2013

Differentiation of voice disorders using objective parameters from harmonic waveform modeling in high-speed digital imaging

Aimee Elaine Perkins
Louisiana State University and Agricultural and Mechanical College

DIFFERENTIATION OF VOICE DISORDERS USING OBJECTIVE PARAMETERS FROM HARMONIC WAVEFORM MODELING IN HIGH-SPEED DIGITAL IMAGING

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Masters of Arts

in

The Department of Communication Disorders

by
Aimee Elaine Perkins
B.A., Louisiana State University, 2011
May 2013

TABLE OF CONTENTS

ABSTRACT

High-speed digital imaging (HSDI) has recently become clinically available for the direct observation of vocal fold movement in the last 20 years. However, before it can become routinely used in the clinical setting, a universal means of objectively analyzing and interpreting the HSDI data must be established.

In this study, preliminary data was gathered for five parameters used to objectively analyze vocal fold vibratory patterns observed with HSDI. The parameters investigated were established by Ikuma, Kunduk, and McWhorter (2012a) and were previously studied with a small sample (N=8) comparing pre and post-phonosurgical removal of benign lesions. The five parameters included fundamental frequency standard deviation (F0SD), harmonics-to-noise ratio (HNR) mean, open quotient (OQ) mean, speed index (SI) mean, and relative glottal gap (RGG) mean.

The current study aimed to statistically and visually analyze measurements of the five objective parameters for differences between pathology groups with different etiologies. High-speed videos (N=50) were divided into five groups based on one of the following medical diagnoses: normal voice, vocal fold nodules, polyps, true vocal fold motion impairment (TVFMI), and adductor spasmodic dysphonia (ADSD). Statistical analysis showed that HNR mean differentiated normal voices from ADSD voices and that F0 mean differentiated ADSD voices from all groups except vocal fold nodules ($p < 0.005$). Visual analysis revealed a strong trend for RGG mean to differentiate vocal fold nodules from all other groups. Less prominent visual trends for OQ mean and SI mean were also noted.

CHAPTER 1. LITERATURE REVIEW

Over the course of the last century, researchers used an array of techniques and instrumentation in the diagnosis of voice disorders. Most recently, high-speed digital imaging (HSDI) technology has become clinically available. Research involving HSDI, however, is not at a consensus about the necessary frame rates required for capturing clinically useful high-speed data. The field also lacks a universally agreed upon process and method for objectively analyzing data obtained through HSDI. Given this, more research is needed to evaluate emerging analysis techniques. The present study aimed to expand on the analysis methods presented by Ikuma, Kunduk, and McWhorter (2012a). The researcher employed several objective parameters, including harmonics-to-noise ratio (HNR), open quotient (OQ), fundamental frequency (F0), relative glottal gap (RGG), and speed index (SI) as measurements to gather preliminary data regarding the distinct vocal fold vibratory patterns among several voice pathologies using 2000 frames per second HSDI data.

This literature review will first provide the reader with a foundational overview of how voice is produced and an introduction to the voice evaluation process. Presented next is a summary of laryngeal imaging, including videostroboscopy, the current imaging gold standard, and the implications and limitations of high-speed digital imaging. Finally, disorders affecting the voice are described, with particular focus on those under current investigation and predictions of how each disorder would affect the objective parameters .

Voice Production

Humans produce speech as a result of several biological processes occurring simultaneously. At the most basic level, lungs provide air that drives the vibration of the vocal folds resulting in sound that is shaped by the vocal tract via articulators and resonant cavities. To produce voice, the vocal folds are first brought together over the airway opening. According to

the myoelastic aerodynamic theory, when air is forced out of the lungs, it forces apart the vocal folds and sets them into vibration as a result of the Bernoulli effect (Colton, Casper, & Leonard, 2011). Regular vibration relies not only on a continuous air supply but also on healthy composition and function of the vocal folds.

The vocal folds are organized into several histological layers, each with varying levels of elasticity. The outermost layer of the vocal folds is covered by stratified squamous cells forming the epithelium. Just beneath the epithelial covering of the vocal folds is the basement membrane zone, which anchors the epithelium to the vocal fold lamina propria (Colton et al., 2011). The lamina propria consists of three distinct layers: the superficial layer of the lamina propria (SLLP), intermediate layer of the lamina propria (ILLP), and the deep layer of the lamina propria (DLLP). The SLLP moves freely during phonation due to its loose, fibrous composition (Colton et al., 2011). The ILLP and DLLP, which comprise the vocal ligament, are situated beneath the SLLP and are stiffer due to increased collagen and elastin (Colton et al., 2011). Finally, at the deepest layer of the vocal folds is the thyroarytenoid muscle, which has its anterior anchors at the thyroid cartilage and is posteriorly manipulated by the arytenoid cartilages. Normal vocal fold vibration relies on this specific balance of free moving tissue around a fibrous and muscular anchor. The mechanics of this behavior were studied and described by Hirano (1974) in the body cover model of the vocal fold structure.

In Hirano's (1974) body cover model, the layers are generalized into two groups with differing mechanical properties: the body and the cover. The cover lies over the body of the vocal fold, which consists of the stiffer thyroarytenoid muscle and collagen fibers in the vocal ligament (Story & Titze, 1995). The cover consists of the epithelium and the lamina propria. The cover is pliable and flexible, allowing for fluid movement around the vocal fold body. The loose movement and shape of the cover creates a propagating wave which is responsible for vocal fold

vibration and voice quality (Colton et al., 2011). Body stiffness is also important for maintaining these oscillations, as it provides a stable mass for the cover to move around.

When structural or functional vocal fold abnormalities develop, however, the shape and tension of the cover may be modified, creating an altered or abnormal voice quality. These structural and functional abnormalities are important because a variety of voice disorders may present with the same or similar voice symptoms. For example, vocal fold nodules and polyps are both capable of causing a hoarse voice quality though they are histologically different – nodules having a thickened basement membrane zone which is unaffected in the case of polyps. Therefore, a thorough voice evaluation is necessary to identify the specific cause for these changes before determining the best management or intervention approach for a voice disorder.

Voice Evaluation Techniques

It is well known among clinical voice specialists that diagnosing a voice disorder relies on integration of qualitative and quantitative information obtained through the voice evaluation (Colton et al., 2011). There are several tools available to fulfill the qualitative portion of an evaluation. Such assessments evaluate the impact of voice auditory perception on individual's daily participation and ability to carry out regular activities (Ma & Yiu, 2001). The quantitative portion of an evaluation, however, requires a combination of acoustic measurements, aerodynamic measurements, and laryngeal imaging to draw conclusions about the structure and function of the vocal folds.

Voice recordings are collected via microphone and acoustic measurements are made using acoustics software to reveal how the voice disorder is altering phonation and creating perturbations in the voice signal. Deliyski and Petrushev (2003) suggested that though acoustic measurements have been widely implemented and shown to be clinically valuable assessment tools, there are still limitations in its application for certain voice types. Aerodynamic assessment

is an *indirect* assessment technique used to evaluate vocal fold vibratory function. It demonstrates how factors such as lung function and phonation threshold pressure are affecting sustained phonation, the vibration of the vocal folds, and ultimately voice quality. Laryngeal imaging, however, allows for *direct* observation of the vocal folds with both subjective (perceptual) and objective analyses. Thus, direct observation of the vocal folds via laryngeal imaging tools gives a more complete and accurate assessment of vocal fold structure and their vibratory function.

Laryngeal Imaging

Currently, diagnosis of a lesion can be done by using videoendoscopy to visualize the voice source at the level of the vocal folds. The vocal folds can be visualized using a halogen light source; however if the goal is to understand how a lesion is affecting vocal function, videostroboscopy is the clinical gold standard used for visualizing the laryngeal structures and vocal fold vibratory patterns (Deliyski, Petrushev, Bonilah, Gerlach, Martin-Harris, Hillman, 2008; Patel, Dailey, & Bless, 2008; Verikas, Uloza, Bacauskiene, Gelzinis, Kelertas, 2009). Though videostroboscopy has greatly enhanced our knowledge of vocal fold vibratory function, this technique has known limitations. For example, videostroboscopy allows us to see an *apparent* motion of the vocal folds by illuminating sequential phases of vibratory cycles across a sustained phonation period (Deliyski et al., 2008; Shaw & Deliyski, 2008; Verikas et al., 2009). The rate of illumination is fast enough that the human eye perceives these fragments as continuous motion. While there are observable parameters (i.e. mucosal wave, amplitude, and phase shift) used to evaluate stroboscopic videos, effective interpretation of data obtained with this clinical tool is highly subjective, relying on the knowledge and skill of the examiner (Rosen, 2005; Colton et al., 1995). This type of *subjective* perceptual analysis is highly susceptible to personal biases and intra and inter-testing reliability issues. Though Woo (1996) investigated

objective parameters to be used with videostroboscopy, such as glottal area waveform, these parameters are limited in their use for severe voice disorders. As a result, researchers have pursued other tools that provide data which can be *objectively* analyzed and compared to a standard, especially in the cases of dysphonic patients (Dollinger, Kunduk, Kaltenbacher, Vondenhoff, Ziethe, Eysholdt, & Bohr, 2012; Ikuma et al., 2012a; Inwald, Dollinger, Schuster, Eysholdt, & Bohr, 2011). For example, in order for videostroboscopy to be used on an individual, the voice must be stable and producing a periodic acoustic signal in order to synchronize the illuminations of the strobe light (Rosen, 2005; Patel et al., 2008; Shaw & Deliyski, 2008; Verikas et al., 2009). This renders videostroboscopy very limited for individuals who have severe dysphonia or any other voice disorder producing an aperiodic signal or multiple tones, such as nodules, polyps, or cysts (Verikas et al., 2009).

Most recently, HSDI has become clinically available as a laryngeal imaging technique to address the shortcomings of videostroboscopy. Because of its ability to capture detailed vibratory information and its applicability to patients of all voice disorder types, it is expected to overcome the present limitations of videostrobsocopy. The high sampling rate used in commercial high-speed systems, between 2000 frames per second to 10,000 frames per second (fps), allows for frame-by-frame analysis of vocal fold vibratory behavior (Shaw & Deliyski, 2008; Verikas et al., 2009; Hertegård, Larsson, & Wittenberg, 2003). This frame rate not only allows for capturing images within each individual vibratory cycle of a sustained phonation period but also for voice onset and offset, which cannot be achieved with videostroboscopy (Hertegård, et al., 2003; Shaw & Deliyski, 2008; Verikas et al., 2009). An illustration of vocal fold vibration and the capturing abilities of videostroboscopy and HSDI is provided in Figure 1 below.

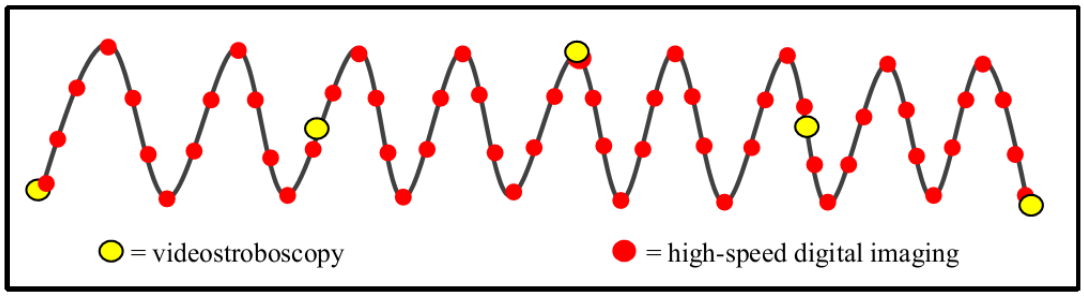

Figure 1. Illustration demonstrating vocal fold vibration in the form of a wave and the capturing abilities for videostroboscopy and high-speed digital imaging.

HSDI, therefore, overcomes stroboscopy in its ability to visualize *actual* vibration at all periods of voice production. In addition, HSDI can be used to observe all voice types because unlike videostroboscopy, its light source is constant and not dependent on the acoustic signal of the voice (Shaw & Deliyski, 2008; Hertegård et al., 2003; Verikas et al., 2009). Studies have also demonstrated that HSDI data is capable of differentiating between normal and pathological voices as well as influencing clinical diagnosis when compared to videostroboscopy (Patel, et al., 2008; Ahmed, Yan, & Bless, 2012; Verikas et al., 2009). Overall, HSDI presents multiple benefits over the current clinical tools used for direct observation of vocal fold vibratory function.

While research is growing to support the clinical implementation of this technology, there are still several significant unresolved issues prohibiting HSDI's widespread use. One concern with using HSDI is the large amount of data that must be stored and analyzed by the user (Ikuma et al., 2012b). Currently, a clinician must have adequate computer storage for the digital files created during recording, as well as adequate time to view the collected images. For example, if we record a one second long video at a rate of 10,000 fps, 10,000 images must be stored on the computer, and then each frame must be reviewed by the clinician or researcher. Because this can use a significant amount of computer storage space, it would behoove a clinician seeing multiple patients to use the lowest frame rate (yielding less frames to be reviewed) that still provides

adequate video detail. In the case of subjective analysis, it would also be extremely time saving to use a lower frame rate camera. If a clinician viewed a video captured at 2000 fps at a viewing rate of one second per frame, for example, simple calculation indicates it would take over 33 minutes to examine a video just one second long. Some researchers, however, do not support the use of this lower frame rate data. One example comes from a study conducted by Shaw and Deliyski (2008), which suggested that subjective analysis of mucosal wave is hindered when HSDIs are captured at 2000 fps. In order for the HSDI technique to become clinically feasible and useable, researchers must agree on a universal set of objective parameters and procedures for video analysis as well as an adequate capturing rate (Dollinger et al., 2012; Inwald et al., 2011; Lohscheller, Toy, Rosanowski, Eysholdt, & Dollinger, 2007; Verikas et al., 2009). The current study aimed to examine an existing set of objective parameters developed for HSDI by Ikuma et al. (2012a).

Objective Analysis in HSDI Using Harmonic Modeling

Recent HSDI research has aimed to parameterize vocal fold vibrations as a way to objectively analyze the video data. In particular, a study by Ikuma et al. (2012a, 2012b) combined the use of common parameters with a novel method for harmonic modeling. In terms of application to HSDI technology, a harmonic model is created from a laryngeal waveform of vocal fold vibratory cycles (Ikuma et al., 2012b). Ikuma and colleagues (2012b) defined a laryngeal waveform as any signal which can be extracted from HSDI data that represents vocal fold vibration, such as a glottal area waveform. This signal is represented by a sinusoidal wave, with minima, maxima, and period defined by the rate and degree of change in area of the glottis. If we calculate the fundamental frequency (F_o) of vocal fold vibration using a glottal area waveform, that F_o signal can be broken down into its component harmonic frequencies by means of a Fourier series (T. Ikuma, personal communication, July 9, 2012). A mathematical model,

which represents this decomposition, is referred to as a harmonic model. A harmonic model is a sum of selected harmonically related sinusoidal waves (i.e. their frequencies are integer multiples of the fundamental frequency) that is used to reconstruct an *ideal* waveform (T. Ikuma, personal communication, July 9, 2012).

Harmonic modeling is beneficial when examining high-speed data because it minimizes effects of naturally occurring phenomena, such as temporal aliasing, that have been shown to influence parameter measurements (Ikuma et al., 2012b). For this reason, harmonic modeling is frequently employed by researchers attempting to develop methods for HSDI analysis (Chodara, Krausert, & Jiang, 2012; Ikuma et al., 2012a, 2012b; Inwald et al., 2011). The harmonic modeling method used by Ikuma and colleagues (2012a, 2012b) is unique for its application over short analysis windows, which accommodate for natural variations in voicing over a period of time and inclusion of selected aliased harmonic frequencies. Because this technique is unique, the reader is referred to the original research article for technical aspects of harmonic model development and selection of aliased frequencies (Ikuma et al., 2012b).

In an additional study, Ikuma and colleagues (2012a) applied their method for harmonic modeling to pre and post-phonosurgical high-speed video data captured at a low frame rate (2000 fps). The study aimed to find significant changes in five objective parameters between the pre-surgery (benign lesions on the vocal folds) and post-surgery (lesion-free vocal folds) conditions as well as differences between the values when three different methods were used to collect measurements: (1) harmonic modeling with aliased harmonics, (2) harmonic modeling without aliased harmonics, and (3) a non-modeling (sample based) condition.

Eight patients participated in the study, all of whom had medical diagnoses made by the same laryngologist and who underwent phonosurgery also performed by the same laryngologist. The authors chose five commonly applied objective feature parameters in signal processing and

videostroboscopy: fundamental frequency (F0), open quotient (OQ), speed index (SI), relative glottal gap (RGG), and harmonics-to-noise ratio (HNR). Their results indicated significant differences between means and standard deviations of each feature parameter in the pre and post-surgical videos in six model-based with aliased harmonics parameter measurements ($p < 0.05$). When the model-based approach was taken without the use of aliased harmonics, five feature parameters showed significant differences between pre and post-surgery ($p < 0.05$). There was a statistically significant difference for only four sample based feature parameter measurements ($p < 0.05$). In addition, the sample-based and model-based measurements were correlated with the Voice Handicap Index (VHI), a valid patient-administered voice rating instrument. Four of the model-based parameters had significant correlations with the VHI (HNR mean, OQ mean, SI SD, RGG mean), while only three sample-based measures (HNR mean, RGG mean, RGG SD) and feature parameter measures without aliased harmonics (HNR mean, OQ mean, RGG SD) resulted in significant correlations ($p<0.05$). Results of the Wilcoxon Signed-Ranks Test performed by Ikuma and colleagues (2012a), including the p-value and effect sizes for each measurement method, are provided in Table 1 below.

Table 1. Effect sizes and significance values (p) for the feature parameters in each measurement method. Adapted from Ikuma, Kunduk, & McWhorter (2012a).

Table 1						
HSDI Feature Parameters	**Model based with aliased harmonics**		**Model based without aliased harmonics**		**Sample based**	
	Effect Size	**P**	**Effect Size**	**P**	**Effect Size**	**P**
F0 Mean	0.28	0.26	0.28	0.26	0.28	0.26
F0 SD	0.53	0.04*	0.53	0.04*	0.25	0.33
HNR Mean	0.63	0.01*	0.63	0.01*	0.21	0.40
HNR SD	0.28	0.26	0.32	0.21	0.11	0.67
OQ Mean	0.60	0.02*	0.55	0.03*	0.56	0.03*
OQ SD	0.08	0.75	0.04	0.87	0.23	0.36
SI Mean	0.11	0.67	0.05	0.83	0.00	1.00
SI SD	0.53	0.04*	0.25	0.33	0.60	0.02*
RGG Mean	0.63	0.01*	0.63	0.01*	0.63	0.01*
RGG SD	0.53	0.04*	0.53	0.04*	0.53	0.04*

Note: Highlighted cells represent greatest effect size for each parameter; (*) Indicates significance at $p<0.05$; SD = standard deviation; F0 = fundamental frequency; HNR = harmonics-to-noise ratio; OQ = open quotient; SI = speed index; RGG = relative glottal gap.

The research conducted by Ikuma and colleagues (2012a, 2012b) provides strong implications for the use of harmonic modeling with aliased harmonics and the chosen feature parameters. However, further research is necessary for the adoption of these methods as a standardized analysis technique. One limitation of their study is the small sample size used to conduct the research (N=8). The current study was performed with 50 previously recorded high-speed videos for greater statistical power. Another limitation is the study's inclusion of only two vocal pathologies, cysts and polyps, as other vocal pathologies arise from differing etiologies. Finally, their use of post-lesion removal videos to represent a "non-pathological" voice was also a limitation because though surgery comes very close to restoring the normal make-up of the vocal folds, it cannot guarantee such results in every patient.

The present study extended the use of the feature parameters with harmonic modeling to four types of vocal pathology types with differing etiologies and compared the measurements to healthy, non-pathological voices. Etiologies included were: adductor spasmodic dysphonia, vocal fold nodules, vocal fold polyps, and true vocal fold motion impairment. The goal of the present study was to increase the body of preliminary data supporting the ability of feature parameters to distinguish between vocal fold vibratory characteristics of voices experiencing common pathologies and non-pathological voices when used in combination with the harmonic modeling method developed by Ikuma and colleagues (2012a, 2012b). This preliminary data was also analyzed for feature parameters that could distinguish between vocal fold pathologies with different etiologies. An overview of voice disorders and an explanation of those featured in the present study, including expected effects on the feature parameters, are provided below.

Voice Disorders

A voice disorder is recognized by the American Speech-Language and Hearing Association (1993) as any alteration or abnormal production of an individual's vocal quality,

pitch, resonance, loudness, and/or duration that is inappropriate for their sex and/or age. In some cases, a voice disorder may be acquired from sudden or habitual vocal misuse or hyperfunction (i.e. excessive throat clearing, speaking over loud background noise, or yelling) while other individuals may live with a voice disorder from the time of infancy (Ramig & Verdolini, 1998). Though studies reporting the prevalence of voice disorders in the general population vary, it is estimated that 3-9% of people in the United States has a diagnosed voice disorder (Ramig & Verdolini, 1998).

Voice disorders have major implications for patients' social-emotional and physical well-being. For individuals whose occupation relies on voice performance (i.e. teachers, lawyers, singers, etc.), a voice disorder could lead to emotional, professional and financial difficulties. In addition, a voice loss could result in an individual missing days of work or changing their career path completely (Roy, Merrill, Gray, & Smith, 2005). For researchers and clinicians who evaluate patients with voice disorders, it is critical to provide the most efficient and accurate diagnosis and treatments for a voice disorder. Timely and accurate diagnosis, along with the appropriate treatment, does not only prevent unnecessary time missed from work but also minimizes financial and emotional distress that the voice disorder can cause for the individual. Thus, it is paramount that we have reliable tools to aid us in clinical evaluation of voice disorders and methods for measuring treatment efficacy to most efficiently restore the voice to patients with a voice disorder.

Some of the major perceptual signs related to voice problems include alterations in pitch, loudness, and quality (Colton et al., 2011). Changes in these facets of the voice can be induced by several factors. In fact, one way Verdolini, Rosen, and Branski (2006) classify voice disorders is based on the physiological changes affecting the vocal folds, including neurological, structural, or functional changes. These physiological changes include neuromuscular

dysfunction (i.e. adductor spasmodic dysphonia), benign lesions (i.e. polyps and nodules), or damage to the nerves innervating the larynx (i.e. vocal fold paralysis). The pathologies chosen for the current study represent these different etiologies, and here each pathology is further described.

Adductor spasmodic dysphonia. In adductor spasmodic dysphonia (ADSD), the vocal folds involuntarily hyperadduct, interfering with speech fluency and causing interruptions in phonation (Crevier-Buchman, Laccourreye, Papon, Nurit, & Brasnu, 1997). This irregular and uncontrollable event causes a "strained and strangled" voice quality (Silverman, Garvan, Shrivastav, & Sapienza, 2012). The etiology of this disorder is still not well understood. However, researchers believe it is the result of a neurogenic disruption, though no specific brain locus has been consistently associated with the disorder (Silverman et al., 2010). An acoustic study conducted by Sapienza, Murray and Brown (1998) reported that among 14 young women with ADSD, the most commonly occurring acoustic event was fragments of aperiodicity between 0-49ms in length. Most of these events occurred within the middle portion of sustained vowel phonations. The least common acoustic event during sustained phonation was phonatory breaks. In another study conducted by Liu, Galatsanos, and Bless (2011), HSDI was used to record and analyze the difference in vibratory patterns for patients with ADSD and those with muscle tension dysphonia. Thirty-six percent of patients with ADSD experienced oscillatory breaks lasting from 8-160ms in length, significantly more than those experienced in the control group ($p=0.009$). Three glottis patterns were identified during these oscillatory breaks, including complete TVF closure, incomplete TVF closure, and complete TVF closure with false vocal fold closure. Based on the findings presented by Liu et al. (2011), it was expected that the open quotient parameter would be most affected by this pathology group.

Nodules. Colton et al. (2011) and Dikkers & Nikkels (1999) each provide the following description of vocal fold nodules. Vocal fold nodules are similar to polyps in etiology, resulting from prolonged vocal misuse or abuse. They differ in that they are localized lesions typically occurring at the middle or anterior 2/3 portion of the vocal fold and become more fibrotic with time (De Bodt et al., 2007). Nodules may occur unilaterally or bilaterally, as is the case with chronic vocal fold nodules, and most often result in hoarseness and breathiness of the voice. Histologically, vocal fold nodules are lesions of the lamina propria (Verdolini et al., 2006). They are associated with a thickened basement membrane zone with a reduced number of anchoring fibers (Verdolini et al., 2006). Nodules typically have a noted effect on vocal fold vibration, causing aperiodic oscillations and breathiness attributed to abnormal closure (Verdolini et al., 2006). The nodules are typically located at the point of maximal vocal fold vibratory amplitude and serve as the point of vocal fold contact, preventing contact along the entire length of the vocal folds (De Bodt et al., 2007). In a study of stroboscopic sign ratings conducted by Colton et al. (1995), results revealed that 12% of patients with nodules were observed to have irregular vocal fold closure and 29% of patients presented with hourglass closure. Based on the expectations of abnormal vocal fold closure, nodules were expected to have the greatest effect on the relative glottal gap parameter.

Polyps. Vocal fold polyps are benign lesions that appear directly on the vocal folds. They are likely the result of prolonged vocal abuse but may also result from a single traumatic event (Colton et al., 2011). Polyps are predominantly unilateral lesions that vary in size, location, depth of attachment, and severity (Colton et al., 2011). The most common perceptual correlate for polyps is hoarseness, while the most commonly affected acoustic parameters are increased jitter and shimmer perturbations (Colton et al., 2011; Petrovic-Lazic, Babac, Vukovic, Kosanovic, & Ivankovic, 2011). Because polyps are highly variable in size and morphology, stroboscopic

features for this disorder group are variable (Verdolini et al., 2006). As opposed to nodules, the basement membrane zone is usually unaffected, thus causing only minor disruptions of the mucosal wave vibration (Verdolini et al., 2006). Dikkers and Nikkels (1999) have suggested that vocal fold polyps may not cause a disruption in vocal fold vibratory function based on the idea that the polyp is squeezed between the vocal folds. However, this has not been objectively investigated using HSDI. If these claims are accurate, it may not expected to observe consistently large disruptions in HSDI parameters in comparison to other voice disorders.

True vocal fold motion impairment. True vocal fold motion impairment (TVFMI) is the direct result of damage to a nerve or part of the brain that controls movement of the thyroarytenoid muscle (Colton et al., 2011). The specific location will determine which intrinsic muscles are affected and subsequently in which position the vocal fold will be paralyzed (Colton et al., 2011). The main perceptual symptoms of TVFMI are breathiness and roughness, while loudness and maximum phonation time are also commonly affected depending on location of the paralyzed vocal fold (Colton, et al., 2011). A lack of movement in the thyroarytenoid muscle often results in atrophy and a decrease in vocal fold mass. When the lungs force air up through the glottic area, a chaotic and unstable movement of all vocal fold layers can potentially disrupt vibration. This unstable movement may lead to asymmetry between the left and right vocal folds and reduced vocal fold contact. Due to this increase in asymmetry and potential lack of contact between the vocal folds, it was expected that this disorder will have a distinctly lower value from the other disorders for the speed index parameter. It was expected that the OQ mean parameter would also have an increased value in the TVFMI group compared to other disorders.

Though the above described voice disorders have been described as very different both in histology and etiology, our understanding of how voice disorders affect vocal fold vibration is still very limited. Currently, HSDI is the most powerful tool available for directly observing

vocal fold vibration. With the recent clinical availability of HSDI, researchers are moving closer to quantifying the components of vocal fold vibration based on direct observation. This would not only improve our knowledge of voice disorders but also theories of voice production, diagnosis and evaluation of voice disorders, and approaches and efficacy for treatment of individuals with voice disorders. Establishing the most informative and sensitive set of objective parameters that can distinguish vibratory features from different voice disorders is essential to accomplishing these future aims. The current study investigated a set of objective parameters proposed by Ikuma et al. (2012a) for their utility in differentiating between voice disorders with varied etiologies.

Purpose of Current Study

Because there are positive implications for using a lower frame rate in HSDI in terms of data saving time, analysis time, and cost, it is in the interests of clinicians and voice patients alike to further evaluate the utility of using 2000 fps capturing rate for HSDI. Ikuma et al. (2012a, 2012b) provided a sufficient research foundation to further investigate harmonic modeling, in conjunction with objective parameters, and lower frame rate data. This study aimed to contribute preliminary data regarding the ability of the objective HSDI parameters to distinguish between vocal fold nodules, polyps, TVFMI, ADSD, and normal voice. The HSDI data was obtained with a capturing rate of 2000fps. The hypothesis was that there would be a significant difference present in the parameter values between the normal voice (control) and the voice disorder groups. This hypothesis was driven by the preliminary results presented in Ikuma et al.'s (2012a) research, showing a significant difference between pre and post phonosurgery groups. It was predicted that the ADSD group would be signified by the OQ parameter based on data previously collected by Liu and colleagues (2011). The nodules group was predicted to have the greatest impact on the RGG parameter due to previous knowledge of vocal fold physiology

outlined by Verdolini and colleagues (2006). Based on suggestions by Dikkers and Nikkels (1999), the polyp group was expected not have a great effect on the HSV parameters. The SI parameter was expected to distinguish the TVFMI group from others in addition to the OQ mean parameter due to increased vocal fold asymmetry and reduced vocal fold contact.

CHAPTER 2. METHODS

Experimental Design

This study employed a between subjects group design using five levels of one independent variable, type of diagnosis, and five dependent variables: HNR mean, OQ, F0, RGG, and SI. The independent variable was type of *voice diagnosis*, which included no disorder, nodules, polyps, true vocal fold motion impairment (TVFMI), and adductor spasmodic dysphonia (ADSD). The data used in the present study was collected at the Voice Center located at Our Lady of the Lake Regional Medical Center. The use of this data has been approved for investigation of voice disorders by the LSU Health Science Center and Our Lady of the Lake Regional Medical Center Internal Review Boards.

Participants

Forty patients seen at the Voice Center at Our Lady of the Lake Regional Medical Center served as the participants in this study. Forty individual high-speed videos from patients seen at the Voice Center were selected from the laryngeal imaging lab database for analysis. Each of the 40 videos was placed into one of four groups: nodules, polyps, TVFMI, and ADSD. Their HSDI recordings were obtained during their initial clinic visit for the evaluation of their voice difficulties. All videos were collected by the same speech-language pathologist, and participants underwent evaluation and diagnosis made by the same laryngologist. Videos from these participants were stored on the laryngeal imaging lab database managed by Louisiana State University's Voice Clinic. In addition, previously recorded HSDI from ten participants with no voice complaints or diagnoses formed the no voice disorder control group data.

Data Collection

The high-speed videos used in the study were collected using a 70° rigid laryngoscope made by KayPENTAX, Model 9106, and a 300-watt cold light source (CLV-U20, Olympus

America Inc., Center Valley, PA). The laryngoscope was paired with a KayPENTAX Model 9700 HSV system which recorded the videos at a rate of 2,000 frames per second (fps). Each video had a spatial resolution of 120 pixels wide by 256 pixels high. The data files were digitally stored on the database as uncompressed 8 bit grayscale videos. For the recording, each patient was instructed to vocalize /i/ in order to visualize structures in the laryngeal cavity.

Video Selection

The videos included in this study were subjectively judged by two researchers with previous experience viewing high-speed videos to meet a set of criteria established to ensure video quality consistency and to eliminate extraneous variables, such as camera motion, that could affect the outcome of the data analysis.

The first criterion for video selection was based on patient voice diagnosis. A total of five data groups were examined, with one group having no pathology (normal voice) and four groups with a voice disorder diagnosis as determined by the same laryngologist. Each group investigated represented a different voice type category with respect to physiological changes in the vocal folds. The physiological differences between these groups were expected to produce different effects on vocal fold vibratory characteristics.

The second criteria for video inclusion in this study were factors of video quality. The researcher subjectively judged each video based on clarity of the image and darkness of the image. Image clarity ensured accurate identification of the glottal edge, which was important for video segmentation and analysis. Video darkness was also an important factor in identifying the glottal edge and will be discussed in further detail in the segmentation portion of this paper.

Videos from the database were viewed in the order they appear in the database until a total of 10 were selected for each of the pathology and normal groups. Because this study was

based on previously collected data, the individuals chosen for each group were not controlled for age, race, gender, or other assigned variables.

Once the videos were chosen, they were edited to a one second segment consisting of 2000 frames for analysis. Due to laryngeal waveform unsteadiness during voice onset and offset periods, segments were chosen so that they consist of only a sustained phonation period. The segments were ultimately chosen if they captured the entire length of the vocal folds, including the anterior and posterior vocal fold commissures. By visualizing the anterior and posterior commissures of the vocal folds, with minimal to no interference of the arytenoids at the posterior, we could consistently analyze the maximum glottal area across all videos. Because the data analysis relied on a pixel count, this eliminated the contribution of lost pixels in the data variance.

Video Analysis

Video analysis for this study began with a process called glottal area segmentation. Glottal area segmentation was accomplished using the computer segmentation software developed through MATLAB 7.13 by Takeshi Ikuma. Segmentation began with the researcher first selecting a single "key frame" from a high speed video. In that frame, the user must determine a pixel in the image that falls within the glottis. Because this area is black and distinct from the vocal folds, the glottis is easily identifiable. When a pixel is selected, the program assigns a threshold value based on grayscale intensity values, where 0 represents the color black and 256 represents white. The value for the point was raised or lowered by the researcher until all of the dark pixels in the entire glottis are selected by the program. The number of pixels selected defined what we call the glottal area. By applying this threshold value to every frame in the video, the program creates a glottal area waveform. The glottal area waveform was defined as a sequence of the glottal areas over time, and the waveform can be used to determine changes

and features within each vibratory cycle of a sustained phonation period. A sample of the high-speed images and glottal area waveform in the segmentation program is provided in Figure 2 below.

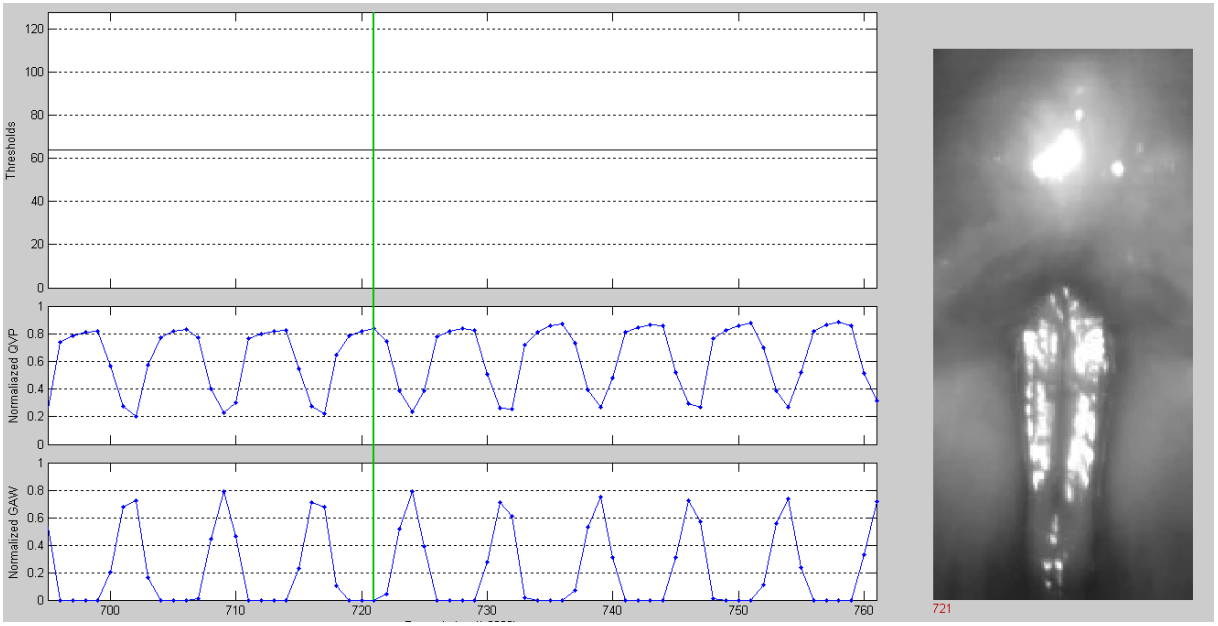

Figure 2. Segmentation program developed by Takeshi Ikuma. The number of pixels in the glottal opening are calculated for each frame. The changing number across each cycle can be plotted as a wave, known as a glottal area waveform (left), and is applied across all frames in the video (Ikuma, Kunduk, & McWhorter, 2012b).

Harmonic Modeling

The next aspect of data analysis is based on a harmonic model derived from the glottal area waveform. As a review, F_o is obtained via the glottal area waveform, and the resulting signal was broken down into its component harmonic frequencies by means of a Fourier series. The frequencies were combined to form a harmonic model. By combining an adequate number of harmonics, the model can come very close to matching the actual signal produced by a normal sustained voice, which is periodic over a short period of time. Formation of the harmonic model followed the guidelines used in Ikuma and colleague's (2012b) research article, which provides a detailed description of how the frequencies and aliased harmonics are chosen.

Data Analysis

In the present study, the F0 standard deviation and means for OQ, SI, RGG, and HNR, were calculated across all analysis windows for each high-speed video. This procedure was carried out by Takeshi Ikuma via the parameter analysis software which he programmed using MATLAB 7.13 (MathWorks, 2012). The OQ, SI, RGG, and HNR means and F0 standard deviation for each disorder type were examined for a statistically significant difference between the groups using a MANOVA procedure. Descriptive statistics and visual inspection of data variance via box plots were also used in analyzing each parameter's mean across all pathologies.

Reliability

In order to determine intra-judge reliability of the segmentation process, 10% of the videos included for analysis were randomly selected for re-segmentation. A Pearson correlation showed that each parameter measurement for re-segmented videos was significantly correlated at the 0.01 level to those used for the study. The correlation coefficients for each parameter are provided in Table 2 below.

Table 2. Correlation coefficients for the objective parameters fundamental frequency standard deviation (F0SD), harmonics-to-noise ratio (HNR) mean, open quotient (OQ) mean, speed index (SI) mean, and relative glottal gap (RGG) mean.

Parameter	Correlation Coefficient
F0SD	1.000
HNR Mean	0.829
OQ Mean	1.000
SI Mean	0.999
RGG Mean	0.995

CHAPTER 3. RESULTS

Due to the study's small sample size, only F0 standard deviation and mean values for SI, RGG, HNR, and OQ were statistically analyzed for between group differences in order to preserve Type I error protection in MANOVA. F0 standard deviation was chosen instead of mean because the mean value will differ greatly between male and female subjects. Therefore, standard deviation allows us to study how well the subjects in each group were able to maintain a consistent F0.

Multivariate MANOVA

The multivariate MANOVA was chosen to test the differences between the disorder groups across each dependent variable simultaneously. Using Pillai's trace, a significant difference was present between all voice groups, $V = 1.34$, $F(20, 176) = 4.43$, $p < .005$. A Bonferroni correction was applied before performing separate univariate ANOVAs to further protect against Type I error, making significance at the level of 0.005. Separate univariate ANOVAs revealed significant differences between disorder groups present for the HNR mean and F0 standard deviation parameters. However, the univariate ANOVAs did not reveal a significant difference between disorder groups for the SI mean, OQ mean, and RGG mean parameters.

Post-hoc Analysis

A Bonferroni post-hoc analysis was performed to investigate pairwise differences between the disorder groups across the significant parameters HNR mean and F0 standard deviation. The HNR mean parameter showed significance between the ADSD and normal groups only ($p < 0.00$). For the F0 standard deviation parameter, the ADSD group was significantly different than the normal group and all other disorders ($p < .005$) except the nodule group.

Results of the Bonferroni post-hoc analysis and the observed power for each parameter in the univariate ANOVAs are presented in Table 3 below.

Table 3. Significance for pairwise comparisons between disorder groups adductor spasmodic dysphonia (ADSD), normal, nodule, polyp, and true vocal fold motion impairment (TVFMI), for each parameter. Between subjects significance levels (p), observed power, and partial eta squared are reported for each parameter.

Comparison	Parameter	Significance (p)	Power	Partial Eta Squared
	F0SD	0.000*	0.998	0.436
ADSD v. Normal		0.000*		
ADSD v. Nodule		0.008		
ADSD v. Polyp		0.001*		
ADSD v. TVFMI		0.000*		
	HNR mean	0.000*	0.995	0.406
ADSD v. Normal		0.000*		
ADSD v. Nodule		0.008		
ADSD v. TVFMI		0.006		
	OQ mean	0.014	0.823	0.237
ADSD v. Nodule		0.023		
	SI mean	0.016	0.826	0.234
Normal v. TVFMI		0.009		
	RGG mean	0.005	0.900	0.236
Nodule v. Normal		0.008		
Nodule v. Polyp		0.036		
Nodule v. TVFMI		0.014		

Note: (*) indicates significance at ($p < 0.005$); HNR = harmonics-to-noise ratio; F0 = fundamental frequency; OQ = open quotient; SI = speed index; RGG = relative glottal gap.

Descriptive Statistics

SI mean. Figure 3 is a standard box plot representing the objective measurement SI mean across all four voice disorder groups and the normal voice group. Whisker ends represent minimum and maximum data values, whiskers represent the first and third quartiles, boxes represent the middle 50% of the data, and the line inside the box represents the median of the data. Based on visual analysis, the normal group had the highest median value of 0.18 with a data range from -0.02 to 0.33. The median value for the TVFMI group was the lowest at -0.15 with a range from -0.24 to 0.25. The measurements in the polyp and nodule groups appear to have a near-normal distribution with their medians at 0.15 and 0.13, respectively, and a data range from approximately 0.00 to 0.3. An outlier is present for the polyp group at approximately -0.18. The

median for the ADSD group is 0.11 with a data range from 0.00 to 0.13. Descriptive statistics for the each disorder group is provided in Table 4.

Table 4. Mean, median, mode and standard deviation (SD) for speed index (SI) mean in each disorder and normal group.

SI Mean			
Voice Group	**Mean**	**Median**	**SD**
ADSD	0.087	0.11	0.21
Normal	0.19	0.18	0.07
Nodule	0.13	0.13	0.01
Polyp	0.12	0.15	0.09
TVFMI	-0.02	-0.15	0.05

Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment.

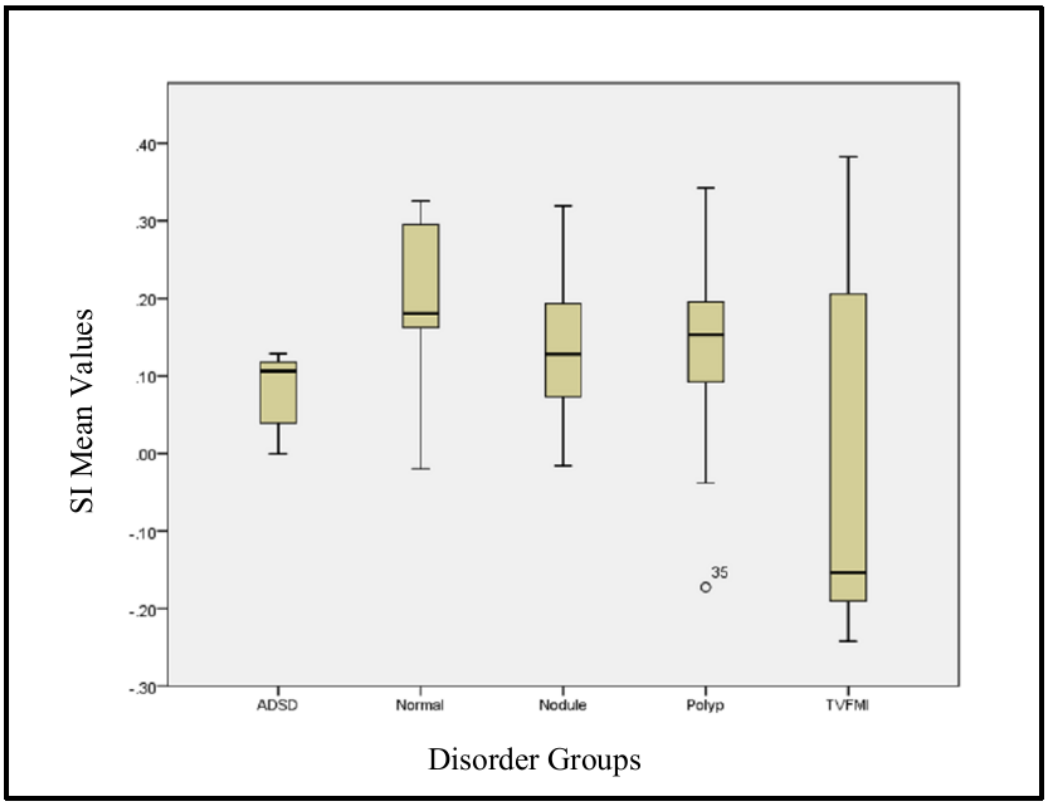

Figure 3. Box plot comparing speed index (SI) mean parameter values across the disorder and normal groups. Possible value range is (-1, 1).
Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment.

RGG mean. Figure 4 is a box plot representing measurements of RGG mean across all voice disorder groups and the normal voice group. Whisker ends represent minimum and maximum data values, whiskers represent the first and third quartiles, boxes represent the middle

50% of the data, and the line inside the box represents the median of the data. The nodule group had the highest RGG median and range values, with the median at 0.30, minimum value of 0.11, and maximum value of 0.42. The distribution of the middle 50% of the data is approximately the same in the ADSD, normal groups, and polyp groups, and data for all groups appear negatively skewed. The median value for the ADSD group is 0.13, and the minimum value in the group is 0.00 and an outlier is present at 0.57. The median value for the normal voice group is 0.11 with minimum and maximum values at 0.00 and 0.30, respectively. The polyp group had a median RGG value at 0.13, and values within the group ranged from 0.00 to 0.47. The value range for the TVFMI group was approximately the same as in the polyp group, and the median RGG value for TVFMI was 0.09. Table 5 lists descriptive statistics for RGG values measured in each disorder group.

Table 5. Mean, median, mode and standard deviation (SD) for relative glottal gap (RGG) mean in each disorder and normal group.

RGG Mean			
Voice Group	**Mean**	**Median**	**SD**
ADSD	0.16	0.13	7.62
Normal	0.11	0.11	0.70
Nodule	0.12	0.30	3.46
Polyp	0.13	0.13	1.76
TVFMI	0.12	0.09	1.28

Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment

OQ mean. Figure 5 is a box plot representing the measurements for OQ mean across all voice disorder groups and the normal voice group. Whisker ends represent minimum and maximum data values, whiskers represent the first and third quartiles, boxes represent the middle 50% of the data, and the line inside the box represents the median of the data. Most of the OQ mean values for participants across the disorder and normal groups were within a similar range. The median values for the nodule, polyp, and TVFMI groups are between 0.92 and 0.93 with their uppermost values ranging from 0.94 to 0.95. The median value for the ADSD group was 8.6, and the normal voice group median was 8.8. The ADSD group had the largest range of

values, from 0.34 to 0.94, however a majority of the group had OQ mean values around the median and upper range. Table 6 provides additional descriptive statistics for the OQ mean measurements in all experimental groups.

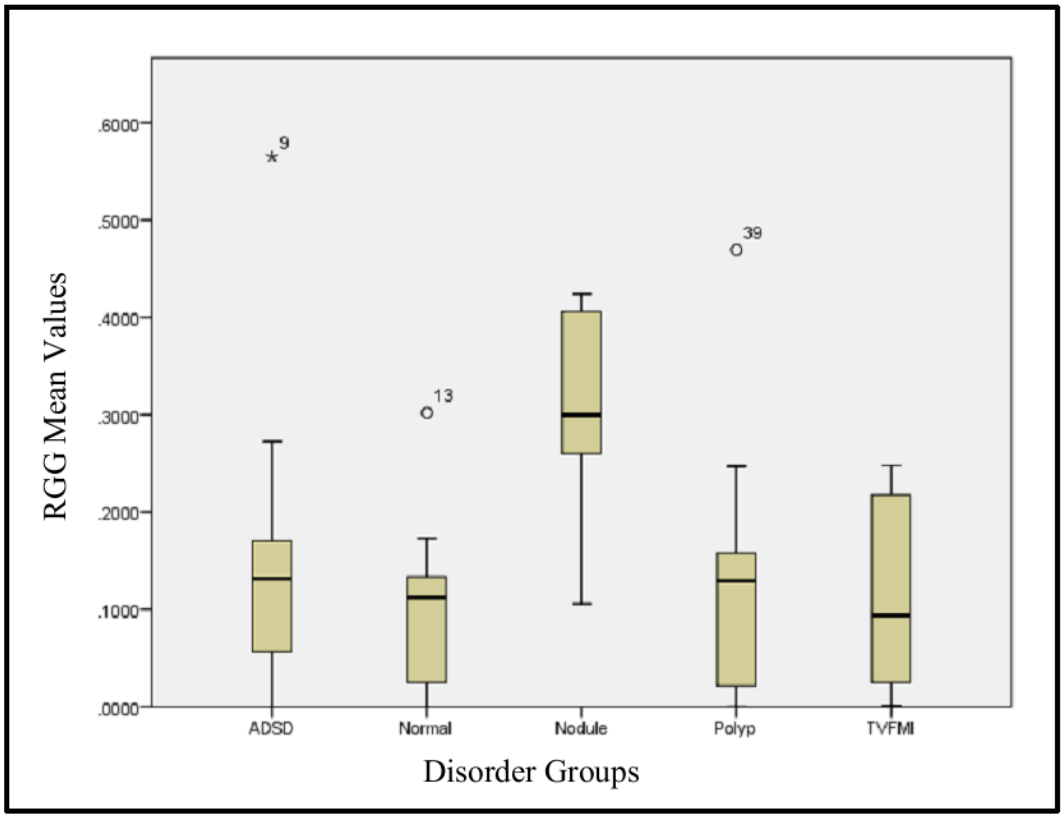

Figure 4. Box plot comparing relative glottal gap (RGG) mean parameter values across the disorder and normal groups. Possible value range is (0, 1).
Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment.

HNR mean. Figure 6 is a box plot representing the values for HNR mean across the disorder groups and the normal group. Whisker ends represent minimum and maximum data values, whiskers represent the first and third quartiles, boxes represent the middle 50% of the data, and the line inside the box represents the median of the data. The group with the lowest median value for HNR mean was the ADSD group (17.75dB), and the normal group had the highest median (30.81) and maximum (33.25dB) values for HNR mean. The distribution for the nodules group was the least spread with the exception of an outlier at 30.38dB. Based on visual

analysis, the polyp group has a nearly normal distribution around the median value at 24.72dB. The TVFMI group had a median HNR mean value of 26.66dB with a range from 16.26dB to 31.76dB. Table 7 provides additional descriptive statistics for the HNR mean measurements in the disorder and normal groups.

Table 6. Mean, median, mode and standard deviation (SD) for open quotient (OQ) mean in each disorder and normal group.

OQ Mean			
Voice Group	**Mean**	**Median**	**SD**
ADSD	0.78	0.88	9.59
Normal	0.83	0.86	2.23
Nodule	0.93	0.93	1.90
Polyp	0.88	0.92	5.52
TVFMI	0.91	0.93	4.36

Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment.

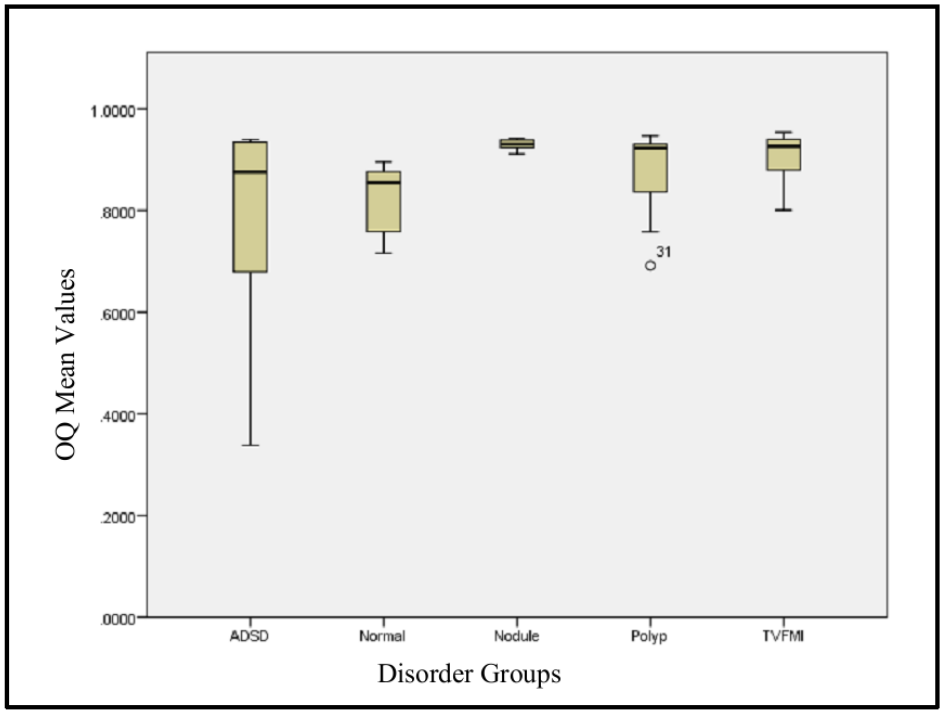

Figure 5. Box plot comparing open quotient (OQ) mean parameter values across the disorder and normal groups. Possible value range is (0, 1).

Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment.

Table 7. Mean, median, mode and standard deviation (SD) for harmonics-to-noise ratio (HNR) mean in each disorder group.

HNR Mean			
Voice Group	**Mean**	**Median**	**SD**
ADSD	17.05	17.75	7.62
Normal	30.22	30.81	0.70
Nodule	25.88	25.79	3.46
Polyp	24.12	24.72	1.76
TVFMI	26.03	26.66	1.28

Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment.

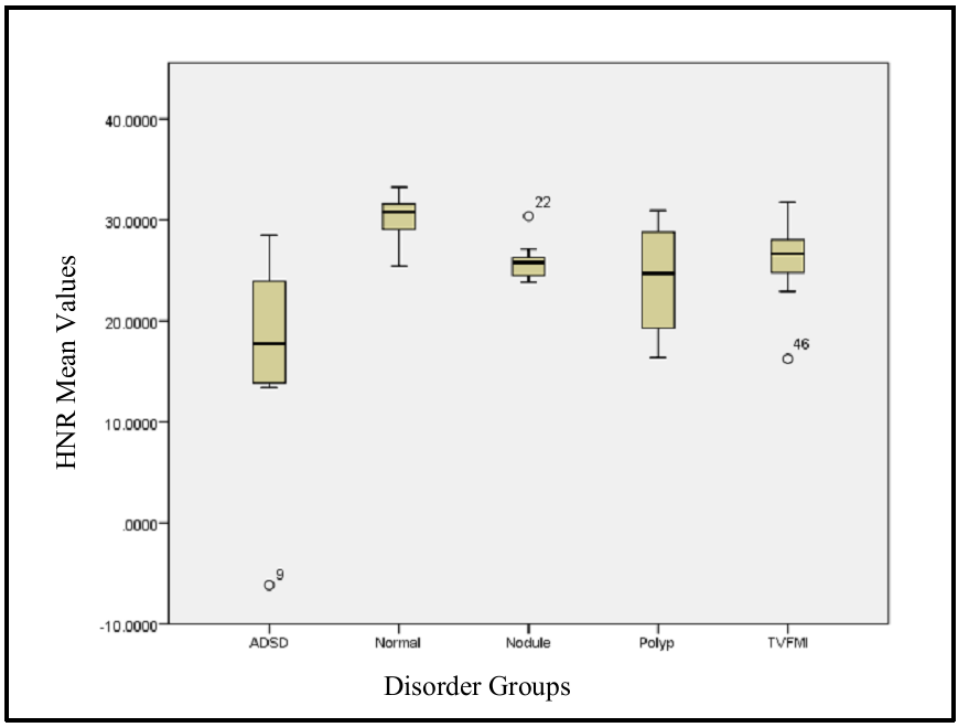

Figure 6. Box plot comparing harmonics-to-noise ratio (HNR) mean parameter values in decibels across the disorder and normal groups.
Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment.

F0 standard deviation. Figure 7 is a box plot representing the values for F0 standard deviation across the disorder groups and normal group. Whisker ends represent minimum and maximum data values, whiskers represent the first and third quartiles, boxes represent the middle 50% of the data, and the line inside the box represents the median of the data. The normal group had the smallest data distribution (min = 0.71Hz, max = 4.58Hz) and the lowest median value (2.15Hz) for F0 standard deviation out of all experimental groups. The ADSD group had the highest median value at 11.12Hz as well as the largest distribution of measurements. The median

values for the nodule, polyp, and TVFMI group were closely related and range from 1.95Hz to 4.14Hz. The nodule group had the second largest distribution ranging from 1.09Hz to 11.53Hz. Table 5 provides additional descriptive statistics for F0 standard deviation measurements across the disorder and normal groups. Table 8 summarizes additional descriptive statistics for F0 standard deviation measures for the disorder and normal groups.

Table 8. Mean, median, mode and standard deviation (SD) for fundamental frequency standard deviation (F0 SD) in each disorder and normal group.

F0 SD			
Voice Group	**Mean**	**Median**	**SD**
ADSD	10.37	11.12	7.62
Normal	1.24	1.07	0.70
Nodule	4.14	4.14	3.46
Polyp	2.76	1.95	1.76
TVFMI	2.33	2.15	1.28

Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment.

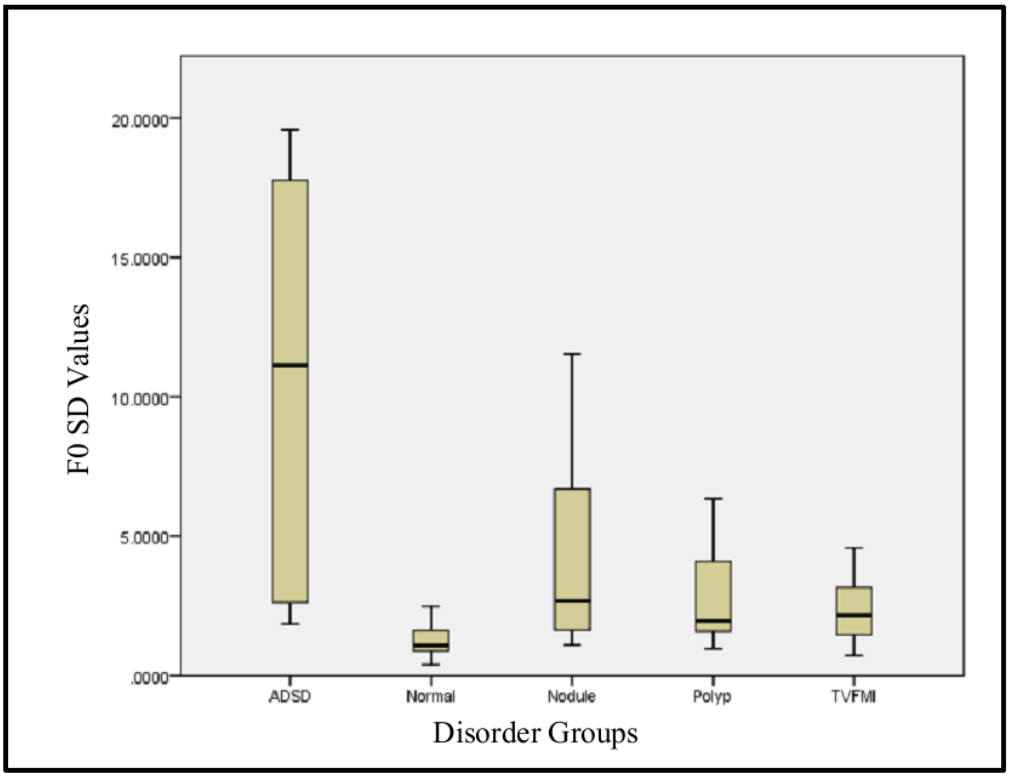

Figure 7. Box plot comparing fundamental frequency standard deviation (F0 SD) parameter values in Hertz across the disorder and normal groups. Possible value range is (0, 1000).
Note: ADSD = adductor spasmodic dysphonia, TVFMI = true vocal fold motion impairment.

CHAPTER 4. DISCUSSION

The purpose of this study was to gather preliminary data regarding the ability of objective measurements in HSDI obtained from harmonic waveform modeling to differentiate between voice disorders with varying etiologies. The current study aimed to gather preliminary data for whether the parameters established by Ikuma et al. (2012a) could distinguish between voice disorders with different etiologies and normal voices. The preliminary findings demonstrated that the HNR mean parameter distinguished between normal and ADSD voices and the F0 SD measure differentiated ADSD from all voice groups except nodules ($p<0.005$). Comparisons between disorder groups for the RGG mean and OQ mean parameters were not significant at present, however the predicted trends in the data were present.

There are studies emerging in the literature attempting to quantify vocal fold vibration in healthy and pathological voices using HSDI (Chodara, Krausert, & Jiang, 2012; Ikuma et al., 2012a, 2012b; Inwald et al., 2011; Krausert, Liang, Zhang, Rieves, Geurink, & Jiang, 2012; Patel, Dixon, Richmond, & Donohue, 2012). Among these, a variety of techniques have been suggested for quantifying vibratory characteristics, yet each still requires further research before they can be regularly used among researchers and clinicians.

Inwald et al. (2011) used a multiparametric approach (i.e. subjective ratings and objective measurements) to classify normal and disordered voices observed with HSDI. Inwald et al.'s (2011) retrospective study included 496 patients with healthy voices and diagnoses of functional dysphonia, bilateral, and unilateral vocal fold nerve paralysis. Subjective ratings were made by a trained and experienced Medical Doctor in order to measure validity for each of the objective measurements. The objective parameters used in Inwald et al.'s study (2011) were perturbation measures used in acoustic analysis and included two that were comparable to the objective parameters used in the present study: jitter area and HNR area. Differing from the present study,

the parameter values for jitter area and HNR area were calculated based on overall sum of segmented glottis pixels, rather than a harmonic model derived from the glottal area waveform. The results of their study showed that jitter area consistently differentiated normal from pathological voices in women and men ($p < 0.001$). This was similar to our findings for the F0SD parameter, in which the ADSD group was significantly different from the normal group. The authors reported that HNR area was able to mostly differentiate between the normal and disorder groups. It was noted, however, that HNR area could not be calculated for eight videos which accounted for 1.7% of the dataset. Similarly, HNR mean was a powerful parameter for differentiating between disorder groups, specifically in the ADSD group.

A second study by Patel and colleagues (2012) aimed to characterize glottic and phase closure patterns in children with normal voices. The study compared HSDI data for 75 children with data from 56 adults for differences in closure pattern. Three trained raters with experience in treating voice disorders assigned a glottal and phase closure label to each video. Though their study was based on subjective ratings of closure patterns, findings from Patel and colleagues (2012) are relevant to the RGG parameter in the present study. While the RGG mean parameter in the present study visually differentiated nodules from all other groups, Patel and colleagues (2012) found that in women, only 21.4% achieved complete glottal closure, 75% had a posterior glottal gap, and 3.6% had an hourglass closure pattern. For men, complete glottal closure was achieved for 39.3%, 53.6% had posterior gap, and 3.6% had hourglass closure. Patel et al.'s study suggested that gender may play an important role in glottal closure patterns in adults. Therefore, gender may have been a contributing to the lack of significance for the RGG parameter in the present study, as the researcher was blind to gender when selecting videos for the sample.

The parameters in the present study were obtained from endoscopic HSDI data and the method used to obtain these objective parameter measurements is unique. Therefore, direct comparisons of the findings for these parameters to those reported in the literature are limited.

F0 Standard Deviation

Fundamental frequency is the acoustic correlate of perceived pitch and measures the number of glottal cycles occurring in one second (Colton et al., 2011). In the current study, fundamental frequency standard deviation (F0SD) was measured to capture the ability of individuals in each disorder group to maintain constant vocal fold vibration during sustained phonation. The highest values measured for both the mean and median F0SD were in the ADSD group. This was an expected trend for the ADSD group given that involuntary, expected vocal fold hyper adduction, which characterizes this disorder, causes the vocal folds to stop and resume vibration. The normal group, as expected, had the lowest median and mean values as well as the lowest group variance. This is likely because there are no functional or structural alterations to the vocal folds, thus no induced changes of vocal fold vibration. The slightly elevated mean value of F0SD in the nodule and polyp groups suggest that alterations to the vocal fold tissue have some impact on the ability of individuals to maintain a consistent rate of vocal fold vibration. Though this finding cannot compare directly with the study of the objective parameters by Ikuma et al. (2012a), their study demonstrated that F0SD was also a significant parameter in distinguishing normal (post-surgery) from tissue-related pathologies (pre-surgery polyps and cysts). Because F0SD appears to be sensitive in measuring neurological and tissue-related pathologies, further investigation of this parameter could eventually reveal F0SD as an important measure of therapy progress when used with HSDI.

HNR Mean

Over the last few decades, HNR has been used in acoustic studies of human voices as an objective measure of hoarseness (Yumoto, Gould, & Baer, 1982). HNR is defined as the ratio of the harmonic content in a signal to the noise content in a signal. Most recently, researchers have applied this parameter in HSDI studies and suggested that it is capable of differentiating normal from pathological voices (Ikuma et al., 2012a; Inwald et al., 2011). For the present study, a mean HNR measurement was extracted from a harmonic waveform model to observe its potential to isolate different types of voice disorders. HNR mean was expected to be the most different in the polyp or nodule group because these disorders are often characterized by hoarseness. Results of the current study showed that HNR mean was able to best differentiate the normal voice group from the ADSD disorder group only.

Based on the perceptual qualities of ADSD, we know that patients with ADSD diagnosis tend to have a strained or strangled voice quality. This could suggest that patients are bringing the vocal folds together more tightly, causing them to produce a forced voice. A low HNR mean value is likely attributed to this increased tightness or spastic closure and irregularity of the vocal fold vibration induced by these features. These suspected changes in vocal fold adduction are also possibly related to increased energy in high frequencies that in turn comprise the noise component in the acoustic and waveform signals (Colton et al., 2011). While the significance in the ADSD group was not expected, this result is of high interest. Several authors have documented the difficulty of differentially diagnosing ADSD because it relies mostly on auditory perceptual features and patient report of onset and course of the problem (Patel et al., 2011; Colton et al., 2011; Tanner, 2012). These preliminary findings demonstrate that HNR mean is able to significantly differentiate ADSD from normal voices. A study including more participants is warranted to determine if this parameter could differentiate ADSD from other

disorder groups. Particular interest lies in the utility of HNR mean in differentiating ADSD from muscle tension dysphonia, as they can mimic each other perceptually, making differential diagnosis difficult.

Polyps and nodules, respectively, had the second and third lowest HNR mean values. Though this did not meet the prediction, it is possible that the size and/or age of these lesions may have a greater effect on the parameters than anticipated. Because these diagnoses were chosen from a database with subjective video quality-related criteria, which did not include disorder-specific information like severity of voice, or size/age of a lesion, individual differences within a disorder group that could be affecting parameter measurements were not controlled. It is also possible that though these pathologies create abnormal vibratory patterns, the vibratory cycles may be periodic. Thus, the HNR parameter may not be affected by this abnormal yet periodic vibration. Further studies are needed to differentiate which of these assumptions are contributing factors of HNR values in these vocal pathologies.

OQ Mean

The OQ mean parameter was employed in the study as a measure of duration that the vocal folds are open and in motion (Ikuma et al., 2012a). It was predicted that the TVFMI group would have the highest value, indicating that the vocal folds are open and in motion longer when compared to other disorder groups, and that the ADSD group would have the lowest value compared to other groups. While there were no significant findings in our statistical analysis, visual analysis of the data supported this expectation (Figure 5, p. 28). This was anticipated because for individuals with TVFMI, the affected vocal fold is unable to move toward midline and make contact with the normal vocal fold to achieve glottic closure (Colton et al., 2011). Thus, where vocal fold motion is briefly halted due to glottic closure upon vocal fold contact in normal voices, the vocal folds may appear constantly in motion or make no complete contact

during phonation for individuals with TVFMI. The ADSD group was expected to have lower values because of the involuntary hyperadductions that characterize the disorder, however, the middle 50% of the data for the ADSD group did not appear to be distinctly different than that of the normal group (Figure 5, page 28).

The results for the TVFMI group were also similar for the vocal fold nodules group. This finding is expected since the nodules project and serve as the point of vocal fold contact, preventing the entire length of the vocal folds from achieving full closure (De Bodt, 2007). Visual analysis of the data suggested that the TVFMI and nodule pathologies, though with different etiologies, create very similar objective measurements for vocal fold motion duration and contact. This might be due to subjects increased effort to adduct vocal folds during phonation to compensate for the lesions resulting in similar vocal fold vibratory patterns. Because these disorders are visually very different when examined with HSDI, the similarity between the two group measurements does not minimize the usefulness of OQ mean as an objective HSDI parameter. For example, the OQ mean parameter could be a beneficial tool for tracking therapy progress rather than assisting in differential diagnosis. Additional studies of OQ mean using larger sample sizes may show significant differences between patients with normal voices, TVFMI, and nodules.

RGG Mean

RGG mean is an objective parameter that compares the minimum and maximum glottal areas during vibration and indicates the degree to which the vocal folds achieve complete or an incomplete closure (Ikuma et al., 2012a). It was anticipated that the nodule group would have the highest measurement for the RGG mean parameter given that the disorder is characterized by irregular incomplete closure patterns. The descriptive statistics demonstrated that as a group, the nodule diagnosis had the highest mean and median measures of RGG mean. Though there were

no statistically significant differences in the mean pairwise comparisons, visual analysis suggested that the nodule group is distinctly different than the other disorder groups (Figure 3, p.26). It is again projected that further investigation of this parameter with larger sample sizes would verify this visual trend. Based on these preliminary findings, the RGG mean parameter could be useful, with further research and establishment of normative data, for objectively measuring therapy effects among individuals with nodules.

SI Mean

Speed index is one of many measurements used to describe glottal area function, which is essentially a graph of glottal opening across time (Baken & Orlikoff, 2000). It has been used in the research as a means of representing opening and closing phase symmetry (Baken & Orlikoff, 2000). It is traditionally calculated as the difference between the opening phase time and the closing phase time relative to the total time the glottis is open (Baken & Orlikoff, 2000). In the present study, the SI calculation is slightly altered. The initiation of the closing phase was slightly adjusted to begin when the glottal area reached < 1% of the maximum glottal opening area, as opposed to the traditional definition, in which the closing phase begins when the glottal area reaches zero (Ikuma et al., 2012a). This adjustment accounts for normally occurring posterior glottal chink, where the vocal folds come into full contact except at the posterior region (Ikuma et al., 2012a). Based on previous studies of glottal symmetry, our expectations were that the TVFMI group would have the lowest SI mean value, indicating that the opening phase for this group is shorter relative to the closing phase than in other voice disorders (Baken & Orlikoff, 2000).

Though our results were not significant at the $p < 0.005$ level, the overall trend across the disorder groups reflected the prediction. The TVFMI group was the only one from the disorder groups with a negative mean and median value, which was in concert with the hypothesis.

Because the affected vocal fold is not able to move toward midline, the functioning vocal fold can move inward a greater distance – in some cases past midline – because the movement is not disrupted by contact with the opposite vocal fold. Thus, the unaffected fold spends more time in motion for the closing phase (Baken & Orlikoff, 2000). Ikuma et al. (2012a) did not find significance for SI mean when comparing pre and post-surgical removal of benign lesions. However findings in the current study, though not significant at the present, may demonstrate the usefulness of SI mean as an objective HSDI parameter for non-tissue related pathologies. While further research is needed to confirm the trend based on visual analysis, it is suspected that additional research could establish the worth of SI mean as a measure for treatment success in patients with TVFMI.

Limitations

Future investigation of the presented objective HSDI parameters is warranted based on the preliminary findings in the current study. However, the present study has some limitations that should be addressed in successive research. Firstly, the sample size (N=50) was minimally large enough to test the hypotheses. Due to the limited video source, it was not possible to tightly control for gender or age effects in the data, thus limiting the generalization of the findings and potentially confound the results. The present study also investigated the disorders based solely on medical diagnosis, thus it was not able to account for individual differences within a disorder group (i.e. size, location, severity, or age of benign lesions). If these individual differences exist within a disorder group, it is possible that they could affect the objective parameter measures.

Future Research

The current study provides useful preliminary data for further investigation of the feature parameters. Significant correlations of the parameters and the VHI presented in Ikuma et al.'s study (2012a) also support the future investigation of the five proposed parameters. Future

research of the objective HSDI parameters used in this study may take several directions. Firstly, future research should seek to replicate our present results with a larger sample size (i.e. N = 150) to verify our significant findings and strong visual trends. It is expected that HNR mean would again be significantly different between ADSD and normal voices and possibly differentiate ADSD from other disorder groups. A larger sample size may also show significance in RGG mean's ability to differentiate nodules from all other voice disorders. Future studies of SI mean could potentially demonstrate a significantly different measurement in the TVFMI group. Finally, it is expected that a larger replication study may show that OQ mean could differentiate the vocal fold nodules and TVFMI groups from the normal and other pathologies.

Because the present study demonstrated significance of the HNR mean and F0 SD parameters in the ADSD group, these particular parameters should be investigated for their ability to differentiate ADSD from other dystonias, such as abductor spasmodic dystonia or dystonia of mixed types. Identifying an objective parameter that could differentiate ADSD and MTD, for example, could greatly improve the timeliness and reliability of differential diagnosis. Because the study conducted by Ikuma et al. (2012) demonstrated that these parameters could also distinguish between tissue-related voice disorders and post-phonogurgical voices, and because they were supported by significant correlations with the VHI, future research could continue to investigate their ability to measure progress for polyps, nodules, and cysts.

Assuming that a replication study supported our visual and statistical analyses, additional studies could also test the parameters' potential as measures for therapy effects. One suggested investigation should observe how OQ mean and SI mean measurements are affected from pre to post-intervention of patients with TVFMI (i.e. thyroplasty or collagen injection). RGG mean could be used in a longitudinal study to track the effects of speech therapy for vocal fold nodules.

These studies would eventually support the collection of normative data for each of the parameters across different types of voice disorders.

The suggested future research paths work to get us closer to a universally accepted set of objective measurements for HSDI. With a standardized method of measuring vocal fold vibrations, the field of speech pathology will gain a more detailed understanding of the physiology of normal and disordered voices. This understanding will aid in development of new therapy techniques and evaluating the efficacy of existing treatments. Overall, pursuing research that works toward establishing the parameters in the present study as a set of universal measurements for HSDI can greatly improve our current understanding of voice production as well as our evidence based practice.

Conclusion

The present study aimed to build evidence for the use of Ikuma et al.'s (2012a) feature objective parameters derived from harmonic waveform modeling in HSDI. The parameters were investigated for their ability to distinguish between voice disorder groups with differing etiologies. The results indicated that F0SD was effective for significantly differentiating ADSD from all voice groups except nodules. The HNR mean parameter differentiated normal voices from the ADSD voice group. Visual analysis suggested that RGG mean may differentiate nodules from all other groups with the elimination of outliers in other disorders. OQ mean and SI mean had weak visual trends, however additional research may show that these parameters are useful for tracking therapy progress in TVFMI, nodule, and polyp groups. Overall, the results of this study move us closer to identifying a set of universal objective measurements and clinical implementation of HSDI.

REFERENCES

Ahmad, K., Yan, Y. L., & Bless, D. M. (2012). Vocal fold vibratory characteristics in normal female speakers from high-speed digital imaging. *Journal of Voice, 26*, 239-253. doi: 10.1016/j.jvoice.2011.02.001

American Speech-Language-Hearing Association. (1993). *Definitions of Communication Disorders and Variations* [Relevant Paper]. Available from www.asha.org/policy.

Baken, R.J. & Orlikoff, R.F (2000). Laryngeal function. *Clinical measurement of speech and voice, second ed.* (pp. 393-451). San Diego, CA: Singular Publishing.

Chodara, A. M., Krausert, C. R., & Jiang, J. J. (2012). Kymographic characterization of vibration in human vocal folds with nodules and polyps. *The Laryngoscope, 122,* 58-65. doi: 10.1002/lary.22324

Colton, R., Casper, J., & Leonard, R. (2011). *Understanding voice problems: A physiological perspective for diagnosis and treatment.* Maryland: Lippincott Williams & Wilkins.

Colton, R. H., Woo, P., Brewer, D. W., Griffin, B., & Casper, J. (1995). Stroboscopic signs associated with benign lesions of the vocal folds. *Journal of Voice*, *9*, 312-325. doi: 10.1016/s0892-1997(05)80240-1

Crevier-Buchman, L., Laccourreye, O., Papon, J. F., Nurit, D., & Brasnu, D. (1997). Adductor spasmodic dysphonia: Case reports with acoustic analysis following botulinum toxin injection and acupuncture. *Journal of Voice*, *11*, 232-237.

De Bodt, M. S., Ketelslagers, K., Peeters, T., Wuyts, F. L., Mertens, F., Pattyn, J., . . . Van de Heyning, P. (2007). Evolution of vocal fold nodules from childhood to adolescence. *Journal of Voice, 21*, 151-156. doi: http://dx.doi.org/10.1016/j.jvoice.2005.11.006

Deliyski D. & Petrushev P. (2003). Methods for objective assessment of high-speed videoendoscopy. *Proceedings of the 6th International Conference: Advances in Quantitative Laryngology, Voice, & Speech Research*, *Germany, 6*, 1–16.

Deliyski, D., Petrushev, P., Bonilah, H., Gerlach, T., Martin-Harris, B., & Hillman, R. (2008). Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution. *Folia Phoniatrica et Logopaedica, 60*, 33-34. doi: 10.1159/000111802

Dikkers, F. G., & Nikkels, P. G. J. (1999). Lamina propria of the mucosa of benign lesions of the vocal folds. *The Laryngoscope, 109*, 1684-1689. doi: 10.1097/00005537-199910000-00025

Döllinger, M., Kunduk, M., Kaltenbacher, M., Vondenhoff, S., Ziethe, A., Eysholdt, U., & Bohr, C. (2012). Analysis of vocal fold function from acoustic data simultaneously recorded with high-speed endoscopy. *Journal of Voice, (26),* 726-733. doi: 10.1016/j.jvoice.2012.02.001

Hertegård, S., Larsson, H., & Wittenberg, T. (2003). High-speed imaging: Applications and development. *Logopedics Phoniatrics Vocology, 28*, 133-139. doi:10.1080/14015430310015246

Hirano, M. (1974). Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatrica et Logopaedica*, *26*, 89-94.

Ikuma, T., Kunduk, M., & McWhorter, A.J. (2012a). *Objective quantification of pre and post phonosurgery vocal fold vibratory characteristics using high-speed videoendoscopy and a harmonic waveform model*. Manuscript submitted for publication.

Ikuma, T., Kunduk, M., & McWhorter, A. J. (2012b). Mitigation of temporal aliasing via harmonic modeling of laryngeal waveforms in high-speed videoendoscopy. *The Journal of the Acoustical Society of America*, *132*, 1636-1645.

Inwald, E. C., Döllinger, M., Schuster, M., Eysholdt, U., & Bohr, C. (2011). Multiparametric analysis of vocal fold vibrations in healthy and disordered voices in high-speed imaging. *Journal of Voice*, *25*, 576-590. doi: 10.1016/j.jvoice.2010.04.004

Krausert, C. R., Liang, Y., Zhang, Y., Rieves, A. L., Geurink, K. R., & Jiang, J. J. (2012). Spatiotemporal analysis of normal and pathological human vocal fold vibrations. *American Journal of Otolaryngology, 33*, 641-649.

Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U., & Dollinger, M. (2007). Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Medical Image Analysis, 11*, 400-413. doi: 10.1016/j.media.2007.04.005

Ma, E. & Yiu, E. (2001). Voice activity and participation profile: Assessing the impact of voice disorders on daily activities. *Journal of Speech Language and Hearing Research, 44*, 511-524.

The MathWorks, Inc. (2012) MATLAB [Statistical analysis software]. Maryland: MathWorks.

Patel, R. R., Dailey, S., & Bless,D. (2008). Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Annals of Otology, Rhinology & Laryngology*, *117*, 413-424.

Patel, R., Dixon, A., Richmond, A., Donohue, K. (2012). Pediatric high speed digital imaging of vocal fold vibration: A pilot normative study of glottal closure and phase closure. *International Journal of Pediatric Otorhinolaryngology, 76*, 954-959.

Patel, R., Liu, L., Galatsanos, N., & Bless, D. M. (2011). Differential vibratory characteristics of adductor spasmodic dysphonia and muscle tension dysphonia on high-speed digital imaging. *The Annals of Otology, Rhinology & Laryngology*, *120*, 21-32.

Petrović-Lazić, M., Babac, S., Vuković, M., Kosanović, R., & Ivanković, Z. (2011). Acoustic voice analysis of patients with vocal fold polyp. *Journal of Voice, 25*, 94-97. doi:10.1016/j.jvoice.2009.04.002

Ramig, L.O. & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research*, *41*, 101-106.

Rosen, C. A. (2005). Stroboscopy as a research instrument: Development of a perceptual evaluation tool. *The Laryngoscope*, *115*, 423-428.

Roy, N., Merrill, R.N., Gray, S.D., & Smith, E.M. (2005). Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *Laryngoscope*, *115*, 1988-1995.

Sapienza, C. M., Murry, T., & Brown Jr, W. S. (1998). Variations in adductor spasmodic dysphonia: Acoustic evidence. *Journal of Voice, 12*, 214-222. doi: 10.1016/s0892-1997(98)80041-6

Shaw, H. & Deliyski, D. (2008). Mucosal wave: A normophonic study across visualization techniques. *Journal of Voice*, *22*, 23-33. doi: 10.1016/j.jvoice.2006.08.006

Silverman, E. P., Garvan, C., Shrivastav, R., & Sapienza, C. M. (2012). Combined modality treatment of adductor spasmodic dysphonia. *Journal of Voice*, *26*, 77-86. doi: 10.1016/j.jvoice.2010.08.004

Story, B. H., & Titze, I. R. (1995). Voice simulation with a body-cover model of the vocal folds. *The Journal of the Acoustical Society of America*, *97*, 1249-1260.

Tanner, K. (2012). Epidemiologic advances in spasmodic dysphonia. *Perspectives on Voice and Voice Disorders, 22*, 104-111.

Verdolini, K., Rosen, C., & Branski, R. (2006). *Classification Manual for Voice Disorders-I* (Vol. I): Psychology Press.

Verikas, A., Uloza, V., Bacauskiene, M., Gelzinis, A., & Kelertas, E. (2009). Advances in laryngeal imaging. *European Archives of Oto-Rhino-Laryngology*, *266*, 1509-1520. doi: 10.1007/s00405-009-1050-4.

Woo, P. (1996). Quantification of videostrobolaryngoscopic findings--measurements of the normal glottal cycle. *The Laryngoscope*, *106*, 1-27. doi: 10.1097/00005537-199603001-00001.

Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*, *71*, 1544-1550.

VITA

Aimee Elaine Perkins was raised in Prairieville, a suburb of Baton Rouge, Louisiana. She graduated from Louisiana State University in 2011 with a Bachelor of Arts in Communication Sciences and Disorders. Immediately following graduation, Aimee enrolled as a master's student again at Louisiana State University in the department of Communication Sciences and Disorders. There, Aimee became interested in the area of voice disorders and the high-speed digital imaging technology. Aimee began her thesis under Dr. Melda Kunduk as a partial fulfillment of the requirements for a Master of Arts degree. Upon graduating, Aimee hopes to obtain a clinical fellowship position as a speech-language pathologist working with patients with voice disorders.