the limitations described above for stroboscopy, it is clear that HSV imaging will provide insights that will assist surgeons in geographically mapping lost phonatory mucosal pliability to target restoration, as well as assess postreconstruction rheologic characteristics of tissue and their impact on enhancing vocal function.

The HSV data in this study were limited to the midportion of the musculomembranous glottis. This is the typical location for performing videokymog-
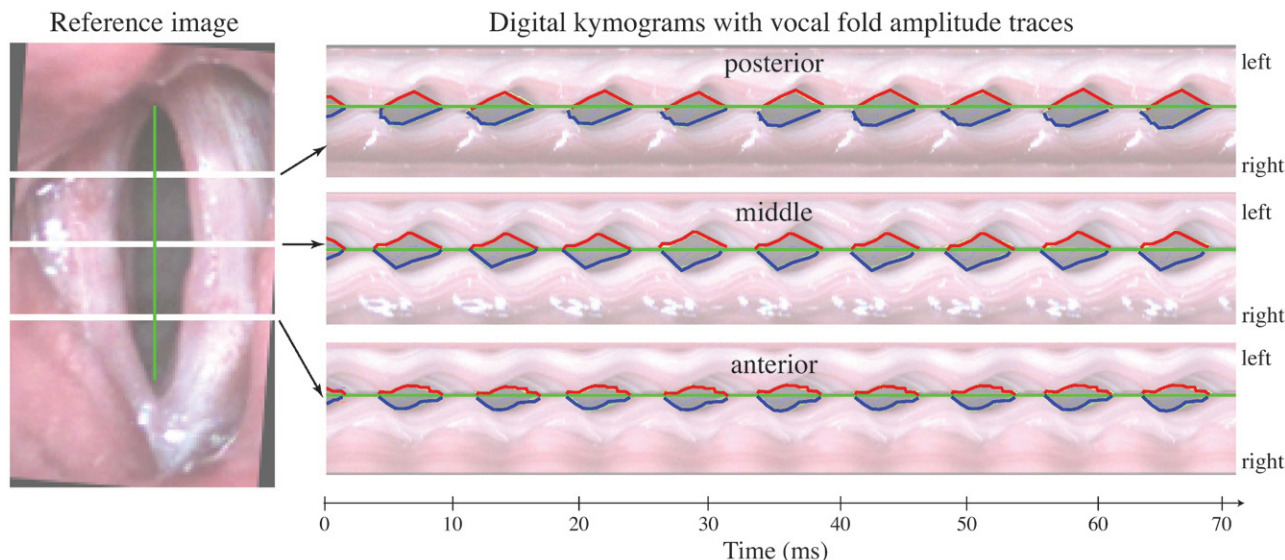
Fig 4. Illustration of derivation of vocal fold amplitude waveforms from digital kymograms extracted from 3 locations along anterior-posterior glottal axis.

raphy, which is only capable of recording activity at 1 location.[13] The differential variation in tissue pliability that can occur within and between the vocal folds after surgical management of glottic cancer can give rise to asymmetries in tissue motion at other points along the glottis. The digital implementation of kymography following post-HSV image capture that was used in this study has the advantage of being able to generate kymographic images (slices) at any location along the entire length of the glottis, thus providing the potential to isolate motion at specific locations of interest. This capability is demonstrated in Fig 4, for which digital kymograms were extracted from 3 glottal locations. Efforts are currently under way to determine how such multiple measures can be combined and/or presented in ways that are meaningful and allow useful comparisons with other simultaneously gathered (time-synchronized) measures of vocal function, including acoustic, electroglottographic, neck skin acceleration, and aerodynamic parameters.

Even though significant correlations were found between the SDs of some HSV measures and acoustic perturbation parameters, the actual correlations were at a moderate level. This means that there is still a sizable amount of variation in acoustic perturbation that is not accounted for by the HSV measures used in this study. It is hoped that a more complete analysis of the HSV imaging data, such as the use of multiple kymograms described above, will increase the amount of acoustic perturbation that can be accounted for. However, it is also possible that there are sources of acoustic perturbation in addition to asymmetries in vocal fold tissue motion, including higher-order aerodynamic phenomena and/or inter-

actions,[19,20] that may play a role during voice production.

CONCLUSIONS

The use of simultaneous, time-synchronized HSV and acoustic recordings can provide new insights into the postoperative voice production mechanisms of patients who have undergone phonosurgical management of vocal disorders, including glottic cancer. In particular, this approach allows for the application of automated digital image and signal processing methods to reveal critical relationships between vocal fold vibratory function and the resulting acoustic characteristics of the voice. Such investigations are not possible with standard clinical stroboscopy methods. This information should substantially assist surgeons in identifying biomechanical phonatory mucosal deficits and assessing the effectiveness of implant reconstruction efforts.

Even though HSV provides much more detailed information about vocal fold phonatory function than does stroboscopy, its eventual adoption into standard clinical practice will depend on the extent to which remaining practical, technical, and methodological challenges can be met. Such challenges include the relatively high cost of HSV systems, management of the large computer files that HSV recordings produce, limitations on the sampling of vocal behaviors due to the brief durations of HSV recordings, and a paucity of solid clinical research that demonstrates that HSV significantly improves the diagnosis and management of voice disorders (ie, clinical trials with controls). Work is already well under way to address many of the remaining barriers to the clinical adoption of HSV.[12]

REFERENCES

1. Lynch RC. Intrinsic carcinoma of the larynx, with a second report of the cases operated on by suspension and dissection. Trans Am Laryngol Assoc 1920;40:119-26.

2. Som ML. Hemilaryngectomy — a modified technique for cordal carcinoma with extension posteriorly. AMA Arch Otolaryngol 1951;54:524-33.

3. Zeitels SM, Burns JA, Lopez-Guerra G, Anderson RR, Hillman RE. Photoangiolytic laser treatment of early glottic cancer: a new management strategy. Ann Otol Rhinol Laryngol Suppl 2008;117(suppl 199):1-24.

4. Zeitels SM. Premalignant epithelium and microinvasive cancer of the vocal fold: the evolution of phonomicrosurgical management. Laryngoscope 1995;105:1-51.

5. Zeitels SM. Phonomicrosurgical treatment of early glottic cancer and carcinoma in situ. Am J Surg 1996;172:704-9.

6. Hartig G, Zeitels SM. Optimizing voice in conservation surgery for glottic cancer. Oper Techn Otolaryngol Head Neck Surg 1998;9:214-23.

7. Zeitels SM, Jarboe J, Franco RA. Phonosurgical reconstruction of early glottic cancer. Laryngoscope 2001;111:1862-5.

8. Zeitels SM. Optimizing voice after endoscopic partial laryngectomy. Otolaryngol Clin North Am 2004;37:627-36.

9. Zeitels SM, Hillman RE, Franco RA, Bunting GW. Voice and treatment outcome from phonosurgical management of early glottic cancer. Ann Otol Rhinol Laryngol Suppl 2002;111(suppl 190):1-20.

10. Bonilha HS, Deliyski DD, Gerlach TT. Phase asymmetries in normophonic speakers: visual judgments and objective findings. Am J Speech Lang Pathol 2008;17:367-76.

11. Deliyski DD. Endoscope motion compensation for laryngeal high-speed videoendoscopy. J Voice 2005;19:485-96.

12. Deliyski DD, Petrushev PP, Bonilha HS, Gerlach TT, Martin-Harris B, Hillman RE. Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution. Folia Phoniatr Logop 2008;60:33-44.

13. Švec JG, Šram F, Schutte HK. Videokymography in voice disorders: what to look for? Ann Otol Rhinol Laryngol 2007;116:172-80.

14. Qiu Q, Schutte HK, Gu L, Yu Q. An automatic method to quantify the vibration properties of human vocal folds via videokymography. Folia Phoniatr Logop 2003;55:128-36.

15. Titze IR. Workshop on acoustic voice analysis: summary statement. Denver, Colo: National Center for Voice and Speech, 1995:1-36.

16. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. Am J Speech Lang Pathol 2009;18:124-32.

17. Nardone M. Analysis of voice perturbations using an asymmetric model of the vocal folds [Thesis]. Bowling Green, Ohio: Bowling Green State University, 2007.

18. Zeitels SM, Blitzer A, Hillman RE, Anderson RR. Foresight in laryngology and laryngeal surgery: a 2020 vision. Ann Otol Rhinol Laryngol Suppl 2007;116(suppl 198):1-16.

19. Murugappan S, Khosla S, Casper K, Oren L, Gutmark E. Flow fields and acoustics in a unilateral scarred vocal fold model. Ann Otol Rhinol Laryngol 2009;118:44-50.

20. McGowan RS, Howe MS. Compact Green's functions extend the acoustic theory of speech production. J Phon 2006;35:259-70.

Chapter 4

Investigating Acoustic Correlates of Human Vocal Fold Vibratory Phase Asymmetry through Modeling and Laryngeal High-Speed Videoendoscopy

Vocal fold vibratory asymmetry is often associated with inefficient sound production through its impact on source spectral tilt. This association was investigated in a computational voice production model and in a group of 47 human subjects. The model provided indirect control over the degree of left-right phase asymmetry within a nonlinear source–filter framework. High-speed videoendoscopy recordings of subjects provided for measurement of vocal fold vibratory asymmetries during phonation. Spectral tilt measures were estimated from the inverse-filtered spectrum of the acoustic voice signal produced. Surprisingly, in both model and subject data, measures of vocal fold vibratory asymmetry did not correlate with spectral tilt measures. Probing further, the glottal area waveform was parameterized to obtain open phase measures (open quotient, plateau quotient) and closing phase measures (speed quotient, closing quotient). In the subject data, left-right phase asymmetry and closing quotient exhibited a mild inverse correlation ($r = -.40$, $p < 0.05$). This result conflicts with model simulations in which the glottal area waveform exhibited less abrupt glottal closure (higher closing quotient) with increasing levels of phase asymmetry. Results call for studies into the applicability of traditional spectral tilt measures and the role of asymmetric vocal fold vibration in efficient voice production.

4.1 Introduction

During sustained vowel phonation, the vocal folds are set into vibration by a combination of muscle tension and aerodynamic forces (van den Berg, 1958). Vocal fold vibratory phase asymmetries have been observed in speakers with normal voices and with voice disorders (Haben et al., 2003; Bonilha et al., 2008), with various factors purported to influence asymmetry within speakers, including subglottal pressure (Berry et al., 1996; Maunsell et al., 2006; Murugappan et al., 2009), phonatory fundamental frequency (Maunsell et al., 2006), vocal fold mass and stiffness (Steinecke and Herzel, 1995), and vocal loading (Lohscheller et al., 2008a).

The goals of the current work were motivated by the clinical need for systematic studies to describe and develop acoustic correlates of vocal fold vibratory asymmetry, potentially aiding clinicians in the effective management of voice disorders. 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 applicable, it is inherently limited to case-by-case observations, and visual judgments of vocal fold vibratory patterns might not adequately reflect changes in objective measures of the acoustic signal (Haben et al., 2003).

A handful of studies have presumably found a strong link between the presence of asymmetric vocal fold vibration and degradations in voice quality. Using acoustic recordings and high-speed videokymography from four voice patients, the auditory perception of roughness was associated with the presence of left-right phase asymmetry (Verdonck-de Leeuw et al., 2001). In a separate study of 22 subjects with voice disorders, ratings of roughness and breathiness were shown to be statistically different for subjects exhibiting symmetric vocal fold vibration versus for subjects exhibiting asymmetric vocal fold vibration (Niimi and Miyaji, 2000). These results suggest that acoustic correlates of vocal fold vibratory phase asymmetry should include acoustic perturbation measures (related to the perception of roughness) and harmonics-to-noise ratio (related to the perception of breathiness).

In a high-speed videoendoscopy study of 14 voice patients, the magnitude of vocal fold vibratory asymmetry did not correlate statistically with levels of acoustic jitter, shimmer, or harmonics-to-noise ratio during sustained vowel phonation (Mehta et al., 2010). Instead, a significant amount of the variation in acoustic jitter was accounted for by the standard deviation of the magnitude of asymmetry over the number of vibratory cycles in the phonatory segment. Furthermore, the harmonics-to-noise ratio was within normal ranges in these subjects due to the

maintenance of glottal closure and vibratory periodicity. Thus, increasing degrees of asymmetric vocal fold vibration were not direct indicators of elevated levels of acoustic perturbation or reductions in harmonics-to-noise ratio.

Laryngeal high-speed videoendoscopy (HSV) methods have enabled the visualization and quantification of vocal fold vibratory asymmetries and the glottal area waveform. Evidence from canine excised larynx experiments suggest that vocal fold vibratory asymmetry is accompanied by an increase in closing quotient, a decrease in closing velocity, and a peak reduction and flattening of the glottal area waveform (Khosla et al., 2008). The presence of a vibratory phase asymmetry, induced via unilateral vocal fold scarring, also accompanied a reduction in flow separation vortices (Murugappan et al., 2009). The modified source excitation produced an increase in spectral tilt of the acoustic pressure waveform at the laryngeal exit.

In the current study, these observations were explored in a large group of subjects and in a computational model of voice production. In a widely applied model of the vocal folds, a set of differential equations simulates self-sustained oscillations by coupling Bernoulli airflow mechanisms to a lumped-element parameterization of the vocal folds (Ishizaka and Flanagan, 1972). Asymmetric vocal fold vibratory regimes have been introduced in adaptations of the two-mass model (Ishizaka and Isshiki, 1976; Steinecke and Herzel, 1995). Each vocal fold is represented by two masses to provide for lower-upper phase differences. Left-right phase vibratory asymmetry is allowed through modifications of the mass, stiffness, and damping properties of each vocal fold. Although the model proposed by Steinecke and Herzel (1995) has been repeatedly used to study asymmetric behavior during phonation (Schwarz et al., 2006; Wurzbacher et al., 2006; Zhang et al., 2006), the model does not account for nonlinear acoustic coupling, which has been shown to affect the source properties (Story and Titze, 1995; Zañartu et al., 2007; Titze, 2008). In the computational model of the current study, nonlinear source–filter coupling between subglottal and supraglottal tracts was incorporated to address this issue.

In a study of high-speed videokymography recordings in a large group of voice patients, vibratory asymmetries were described qualitatively from the lateral displacement exhibited by the left and right vocal folds (Švec et al., 2007). Categories of vocal fold vibratory asymmetries included left-right amplitude asymmetry, left-right phase asymmetry, frequency differences, and axis shifts during glottal closure. Objective methods for quantifying the various types of asymmetry have been suggested in previous work (Deliyski and Petrushev, 2003; Qiu et al., 2003; Bonilha et al., 2008; Mehta et al., 2010).

The purpose of the current study was to better understand the relationships between asymmetric vocal fold vibration and spectral measures of the radiated acoustic pressure waveform. Simulations of asymmetric vocal fold vibration were produced using a mathematical model of the vocal folds that offered indirect control over the phase asymmetry between the left and right vocal folds. Spectral tilt measures were correlated to vocal fold vibratory asymmetry and properties of the glottal area waveform. Data from HSV recordings of human subjects were compared with outputs of the model simulations. Correlational analysis, in addition to illustrative case studies, was performed to document the impact of phase asymmetry on characteristics of the glottal area waveform and acoustic spectral measures of the radiated acoustic pressure waveform.

4.2　Methods

4.2.1　Computational model of voice production

Figure 4.1 schematizes the proposed computational model that extends the original Steinecke and Herzel (1995) model to allow for nonlinear acoustic coupling. Each vocal fold (subscripts $\alpha = l$ for left and $\alpha = r$ for right) was represented by lower and upper coupled oscillators with mass $m_{1\alpha}$ and $m_{2\alpha}$, respectively. Mechanical spring constants $k_{1\alpha}$, $k_{2\alpha}$, and $k_{c\alpha}$, and damping parameters $r_{1\alpha}$ and $r_{2\alpha}$, described viscoelastic tissue properties. The vocal fold displacements from rest $x_{1\alpha}(t)$ and $x_{2\alpha}(t)$ were restricted to the horizontal (mediolateral) axis. Additional spring constants $c_{1\alpha}$ and $c_{2\alpha}$ (not shown in Figure 4.1) were activated during collision of the lower and upper elements, respectively.

Right Vocal Fold Left Vocal Fold

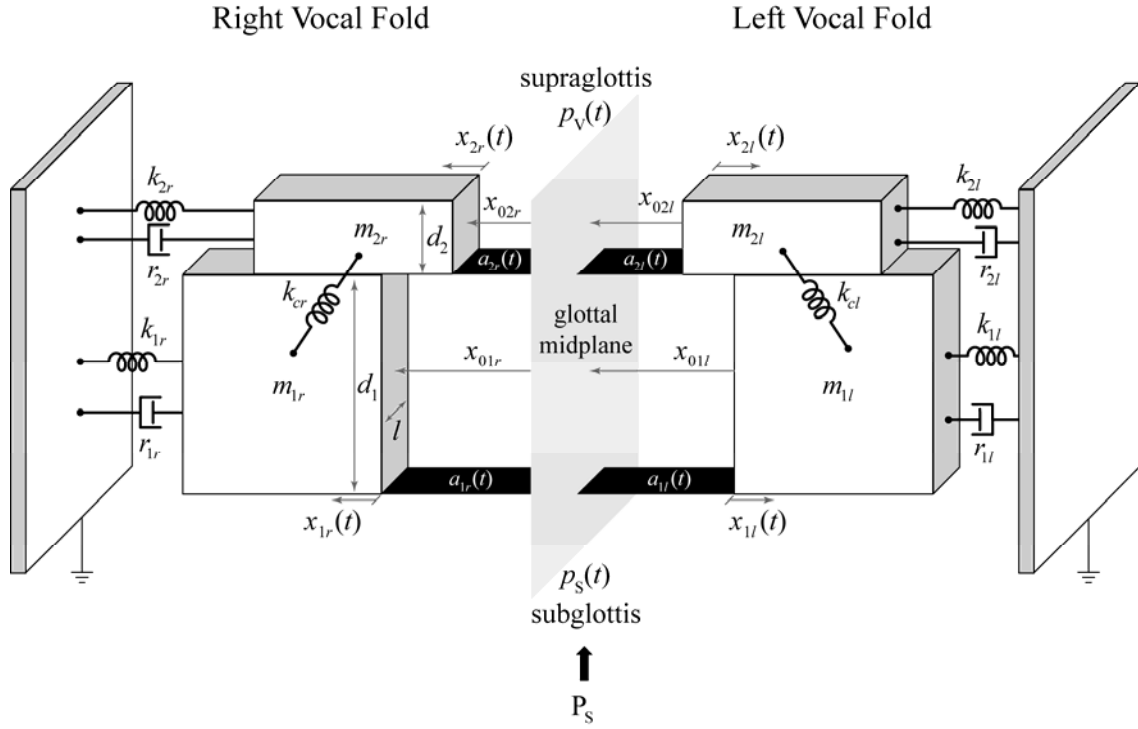**Figure 4.1   Schematic diagram of the lumped-element vocal fold model. See text for details.**

Table 4.1 lists baseline model values that were applied to the left vocal fold parameters:

$$
\begin{aligned}
m_{1l} &= m_1, & m_{2l} &= m_2, \\
k_{1l} &= k_1, & k_{2l} &= k_2, \\
c_{1l} &= c_1, & c_{2l} &= c_2, \\
k_{cl} &= k_c, \\
r_{1l} &= r_1, & r_{2l} &= r_2, \\
x_{01l} &= x_{01}, & x_{02l} &= x_{02}.
\end{aligned}
\tag{4.1}
$$

Table 4.1   Baseline parameter values for the vocal fold model (Steinecke and Herzel, 1995).

Description	Symbol	Units	Baseline value
Mass of lower portion of vocal fold	m_1	g	0.125
Mass of upper portion of vocal fold	m_2	g	0.025
Damping constant of lower element	r_1	$g \cdot s^{-1}$	20
Damping constant of upper element	r_2	$g \cdot s^{-1}$	20
Spring constant of lower element	k_1	$g \cdot cm \cdot s^{-2}$	80 000
Spring constant of upper element	k_2	$g \cdot cm \cdot s^{-2}$	8000
Spring constant between elements	k_c	$g \cdot cm \cdot s^{-2}$	25 000
Thickness of lower element	d_1	cm	0.25
Thickness of upper element	d_2	cm	0.05
Spring constant during collision of lower elements	c_1	$g \cdot cm \cdot s^{-2}$	240 000
Spring constant during collision of upper elements	c_2	$g \cdot cm \cdot s^{-2}$	24 000
Anterior-posterior length of glottis	l	cm	1.4
Rest displacement of lower element	x_{01}	cm	$0.025 / l \approx 0.018$
Rest displacement of upper elements	x_{02}	cm	$0.025 / l \approx 0.018$
Subglottal pressure	P_S	$g \cdot cm^{-1} \cdot s^{-2}$	8000 (≈ 8 cm/H_2O)
Value for measuring gradation of Θ function	x_0	cm^2	0.05

As in the original model, asymmetric vocal fold vibration was produced by applying the factor Q to scale the parameter values of the opposite vocal fold (Steinecke and Herzel, 1995):

$$
\begin{aligned}
m_{1r} &= m_1 / Q, & m_{2r} &= m_2 / Q, \\
k_{1r} &= Q \cdot k_1, & k_{2r} &= Q \cdot k_2, \\
c_{1r} &= Q \cdot c_1, & c_{2r} &= Q \cdot c_2, \\
k_{cr} &= Q \cdot k_c, & & \\
r_{1r} &= r_1, & r_{2r} &= r_2, \\
x_{01r} &= x_{01}, & x_{02r} &= x_{02}.
\end{aligned}
\tag{4.2}
$$

This configuration was used to model a superior laryngeal nerve paralysis and was applied in the current study with the intent of representing imbalances in vocal fold tissue properties due to pathologies such as polyps and glottic cancer. In subsequent model analysis, only Q was changed to simulate the effects of vocal fold vibratory asymmetries.

Four equations of motion described the oscillation amplitudes of the mass elements (time dependencies omitted for clarity):

$$m_{1\alpha}\ddot{x}_{1\alpha} + r_{1\alpha}\dot{x}_{1\alpha} + k_{1\alpha}x_{1\alpha} + \Theta(-a_1)c_{1\alpha}[a_1/(2l)] + k_{c\alpha}(x_{1\alpha} - x_{2\alpha}) = ld_1P_1$$
$$m_{2\alpha}\ddot{x}_{2\alpha} + r_{2\alpha}\dot{x}_{2\alpha} + k_{2\alpha}x_{2\alpha} + \Theta(-a_2)c_{2\alpha}[a_2/(2l)] + k_{c\alpha}(x_{2\alpha} - x_{1\alpha}) = ld_2P_2,$$
(4.3)

where P_1 and P_2 represented the time-varying intraglottal pressures and $a_1 = a_{1l} + a_{1r}$ and $a_2 = a_{2l} + a_{2r}$ represented the time-varying area functions between the lower and upper mass elements, respectively. $\Theta(x) = \tanh[50(x/x_0)]$ when $x > 0$ and $\Theta(x) = 0$ otherwise. Therefore, when a_1 and a_2 were negative (ie, during collision), the additional spring constants $c_{1\alpha}$ and $c_{2\alpha}$ were included in the equations of motion.

To allow for nonlinear acoustic coupling, acoustic pressures were added to the assumed Bernoulli regime below the narrowest part of the glottis and to the jet regime above the glottal narrowing (Story and Titze, 1995). Thus, the equations for the intraglottal pressures P_1 and P_2 are the following:

$$P_1 = (P_S + p_S)\left(1 - \frac{a_{min}}{a_1}\right)^2 \Theta(a_1)$$
$$P_2 = p_V\Theta(a_2),$$
(4.4)

where P_S was the static subglottal pressure, p_S was the acoustic subglottal pressure, p_V was the acoustic supraglottal pressure, and a_{min} was obtained by consecutively computing $a_{min} = \min(a_{1l}, a_{2l}) + \min(a_{1r}, a_{2r})$ and $a_{min} = \max(0, a_{min})$. The oscillation amplitudes $x_{1\alpha}$ and $x_{2\alpha}$ in Eq. (4.3) were solved for using the explicit Runge-Kutta RK5(4) formula with a step size corresponding to a 44.1 kHz sampling rate (Dormand and Prince, 1980). The applied initial conditions forced non-trivial solutions: $x_{1\alpha}(0) = \dot{x}_{1\alpha}(0) = 0.1$, $x_{2\alpha}(0) = \dot{x}_{2\alpha}(0) = 0$.

Nonlinear source–filter interactions were incorporated by coupling vocal fold dynamics to wave reflection analog models of the supraglottal and subglottal tracts. The acoustic pressures were allowed to interact with both glottal airflow (Level 1 interaction) and tissue dynamics (Level 2 interaction) (Titze, 2008). The wave reflection analog model is a time-domain description of the propagation of one-dimensional planar acoustic waves through a collection of uniform cylindrical tubes. The implementation included a radiation impedance (Story and Titze, 1995) and different loss factors for the subglottal and supraglottal tracts (Zañartu et al., 2007).
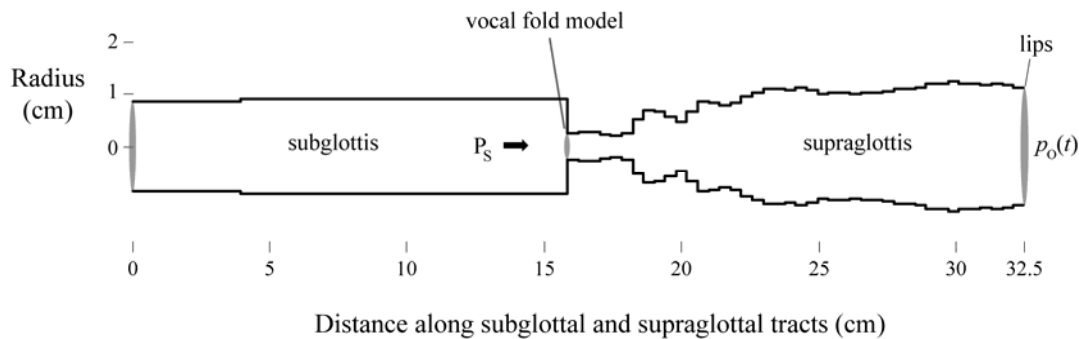
Figure 4.2 **Subglottal and supraglottal tract geometry in the voice production model. For illustration, the lower half of the plot mirrors the top half.**

Figure 4.2 illustrates the geometry of the voice production system. The vocal tract area function was estimated from magnetic resonance images of an adult male sustaining the /ae/ phoneme (Story et al., 1996). The subglottal area function was adapted from respiratory system measurements of human cadavers (Weibel, 1963) and included the trachea, bronchi, and a resistive termination impedance. The interactive airflow model at the glottis was obtained according to Titze (1984), which was originally designed to describe nonlinear coupling using a wave reflection analog scheme.

4.2.2   Human subject data

Subject selection

Data were collected from 47 subjects, 40 individuals with voice disorders (24 male, 16 female) and 7 speakers with no history of voice disorders (4 male, 3 female). Six of the subjects with voice disorders participated before and after voice surgery, yielding a total of 53 recordings across both subject groups. The average age of subjects with voice disorders was 55 years, with a range from 19 years to 85 years. Speakers with normal voices were 33 years old on average, with a range from 20 years to 52 years.

Data acquisition

Subjects underwent laryngeal HSV using transoral rigid endoscopy and were instructed to sustain the vowel /i/ at comfortable pitch and loudness for four seconds. HSV recordings were acquired with the Phantom v7.3 high-speed video camera (Vision Research, Inc., Wayne, NJ),

enabling color imaging at 14-bit quantization with a CMOS image sensor consisting of a Bayer filter array (Deliyski et al., 2008). The image sensor was optically coupled to a 70-degree transoral rigid endoscope having a 10 mm diameter (JEDMED, St. Louis, MO). Coupling was provided by a C-mount lens adapter with a 45 mm focal length (Model 9117, KayPENTAX, Lincoln Park, NJ). Video frame rates were set between 4000 and 6250 images per second with maximum integration time. Spatial resolution was approximately 0.05 mm per pixel. The light source (Model 7152A, KayPENTAX) consisted of a 300 W Xenon lamp in a fan-cooled housing that generated a collimated beam of light with a color temperature of over 6000 K. Three glass infrared filters (two dichroic and one absorbing) served to block infrared light and reduce thermal energy buildup during endoscopy.

The acoustic signal was recorded using a head-mounted condenser microphone with a cardioid pattern and 20 kHz bandwidth (MKE104, Sennheiser electronic GmbH, Wennebostel, Germany). The microphone was positioned about 4 cm from the lips at a 45-degree azimuth. Preamplifier circuitry (302 Dual Microphone Preamplifier, Symetrix, Inc., Mountlake Terrace, WA) offered a low-noise, low-distortion input signal to additional preconditioning electronics that provided for anti-aliasing with low-pass filtering at a 3 dB cutoff frequency of 30 kHz and pre- and post-filter gain control (CyberAmp 380, Danaher Corporation, Washington, DC). The conditioned analog signals were digitized at 100 kHz sampling rate, 16-bit quantization, and ± 10 V dynamic range by a PCI data acquisition board (6259 M series, National Instruments, Austin, TX).

Time synchronization of the HSV recordings and the acoustic signal was enabled by generating a clock source common to the camera and data acquisition board. Hardware clock division and signal acquisition settings were controlled in software (MiDAS DA, Xcitex Corporation, Cambridge, MA). A camera control signal marked the time of a recorded image to within 11 µs. The microphone signal was time-shifted into the past by 600 µs relative to the HSV recording to compensate for the larynx-to-microphone acoustic propagation time. Precise compensation for this delay was not necessary for the purposes of computing average values over several glottal cycles.
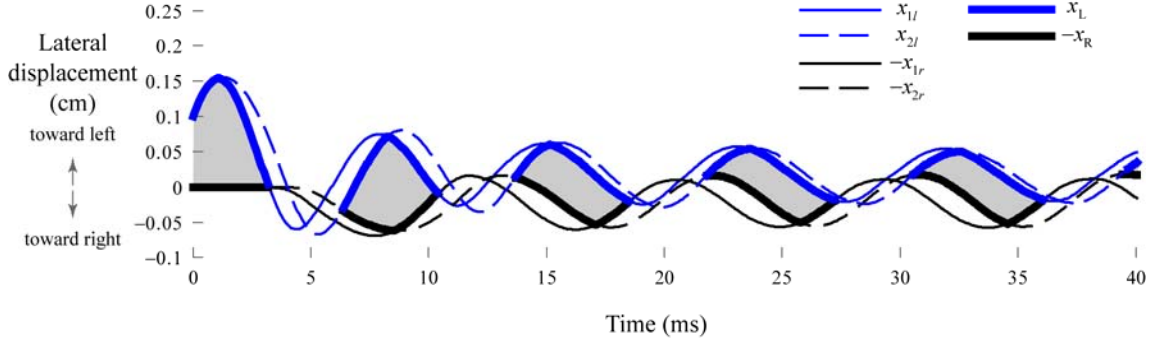
Figure 4.3   Illustration of how lateral displacement waveforms of the left and right vocal folds (x_L and x_R, respectively) are derived from a model simulation with $Q = 0.75$. Gray regions indicate times of positive glottal area.

4.2.3   Data analysis

Pre-processing of model waveforms

The duration of each model simulation was 500 ms. The initial 250 ms segment was discarded to neglect transient effects. Waveforms were resampled to an 8 kHz sampling rate. Figure 4.3 illustrates the derivation of lateral displacement waveforms x_L and x_R, which were only defined when a_min was positive:

$$
\begin{aligned}
x_\mathrm{L} &= \begin{cases} \min(x_{1L},\, x_{2L}), & a_\mathrm{min} > 0 \\ \text{undefined}, & \text{otherwise} \end{cases} \\
x_\mathrm{R} &= \begin{cases} \min(x_{1R},\, x_{2R}), & a_\mathrm{min} > 0 \\ \text{undefined}, & \text{otherwise,} \end{cases}
\end{aligned}
\tag{4.5}
$$

where

$$
\begin{aligned}
x_{1L} &= \frac{a_{01}}{2l} + x_{1l}, & x_{2L} &= \frac{a_{02}}{2l} + x_{2l} \\
x_{1R} &= \frac{a_{01}}{2l} + x_{1r}, & x_{2R} &= \frac{a_{02}}{2l} + x_{2r},
\end{aligned}
\tag{4.6}
$$

and l was the anterior-posterior length of the glottis. The onsets of glottal opening and closure do not necessarily occur at the glottal midplane. (Figure 4.1). The glottal area waveform was equal to

a_{min} and assumed a rectangular glottal aperture. The acoustic voice signal was taken as the radiated acoustic pressure p_O at the "lips" of the model (Figure 4.2).

Pre-processing of subject data

Stable phonatory segments between 320 ms to 400 ms in duration were selected from each HSV recording and corresponding acoustic voice signal. Motion artifacts were compensated for so that edge detection was equivalent to motion tracking of the vocal fold edge closest to the midline (Deliyski, 2005). The glottal midline (anterior-posterior axis) was defined interactively on the first image capturing maximum vocal fold abduction (Mehta et al., 2010). Endpoints of the midline indicated the anterior commissure and the posterior end of the musculo-membranous glottis. All HSV images were rotated such that the glottal midline was oriented vertically. A cropping stage framed the glottis to reduce processing time.

Lateral displacement waveforms x_L and x_R were derived from a single digital kymogram (DKG) taken halfway between the endpoints of the glottal midline. The DKG was converted to a monochromatic color space by keeping red-channel information. A user-defined intensity threshold segmented pixel regions of the relatively dark glottis from tissue region. Upper and lower edges of the glottal segmentation defined x_L and x_R, respectively. As in the model, the lateral displacement waveforms were only defined when glottal closure was exhibited. See Appendix B for details of the threshold-based segmentation. The glottal area waveform a_{min} was derived from the full-frame HSV images through threshold-based glottal segmentation using an intensity threshold for the entire phonatory segment.

A phonovibrogram (PVG) was generated to aid in visualizing vocal fold displacement at different positions along the anterior-posterior length of the glottis (Lohscheller et al., 2008b). The PVG was used to provide qualitative information regarding the uniformity of left-right phase and amplitude asymmetries in the subject data. PVGs of the model outputs were not necessary to compute due to the assumption of a rectangular glottal aperture.
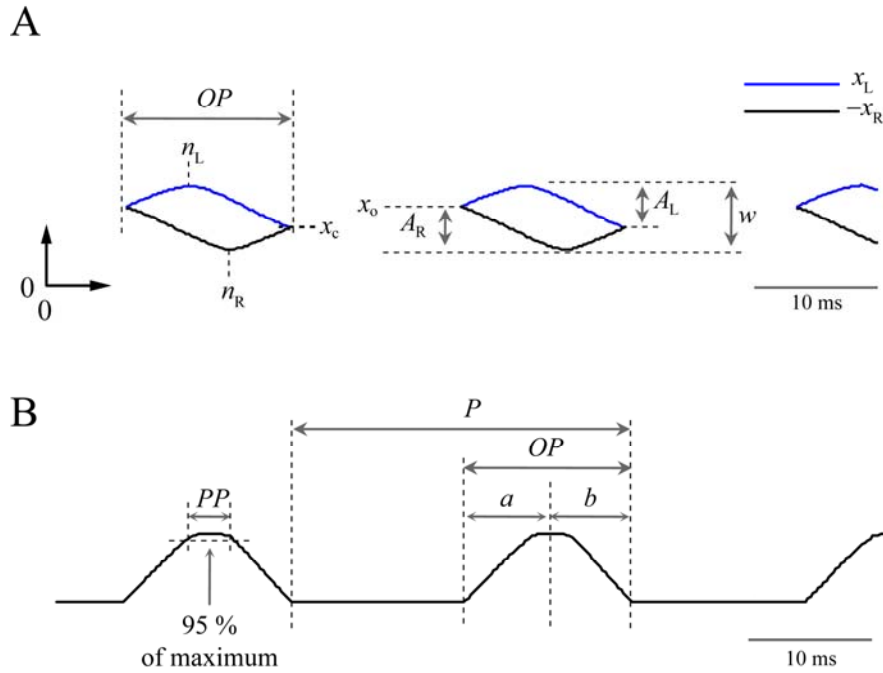
Figure 4.4 **Parameterization of (A) lateral displacement waveforms and (B) glottal area waveforms to obtain measures of vocal fold vibratory characteristics. See text for details.**

Measures of vocal fold vibratory asymmetry

Figure 4.4A illustrates the parameterization of the lateral displacement waveforms to obtain measures of left-right phase asymmetry, left-right amplitude asymmetry, and axis shift during closure. Left-right phase asymmetry PA quantified the normalized phase delay between the left and right vocal folds (Bonilha et al., 2008; Lohscheller et al., 2008b):

$$PA = \frac{n_R - n_L}{OP},$$
(4.7)

where zero indicated phase symmetry; positive values indicated that the left vocal fold led the right vocal fold in phase, and vice versa. Left-right amplitude asymmetry AA quantified the relative peak-to-peak displacements of the left and right vocal folds (Qiu et al., 2003):

$$AA = \frac{A_{\mathrm{L}} - A_{\mathrm{R}}}{A_{\mathrm{L}} + A_{\mathrm{R}}}, \tag{4.8}$$

where zero indicated amplitude symmetry; positive values reflected a relative reduction of the right vocal fold displacement, and vice versa. Finally, the axis shift *AS* quantified the distance traveled by the vocal folds during glottal closure (Švec et al., 2007):

$$AS = \frac{x_O - x_C}{w}, \tag{4.9}$$

where zero indicated no axis shift; positive values indicated that the axis shifted toward the left vocal fold during closure, and vice versa.

Glottal area waveform measures

Figure 4.4B illustrates the parameterization of the glottal area waveform that was motivated by studies on acoustic correlates of glottal characteristics using inverse-filtered airflow waveforms (Holmberg et al., 1988). Open quotient *OQ* (ratio between open phase duration and period), speed quotient *SQ* (ratio between opening and closing phase durations), and closing quotient *CQ* (ratio between closing phase duration and period) were computed from each cycle:

$$OQ = \frac{OP}{P}, \; SQ = \frac{a}{b}, \; CQ = \frac{b}{P}. \tag{4.10}$$

An additional measure, termed plateau quotient *PQ*, described the peakiness of the glottal area waveform:

$$PQ = \frac{PP}{OP}, \tag{4.11}$$

where higher values corresponded to flatter waveform peaks. The inclusion of *PQ* was motivated by observations of flattened peaks of the glottal area waveform when vibration of excised canine larynges exhibited asymmetry (Khosla et al., 2008; Murugappan et al., 2009).

Spectral measures of acoustic pressure

Commonly used acoustic measures related to voice quality were estimated from the magnitude spectrum of the radiated acoustic pressure waveform. The waveform was first downsampled to an 8 kHz sampling rate with appropriate prefiltering. An inverse filtering method was applied to compensate for the effects of the vocal tract transfer function to obtain source-related measures. Estimates of the frequency and bandwidth of the first three formants (Boersma and Weenink, 2009) were used in compensating the harmonic magnitudes for the vocal tract transfer function (Iseli et al., 2007). Modified magnitudes (denoted by asterisks) were labeled either by harmonic number (eg, *H1**) or by nearest formant number (eg, *A1**).

The following acoustic spectral measures were obtained due to their links to source-related characteristics: *H1*−H2**, a correlate of open quotient (Holmberg et al., 1995); *H1*−A1**, a correlate of the first formant bandwidth and energy efficiency (Hanson and Chuang, 1999); and *H1*−A3**, a correlate of source spectral slope (Hanson and Chuang, 1999). An alternative measure of spectral tilt, *TL**, was computed as the linear regression slope over the first eight modified harmonic peaks in the octave frequency domain. This tilt measure was applied to mirror a spectral slope measure used in an excised larynx study (Murugappan et al., 2009). It was noted that these spectral measures assumed a linear source-filter framework, in which source harmonic magnitudes typically exhibit a monotonic decay with increasing frequency.

4.2.4   Statistical analysis

Measures of the lateral displacement and glottal area waveforms were averaged over all the cycles in the phonatory segment. Pearson's correlation coefficient analyzed the pairwise relationships among the measures computed on the lateral displacement waveforms, the glottal area waveform, and the radiated acoustic pressure waveform. Correlations were statistically significant at 95 % confidence levels. Scatter plots were generated to visually analyze marginal and joint distributions.
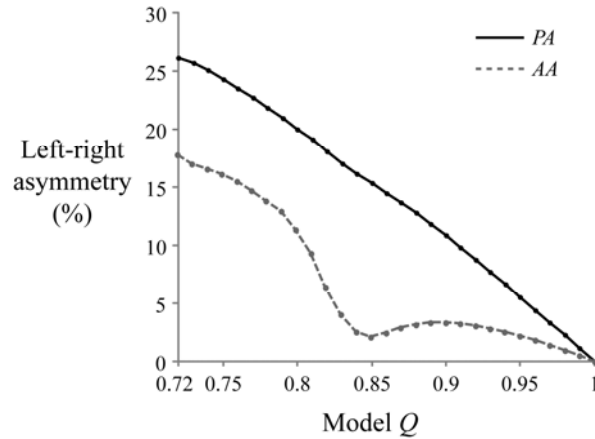
Figure 4.5   Relationships between model parameter *Q* and left-right phase asymmetry *PA* (solid line) and left-right amplitude asymmetry *AA* (dashed line).

4.3   Results

4.3.1   Model simulations

Control over left-right phase asymmetry

Figure 4.5 shows the effects that changing Q has on left-right phase asymmetry *PA* and left-right amplitude asymmetry *AA* over the range $0.72 \leq Q \leq 1$. Given the linear relationship between Q and *PA*, subsequent model analysis was restricted to this region so that changes in Q mapped to changes in *PA*. The model exhibited a nonlinear relationship between Q and *PA* when Q was less than 0.72. *AA* had a nonlinear relationship with Q but remained low (less than 18 %) for all values of Q in the range of interest.

Figure 4.6 displays waveforms from two model simulations showing extreme Q cases. With $Q = 1$, $PA = 0$ %, $AA = 0$ %, $CQ = 33.5$ %, $TL* = -9.6$ dB. Clear notches in the acoustic spectrum affected the value of $TL*$. A decrease in Q increased mass and decreased stiffness parameters of the right vocal fold relative to the left vocal fold. With $Q = 0.72$, the resonant frequency of the right vocal fold decreased, lowering the overall fundamental frequency of phonation. The imbalance in tissue properties induced a phase offset between the lateral displacement waveforms of the left and right vocal folds, with the right vocal fold lagging the left vocal fold in phase. With $Q = 0.72$, $PA = 26.1$ %, $AA = 17.7$ %, $CQ = 42.3$ %, $TL* = -6.8$ dB. Although a steeper tilt was observed over the first four harmonics in the asymmetric case, $TL*$ yielded a similar measure of tilt for this asymmetric case. The measures of spectral tilt were highly sensitive to spectral notches generated by the nonlinear model.
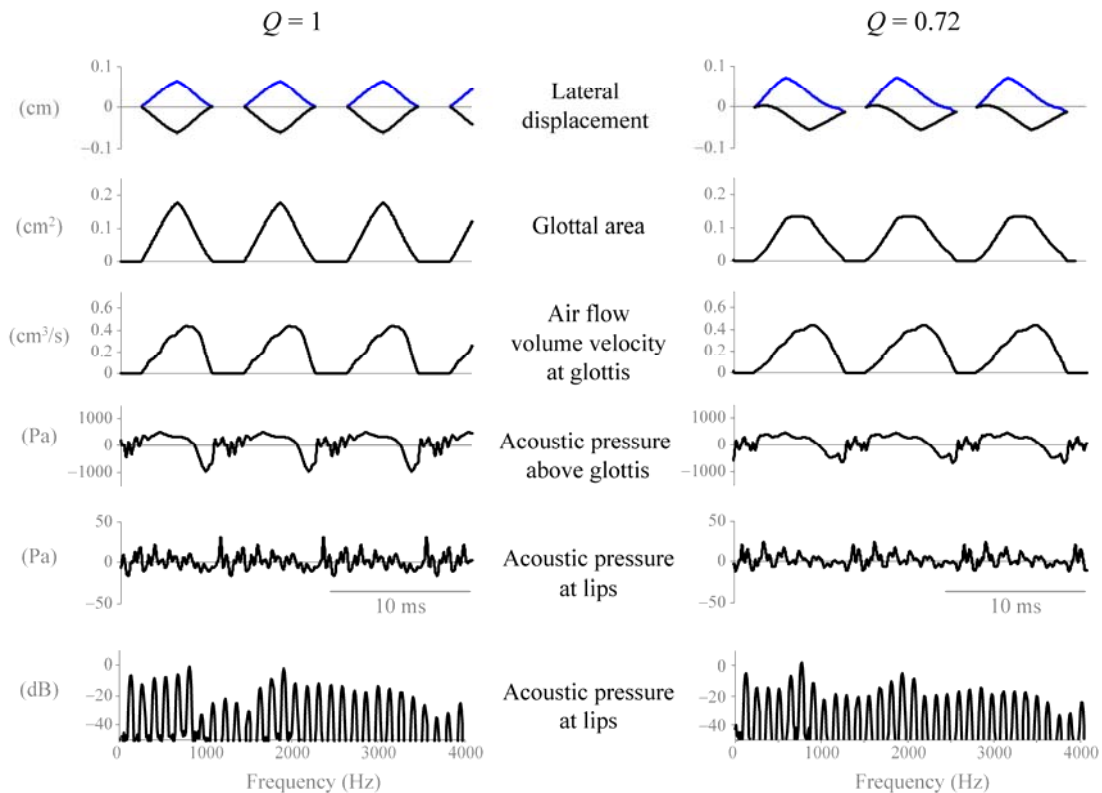
Figure 4.6   Waveforms from model simulations with $Q = 1$ (left column) and $Q = 0.72$ (right column).

Effects of left-right phase asymmetry

Over certain ranges, there were linear relationships between left-right phase asymmetry and both the plateau quotient (Figure 4.7A) and axis shift during closure (Figure 4.7B). Figure 4.7C indicated that increasing left-right phase asymmetry corresponded to a nonlinear, but monotonic, increase in the closing quotient. This relationship pointed to less abrupt closure during simulations with increasing left-right phase asymmetry. However, there were highly nonlinear relationships between left-right phase asymmetry and the acoustic spectral measures. Figure 4.7D shows the overall trend between left-right phase asymmetry and spectral tilt measure TL^*. These relationships were explored further in the subject data. The applicability of TL^* depended on a monotonic source spectrum, which was not necessarily observed in the model behavior due to notches in the acoustic spectra (Figure 4.6).

Figure 4.7   Model-based covariations of left-right phase asymmetry with (A) plateau quotient PQ, (B) axis shift during closure AS, (C) closing quotient CQ, and (D) spectral tilt TL^*.

Table 4.1   Summary statistics of HSV-based asymmetry measures and spectral measures of the radiated acoustic pressure waveform. Minimum (Min), maximum (Max), average (Mean), and standard deviation (SD) are given for each measure (N = 53).

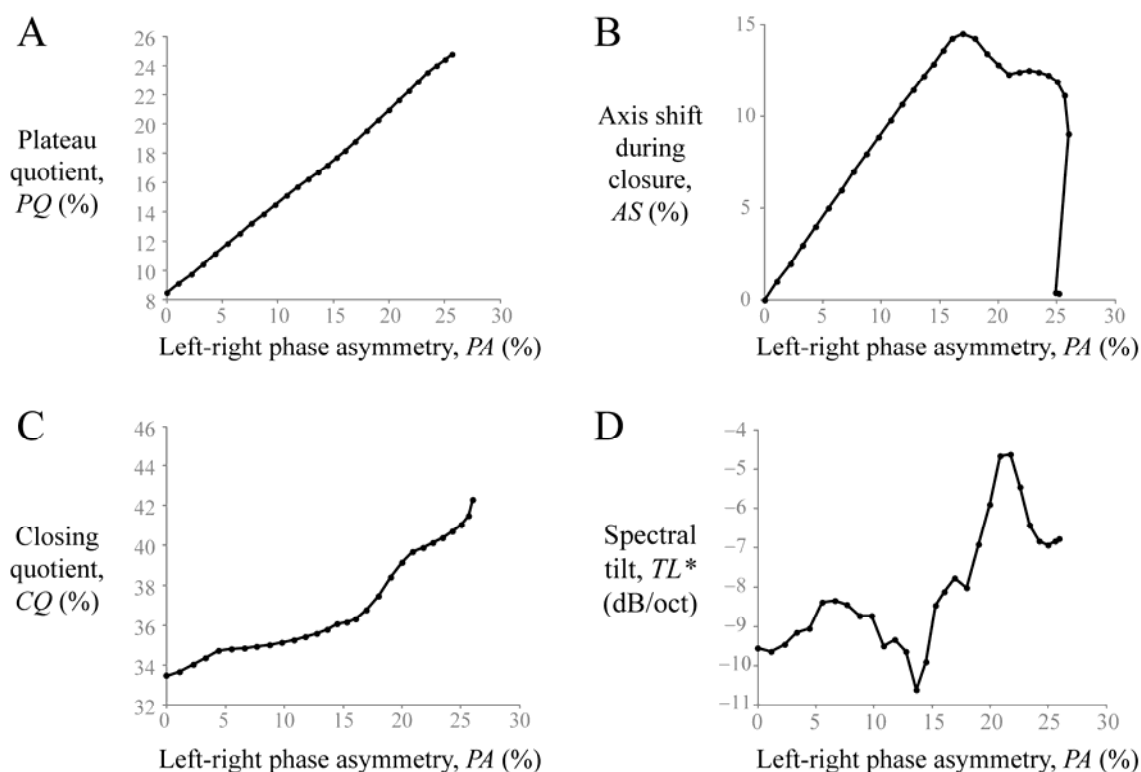| Description	Symbol	Units	Min	Max	Mean	SD
Left-right phase asymmetry†	*PA*	%	0	56	18	14
Left-right amplitude asymmetry†	*AA*	%	0	40	11	10
Axis shift during closure†	*AS*	%	1	50	14	13
Open quotient	*OQ*	%	36	90	65	15
Speed quotient	*SQ*	%	51	301	131	49
Closing quotient	*CQ*	%	11	55	28	9
Plateau quotient	*PQ*	%	8	29	15	4
Relative magnitude of second harmonic	*H1*–H2**	dB	−41.5	21.4	6.54	8.21
Relative magnitude of harmonic closest to first formant	*H1*–A1**	dB	−12.9	27.3	13.2	7.06
Relative magnitude of harmonic closest to third formant	*H1*–A3**	dB	−8.40	40.3	19.7	8.73
Spectral tilt over the first 8 harmonics	*TL**	dB/oct	−14.8	0.817	−8.70	3.13

†Directionality of this measure was removed by computing the absolute value to quantify the absolute degree of the measure.

4.3.2   Human subject data

Prevalence of vocal fold vibratory asymmetry

Table 4.2 displays summary statistics of the HSV-based measures and acoustic measures from all the subject trials (N = 53). Vocal fold vibration exhibited glottal closure with the absolute value of average left-right phase asymmetry approaching 50 %. This observation was in contrast to the limitations of the model implementation, which allowed for glottal closure for levels of left-right phase asymmetry up to 26 %.

Case studies

Figure 4.8 displays a snapshot of HSV and acoustic descriptors for a male subject with no history of voice disorders. The acoustic fundamental frequency was 136 Hz. Left-right phase asymmetry *PA* measured from the medial digital kymogram was 6 %. The measure of axis shift during closure *AS* was −2 %, and left-right amplitude asymmetry *AA* was 0 %. The spectral tilt *TL** of the acoustic magnitude spectrum was −6.4 dB/oct. Glottal area waveform measures obtained: open quotient *OQ* = 56 %, speed quotient *SQ* = 86 %, closing quotient *CQ* = 29 %, and plateau quotient *PQ* = 12 %.

Figure 4.8   Case study of Subject N1 with low left-right phase asymmetry *PA* (6 %). Plots display the lateral displacement waveforms, phonovibrogram, glottal area waveform, radiated acoustic voice signal, and magnitude frequency spectra of the glottal area and acoustic voice signal.
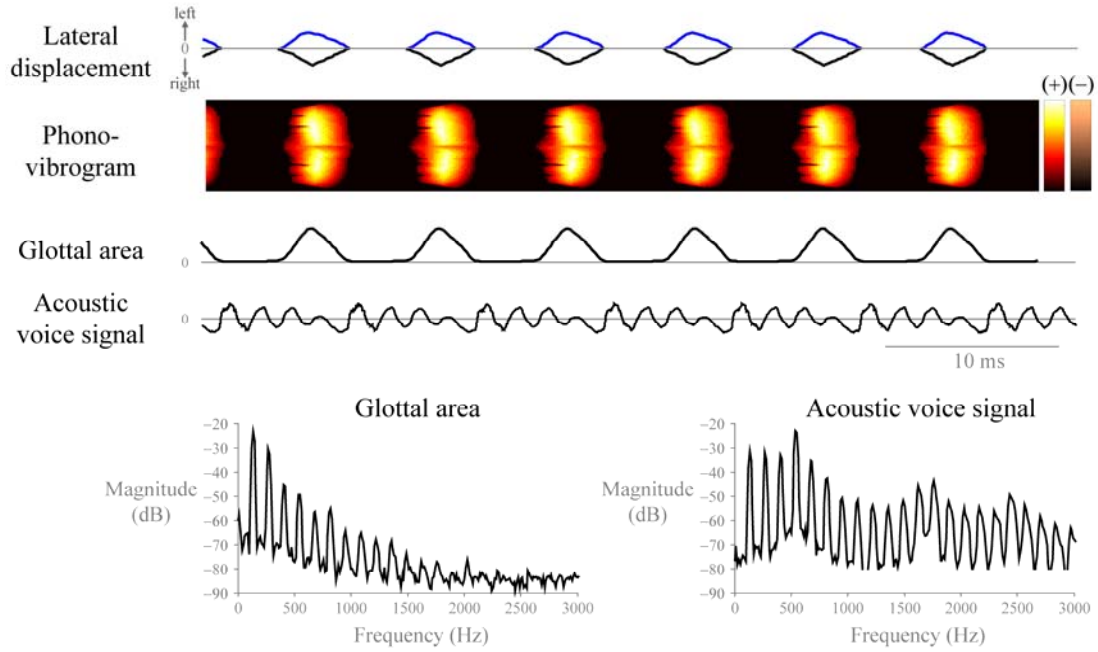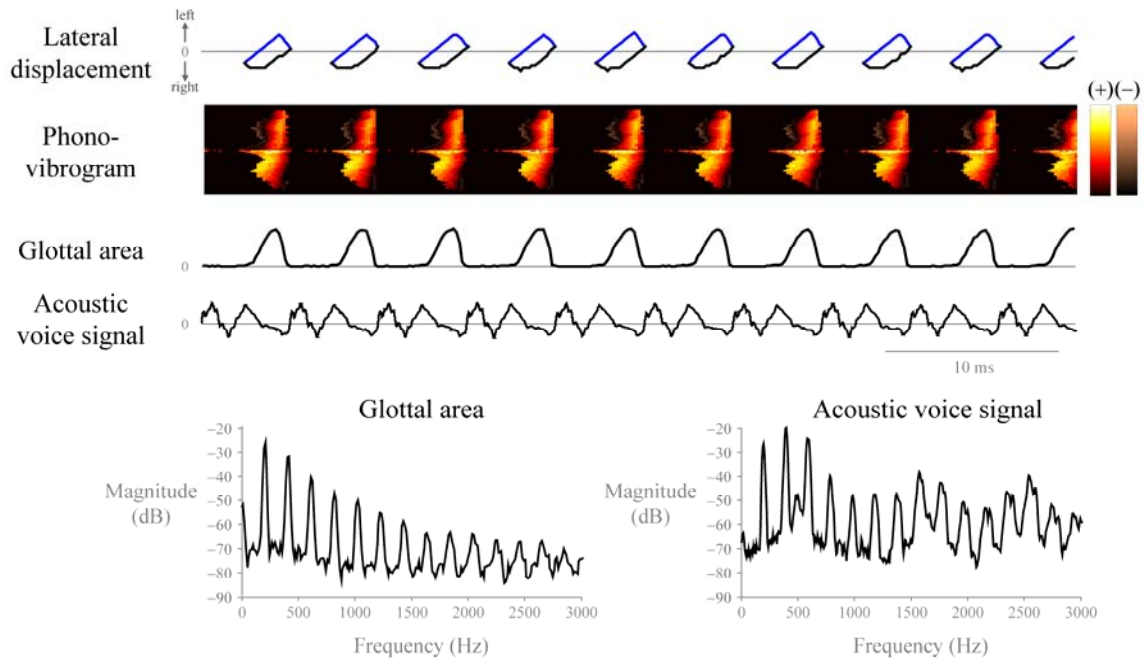

Figure 4.9   Case study of Subject P13 with high left-right phase asymmetry *PA* (−51 %). Plots display the lateral displacement waveforms, phonovibrogram, glottal area waveform, radiated acoustic voice signal, and magnitude frequency spectra of the glottal area and radiated acoustic voice signal.

In contrast, Figure 4.9 presents data from a male subject who underwent surgical treatment for glottic cancer of the right vocal fold. The acoustic fundamental frequency was 197 Hz. Due to the surgical procedure, the tissue on the right vocal fold exhibited a reduction in mass and an increase in stiffness due to scar tissue and loss of the pliable superficial lamina propria. The left-right imbalance produced asymmetric vocal fold vibration. Agreeing with model-based predictions, the stiffer (right) vocal fold led the opposite vocal fold in phase, yielding a negative measure of left-right phase asymmetry PA (−51 %). The relative amplitude of the right vocal fold was reduced (AA = 11 %). The speed quotient SQ and closing quotient CQ were 300 % and 11 %, respectively, signifying abrupt closure during the closing phase relative to that in the normal case above. The spectral tilt $TL*$ of the disordered voice was −6.3 dB/oct, which was similar to the normal case.

Acoustic correlates of glottal characteristics

Table 4.3 displays the intercorrelation matrix for the HSV-based measures and acoustic spectral measures. In particular, none of the measures of vibratory asymmetry correlated to a statistically significant degree with any of the acoustic spectral measures. Figure 4.10 displays scatter plots for statistically significant pairwise correlations. Figure 4.10A displays the variation in AA accompanying different values of PA, which was not modeled in the same manner. Figure 4.10B shows the linear relationship between PA and AS over a larger range than exhibited in the model. Figure 4.10C displays the inverse relationship between CQ and PA. Figure 4.10D indicates that closing quotient correlated statistically with $TL*$ to a moderate degree.

Table 4.1 Pearson's correlation coefficient for significant pairwise relationships among average values of HSV-based asymmetry measures and spectral measures of the radiated acoustic pressure waveform ($N = 53$). Correlations shown are statistically significant at a 95 % confidence level.

	$AA^\dagger$	$AS^\dagger$	OQ	SQ	CQ	PQ	$H1–H2^*$	$H1–A1^*$	$H1–A3^*$	TL^*
$PA^\dagger$	−.42	.93	—	.45	−.40	—	—	—	—	—
$AA^\dagger$		−.36	—	.28	—	—	—	—	—	—
AS			—	.44	−.41	—	—	—	—	—
OQ				—	.78	−.35	.28	.27	.40	−.31
SQ					−.63	—	—	—	−.28	—
CQ						—	.32	.31	.48	−.41
PQ							—	—	—	—
$H1–H2^*$								.69	.58	−.52
$H1–A1^*$									.67	−.65
$H1–A3^*$										−.85

†Directionality was removed by computing the absolute value of this measure for all correlations except for PA–AA, PA–AS, and AA–AS.
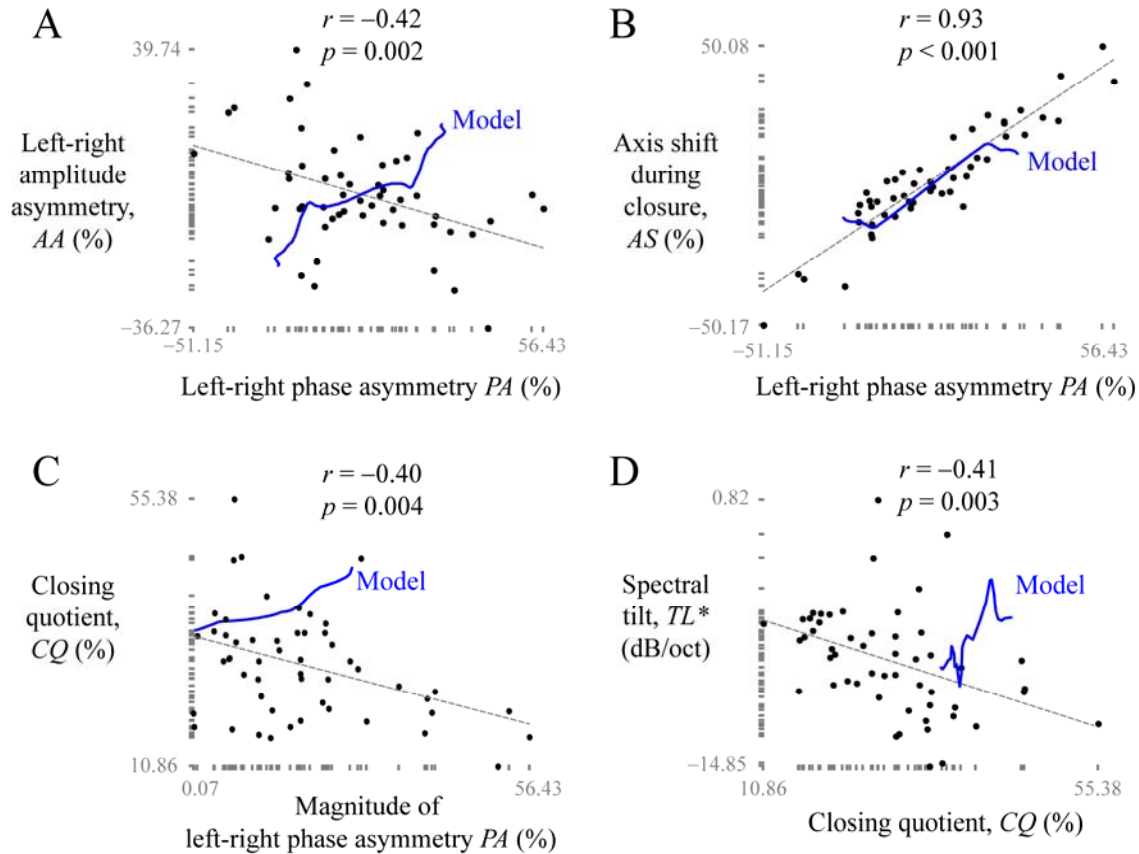
Figure 4.10 Scatter plots of statistically significant relationships in the subject data. **(A)** *PA* and *AA*, **(B)** *PA* and *AS*, **(C)** *PA* and *CQ*, and **(D)** *CQ* and *TL**. Pearson's *r* and *p*-value are indicated. Pairwise relationships obtained from the model simulations are overlaid on the respective plot.

4.4 Discussion

Left-right phase asymmetry was simulated in a computational voice production model. Effects of asymmetric vocal fold vibration on the glottal area and acoustic pressure waveforms were hypothesized using the model. Model-based hypotheses were compared with human subject results obtained from time-synchronized measures of laryngeal high-speed videoendoscopy and the acoustic voice signal. The model-based simulations were performed with a four-mass model of vocal fold vibration that was extended to allow for glottal source interaction with subglottal and supraglottal tracts.

In both model and subject data, measures of vocal fold vibratory asymmetry did not correlate with any of the acoustic spectral measures. However, in the subject data, left-right phase asymmetry and closing quotient exhibited a mild inverse correlation ($r = -.40$, $p < 0.05$). This result conflicted with model simulations, in which the glottal area waveform exhibits higher closing quotients (less abrupt glottal closure) with increasing levels of phase asymmetry. This preliminary result was unexpected because prevailing theories have hypothesized that asymmetric vocal fold vibration is associated with a less abrupt glottal closure and, thus, less efficient sound production (Khosla *et al.*, 2008; Murugappan *et al.*, 2009). Figure 4.11 illustrates how the presence of left-right amplitude asymmetry can affect the ability of left-right phase asymmetry to influence the closing phase duration of the glottal area waveform. In this example, two subjects exhibited similar levels of left-right phase asymmetry but different degrees of amplitude asymmetry. Thus, the closing phase duration actually shortened in the presence of both left-right phase and amplitude asymmetry.
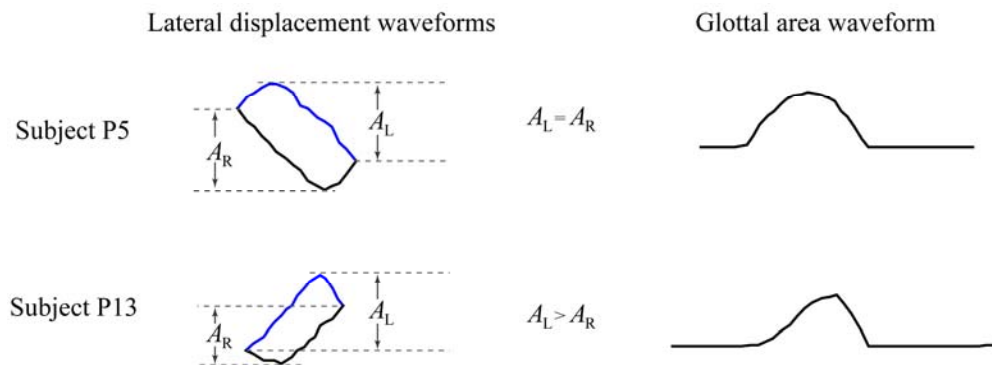

Figure 4.11 Illustration of the impact of left-right amplitude asymmetry on the shape of the glottal area waveform, given similar values of left-right phase asymmetry. Lateral displacements and corresponding glottal area waveforms for two subjects were obtained from a representative cycle of the medial digital kymogram.

The range of values chosen for the Q factor that created left-right phase asymmetries was selected to maintain periodicity and glottal closure during voice synthesis, in contrast with the use of the model to analyze parameters creating bifurcations and irregular behavior (Steinecke and Herzel, 1995). Cycle-to-cycle perturbations of the glottal area waveform were not modeled in the current study. Glottal correlates of acoustic shimmer and jitter have proven difficult to investigate, possibly owing to the sensitivity of the measures to turbulent noise in the signal and to the lack of understanding exactly how to incorporate shimmer, jitter, and turbulent noise in such a voice production model (Jiang and Zhang, 2002). Also, future modifications of vocal fold models should reflect realistic magnitudes of asymmetry that occur in subjects during sustained phonation. In addition, the model assumption of a uniform rectangular glottis was violated in many of the subjects with voice disorders (exemplified in Figure 4.9), in which the glottal aperture in reality was not rectangular.

The model predicted the degree and polarity of the relationship between left-right phase asymmetry and axis shift during closure to a large extent. However, the model results disagreed with relationships between left-right phase asymmetry and measures of left-right amplitude asymmetry, plateau quotient, closing quotient, and spectral tilt in the subject data. The spectral tilt measures were possibly not applicable to acoustic pressure waveforms in the model due to the non-monotonic decay of source harmonics. Even in the symmetric case, notches were observed in the acoustic spectrum (Figure 4.6). This points to the need for alternative measures that can handle non-linear properties as exhibited in the model.

The effect of left-right amplitude asymmetry was not controlled for when obtaining the correlation between left-right phase asymmetry and spectral tilt. Thus, changes in Q might not adequately represent the pathological cases described. The addition of nonlinear coupling effects further emphasized the disadvantages of using linear measures such as spectral tilt to describe model behavior. Spectral tilt, however, tended to perform more predictably in the human data.

In a recent study, vibratory asymmetry was induced in a silicone model of the vocal folds by varying the Young's modulus of the material (Pickup and Thomson, 2009). In particular, results showed that increasing stiffness asymmetry was linked to overall decreases in the glottal area, higher phonation threshold pressures, incomplete glottal closure, and glottal jet skewing toward the stiffer vocal fold. The authors speculated that the vibratory asymmetry could possibly result in a qualitatively "unnatural-sounding voice." Evidence supporting this statement was found in the analysis of human subject data and the computational model, in which the acoustic spectrum was

71

affected through changes in the glottal area waveform. Statements regarding effects of a single type of vibratory asymmetry, however, cannot be made due to the complex dynamics of vocal fold vibration.

Future studies would benefit from measures of the glottal airflow that would provide better estimates of source-related properties that affect acoustic spectrum characteristics. With access to glottal airflow information through inverse filtering or hot-wire anemometry, correlations with acoustic spectral measures would presumably improve. Statistically significant correlations were found between closing quotient and spectral tilt although glottal area measures were applied rather than airflow–based measures (Holmberg et al., 1995), suggesting that the properties of the acoustic waveform could be traced to vocal fold tissue motion.

4.5 Conclusions

Voice quality perception is dependent on spectral and temporal variations of the radiated pressure waveform, which is a result of vocal fold dynamics, noise turbulence, tract geometries, and nonlinear aerodynamic interactions between the source and tracts. Phase asymmetry between the displacements of the left and right vocal folds plays only one role in shaping the modulation of the glottal airflow and, ultimately, the spectral characteristics of the acoustic voice signal. Laryngeal high-speed videoendoscopy recordings of human subjects revealed that there was no direct correlation between vocal fold vibratory asymmetries and acoustic spectral measures. This result calls for investigations into the development of better acoustic measures and the role of asymmetric vocal fold vibration in efficient voice production.

References

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. (**2009**). *Praat: Doing phonetics by computer* (Amsterdam, The Netherlands). Ver. 5.1.07. Available at http://www.praat.org.

Bonilha, H. S., Deliyski, D. D., and Gerlach, T. T. (**2008**). "Phase asymmetries in normophonic speakers: Visual judgments and objective findings," Am. J. Speech Lang. Pathol. **17**(4), 367–376.

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.

Dormand, J. R., and Prince, P. J. (**1980**). "A family of embedded Runge-Kutta formulae," J. Comput. Appl. Math. **6**(1), 19–26.

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.

Hanson, H. M., and Chuang, E. S. (**1999**). "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," J. Acoust. Soc. Am. **106**(2), 1064–1077.

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., Perkell, J. S., Guiod, P. C., and Goldman, S. L. (**1995**). "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," J. Speech Hear. Res. **38**(6), 1212–1223.

Iseli, M., Shue, Y.-L., and Alwan, A. (**2007**). "Age, sex, and vowel dependencies of acoustic measures related to the voice source," J. Acoust. Soc. Am. **121**(4), 2283–2295.

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.

Ishizaka, K., and Isshiki, N. (**1976**). "Computer simulation of pathological vocal-cord vibration," J. Acoust. Soc. Am. **60**(5), 1193–1198.

Jiang, J. J., and Zhang, Y. (**2002**). "Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds," J. Acoust. Soc. Am. **112**(5 Pt 1), 2127–2133.

Khosla, S., Murugappan, S., and Gutmark, E. (**2008**). "What can vortices tell us about vocal fold vibration and voice production," Curr. Opin. Otolaryngol. Head Neck Surg. **16**(3), 183–187.

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.

Mehta, D. D., Deliyski, D. D., Zeitels, S. M., Quatieri, T. F., and Hillman, R. E. (**2010**). "Voice production mechanisms following phonosurgical treatment of early glottic cancer," Ann. Otol. Rhinol. Laryngol. **119**(1), 1–9.

Murugappan, S., Khosla, S., Casper, K., Oren, L., and Gutmark, E. (**2009**). "Flow fields and acoustics in a unilateral scarred vocal fold model," Ann. Otol. Rhinol. Laryngol. **118**(1), 44–50.

Niimi, S., and Miyaji, M. (**2000**). "Vocal fold vibration and voice quality," Folia Phoniatr. Logop. **52**(1-3), 32–38.

Pickup, B. A., and Thomson, S. L. (**2009**). "Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models," J. Biomech. **42**(14), 2219–2225.

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.

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.

Steinecke, I., and Herzel, H. (**1995**). "Bifurcations in an asymmetric vocal-fold model," J. Acoust. Soc. Am. **97**(3), 1874–1884.

Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**(2), 1249–1260.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**(1), 537–554.

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.

Titze, I. R. (**1984**). "Parameterization of the glottal area, glottal flow, and vocal fold contact area," J. Acoust. Soc. Am. **75**(2), 570–580.

Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**(5), 2733–2749.

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.

Weibel, E. R. (**1963**). *Morphometry of the Human Lung* (Springer, New York).

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.

Zañartu, M., Mongeau, L., and Wodicka, G. R. (**2007**). "Influence of acoustic loading on an effective single mass model of the vocal folds," J. Acoust. Soc. Am. **121**(2), 1119–1129.

Zeitels, S. M., Blitzer, A., Hillman, R. E., and Anderson, R. R. (**2007**). "Foresight in laryngology and laryngeal surgery: A 2020 vision," Ann. Otol. Rhinol. Laryngol. Suppl. **116(suppl 198)**, 1–16.

Zhang, Y., Tao, C., and Jiang, J. J. (**2006**). "Parameter estimation of an asymmetric vocal-fold system from glottal area time series using chaos synchronization," Chaos **16**(2), 023118.

Chapter 5

Overall Conclusions

The three specific aims of this thesis project were addressed, respectively, in Chapters 2–4. Overall conclusions are summarized here with a look to future directions.

5.1   Summary

In Chapter 2, an automated framework was developed to obtain measures of vocal fold vibratory asymmetry from laryngeal high-speed videoendoscopy (HSV) recordings. Measurements were designed to provide objective quantification of clinically based visual-perceptual judgments and to also serve as a platform for investigating the effects of asymmetric vocal fold vibration on voice quality.

In Chapter 3, measures of vocal fold vibratory asymmetry were correlated with acoustic perturbations measures using time-synchronized measures of HSV and the acoustic voice signal. The results of pairwise correlations revealed that variations in levels of acoustic jitter and shimmer were not related to within-cycle asymmetries in tissue vibration. Instead, a significant amount of the variation in acoustic jitter was accounted for by the changes (standard deviations) in the symmetry of the phase and amplitude of tissue motion across vibratory cycles, and a significant amount of the shimmer variation was correlated with the standard deviation of glottal closure durations across phonatory cycles. Thus, asymmetries that appear to affect acoustic perturbation measures to a greater degree are less likely to be captured by standard stroboscopic visualization that requires a fairly periodic vibratory pattern.

In Chapter 4, the spectral effects of vocal fold vibratory asymmetries were investigated. This association was investigated in a computational voice production model and in a group of 47 human subjects. Surprisingly, in both model and subject data, measures of vocal fold vibratory asymmetry did not correlate with spectral tilt measures. Probing further, the glottal area waveform was parameterized to obtain open phase measures (open quotient, plateau quotient) and closing phase measures (speed quotient, closing quotient). In the subject data, left-right phase asymmetry and closing quotient exhibited a mild inverse correlation ($r = -.40$, $p < 0.05$). This result conflicts with model simulations in which the glottal area waveform exhibited less abrupt glottal closure (higher closing quotient) with increasing levels of phase asymmetry. Results call for studies into the applicability of traditional spectral tilt measures and the role of asymmetric vocal fold vibration in efficient voice production.

5.2 Future Work

In Chapter 3, even though statistically significant correlations were found between HSV-based asymmetry and acoustic perturbation, the actual correlations were at moderate levels. A significant amount of variation in acoustic perturbation was still unaccounted for by the HSV measures used in this study. Furthermore, correlations that were performed in Chapter 4 could only include glottal characteristics related to the shape of the glottal area waveform.

One missing link in these studies was an analysis of the glottal airflow to help complete the picture and investigate the main source of excitation for voice. As was found in the model simulations in Chapter 4, the acoustic sound pressure at the lips is a result of complex and highly nonlinear aerodynamic interactions throughout the voice production system. Attempts at estimating the airflow signal simultaneously with HSV recordings are under way at the MGH Voice Center.

Figure 5.1 depicts our setup for collecting measures of airflow using a modified Rothenberg mask. Since it is critical to maintain an aerodynamic seal around the mouth, HSV recording is accomplished with a transnasal flexible fiberscope. With the airflow signal, inverse filtering approaches similar to the method in Chapter 4 can be applied to characterize the airflow shape with particular interest in tracking changes due to asymmetric vocal fold vibration.

Additionally, two-dimensional imaging of the vocal folds has inherent technical limitations. In this thesis, asymmetries were measured as if they only existed in the mediolateral direction. In reality, the three-dimensional structure of the vocal folds behaves in a complex manner with

asymmetries possible in the medial-lateral, inferior-superior, and anterior-posterior directions. Approaches to stereoscopic imaging show promise to capture more of the variations exhibited during vocal fold vibration (Schuberth *et al.*, 2002; George *et al.*, 2008), especially for understanding abnormal voice production mechanisms.
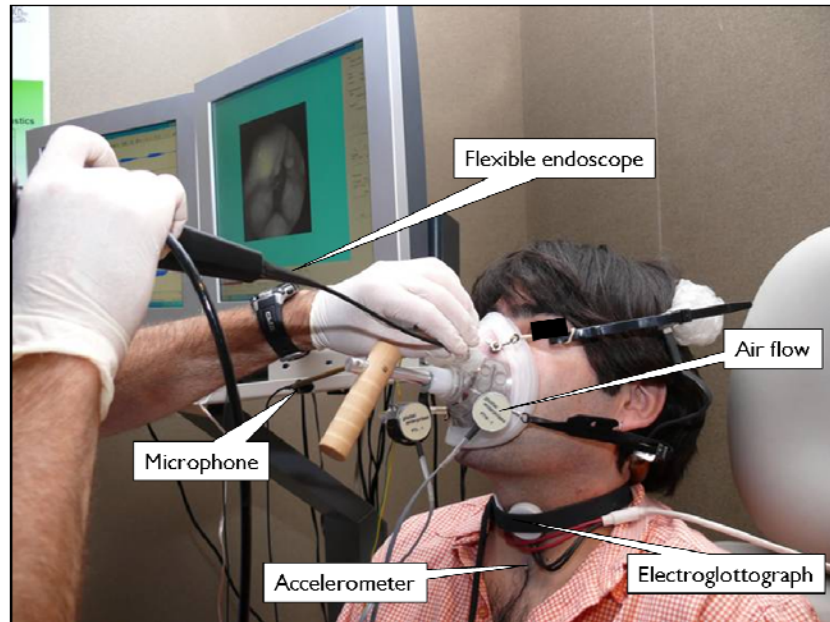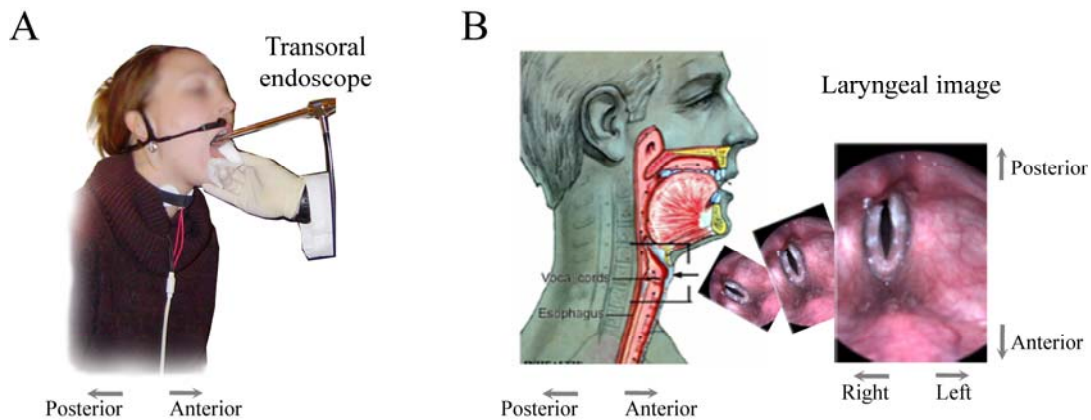

Figure 5.1   Monochromatic HSV with transnasal endoscopy to provide estimates of glottal airflow.

Appendix A

Laryngeal Videostroboscopy

Figure A.1   Endoscopic imaging setup. (A) Typical positioning for transoral endoscope. (B) Illustration of the position of the larynx and the orientation of the larynx in a recorded image.

In any laryngeal imaging technology, the larynx must be viewed using an endoscope that is passed through the nasal or oral cavity. Figure A.1 illustrates the imaging setup and the orientation of a laryngeal image recorded using transoral endoscopy. The maximum frame rate of standard video cameras 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). During a sustained vowel, the vocal folds often open and close in a regular pattern over time, and the presumed temporal redundancy may be exploited by sampling the pattern in a stroboscopic manner. Figure A.2 illustrates the stroboscopic principle by displaying a continuous periodic waveform that is too fast to sample adequately (the light bulbs cannot occur often enough). Only a

composite video sequence of one period of the waveform can be reconstructed from samples taken from several non-consecutive cycles. Thus, intracycle information is not captured and the entire process requires near-periodic behavior.
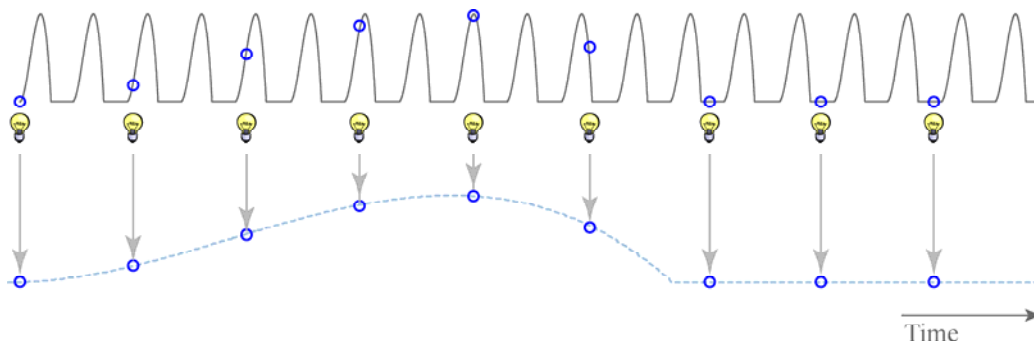
Figure A.2. An illustration of how stroboscopic principles sample a periodic waveform. From (Hillman and Mehta, 2010: in press). Used with permission.

Laryngeal stroboscopy applies this sampling technique to two-dimensional images captured during vocal fold vibration. Cycle-to-cycle regularity is assumed during sustained phonation, and a strobe light flashes at successive phases of the repeating vocal fold cycle. In the voice clinic today, the inclusion of digital video camera technology necessitates the term *videostroboscopy*, in which camera frame rates are limited to about 30 Hz (The Society of Motion Picture and Television Engineers, 2004). To prevent multiple exposures during one video field duration, the timing of the strobe light must be regulated so only one flash occurs during each video field. Finally, a composite video sequence of the vocal fold vibratory cycle is reconstructed from images strobed from several non-consecutive cycles.

How much information is missed using laryngeal videostroboscopy? Figure A.3 shows a direct comparison of videostroboscopic sampling and the true motion of the vocal folds captured using high-speed videoendoscopy (Hillman and Mehta, 2010: in press). High-speed video data were obtained from a male adult with no history of voice disorders. Two transoral rigid endoscopes were simultaneously positioned to view the vocal folds while the subject produced a sustained vowel. One endoscope provided continuous illumination for capturing video images from the attached high-speed camera. The second endoscope was connected to the light source of a clinical stroboscopy system that delivered strobe flashes that were triggered off of the voice signal obtained from a neck-mounted contact microphone (KayPENTAX, 2008).
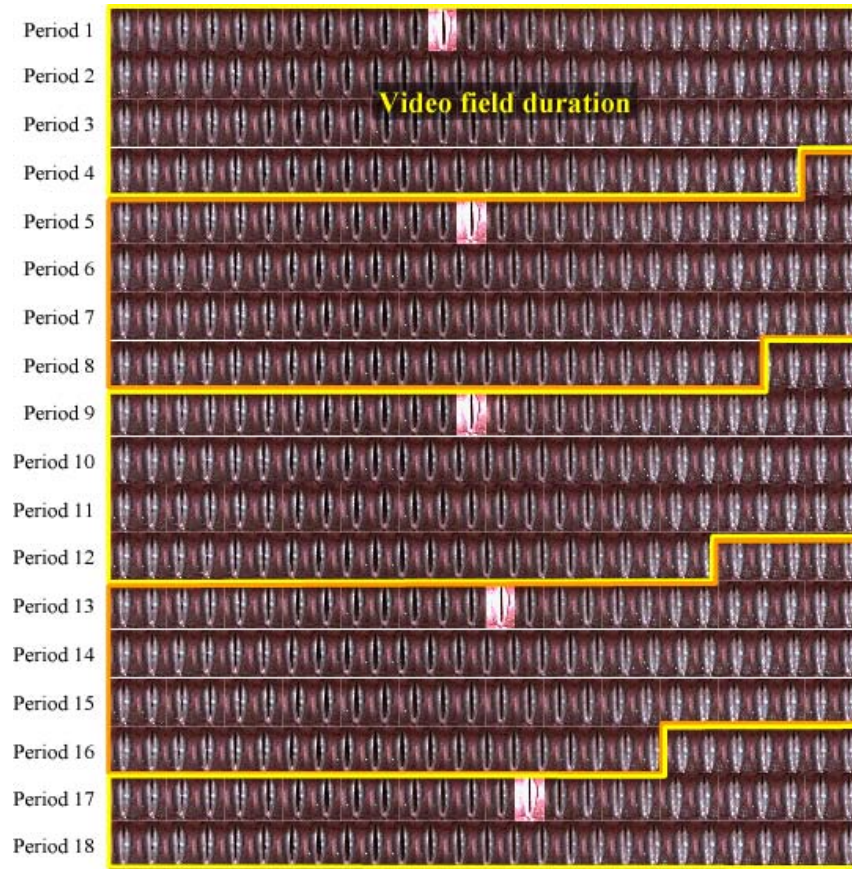
Figure A.3   Comparison of videostroboscopy and high-speed videoendoscopy. From (Hillman and Mehta, 2010: in press). Used with permission.

The fundamental frequency during the phonatory segment was estimated to be approximately 236 Hz (period = 4.23 milliseconds), yielding 26.44 high-speed video frames per vocal fold vibratory cycle [(4.23 milliseconds per period) / (0.16 milliseconds per frame) = 26.44 frames per cycle].

In this example, the slow-motion rate of the strobe light was set to 1.5 Hz, resulting in the display of 1.5 periods of vibration per second (a common setting used on clinical strobe units). The strobe rate must be close to 60 Hz to provide exposure (exactly one strobe flash) for each video field. The strobe rate can be estimated by simply dividing the fundamental frequency (236 Hz) by the video field rate, in this example 236/60 ≈ 3.9, indicating that the strobe would flash approximately once every 4 cycles or periods of vocal fold vibration.

Appendix B

Chapter 2 Supplement

This section details some of the "under-the-hood" image processing that was performed using custom MATLAB functions. In particular, threshold-based segmentation is described to obtain lateral displacement waveforms from a digital kymogram.

Mathematical Framework

The high-speed video sequence represents a time sampling of the true motion of the vocal fold tissue, yielding a two-dimensional intensity image at each sampling point. In reality, an image is created over a finite exposure time, during which the camera sensor integrates intensity information from the illuminated target (similar to a sample-and-hold analog-to-digital converter).

Each image has horizontal and vertical spatial dimensions that are discretely sampled, with indices x and y, respectively. The image has a width W and height H associated with it. Specifically, $x = 1, 2, ..., W$ and $y = 1, 2, ..., H$. The coordinate $[x, y]$ defines the location of a pixel that has an intensity level—$I[x, y] = 0, 1, ..., 255$—indicating luminance, or grayness, at the corresponding location. Black is defined as $I = 0$, and white is defined as $I = 255$ for 8-bit quantization. Adding the discrete temporal dimension $n = 1, 2, ..., N$, the definition of the image sequence $I[x, y, n]$ is complete for a given number of images N. Four "blank" images are shown in Figure B.1 to illustrate the spatial and temporal indexing.
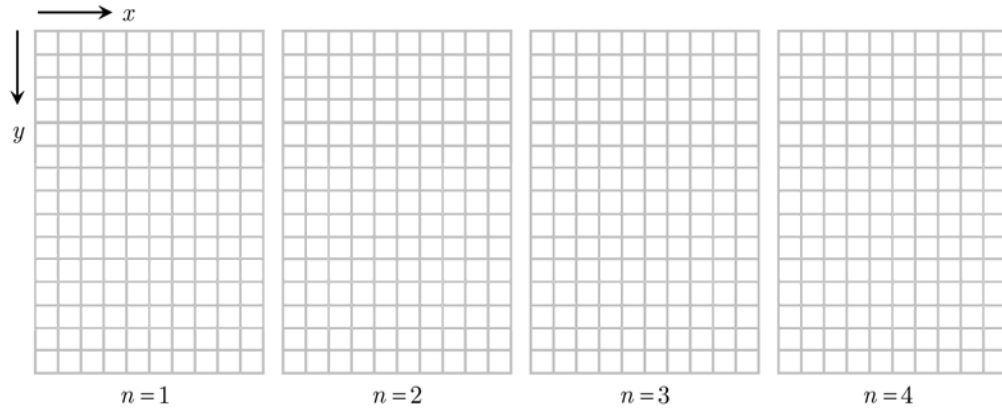
Figure B.1   **Indexing of pixels in blank images.**

Of particular relevance to Aim 1 was the construction of kymograms from the full-frame images of the high-speed video. A kymogram is an image that essentially suppresses one spatial dimension in the video sequence to create a two-dimensional image with time on the horizontal axis and space on the vertical axis. A digital kymogram was generated from vocal fold images by suppressing the anterior-posterior dimension and compactly displaying visual information related to the left-right motion of the vocal folds that is of primary interest. In the present framework, x indexes the left-right glottal axis, and y indexes the anterior-posterior axis. Consequently, the digital kymogram is defined as $K_i[n,x] = I[x,y,n]\big|_{y=i}$. One common row of pixels is extracted from each full-frame image. On the digital kymogram, the horizontal dimension represents the temporal axis, and the vertical dimension represents the left-right glottal axis (the horizontal dimension in the full-frame image). Figure B.2 shows how the common rows (row 7 in the figure) from the video images become the columns of the kymographic image.
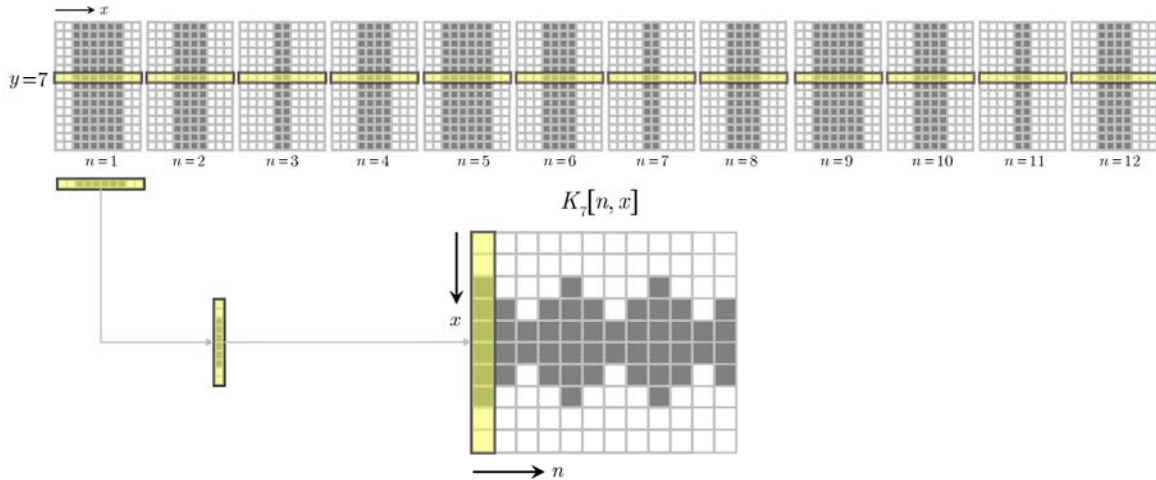
Figure B.2   Generation of a kymographic image (lower, enlarged for illustration) from a series of full-frame images (upper).

Threshold-based segmentation

This section details the image processing algorithms applied to the kymograms to define edges that aided in quantitatively describing kymographic properties such as vibratory asymmetry. Parameters of the rotation and cropping stages described in Chapter 2 were applied to all the images in a motion-compensated video sequence. The resulting three-dimensional data set was $I[x, y, n]$.

Prior to edge detection, the kymogram is segmented into two regions: glottis and non-glottis. The lighting situation during endoscopy is typically such that the vocal fold tissue reflects a substantial amount of light relative to the light reflected by the glottal gap. (In some cases, the light is so intense that the subglottal tracheal shelf is illuminated; these situations were not observed in the database.) Therefore, darker pixels most likely make up the glottis, while lighter pixels reflect non-glottal vocal fold tissue. Figure B.3 illustrates a digital kymogram $K_i[n, x]$ displaying information from 256 video images $(N = 256)$ exhibiting four open phase segments of a phonatory segment, accompanied by its pixel intensity histogram.
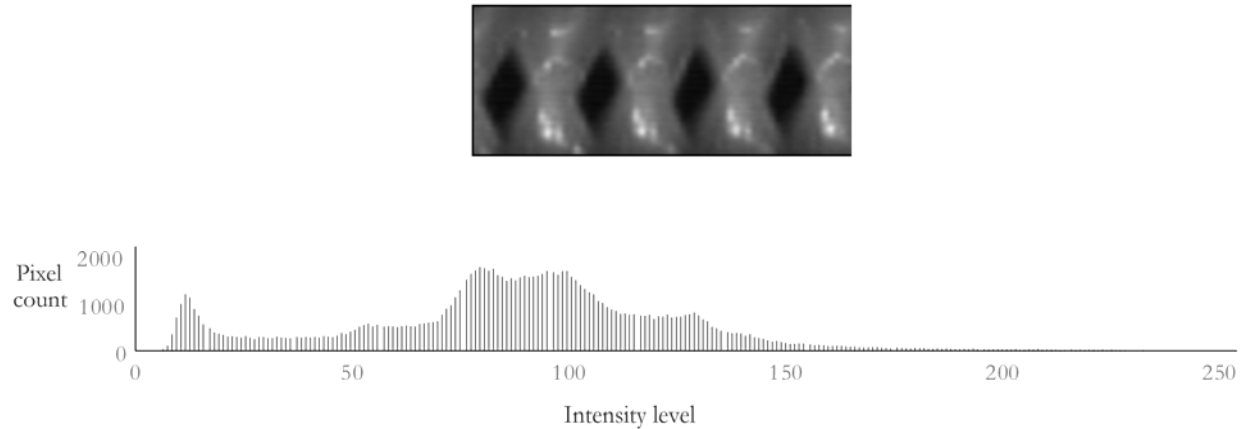
Figure B.3   Example digital kymogram with its corresponding intensity histogram.

The histogram of the kymogram gives information regarding the number of pixels (*y*-axis) exhibiting a certain intensity level K (*x*-axis range: 0–255). Lower intensity levels correspond to dark pixels in the kymogram, and higher intensity levels indicate lighter pixels. Threshold-based glottal segmentation means that the boundary between glottis and tissue is defined by an intensity threshold. Pixels exhibiting intensity levels less than the threshold are classified as glottis, while the remaining pixels are classified as tissue, or non-glottis. In the histogram above, several peaks are observed, and of most interest, is the peak at lowest intensity levels. This peak reflects the number of pixels of darkest intensity that is assumed to be in the glottis. To account for glottal pixels with intensity levels just above the threshold, yielding unwanted gaps in the glottis, and for stray tissue-related pixels that are errantly labeled as glottis, the segmented image undergoes *erosion* and *dilation* operations. Erosion removes stray pixels not in the neighborhood of other glottal pixels. Dilation detects groups of neighboring pixels and fills in any gaps that exist that prevent a contiguous area.
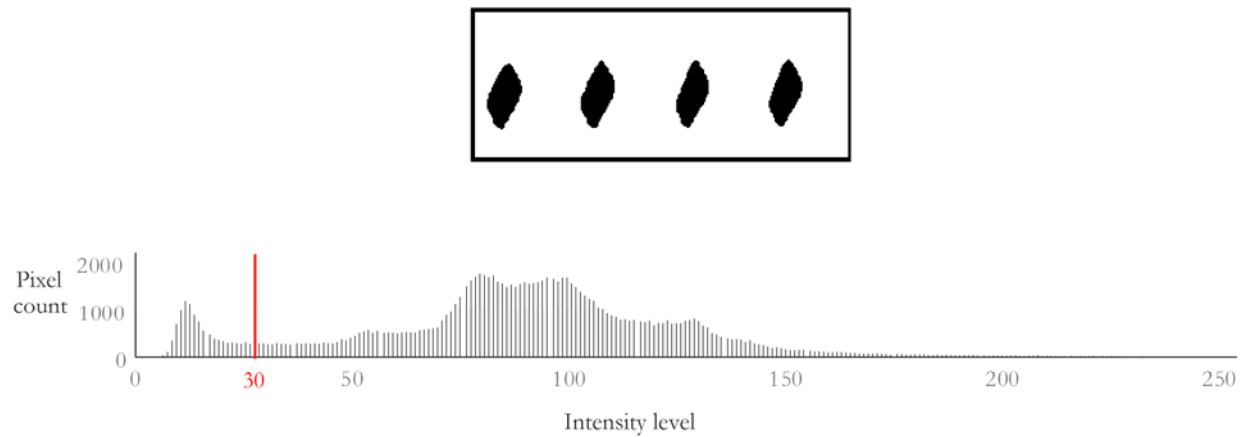
Figure B.4   Glottal segmentation output (upper) with intensity threshold set 30. Intensity histogram of the original image also shown (lower).

For example, Figure B.4 illustrates segmentation with an intensity threshold (red line) of 30. Glottal segmentation results in the binary image, where glottal pixels are set to black, and non-glottal pixels to white: The selection of the threshold is performed interactively to take into account various lighting situations. Several thresholds can be set, and the results visually inspected for the most accurate segmentation. A single threshold is set for each kymogram. The next section explores the effects of different thresholds on the glottal segmentation, as well as on the edges that are subsequently estimated.

Prior to interactive segmentation, an initial threshold is determined automatically from the kymogram's histogram. A smoothed estimate of the envelope of the histogram is calculated, and the initial threshold is chosen to be the intensity level at the first valley in the histogram. Some videos yield kymograms whose histograms exhibit a flatter shape, thus preventing automatic detection of a valley. The outline, or edge, is defined to be one pixel outside of the pixels denoted as glottis.
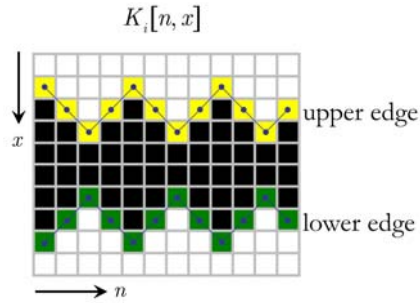
**Figure B.5** **Segmented schematic kymogram with upper (yellow) and lower (green) edges color coded.**

In Figure B.5, the upper (yellow) and lower (green) pixels are colored differently to distinguish between tissue motion in the upper and lower parts of the kymogram, respectively. Depending on the orientation of the original video images, the upper and lower edges in the kymogram correspond to lateral tissue motion of the left and right vocal folds or vice versa. Note that the quantization of the tissue displacement is directly related to the spatial resolution achieved during video capture. The maximum displacement of the vocal folds is typically sampled by about 25 pixels. The first vertical column in which an edge is detected exhibits two edge pixels—one for the upper edge and the other for the lower edge.

For fine-tuning of the initial glottal segmentation, the user is given feedback consisting of the original kymogram overlaid with the upper and lower edges outlining the glottal pixels. The effect of incorrect thresholds is illustrated in Figure B.6. Intensity thresholds of 10 and 50 (on the 256-point scale) were applied. The segmentation with a threshold of 10 (Figure B.6A) exhibits edges that miss many pixels that are visually perceived to be part of the glottis. The corresponding histogram corroborates this observation by showing the threshold within the peak around the darkest intensity levels. At a threshold value of 50 (Figure B.6B), too many pixels are classified as glottis, particularly in the lower part of the kymogram, yielding a lower edge that does not accurately follow the tissue motion. An intensity of 30 would be considered a reasonable threshold for the glottal segmentation of this digital kymogram.

A

Glottal segmentation · User feedback

B

Glottal segmentation · User feedback
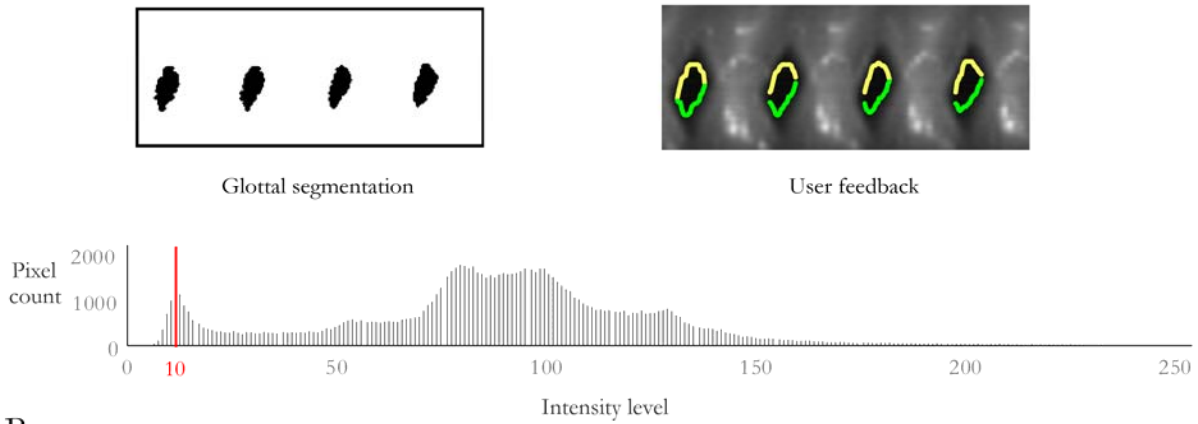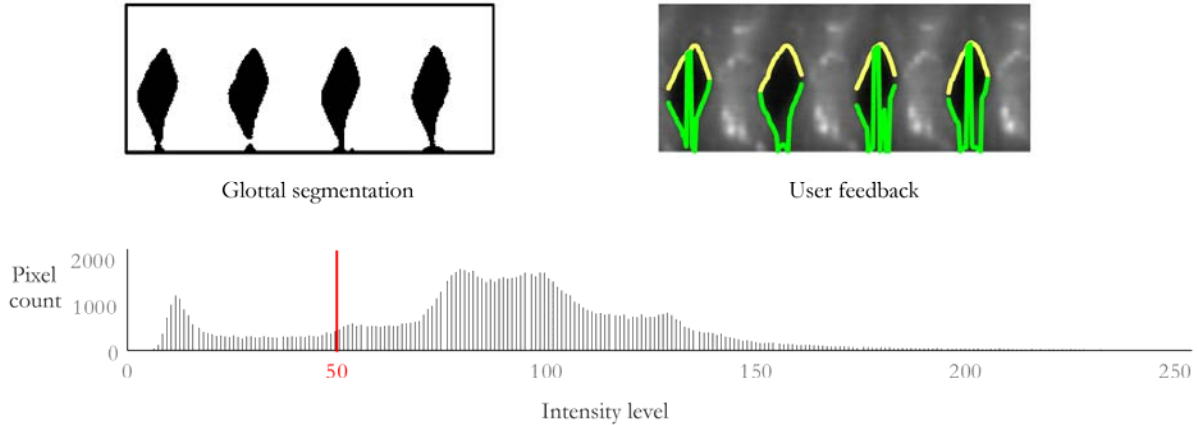
Figure B.6   Visual feedback to evaluate glottal segmentation on a digital kymogram with intensity threshold equal to (A) 10 and (B) 50. Shown are the segmentation output, user feedback with edges on the original kymogram, and intensity histogram of the original kymogram.

Once an appropriate threshold is determined, the edges are forced to be coincidental at the instants of glottal opening and closing, one column to the left and right of the segmented glottis of each cycle in the digital kymogram, as shown in Figure B.7. The yellow and green edges in Figure B.7 are referred to in this thesis as *lateral displacement waveforms*. As a function of column number (frame number), the lateral displacement waveforms are $x_L[n]$ and $x_R[n]$ for the left (top edge) and right (bottom edge) vocal folds, respectively. Note that the waveforms are undefined (NaN in MATLAB) for frames exhibiting glottal closure.
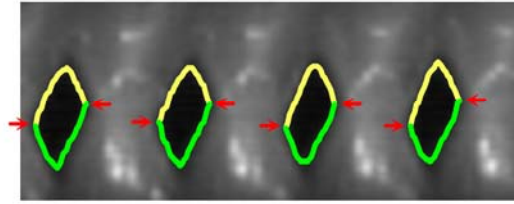
Figure B.7 Lateral displacement waveforms forced to be coincidental at instants (red arrows) of glottal opening and closure.

Appendix C

Chapter 3 Supplement

This section contains information supplementary to the study presented in Chapter 3.

Overview of the endoscopy setup

Figure C.1 shows a subject undergoing laryngeal high-speed videoendoscopy (HSV). In addition to the acoustic microphone, an electroglottograph and neck-mounted accelerometer were used to document voice-related activity for future studies.

Figure C.1   Subject undergoing laryngeal high-speed videoendoscopy.
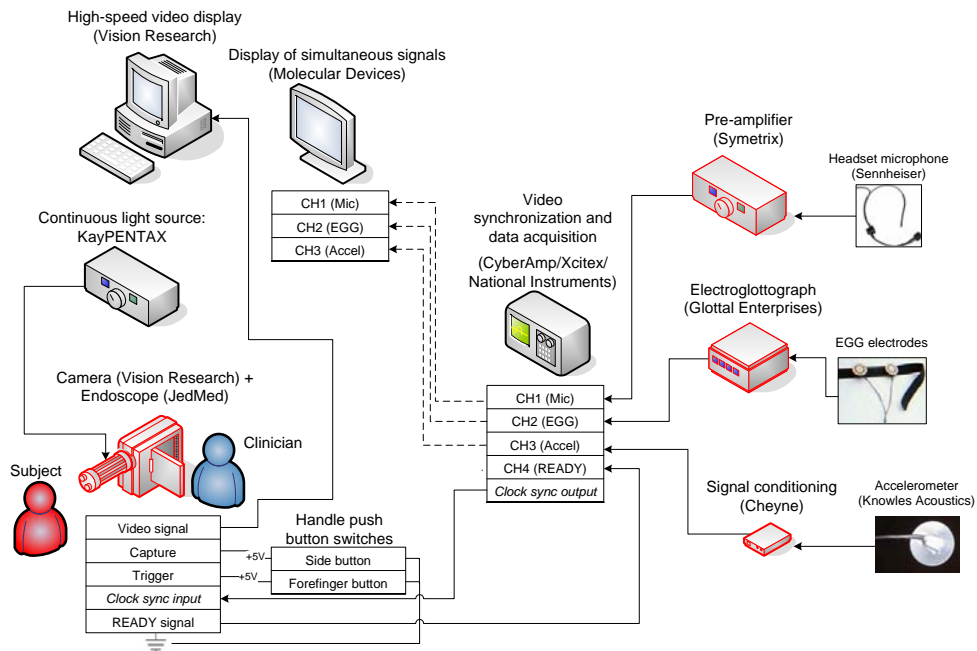
Figure C.2 Wiring diagram of the data acquisition system. Signals recorded (in red): high-speed video, acoustic microphone signal, electroglottogram, and neck-skin acceleration.

Figure C.2 displays the wiring diagram of the system built for time-synchronized recordings of laryngeal HSV and the three data channels. The data acquisition system captures four separate voice-related signals: high-speed videoendoscopic images, the acoustic microphone signal, electroglottography, and neck-skin acceleration. For the current investigation, the high-speed video and acoustic data were primarily used, with the additional signals providing supplementary information.

Time-Synchronized Signal Acquisition

Time-synchronization of the HSV data with the other sensor signals was critical for enabling correlations between HSV-based voice production measures and signal characteristics from the same phonatory segment and even from the same cycle. A simple trigger signal has been used in previous studies to align signals and video; but error propagation due to asynchronous crystal oscillators in multiple devices prevented analysis from segments far from the trigger event (Larsson *et al.*, 2000; Granqvist *et al.*, 2003). In our HSV system, the camera's clock was supplied by the National Instruments board's clock source that was synchronized to the sampling of the multiple

channels. The hardware clock division and data acquisition settings were controlled by MiDAS DA software (Xcitex Corporation, Cambridge, MA).

Alignment of the HSV data and the sensor data was accomplished by recording an analog Ready signal from the camera that exhibiting a falling TTL edge to precisely indicate the time of a recorded image. The 3 dB antialiasing cutoff frequency used is at the highest setting (30 kHz) to follow this falling edge as rapidly as possible to achieve a synchronization precision to within 11μs, which is on the order of one sample (sampling rate of 100 kHz). Alignment was validated with a simple light-emitting diode circuit. An error was found in the instruction manual of the high-speed camera. Instead of exhibiting a falling edge at the end of the last image's exposure, the Ready signal instead exhibited this falling edge one frame period later than the recorded frame.

With the sampling rates of the signals and HSV images synchronized, the ratio between the frame interval and sampling period is always a whole number. Thus, there is a finite set of sampling rates that can be used so that sampling rates of video and data channels are in the correct ratio. With video sampling at 6250 Hz and channel sampling at 100 kHz, there would be 16 audio samples per frame. With video sampling at 4000 Hz and channel sampling at 100 kHz, there would be 25 samples per frame.

Electroglottography

The electroglottograph transducer EG2-PC (Glottal Enterprises, Syracuse, NY) utilizes a high-pass filter to offer noise immunity and gross conductance changes due to neck motion. One of the disadvantages of including this high-pass filter is the presence of low-frequency amplitude distortion. Using the Larynx Simulator LS-1 to input a square wave into the EG2-PC, the distorting high-pass filter was estimated to be a fourth-order Butterworth filter with a 3 dB cutoff frequency of 18.5 Hz. Custom code compensated for the estimated distortion by convolving the distorted waveform by the time-reversed impulse response of the estimated high-pass filter (Rothenberg, 2002). The compensation algorithm essentially filtered the signal with a high-pass filter with the same magnitude, but reversed phase, spectrum, yielding an overall high-pass filter with zero phase. This compensation is not a canonical inverse filter, which aims to undo both magnitude and phase effects of a distortion filter.

Acceleration of the neck skin above the sternal notch

Neck skin acceleration is a non-invasive measure of vocal fold function that can provide indirect acoustic information from the subglottal system when placed appropriately. In our HSV system, the accelerometer (BU-7135, Knowles, Itasca, IL) was enclosed in an epoxy and mounted onto the subject's neck using Double Stick Discs (Model 2181, 3M, Maplewood, MN). It was positioned about 1 cm above the sternal notch. Custom-built circuitry provided pre-amplification and signal conditioning, as well as a knob to vary the DC level to maximize the dynamic range (Cheyne, 2002). The factor relating voltage to acceleration was derived in (Cheyne, 2002) and was approximately constant at 90 dB re 1 cm $\cdot s^2 \cdot V^{-1}$ (31 622 cm/s^2 per V) for frequencies between 70 Hz and 3000 Hz.

Graphical user interface

A graphical user interface was developed for aiding in the evaluation of high-speed video and audio data in an integrated playback format. A screen shot of the MATLAB-based Synchronous Signal View GUI is displayed in Figure C.3. The acoustic waveform, electroglottographic signal, and neck skin 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 the video playback rate and the acoustic propagation time between the larynx and microphone. With this integrated view, a clinician 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.
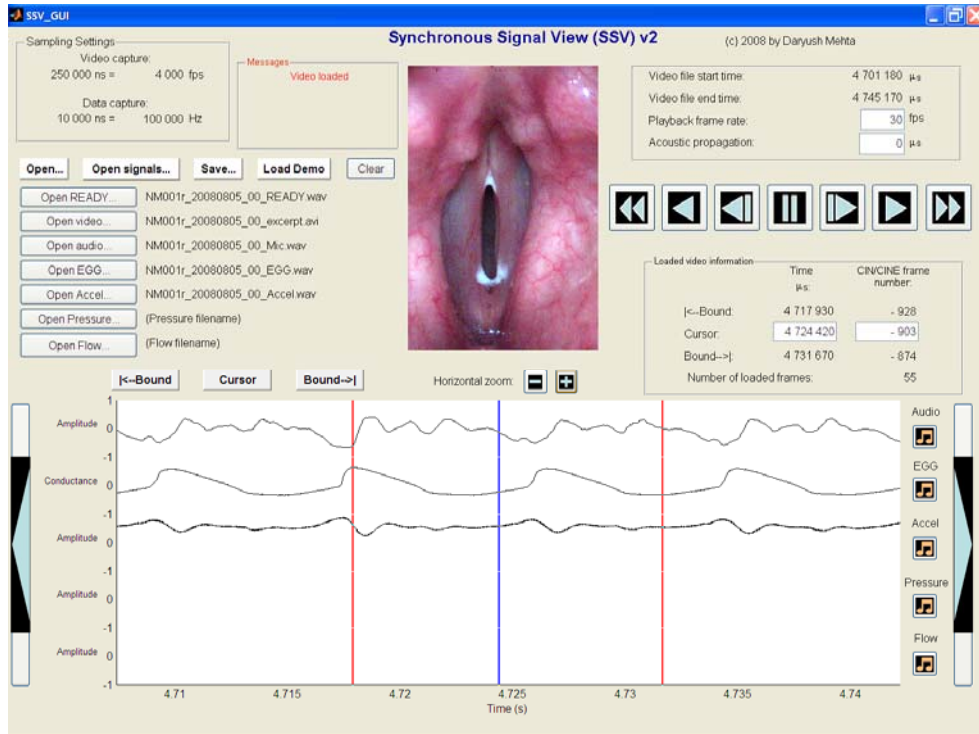
Figure C.3 Synchronous Signal View software for facilitating synchronous playback of high-speed videoendoscopic images and time-synchronized data channels.

Appendix D

Chapter 4 Supplement

This section contains information supplementary to the study presented in Chapter 4.

Acoustic perturbation

Voice quality depends on glottal characteristics and spectral tilt measures that reflect long-term phonatory characteristics across multiple glottal cycles and also irregularities or perturbations in the timing and amplitude of the source excitation that might give rise to creaky, rough, or hoarse voice perceptions. Average measures of the vocal fold vibratory asymmetry did no significantly correlate with acoustic jitter. However, magnitudes of left-right phase asymmetry and axis shift did correlate significantly ($p < 0.01$) with acoustic jitter, although the strength of correlation was low at about +0.37. Variations in left-right phase asymmetry and axis shift explain less than 15 % of the variance in acoustic shimmer. Special care must be taken when interpreting correlation results on perturbation measures due to their inherent sensitivity to noise and the risk of outliers inappropriately skewing the outcome measures.

A previous investigation into sources of acoustic perturbation found that standard deviations of vocal fold asymmetry measures correlated to a higher degree to jitter and shimmer than average measures of asymmetry (Mehta *et al.*, 2010). The subjects in that study formed a subset of the subjects in Chapter 4. Consistent with the previous work, the standard deviation metrics of left-right phase asymmetry ($r = 0.52$, $p < 0.001$), speed quotient ($r = 0.50$, $p < 0.001$), and closing quotient ($r = 0.36$, $p < 0.001$) correlated weakly to moderately with acoustic jitter. These measures were computed on medial digital kymograms due to robust edge detection relative to the full-frame

images. In addition, the left-right phase asymmetry measure was modified to normalize by the period duration instead of the open phase duration. None of the average or standard deviation metrics of asymmetry correlated significantly with harmonics-to-noise ratio.

Validation of acoustic spectral measures

The open quotient has been shown to vary linearly with $H1^*-H2^*$ in acoustic-based (Hanson and Chuang, 1999; Iseli *et al.*, 2007) and airflow-based (Holmberg *et al.*, 1995) studies. As expected, there is a positive, albeit weak, correlation between open quotient and $H1^*-H2^*$ ($r = 0.28$, $p < 0.05$). This result is consistent with previous findings comparing the two measures in normal speakers and subjects with vocal fold nodules (Kuo, 1998). In that study, whereas there was a strong correlation ($\varrho = 0.77$) between the two measures in the normal speaker cohort, the correlation disappeared in the group with nodules.

Interestingly, closing quotient performed better than speed quotient when correlated with the spectral tilt of the radiated pressure waveform. Speed quotient was expected to correlate negatively with $H1^*-A3^*$ and positively with spectral tilt over the first eight harmonics (TL^*), owing to the fact that area (and thus airflow) skewing excites higher-frequency harmonics in the magnitude spectrum. Speed quotient did, in fact, correlate negatively with $H1^*-A3^*$ ($r = -0.28$, $p < 0.05$), although weakly, but did not correlate significantly with TL^*. In comparison, the measure of closing quotient CQ correlated moderately with both $H1^*-A3^*$ ($r = +0.48$, $p < 0.01$) and TL^* ($r = -0.40$, $p < 0.01$) with expected directionalities. Perhaps the better performance of closing quotient measure was due to its incorporation of timing information, via the period, which reflected rapid closure better than speed the speed quotient metric, which normalizes skewness by the open phase duration.

Appendix E

Informed Consent Form

Included here is the informed consent form approved by the MGH Institutional Review Board to provide to prospective subjects participating in the studies described in Chapters 3 and 4. Only the sections of the consent form pertaining to rigid endoscopy were relevant to the thesis project (the flexible endoscopy setup was used for other purposes).

Subject Identification

General Template
Version Date: November 2005

Protocol Title: Efficacy of Laryngeal High-Speed Videoendoscopy

Principal Investigator: Robert E. Hillman, Ph.D.

Site Principal Investigator:

Description of Subject Population: Adults with voice disorders and adults with normal voices

About this consent form

Please read this form carefully. It tells you important information about a research study. A member of our research team will also talk to you about taking part in this research study. People who agree to take part in research studies are called "subjects." This term will be used throughout this consent form. If you have any questions about the research or about this form, please ask us. If you decide to take part in this research study, you must sign this form to show that you want to take part. We will give you a copy of this form to keep.

Why is this research study being done?

The purpose of this research study is to develop better ways to use high-speed camera pictures of vocal cords. Vocal cords are stretchy tissues in the throat that allow us to make sounds. We hope this new technology will help doctors and scientists better understand the human voice. We are testing a new and improved version of high-speed camera technology. Older versions of high-speed cameras have been used in the same way in clinical settings for over 10 years to take pictures of vocal cords.

We are asking you to take part in this study either because you have a voice disorder or because you have a normal voice. We need to collect high-speed recordings, and other measurements from a large group of adults with voice problems, as well as from a few adults with normal voices. Our study testing involves use of an instrument called an endoscope: a thin lighted tube that is inserted in your nose or mouth and allows us to see the details of the larynx (voicebox) and vocal cords.

Subject Population: Adults with voice disorders and adults with normal voices	
IRB Protocol No.: 2008p000652	**Sponsor Protocol No.:** Sub 2RO1DK050623-10A2
Consent Form Valid Date: 04/29/2009	**IRB Amendment No.**: 3 **Sponsor Amendment No.**: N/A
IRB Expiration Date: 02/19/2010	**IRB Amendment Approval Date:** 04/24/2009

Subject Identification

We will ask about 12 subjects with voice disorders and 16 subjects with normal voices to take part in some extra testing for this study. Using several different devices, we will measure air flow and record other voice qualities of these subjects.

We are doing this study with researchers at the University of South Carolina. Over the next few years, we hope to enroll 462 patients with voice disorders in the study at MGH, as well as 40 subjects with normal voices.

How long will I take part in this research study?

If you have been diagnosed with a voice disorder that requires surgery, we will ask you to spend an extra 30 minutes today (before your surgery) having study tests. We will repeat this same testing again when you return to the clinic for your regular checkup after surgery. We will ask subjects with normal voices to undergo this testing only once. Subjects with normal voices will not be having surgery. These tests will again take about 30 extra minutes of your time.

About 12 subjects with voice disorders, and 24 subjects with normal voices will have extra testing that includes air flow measurements. If you are in this group, we will schedule an extra session at your convenience. This test session will last about 1 hour. Subjects with normal voices will not be having surgery, and some subjects diagnosed with certain kinds of voice disorders may not have surgery.

What will happen in this research study?

If you have been diagnosed with a voice disorder that requires surgery, you will have two sets of tests. We will do one set of tests today, before your surgery. The second set of tests will happen at your regular office visit after your surgery. If you have a normal voice, you will have one set of tests.

We will do testing as you make and hold a vowel sound (like "eeeeee") for 3-5 seconds several times. You will need to make the sound at different vocal pitch (high or low) and loudness levels. You will have the following study tests:

- **A laryngeal (larynx) examination.** We will use the same type of instrument (endoscope) that was used to examine your larynx in the clinic. The endoscope will

Subject Population: Adults with voice disorders and adults with normal voices

IRB Protocol No.: 2008p000652	**Sponsor Protocol No.:** Sub 2RO1DK050623-10A2
Consent Form Valid Date: 04/29/2009	**IRB Amendment No.**: 3 **Sponsor Amendment No.**: N/A
IRB Expiration Date: 02/19/2010	**IRB Amendment Approval Date:** 04/24/2009

attaches to a special high-speed camera, instead of to the camera used in the clinic. In order to use the special camera, we need to also use a very bright light to make pictures. There may be a slight risk of tissue overheating associated with use of the high intensity light source which could cause a burning sensation. Please tell us immediately if you feel a burning sensation, and we will stop the testing.

If you tend to gag during this type of examination, we will first spray your throat with a topical anesthetic (numbing medicine.) This will help control any gagging. Then we will pass the endoscope through your mouth to the back of your throat.

- **Audio recordings of your speech**. We will use a small microphone suspended from headgear that will be adjusted to fit comfortably on your head.

- **Electroglottography.** We will make recordings from two small wires (electrodes) placed on your neck over your voice box. A Velcro strap will hold the electrodes in place. These electrodes check how your vocal cords touch each other when you produce a voice. This is a standard test used in our clinic.

- **Accelerometer recordings.** An accelerometer is a small device attached to the base of your neck (in front) with sticky tape. This device measures the vibrations of the skin that are caused when your vocal cords vibrate to make a sound.

If you are having an operation, we will measure the height, width, and length of the scar that we remove from your vocal cord. We measure this scar because we want to study how the size of the scar effects how your vocal cords work.

If You Are Having the Extra Study Testing

A total of 24 subjects with normal voices will take part in this study session, which will take about 1 hour to complete. If you have a voice disorder that requires surgery, and agree to have the extra study testing, we will schedule a separate study visit for these tests prior to your surgery.

We will make air flow measurements and use a flexible endoscope to examine your larynx. A flexible endoscope goes in through the nose and down the throat, instead of going into the mouth.

Subject Population: Adults with voice disorders and adults with normal voices	
IRB Protocol No.: 2008p000652	**Sponsor Protocol No.:** Sub 2RO1DK050623-10A2
Consent Form Valid Date: 04/29/2009	**IRB Amendment No.**: 3 **Sponsor Amendment No.**: N/A
IRB Expiration Date: 02/19/2010	**IRB Amendment Approval Date:** 04/24/2009

We will have you repeat some speech sounds (p-sounds + vowel-sounds) several times at different vocal pitch and loudness levels. While you make the sounds, we will do the following procedures:

- **Record the air flow.** A specially designed face mask that is held in place over your nose and mouth by elastic straps that fit around your head will measure the air flow.

- **Record the air pressure** in your mouth while you are making p-sounds. To do this, we use a small tube that fits between your lips (like a small straw).

- **A laryngeal examination.** We will use a standard flexible endoscope fit through a hole in the air flow face mask. The flexible endoscope will pass into one of your nostrils and nasal cavity, and then into the back of your throat.

  Before putting in the instrument, we will spray your nostril with a numbing medicine to help reduce any possible discomfort. For this testing, we will attach the flexible endoscope to a high-speed camera, instead of to the clinic camera we routinely use for these exams.

- **Audio recordings of your speech.** We will use a small microphone that is attached to the air flow face mask.

- **Electroglottography**, as described above.

- **Accelerometer recordings**, as described above.

Study information will become part of your medical record and will be shared with your doctor.

What are the risks and possible discomforts from being in this research study?

Risks of Standard Endoscopy

If you have been diagnosed with a voice disorder that requires surgery, the risks of taking part in this study are the same as for the examination that you had at the clinic. We use the same type of endoscope for the study procedure.

Subject Identification

You may cough or gag when we place the endoscope into your throat. We will use numbing medicine (Cetacaine or Lidocaine) to lower the risk of these reactions. Before the endoscope goes into your throat, we will spray your throat with numbing medicine. We do laryngeal endoscopies routinely in the MGH Voice Center, so we are prepared to deal with these types of reactions.

Risks of the Flexible Endoscopy

If you are having the extra tests that include air flow measurements, this testing will involve the use of a flexible endoscope, you may feel some pressure or fullness as the endoscope passes through your nose. There is also a slight risk of a nose bleed. We will stop putting the flexible endoscope into your nose if your nasal passages are too narrow. This will reduce the chances of nose bleed. To help reduce any possible discomfort, we will spray the nostril receiving the endoscope with a topical anesthetic (Lidocaine).

You may also experience some coughing or gagging when the flexible endoscope is passing into your throat. Because we are very experienced with laryngeal endoscopy in the MGH Voice Center, we know how best to avoid or treat these types of reactions. We can carefully control the endoscope placement.

Risks of Topical Anesthesia (numbing spray)

There is the risk of an allergic reaction to the topical anesthesia. This would typically cause irritation, itching, and/or swelling (edema) in the mouth and throat. We will ask if you have a history of such reactions. If an allergic reaction does occur, we will stop the testing and will give you medical attention. If you have not had problems with earlier anesthesia or exams of your larynx, the risk of an allergic reaction is very small.

Important Note: If you receive a topical anesthetic, <u>you should not eat or drink until the effect of the anesthetic has worn off, to avoid choking. This typically takes about 30 minutes</u>. A study staff member will stay with you until the effects of the anesthetic have worn off.

Risks of High Intensity Light Source

Subject Population: Adults with voice disorders and adults with normal voices	
IRB Protocol No.: 2008p000652	**Sponsor Protocol No.:** Sub 2RO1DK050623-10A2
Consent Form Valid Date: 04/29/2009	**IRB Amendment No.**: 3 **Sponsor Amendment No.**: N/A
IRB Expiration Date: 02/19/2010	**IRB Amendment Approval Date:** 04/24/2009

For all subjects, there may be a slight risk of tissue overheating associated with use of the high intensity light source which could cause a burning sensation. Please tell us immediately if you feel a burning sensation, and we will stop the testing. This special light is necessary to make recordings with the high speed camera. To reduce this risk, we will limit your exposure to the light source to less than 1 minute each time you have a laryngeal exam.

Other Risks

For all subjects, removing the tape that holds the accelerometer on the neck may cause you brief discomfort similar to removing a Band-Aid. You may have a tiny bit of skin irritation where the sticky tape came off.

It is possible that you may feel some slight discomfort from the mask or from the Velcro straps around your head and neck that we use to hold the devices in place during testing. Please tell us if you feel discomfort, so we can adjust the equipment.

There may be other risks or side effects that are not known at this time.

What are the possible benefits from being in this research study?

You may not benefit from participating in this study.

If you have a voice disorder, you may benefit by getting more detailed information about how your voice is functioning. This might be useful in managing your disorder.

We hope to learn new and better ways for using high-speed cameras to understand how we make voices. We hope this will lead to treatments for voice disorders, and reveal new information about the human voice. This may benefit future patients with voice disorders.

What other treatments or procedures are available for my condition?

This is not a treatment study. We are taking measurements of voices to learn more about special cameras to help diagnose and treat voice disorders. If you have a voice disorder, you can receive standard treatment at the MGH Voice Center for your condition without being part of this study.

Subject Identification

Can I still get medical care within Partners if I don't take part in this research study, or if I stop taking part?

Yes. Your decision won't change the medical care you get within Partners now or in the future. There will be no penalty, and you won't lose any benefits you receive now or have a right to receive.

Taking part in this research study is up to you. You can decide not to take part. If you decide to take part now, you can change your mind and drop out later. We will tell you if we learn new information that could make you change your mind about taking part in this research study.

If you take part in this research study, and want to drop out, you should tell us. We will make sure that you stop the study safely. We will also talk to you about follow-up care, if needed.

It is possible that we will have to ask you to drop out before you finish the study. If this happens, we will tell you why. We will also help arrange other care for you, if needed.

Will I be paid to take part in this research study?

If you have a normal voice and participate in the testing, you will be paid $25. You will be paid this entire amount even if you are unable to complete the entire testing session.

If you participate in more extensive testing that includes air flow measurements and the use of a flexible endoscope, you will be paid $50. If you have a normal voice, this amount is in addition to the amount for the standard testing. You will be paid this entire amount even if you are unable to complete the entire testing session.

What will I have to pay for if I take part in this research study?

Study funds will pay for all procedures done only for the study. This includes everything at the extra testing session, if you take part, and for subjects with normal voices that will take part in the initial testing session.

Study funds will pay for all the study tests done before and after your surgery, if you have surgery for your voice disorder. The costs of your regular medical care will be billed to you or your insurance company in the usual way.

What happens if I am injured as a result of taking part in this research study?

We will offer you the care needed to treat any injury that directly results from taking part in this research study. We reserve the right to bill your insurance company or other third parties, if appropriate, for the care you get for the injury. We will try to have these costs paid for, but you may be responsible for some of them.

Giving you care does not mean that Partners hospitals or researchers are at fault, or that there was any wrongdoing. There are no plans for Partners to pay you or give you other compensation for the injury. However, you are not giving up any of your legal rights by signing this form.

If you think you have been injured or have experienced a medical problem as a result of taking part in this research study, tell the person in charge of this study as soon as possible. The researcher's name and phone number are listed in the next section of this consent form.

If I have questions or concerns about this research study, whom can I call?

You can call us with your questions or concerns. Our telephone numbers are listed below. Ask questions as often as you want.

Robert E. Hillman, Ph.D. is the person in charge of this research study. You can page him at 617-644-0648, Monday – Friday 7 am – 10 pm. You can also call Steven Zeitels, M.D. or James Burns, M.D. at 617-726-1444, Monday – Friday 8 am – 5 pm with questions about this research study.

If you have questions about the scheduling of appointments or study visits, call Jennifer Bourque at 617-643-2466.

Subject Population: Adults with voice disorders and adults with normal voices	
IRB Protocol No.: 2008p000652	**Sponsor Protocol No.:** Sub 2RO1DK050623-10A2
Consent Form Valid Date: 04/29/2009	**IRB Amendment No.**: 3 **Sponsor Amendment No.**: N/A
IRB Expiration Date: 02/19/2010	**IRB Amendment Approval Date:** 04/24/2009

If you want to speak with someone **not** directly involved in this research study, please contact the Partners Human Research Committee office. You can call them at 617-424-4100.

You can talk to them about:
▪ Your rights as a research subject
▪ Your concerns about the research
▪ A complaint about the research

Also, if you feel pressured to take part in this research study, or to continue with it, they want to know and can help.

If I take part in this research study, how will you protect my privacy?

Federal law requires Partners (Partners HealthCare System and its hospitals, health care providers and researchers) to protect the privacy of health information that identifies you. This information is called Protected Health Information. In the rest of this section, we refer to this simply as "health information."

If you decide to take part in this research study, your health information may be used within Partners and may be shared with others outside of Partners, as explained below.

We have marked with a ⊠ how we plan to use and share your health information. If a box is not checked ☐, it means that type of use or sharing is not planned for in this research study.

We will also give you the **Partners Notice for Use and Sharing of Protected Health Information.** The Notice gives more details about how we use and share your health information.

▪ **Health Information About You That Might be Used or Shared During This Research**

⊠ Information from your hospital or office health records within Partners or elsewhere, that may be reasonably related to the conduct and oversight of the research study. If health information is needed from your doctors or hospitals outside Partners, you will be asked to give permission for these records to be sent to researchers within Partners.

Subject Population: Adults with voice disorders and adults with normal voices		
IRB Protocol No.: 2008p000652	**Sponsor Protocol No.:** Sub 2RO1DK050623-10A2	
Consent Form Valid Date: 04/29/2009	**IRB Amendment No.:** 3	**Sponsor Amendment No.:** N/A
IRB Expiration Date: 02/19/2010	**IRB Amendment Approval Date:** 04/24/2009	

⊠　New health information from tests, procedures, visits, interviews, or forms filled out as part of this research study

- **Why Health Information About You Might be Used or Shared with Others**

  The reasons we might use or share your health information are:

  - To do the research described above

  - To make sure we do the research according to certain standards - standards set by ethics and law, and by quality groups

  - For public health and safety - for example, if we learn new health information that could mean harm to you or others, we may need to report this to a public health or a public safety authority

  - For treatment, payment, or health care operations

- **People and Groups That May Use or Share Your Health Information**

  1. **People or groups within Partners**

     ⊠　Researchers and the staff involved in this research study

     ⊠　The Partners review board that oversees the research

     ⊠　Staff within Partners who need the information to do their jobs (such as billing, or for overseeing quality of care or research)

  2. **People or groups outside Partners**

     ⊠　People or groups that we hire to do certain work for us, such as data storage companies, our insurers, or our lawyers

     ⊠　Federal and state agencies (such as the U.S. Department of Health and Human Services, the Food and Drug Administration, the National Institutes of Health, and/or the Office for Human Research Protections) and other U.S. or foreign government bodies, if required by law or involved in overseeing the research

     ⊠　Organizations that make sure hospital standards are met

     ☐　The sponsor(s) of the research study, and people or groups it hires to help perform this research study

     ⊠　Other researchers and medical centers that are part of this research study

Subject Population: Adults with voice disorders and adults with normal voices	
IRB Protocol No.: 2008p000652	**Sponsor Protocol No.:** Sub 2RO1DK050623-10A2
Consent Form Valid Date: 04/29/2009	**IRB Amendment No.**: 3　　**Sponsor Amendment No.**: N/A
IRB Expiration Date: 02/19/2010	**IRB Amendment Approval Date:** 04/24/2009

☐ A group that oversees the data (study information) and safety of this research study

☐ Other:

Some people or groups who get your health information might not have to follow the same privacy rules that we follow. We share your health information only when we must, and we ask anyone who receives it from us to protect your privacy. However, once your information is shared outside Partners, we cannot promise that it will remain private.

▪ **Time Period During Which Your Health Information Might be Used or Shared With Others**

- Because research is an ongoing process, we cannot give you an exact date when we will either destroy or stop using or sharing your health information.

▪ **Your Privacy Rights**

- You have the right **not** to sign this form permitting us to use and share your health information for research. If you don't sign this form, you can't take part in this research study. This is because we need to use the health information of everyone who takes part in this research study.

- You have the right to withdraw your permission for us to use or share your health information for this research study. If you want to withdraw your permission, you must notify the person in charge of this research study in writing.

  If you withdraw your permission, we will not be able to take back information that has already been used or shared with others. This includes information used or shared to carry out the research study or to be sure the research is safe and of high quality.

  If you withdraw your permission, you cannot continue to take part in this research study.

- You have the right to see and get a copy of your health information that is used or shared for treatment or for payment. To ask for this information, please contact the person in charge of this research study.

▪ **If Research Results Are Published or Used to Teach Others**

Subject Population: Adults with voice disorders and adults with normal voices		
IRB Protocol No.: 2008p000652	Sponsor Protocol No.: Sub 2RO1DK050623-10A2	
Consent Form Valid Date: 04/29/2009	IRB Amendment No.: 3	Sponsor Amendment No.: N/A
IRB Expiration Date: 02/19/2010	IRB Amendment Approval Date: 04/24/2009	

The results of this research study may be published in a medical book or journal, or used to teach others. However, your name or other identifying information **will not** be used for these purposes without your specific permission.

Consent/Assent to take part in this research study, and authorization to use or share your health information for research

Statement of Subject or Person Giving Consent/Assent

- I have read this consent form.
- This research study has been explained to me, including risks and possible benefits (if any), other options for treatments or procedures, and other important things about the study.
- I have had the opportunity to ask questions.

If you understand the information we have given you, and would like to take part in this research study, and also agree to allow your health information to be used and shared as described above, then please sign below:

Signature of Subject:

_____   _____

Adults <u>or</u> Minors, ages 14-17   Date/Time

OR

If you understand the information we have given you, and would like to give your permission for your child/the person you are authorized to represent to take part in this research study, and also agree to allow his/her health information to be used and shared as described above, then please sign below:

Signature of Parent(s)/Guardian or Authorized Representative:

_____	_____
Parent(s)/Guardian of Minor	Date/Time

OR

_____	_____
Court-appointed Guardian or Health Care Proxy	Date/Time

OR

_____	_____
Family Member/Next-of-Kin	Date/Time

Relationship to Subject:_____

Signature of a Witness:

_____	_____
Witness (when required by the PHRC <u>or</u> sponsor)	Date/Time

Statement of Study Doctor or Person Obtaining Consent

- ▪ I have explained the research to the study subject, and
- ▪ I have answered all questions about this research study to the best of my ability.

Subject Population: <u>Adults with voice disorders and adults with normal voices</u>	
IRB Protocol No.: <u>2008p000652</u>	**Sponsor Protocol No.:** <u>Sub 2RO1DK050623-10A2</u>
Consent Form Valid Date: <u>04/29/2009</u>	**IRB Amendment No.**: <u>3</u> **Sponsor Amendment No.**: <u>N/A</u>
IRB Expiration Date: <u>02/19/2010</u>	**IRB Amendment Approval Date:** <u>04/24/2009</u>

Subject Identification

_____   _____

Study Doctor or Person Obtaining Consent   Date/Time

In certain situations, the Partners Human Research Committee (PHRC) will require that a subject advocate also be involved in the consent process. The subject advocate is a person who looks out for the interests of the study subject. This person is not directly involved in carrying out the research. By signing below, the subject advocate represents (or "says") that the subject has given meaningful consent to take part in the research study.

Statement of Subject Advocate Witnessing the Consent Process

- I represent that the subject or authorized individual signing above has given meaningful consent.

_____   _____

Subject Advocate (when required by the PHRC or sponsor)   Date/Time

Consent Form Version Date: 04/22/09

Subject Population: Adults with voice disorders and adults with normal voices	
IRB Protocol No.: 2008p000652	**Sponsor Protocol No.:** Sub 2RO1DK050623-10A2
Consent Form Valid Date: 04/29/2009	**IRB Amendment No.**: 3 **Sponsor Amendment No.**: N/A
IRB Expiration Date: 02/19/2010	**IRB Amendment Approval Date:** 04/24/2009

Appendix F

Experimental Protocol for Subject Data Collection

This section presents the protocol used at the MGH Voice Center to record and organize data from subjects in the studies described in Chapters 3 and 4. Settings were specified for the high-speed camera and the acquisition board responsible for recording the time-synchronized data channels.

Preparation before subject enters

File organization
- [ ] If new subject, assign subject ID XY001z:
  - X = N, P (normal, path)
  - Y = M, F (male, female)
  - 00 = 2-digit sequential numbering
  - z = r, f (rigid, flex)

  For example, ID for male path subject with rigid endoscopy would be *PM001r*.
- [ ] Record subject ID, subject type (normal/path), gender (male/female), type of endoscopy (rigid/flex), name, date, diagnosis (if applicable), and this unique identifier in *Desktop\High-speed subject info\subject_info.xls*.
- [ ] Place copy of *Desktop\High-speed subject info\zzzz NEW folder container* into *Desktop\High-speed subject info*.
- [ ] Edit README.txt file for new subject.

Hardware setup
- [ ] Weights required to counterbalance the color camera, in order from inside to outside: 2.5 lb, 2.5 lb, 5 lb, 10 lb, 10 lb, 10 lb, 5 lb, 10 lb (one more than mono camera)
- [ ] Turn on CyberAmp 380, if necessary (green light indicator).
- [ ] Turn on Digidata 1440A if necessary (green light indicator).
- [ ] Turn EGG box knob to A or B for power.
- [ ] Turn on phantom power on Symetrix microphone preamplifier (left channel used).
- [ ] Turn on Xenon light source to test. Turn off for now. Angle box toward chair for closer fiber optic cable proximity.
- [ ] Plug in power for Phantom camera. Keep camera fan on for now.
- [ ] Obtain and sterilize rigid endoscope. Fits into Storz port on Xenon light source.
- [ ] Get hot water and gauze ready near chair.

Channel Name	MiDAS DA hardware (software) channel	Digidata (CyberAmp) channel	CyberAmp gain (pre/total/ EGG box)	Lowpass filter (Hz)	Comments
Mic	0 (1)	1 (1)	10/20	30 000	Acoustic signal
EGG	1 (2)	2 (2)	1/2/high	30 000	EGG box has high/low gain
Accel	2 (3)	3 (3)	10/500	30 000	Autozero the neck accelerometer signal
READY	3 (4)	4 (4)	1/1	30 000	Necessary for video synchronization
FSYNC	6 (7)	8 (8)	1/1	30 000	Display only
TRIGGER	N/A	14 (N/A)	N/A	N/A	Display only
STROBE	7 (8)	15 (N/A)	N/A	N/A	Display only

Software setup

AxoScope for displaying signals and setting gain/filter settings of CyberAmp
☐ Open AxoScope: *AxoScope 10.2* (wireless mouse!)
☐ Default settings: Click Acquire, Open Protocol…,
Desktop\AxoScope High-speed protocols\HSV protocol_rigid.pro
☐ Click Configure, Lab Bench…, Verify that CyberAmp settings for each channel assign DC to positive end and GND to negative end of sensor. DC offsets should be 0.00 volts.
☐ Click play to verify that all sensors are generating signals.

MiDAS DA for acquiring data signals
☐ Open *MiDAS DA*
☐ Default settings for 4000, 6250, 8000, and 10000 frames per second.
Click Open, *Desktop\MiDAS DA High-speed protocols\HSV protocol_rigid**
        * = _4000fps.drc, _6250fps.drc, _8000fps.drc, and *_10000fps.drc*
☐ Make sure check mark is on all channels
☐ Default settings:
        Sample Rate (Hz): 100000 [4000 fps], 100000 [6250 fps], 160000 [8000 fps], 100000 [10 000 fps]
        Buffer Size (sec): 5
        sync out—Rate (Hz): 4000 [actual: 4000], 6250 [actual: 6250], 8000 [actual: 8000], or 10 000 [actual: 10 000]
        Trigger Percentage: 99
        Trigger Type: Falling Edge
☐ Click Green Circle to send external clock to Phantom camera

Phantom software for acquiring high-speed video
☐ Open Phantom software: *Phantom 640*
        • From menu bar, click Acquisition, Setup and Recording...
☐ Default setting for 4000, 6250, 8000, and 10 000 frames per second. Click Open…,
Desktop\Phantom High-speed protocols\HSV protocol_rigid*
        * = _4000fps.stg, _6250fps.stg, _8000fps.stg, and *_10000fps.stg*
☐ Default settings:
        Resolution: 304 x 352
        Sample rate: (external clock)
        Exposure: 247.5 μs (4000 fps), 157.5 μs (6250 fps), 122.5 μs (8000 fps), 97.5 μs (10 000 fps)
        EDR Exp.: 0 μs
        PostTrig: 1 p
        Mode—Auto exposure (unchecked), Sync: External
        White balance (Click Adjust…): Red–1.175; Blue–0.933
        Bit range (Click Adjust…): 0–4511
☐ Verify memory is partitioned to desired segments. See below for experimental protocol. MiDAS DA buffer duration should be larger than the largest Phantom memory partition.
☐ Click *Capture* to start saving to camera memory buffer.
☐ Black reference calibration—remember to calibrate in External Sync mode—cover lens:
Click *Current Session Reference…*

☐ If there are multiple partitions, click Cine, Set to all. This sets black reference to all partitions.

Experiment protocol with subject

☐ Consent form and, for normals only, payment form.
☐ Determine phonatory pitch range to decide on frame rate. Use second microphone (AKG C420) input into camera PC. Open Wavesurfer and new file with saved configuration *real-time_pitch_tracking*. Record a few seconds of phonation, click stop, and read off estimated pitches.
☐ Turn camera fan off.
☐ Focusing: In Phantom software, Open protocol for focus adjustment. Only low light intensity required.

For 45 mm (normal) lens, rotate counter-clockwise fully (facing subject) and rotate clockwise ½ to 1 notch for ballpark focal lengths (3.5–5.0 cm):
- to move focus to a farther object, rotate focus knob clockwise (facing subject)
- to move focus to a closer object, rotate counter-clockwise (facing subject)

For 80 mm (zoom) lens, rotate clockwise fully (facing subject) and rotate counter-clockwise 4 ¼ to 5 notches for ballpark focal lengths (3.5–5.0 cm):
- to move focus to a farther object, rotate focus knob counter-clockwise (facing subject)
- to move focus to a closer object, rotate clockwise (facing subject)

☐ Light less than maximum on oral entry to prevent heat sensation.

TRIALS
Normals and disordered (pre-/post-surgery)

☐ 1. Normal pitch and loudness, vowel with onset/offset or sustained phonation, 6250 fps, full partition but trim to phonation duration.
☐ 2. Higher pitch, vowel with onset/offset or sustained phonation, 6250 fps, full partition but trim to phonation duration
For singers at very high pitches, consider higher frame rate

☐ After subject performs task, press foot pedal or forefinger button to trigger signal acquisition.
☐ Set clock to *Internal* in Phantom to enable live preview.
☐ Save MiDAS DA .dap file after every trial (files renamed later anyway): *Subject_date_0000.dap*
☐ Re-arm MiDAS DA acquisition: File, Close Project, OK, click Green Circle.
☐ Save CyberAmp settings in AxoScope. Click File, Properties, and copy rows from "ADC channel" to "Lowpass (Hz)." Paste into README.txt file.
☐ Set clock to *External* in Phantom to re-arm acquisition.

☐ After all trials complete, turn camera fan on to cool camera.
☐ Save all .cine files at once with automatic incrementing. Click File, Save as…, All cines from the RAM of the selected camera… (Check box for "No range (full cine always)"): *Subject_date*, will output *Subject_date_Cine1.cine*, *Subject_date_Cine2.cine*, etc. Monochromatic camera takes about 10 minutes to download entire 4 GB memory buffer. Color camera takes about 4 minutes to download entire 4 GB memory buffer.

Calibration

Microphone calibration
☐ Obtain Cooper-Rand EL synthesizer to output reference tone and sound level meter.
☐ Default settings: Click File, Open Recording Configuration…, *Desktop:\MiDAS DA High-speed protocols\HSV protocol_calibration_MIC.drc*
☐ Make sure check mark is on all channels.

Channel Name	MiDAS DA hardware channel	Comments
Mic	0	Acoustic signal

☐ Default settings:
  Sample Rate (Hz): 100000
  Buffer Size (sec): 10 [modify if desired]
  sync out—Rate: 4000
  Trigger Percentage: 99
  Trigger Type: Falling Edge
☐ Click Green Circle.
☐ Mount sound level meter on tripod, attach microphone to meter transducer at same angle, and place rubber band on EL with volume low. Turn volume up on box and verify that signal does not saturate.
☐ Press foot pedal to trigger signal acquisition. Keep vowel steady for a ~5 s, turn off, and call out sound level, in dB SPL, from sound level meter.

EGG calibration
  None, see Cheyne's thesis for reasons.

Accelerometer
  None, see Cheyne's thesis for conversion from volts to cm/s^2.

File formatting

☐ Set permissions on all .cine and .dap files (not directories) to *Deny Delete* to prevent deletion of raw files. (Right click, Properties, Security tab, Advanced, Permissions tab, Edit…, tick checkbox under *Deny* for Delete permissions.)
☐ Rename raw files (.cine and .dap) as SubjectID_date_XX. Increment XX, starting with 00, for each trial in order of time file created.
☐ Directory structure. Place .cine files into *VIDEO raw* folder and .dap files into *DATA raw* folder.
  SubjectID_date\
    DATA converted\
      MAT\
      TXT\
      WAV\
    DATA raw\
    VIDEO converted\
    VIDEO processed\
    VIDEO raw\

README.txt

☐ Edit README.txt file for protocol and settings.

☐ Export MiDAS DA .dap files as .txt with same filename.

☐ Run *run_MiDAS_DA_txt2mat_batch* and *run_MiDAS_DA_txt2wav_batch* to convert MiDAS DA .txt data to .mat and .wav files.

Explanation of file formats and directory organization

SubjectID — Directory naming convention is AX,
  A = N (normal) or P (path); X = sequential numbering
  AXpre indicates pre-surgery data collection
  AXpost indicates post-surgery data collection
File naming convention — SubjectID_XX, XX = trial number starting with 00; files with the same root name indicate that data was obtained from the denoted trial
DATA converted\ — directory contains data signals (not video) that have been converted from raw formats into other file formats
MAT\ — directory containing data in Matlab MAT file format, converted from TXT data format using run_MiDAS_DA_txt2mat_batch.m. Matlab variables:
  Accel — structure holding accelerometer data with the following fields:
    Units — should be 'volts'
    Scale — should be 1 always
    Offset — should be 0 always
    Data — vector of raw data (in volts) from the digitizer
  EGG — structure holding electroglottography data with the following fields:
    Units — should be 'volts'
    Scale — should be 1 always
    Offset — should be 0 always
    Data — vector of raw data (in volts) from the digitizer
  Flow (if applicable) — structure holding air flow volume velocity data with the following fields:
    Units — should be 'volts'
    Scale — should be 1 always
    Offset — should be 0 always
    Data — vector of raw data (in volts) from the digitizer
  Mic — structure holding the microphone's acoustic data with the following fields:
    Units — should be 'volts'
    Scale — should be 1 always
    Offset — should be 0 always
    Data — vector of raw data (in volts) from the digitizer
  Pressure (if applicable) — structure holding the intraoral pressure data with the following fields:
    Units — should be 'volts'
    Scale — should be 1 always
    Offset — should be 0 always
    Data — vector of raw data (in volts) from the digitizer
  READY — structure holding the Phantom READY signal data, necessary for video/data synchronization, with the following fields:

Units — should be 'volts'
Scale — should be 1 always
Offset — should be 0 always
Data — vector of raw data (in volts) from the digitizer
buffer_dur — duration of data collected, in seconds
fs — sampling rate of data channels, in samples per second
numChannels — number of channels recorded simultaneously
numSamples — number of samples recorded per channel
posttriggerSamples — number of data samples recorded after the trigger was set
pretriggerPercent — pretriggerSamples divided by numSamples
pretriggerSamples — number of data samples recorded before the trigger was set
sample_vector — vector of sample indices, where 0 indicates time of trigger (not used for synchronization because READY signal is direct camera feedback)
time_vector — vector of time indices, where 0 indicates time of trigger (not used for synchronization because READY signal is direct camera feedback)
triggerType — polarity of the trigger signal; should be 'FallingEdge'

TXT\ — directory containing raw data readable with notepad. This is the only file format that can be exported from the MiDAS DA proprietary data acquisition format. Other file formats for the data signals are converted from this format.

WAV\ — directory containing data in Microsoft WAV file format, converted from TXT data format using run_MiDAS_DA_txt2wav_batch.m. Signals are normalized by their maximum absolute values to fill the dynamic range. Sampling rate equals the sampling rate of the data channel, in samples per second.
File naming:
_Accel.wav — accelerometer signal
_EGG.wav — electroglottography signal
_EGGc.wav — electroglottography signal compensated for low-frequency amplitude distortion caused by phase distortion around the cutoff frequency of the high-pass filter in the EGG transducer (Glottal Enterprises EG2-PC). See Matlab code: EGG_compensation.m, run_EGG_compensation_batch.m.
_Flow.wav — air flow volume velocity signal
_Mic.wav — microphone signal
_Pressure.wav — intraoral air pressure signal
_READY.wav — READY signal from camera
*_10000_Hz.wav — indicates that given signal has been resampled to 10 kHz for easy viewing (eg, in Wavesurfer) and playback. These files should not be used for high-bandwidth analysis. See Matlab code: decimate2fs.m, run_decimate_batch.m.

DATA raw\ — directory containing raw data signals for each trial in the native format (DAP) of the MiDAS DA data acquisition software. Proprietary format. The software allows for export to the TXT format, which is then readable using Matlab.

VIDEO converted\ — directory intended for video formats useful for playback, such as AVI, MPEG, MOV, etc.

VIDEO processed\ — directory intended for outputs from video processing algorithms

VIDEO raw\ — directory containing raw video data (CIN or CINE formats). Readable using Phantom playback software or Matlab (cineinfo.m, cineread.m)

README.txt — subject-specific data acquisition settings during each trial.

Table F.1   Minimum video frame rate given phonatory pitch (minimum 15 frames per cycle).

Pitch (Hz)	Minimum camera frame rate (frames per second)
100	1 500
150	2 250
200	3 000
250	3 750
300	4 500
350	5 250
400	6 000
450	6 750
500	7 500
550	8 250
600	9 000
650	9 750
700	10 500
750	11 250
800	12 000
850	12 750
900	13 500
950	14 250
1 000	15 000

Table F.2   Maximum data/video sampling rates given number of channels recorded.

Number of channels	Minimum sampling period per channel	Maximum sampling rate per channel (approximate)	Maximum camera frame rate (approximate)
1	0.8 μs	1 250 000 Hz	312 500 Hz
2	3.0 μs	333 333 Hz	83 333 Hz
3	2.5 μs	222 222 Hz	55 555 Hz
4	6.0 μs	166 666 Hz	41 666 Hz
5	7.5 μs	133 333 Hz	33 333 Hz
6	9.0 μs	111 111 Hz	27 777 Hz
7	10.5 μs	95 238 Hz	23 809 Hz
8	12.0 μs	83 333 Hz	20 833 Hz
9	13.5 μs	74 074 Hz	18 518 Hz
10	15.0 μs	66 667 Hz	16 666 Hz
11	16.5 μs	60 606 Hz	15 151 Hz
12	18.0 μs	55 556 Hz	13 888 Hz
13	19.5 μs	51 282 Hz	12 820 Hz
14	21.0 μs	47 619 Hz	11 904 Hz
15	22.5 μs	44 444 Hz	11 111 Hz
16	24.0 μs	41 667 Hz	10 416 Hz

All Cited References

Berry, D. A., Herzel, H., Titze, I. R., and Story, B. H. (**1996**). "Bifurcations in excised larynx experiments," J. Voice **10**(2), 129–138.

Bless, D. M., Hirano, M., and Feder, R. J. (**1987**). "Videostroboscopic evaluation of the larynx," Ear. Nose. Throat J. **66**(7), 289–296.

Boersma, P., and Weenink, D. (**2009**). *Praat: Doing phonetics by computer* (Amsterdam, The Netherlands). Ver. 5.1.07. Available at http://www.praat.org.

Bonilha, H. S., Deliyski, D. D., and Gerlach, T. T. (**2008**). "Phase asymmetries in normophonic speakers: Visual judgments and objective findings," Am. J. Speech Lang. Pathol. **17**(4), 367–376.

Cheyne, H. A. (**2002**). "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," Doctor of Philosophy thesis, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, 2002.

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.

Dormand, J. R., and Prince, P. J. (**1980**). "A family of embedded Runge-Kutta formulae," J. Comput. Appl. Math. **6**(1), 19–26.

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.

George, N. A., de Mul, F. F., Qiu, Q., Rakhorst, G., and Schutte, H. K. (**2008**). "Depth-kymography: High-speed calibrated 3D imaging of human vocal fold vibration dynamics," Phys. Med. Biol. **53**(10), 2667–2675.

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.

Hanson, H. M., and Chuang, E. S. (**1999**). "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," J. Acoust. Soc. Am. **106**(2), 1064–1077.

Hartig, G., and Zeitels, S. M. (**1998**). "Optimizing voice in conservation surgery for glottic cancer," Oper. Techn. Otolaryngol. Head Neck Surg. **9**, 214–223.

Hillman, R. E., and Mehta, D. D. (**2010: in press**). "The science of stroboscopic imaging," in *Laryngeal examination: Indirect laryngoscopy to high-speed digital imaging*, edited by K. A. Kendall, and R. J. Leonard (Thieme Medical Publishers, Inc., New York, NY).

Hillman, R. E., Montgomery, W. W., and Zeitels, S. M. (**1997**). "Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders," Curr. Opin. Otolaryngol. Head Neck Surg. **5**, 172–175.

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., Perkell, J. S., Guiod, P. C., and Goldman, S. L. (**1995**). "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," J. Speech Hear. Res. **38**(6), 1212–1223.

Iseli, M., Shue, Y.-L., and Alwan, A. (**2007**). "Age, sex, and vowel dependencies of acoustic measures related to the voice source," J. Acoust. Soc. Am. **121**(4), 2283–2295.

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.

Ishizaka, K., and Isshiki, N. (**1976**). "Computer simulation of pathological vocal-cord vibration," J. Acoust. Soc. Am. **60**(5), 1193–1198.

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., and Zhang, Y. (**2002**). "Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds," J. Acoust. Soc. Am. **112**(5 Pt 1), 2127–2133.

KayPENTAX (**2008**). *Instruction manual: Stroboscopy systems and components* (Lincoln Park, NJ).

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., and Hillman, R. E. (**2009**). "Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol," Am. J. Speech Lang. Pathol. **18**(2), 124–132.

Khosla, S., Murugappan, S., and Gutmark, E. (**2008**). "What can vortices tell us about vocal fold vibration and voice production," Curr. Opin. Otolaryngol. Head Neck Surg. **16**(3), 183–187.

Kreiman, J., Gerratt, B. R., and Ito, M. (**2007**). "When and why listeners disagree in voice quality assessment tasks," J. Acoust. Soc. Am. **122**(4), 2354–2364.

Kuo, H.-K. J. (**1998**). "Voice source modeling and analysis of speakers with vocal-fold nodules," Doctor of Philosophy thesis, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, 1998.

Larsson, H., Hertegård, S., Lindestad, P. A., and Hammarberg, B. (**2000**). "Vocal fold vibrations: High-speed imaging, kymography, and acoustic analysis: A preliminary report," Laryngoscope **110**(12), 2117–2122.

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., and Eysholdt, U. (**2008**). "Phonovibrogram visualization of entire vocal fold dynamics," Laryngoscope **118**(4), 753–758.

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.

Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U., and Döllinger, M. (**2007**). "Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos," Med. Image Anal. **11**(4), 400–413.

Lynch, R. C. (**1920**). "Intrinsic carcinoma of the larynx, with a second report of the cases operated on by suspension and dissection," Trans. Am. Laryngol. Assoc. **40**, 119–126.

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.

McGowan, R. S., and Howe, M. S. (**2006**). "Compact Green's functions extend the acoustic theory of speech production," J. Phon. **35**(2), 259–270.

Mehta, D. D., Deliyski, D. D., Zeitels, S. M., Quatieri, T. F., and Hillman, R. E. (**2010**). "Voice production mechanisms following phonosurgical treatment of early glottic cancer," Ann. Otol. Rhinol. Laryngol. **119**(1), 1–9.

Mehta, D. D., and Hillman, R. E. (**2008**). "Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods," Curr. Opin. Otolaryngol. Head Neck Surg. **16**(3), 211–215.

Moukalled, H. J., Deliyski, D. D., Schwarz, R. R., and Wang, S. (**2009**). "Segmentation of laryngeal high-speed videoendoscopy in temporal domain using paired active contours," in *Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications MAVEBA*, edited by C. Manfredi (Firenze University Press, Firenze, Italy), p. 137–140.

Murugappan, S., Khosla, S., Casper, K., Oren, L., and Gutmark, E. (**2009**). "Flow fields and acoustics in a unilateral scarred vocal fold model," Ann. Otol. Rhinol. Laryngol. **118**(1), 44–50.

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.

Niimi, S., and Miyaji, M. (**2000**). "Vocal fold vibration and voice quality," Folia Phoniatr. Logop. **52**(1–3), 32–38.

Pickup, B. A., and Thomson, S. L. (**2009**). "Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models," J. Biomech. **42**(14), 2219–2225.

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.

Rothenberg, M. (**2002**). "Correcting low-frequency phase distortion in electroglottograph waveforms," J. Voice **16**(1), 32–36.

Schuberth, S., Hoppe, U., Döllinger, M., Lohscheller, J., and Eysholdt, U. (**2002**). "High-precision measurement of the vocal fold length and vibratory amplitudes," Laryngoscope **112**(6), 1043–1049.

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., 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.

Som, M. L. (**1951**). "Hemilaryngectomy - A modified technique for cordal carcinoma with extension posteriorly," Arch. Otolaryngol. **54**(5), 524–533.

Steinecke, I., and Herzel, H. (**1995**). "Bifurcations in an asymmetric vocal-fold model," J. Acoust. Soc. Am. **97**(3), 1874–1884.

Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**(2), 1249–1260.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**(1), 537–554.

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.

Titze, I. R. (**1984**). "Parameterization of the glottal area, glottal flow, and vocal fold contact area," J. Acoust. Soc. Am. **75**(2), 570–580.

Titze, I. R. (**1995**). "Workshop on acoustic voice analysis: Summary statement," (National Center for Voice and Speech, Denver, CO), 1–36.

Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**(5), 2733–2749.

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.

Weibel, E. R. (**1963**). *Morphometry of the Human Lung* (Springer, New York).

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., Damrose, E., and Bless, D. (**2007**). "Functional analysis of voice using simultaneous high-speed imaging and acoustic recordings," J. Voice **21**(5), 604–616.

Zañartu, M., Mongeau, L., and Wodicka, G. R. (**2007**). "Influence of acoustic loading on an effective single mass model of the vocal folds," J. Acoust. Soc. Am. **121**(2), 1119–1129.

Zeitels, S. M. (**1995**). "Premalignant epithelium and microinvasive cancer of the vocal fold: The evolution of phonomicrosurgical management," Laryngoscope **105**(3 Pt 2), 1–51.

Zeitels, S. M. (**1996**). "Phonomicrosurgical treatment of early glottic cancer and carcinoma in situ," Am. J. Surg. **172**(6), 704–709.

Zeitels, S. M. (**2004**). "Optimizing voice after endoscopic partial laryngectomy," Otolaryngol. Clin. North Am. **37**(3), 627–636.

Zeitels, S. M., Blitzer, A., Hillman, R. E., and Anderson, R. R. (**2007**). "Foresight in laryngology and laryngeal surgery: A 2020 vision," Ann. Otol. Rhinol. Laryngol. Suppl. **116(suppl 198)**, 1–16.

Zeitels, S. M., Burns, J. A., Lopez-Guerra, G., Anderson, R. R., and Hillman, R. E. (**2008**). "Photoangiolytic laser treatment of early glottic cancer: A new management strategy," Ann. Otol. Rhinol. Laryngol. Suppl. **117(suppl 199)**(7 Part 2), 1–24.

Zeitels, S. M., Hillman, R. E., Franco, R. A., and Bunting, G. W. (**2002**). "Voice and treatment outcome from phonosurgical management of early glottic cancer," Ann. Otol. Rhinol. Laryngol. Suppl. **111(suppl 190)**(12 Part 2), 1–20.

Zeitels, S. M., Jarboe, J., and Franco, R. A. (**2001**). "Phonosurgical reconstruction of early glottic cancer," Laryngoscope **111**(10), 1862–1865.

Zhang, Y., Bieging, E., Tsui, H., and Jiang, J. J. (**2010**). "Efficient and effective extraction of vocal fold vibratory patterns from high-speed digital imaging," J. Voice **24**(1), 21–29.

Zhang, Y., Tao, C., and Jiang, J. J. (**2006**). "Parameter estimation of an asymmetric vocal-fold system from glottal area time series using chaos synchronization," Chaos **16**(2), 023118.