

TOWARD VALIDATION OF AN ACOUSTIC INDEX  
OF DYSPHONIA SEVERITY

by

Elizabeth Anne Peterson

A thesis submitted to the faculty of  
The University of Utah  
in partial fulfillment of the requirements for the degree of

Master of Science

in

Speech-Language Pathology

Department of Communication Sciences and Disorders

The University of Utah

August 2012

Copyright © Elizabeth Anne Peterson 2012

All Rights Reserved

**The University of Utah Graduate School**

**STATEMENT OF THESIS APPROVAL**

The thesis of Elizabeth Anne Peterson

has been approved by the following supervisory committee members:

Nelson Roy, Chair 04/27/2012  
Date Approved

Bruce Smith, Member 04/27/2012  
Date Approved

Michael Blomgren, Member 04/27/2012  
Date Approved

and by Michael Blomgren, Chair of  
the Department of Communication Sciences and Disorders

and by Charles A. Wight, Dean of The Graduate School.

## ABSTRACT

This study investigated the relationship between a cepstral/spectral index of dysphonia severity (i.e., the CSID) and listener severity ratings of disordered voices. To assess the value of the CSID as a potential objective treatment outcomes tool, pre- and posttreatment samples of continuous speech and sustained vowel /ɑ/ productions were elicited from 112 patients (with varying degrees of dysphonia) from six diagnostic categories: (1) unilateral vocal fold paralysis (UVFP), (2) adductor spasmodic dysphonia (ADSD), (3) primary muscle tension dysphonia (PMTD), (4) benign vocal fold lesions (BVFL), (5) presbylaryngis, and (6) mutational falsetto. Perceptual ratings of dysphonia severity in continuous speech were compared to acoustically-derived severity estimates using a three factor CSID model consisting of the cepstral peak prominence (CPP), the ratio of low-to-high spectral energy, and its standard deviation. A five factor CSID model incorporating all acoustic variables as well as gender and the CPP standard deviation was used to estimate severity in sustained vowel samples. Results showed strong relationships between perceptual and acoustic estimates in dysphonia severity in connected speech ( $r = 0.72, p < 0.0001$ ) and sustained vowels ( $r = 0.836, p < 0.0001$ ). A strong relationship between the perceived and predicted change in dysphonia severity from pre- to posttreatment was also observed for connected speech ( $r = 0.77, p < 0.001$ ) and sustained vowels ( $r = 0.81, p < 0.0001$ ). Spectrum effects were also examined, and overall severity (mild, moderate, or severe) did not influence the relationship between perceived and

estimated severity ratings in connected speech ( $F[1, 2] = 0.58, p = 0.56$ ); however, dysphonia severity did influence the relationship in sustained vowels ( $F[1, 2] = 6.22, p = 0.002$ ). In general, the results confirm a robust relationship between listener perceived and acoustically-derived estimates of severity within the contexts of connected speech and sustained vowels across diverse diagnostic categories and varying degrees of dysphonia severity. As such, the CSID shows considerable promise as an objective treatment outcomes measure.

## TABLE OF CONTENTS

ABSTRACT .....	v
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
ACKNOWLEDGMENTS .....	viii
INTRODUCTION .....	1
METHODS .....	7
Speech Samples .....	7
Acoustic Analyses .....	8
Auditory-Perceptual Ratings .....	13
RESULTS .....	16
Data Analysis .....	16
Predicted Severity: Connected Speech .....	17
Predicted Severity: Sustained Vowel .....	20
Effect of Dysphonia Severity .....	24
DISCUSSION .....	30
Performance of the CSID: Connected Speech Versus Sustained Vowels .....	31
Performance of CSID Across Diagnostic Category .....	33
Performance of the CSID Across the Severity Spectrum .....	36
CONCLUSION .....	40
APPENDIX .....	41
REFERENCES .....	45

## LIST OF TABLES

### Tables

1. Participant characteristics .....	9
2. Intrarater reliability .....	14
3. Interrater reliability with all 95% confidence intervals expressed parenthetically .....	15
4. Correlation coefficients between mean perceived and mean predicted ratings in connected speech within each diagnostic category for pre- and posttreatment samples and the change in dysphonia (pretreatment – posttreatment) .....	18
5. Correlation coefficients between mean perceived and mean predicted ratings in sustained vowel within each diagnostic category for pre- and posttreatment samples and the change in dysphonia (pretreatment – posttreatment) .....	22
6. Interaction effect of severity within each diagnostic category for connected speech and sustained vowel contexts.....	25

## LIST OF FIGURES

### Figures

1. Mean perceived versus predicted change in severity by diagnostic category ..... 19
2. Scatter plot of mean perceived and predicted severity ratings for all connected speech samples..... 19
3. Mean perceived versus predicted severity ratings for connected speech by diagnostic category ..... 20
4. Scatter plot of mean perceived and predicted severity ratings for all sustained vowel samples..... 21
5. Mean perceived versus mean predicted severity ratings for sustained vowel by diagnostic category ..... 23
6. Mean acoustic severity estimates versus mean listener ratings per sample for all pre- and posttreatment samples combined for connected speech ..... 26
7. Mean acoustic severity estimates versus mean listener ratings per sample for all pre- and posttreatment samples combined for sustained vowel..... 26
8. Mean perceived versus predicted size of change in dysphonia severity for connected speech ..... 28
9. Mean perceived versus predicted size of change in dysphonia severity for sustained vowel..... 28



## ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. Nelson Roy for his patience and encouragement throughout this great project. His excellent advice impacted many facets of my educational career, and his example of excellence represents a standard that I hope to emulate throughout my life. I would also like to thank those with whom I had the privilege to work while completing this research project: Dr. Kristine Tanner and Dan Houtz for contributing their invaluable skills and participating in data acquisition, and Dr. Ray Merrill, who greatly contributed to the statistical analyses. I would also like to express my love and gratitude to my husband and family. Their support and strength sustained me through graduate school, and my completion of this degree would not have been possible without their continual encouragement.

## INTRODUCTION

The professions of speech-language pathology and laryngology have entered into an age where clinicians are faced with increased pressure to practice evidence-based treatments, and to demonstrate positive and objective outcomes associated with their interventions (Frattali, 1998). This growing pressure has stimulated a search for objective, reliable, and valid methods to measure treatment effects within the area of voice disorders. With the advent of relatively inexpensive personal computers, low cost analysis software, and increased availability of digital audio recording systems, acoustic analysis of voice has become an increasingly popular option for tracking intervention outcomes.

Early acoustic analysis methods used various time-based measures to estimate the severity of dysphonia in sustained vowels only. While time-based measures such as jitter and shimmer are useful in acoustic analysis of sustained vowels, these measures have limitations when applied to connected speech, and severely disordered voices (Maryn, Roy, De Bolt, Van Cauwenberge, & Corthals, 2009). For instance, continuous speech, as opposed to sustained vowels, contains rapid onsets and offsets, voiced and voiceless phonemes, amplitude variation, fundamental frequency variation related to prosody, as well as speech rate, phonetic contexts, vocal pauses, and stress (Maryn et al., 2009). Time-based measures require cycle boundary identification to determine fundamental frequency and ultimately aperiodicity, and these features make time-based measures

inadequate for quantifying dysphonia severity in connected speech. Awan, Roy, Jette, Meltzner, and Hillman (2010) asserted that time-based analysis measures such as jitter and shimmer, when applied to connected speech, may falsely inflate acoustically predicted dysphonia severity ratings because of the influence of voiceless phonemes and prosodic variations (such factors make it difficult to identify the period, and thus, the fundamental frequency). The authors also suggested that vowel duration, which is much shorter in connected speech than in sustained vowels, negatively impacts the ability of time-based measures to accurately track aperiodicity and dysphonia severity. Therefore, in order to be able to acoustically estimate the severity of dysphonia in connected speech, measures other than those that are time-based are required.

In addition to difficulty analyzing connected speech, time-based measures are especially problematic when analyzing the voices of individuals with severe dysphonia, which causes considerable aperiodicity in the voice signal. The validity of such time-based measures (like jitter and shimmer), when applied to moderately to severely disordered voices, has been recently called into question because cycle boundary identification can be exceptionally difficult (Awan & Roy, 2005). Thus, in order to reliably analyze voice in connected speech, sustained vowels, and across a continuum of dysphonia severity, acoustic measurements other than those that are time-based are necessary.

In this regard, Maryn et al. (2009) reported the results from a meta-analysis of the assessment of overall voice quality (i.e., dysphonia severity) and the relationship between perceptual ratings and acoustic measures. The meta-analysis reviewed a total of 25 studies; 21 studies examined sustained vowels using 69 acoustic markers and seven

studies examined connected speech using 26 acoustic markers. The meta-analysis identified six acoustic parameters that were determined to correlate reasonably well with listener ratings: (1) Pearson  $r$  at autocorrelation peak, (2) spectral flatness of residue signal, (3) pitch amplitude, (4) cepstral peak prominence (CPP), (5) smoothed cepstral peak prominence, and (6) signal-to-noise ratio from Qi (Maryn et al., 2009). Most of these measures are not time-based; thus they do not require cycle boundary identification to determine fundamental frequency and estimate aperiodicity. Two of these measures, the CPP and the smoothed CPP, were found to be the best predictors of dysphonia severity as compared to listener ratings. These cepstral-based measures can also be used to evaluate the severity of voice for both continuous speech and sustained vowels, thus they appear to be ideal for evaluating overall severity.

Recently, an acoustic analysis program called the Analysis of Dysphonia in Speech and Voice (ADSV) was developed and tested by Awan et al. (2010). Within this program, an acoustic estimate of dysphonia severity, known as the Cepstral/Spectral Index of Dysphonia (CSID) is generated. The CSID uses spectral and cepstral-based measures to predict dysphonia severity and evaluate treatment outcomes. The spectral and cepstral-based measures included in the CSID were based upon earlier work by Awan and colleagues, who confirmed that these acoustic parameters related well to listener ratings of dysphonia derived from continuous speech and sustained vowels (Awan & Roy, 2005, 2006, 2009; Awan et al., 2010; Hillenbrand, Cleveland, & Erickson, 1994; Watts & Awan, 2011).

The development of the CSID was based upon a number of earlier studies that employed stepwise multiple regression to determine which acoustic variables accounted

for most of the variance in listener judgments of dysphonia severity (Awan & Roy 2005, 2006, 2009). This iterative and empirically driven process produced an acoustic algorithm (model) for predicting/estimating dysphonia severity. The acoustic variables included in the final CSID algorithm for analysis of connected speech are the CPP, the low/high (L/H) spectral ratio, and the L/H spectral ratio standard deviation (Awan et al., 2010). The acoustic variables for the sustained vowel model are the same as those used in the connected speech model, but also included gender and the CPP standard deviation as variables. The ADSV, Model 5109, was recently commercialized by the KayPENTAX Corporation, and can be purchased as an option with Multi-Speech, the Computerized Speech Lab (CSL), Visi-Pitch, or Sona-Speech programs. Within the ADSV program are the routines that generate the CSID, which is the quantitative, dysphonia summary tool that reflects the spectral and cepstral measures that can be extracted from a speech or voice sample. For the purpose of this study, the CSID measurement, generated by the ADSV, was used as the acoustic measurement to estimate dysphonia severity.

In an age of regulatory agency demands for data, an objective measure that accurately quantifies dysphonia severity would be particularly useful to evaluate the effects of behavioral, medical, or surgical intervention. Such a measure could be used by otolaryngologists and speech-language pathologists to provide objective data for patients, third-party payers, and other stakeholders. A valuable outcomes tool, however, must be sensitive to changes following treatment, and it must be sensitive to heterogeneous voice qualities related to a variety of pathologies and ranges of severity. Furthermore, a valuable outcomes tool must be sensitive across a range of severity, and it must capture varying degrees of change in dysphonia severity. Acoustic analysis and its potential to

accurately quantify dysphonia severity offers great promise as a potential treatment outcomes measure. Specifically, spectral and cepstral-based measures present the possibility of such a tool; however, to date, there have been a limited number of published studies that examine such a tool's potential.

Previous studies have examined the validity of the acoustic parameters used in the CSID. The spectral and cepstral-based measures used in the CSID have been found to have high sensitivity and specificity as a diagnostic tool within the contexts of connected speech and sustained vowels in hypofunctional voices (Watts & Awan, 2011). In addition, these parameters have been found to be sensitive to listener perceived changes in voice quality pre- and postthyroidectomy in connected speech samples (Awan, Helou, Stojadinovic, & Solomon, 2011), indicating that these acoustic measures are sensitive to change in specific types of dysphonia. In one previous study, an algorithm comprised of spectral and cepstral-based measures was used to measure treatment outcomes in a disorder known as Muscle Tension Dysphonia (MTD) (Awan & Roy, 2009). The results indicated that spectral and cepstral-based measures were strongly associated with perceptual ratings of dysphonia severity. The authors concluded that the acoustic algorithm seemed to be a sensitive treatment outcomes measure for MTD. However, these measures and their utility as a treatment outcomes tool have not been examined across a variety of diagnostic categories, and it is possible that idiosyncratic features associated with other vocal pathologies may attenuate the spectral and cepstral-based measures' performance. For example, the vocal qualities characteristic of unilateral vocal fold paralysis (UVFP), spasmodic dysphonia (SD), and mutational falsetto are potentially dissimilar to MTD, and it is unknown how well spectral and cepstral-based measures will

perform as an outcomes measure when applied to these pathologies. Thus, the primary purpose of this study was to determine the validity of the CSID, an acoustic analysis algorithm comprised of spectral and cepstral-based measures, as a possible objective treatment outcomes measure. The strength of the relationship between perceptual ratings and acoustic estimates of dysphonia severity across various diagnostic categories (pre- and posttreatment) and severities was investigated. A secondary purpose of this study was to determine the validity of the CSID across a spectrum of severity to identify possible spectrum effects.

## METHODS

### Speech Samples

Pre- and posttreatment voice/speech samples were selected from a database of recordings of patients who attended the University of Utah Voice Disorders Center, Salt Lake City, Utah. These audio recordings were collected in a quiet environment by speech-language pathologists at the Center as part of routine, standard care, using research-quality recording instrumentation. One hundred twelve patients were selected from six diagnostic categories, including (1) unilateral vocal fold paralysis (UVFP), (2) adductor spasmodic dysphonia (ADSD), (3) primary muscle tension dysphonia (PMTD), (4) benign vocal fold lesions (BVFL), (5) presbylaryngis, and (6) mutational falsetto. These diagnostic categories were selected because they are: (1) encountered frequently in multidisciplinary voice clinics, and (2) characterized by heterogeneous and possibly idiosyncratic voice qualities and severities.

Individuals who had been diagnosed with one of the six diagnoses of interest were reviewed sequentially beginning in August 2011 and moving backward until 20 consecutive participants in each diagnostic category, who had a complete and analyzable dataset, were selected for inclusion. The review for the sixth diagnostic category (i.e., mutational falsetto) identified only 12 participants who had complete datasets.

Sample inclusion was based upon several factors. First, the voice/speech samples were selected based upon their primary voice disorder diagnosis, as determined by an



otolaryngologist and a speech-language pathologist who specialize in voice disorders. Second, patients underwent some form of intervention and a follow-up voice/speech sample was available. Thus, to be included in the voice/speech samples, a second, follow-up posttreatment sample was required, as the intent of this study was to assess the CSID's sensitivity to pre- and posttreatment changes in dysphonia severity. The specific intervention technique was of no particular consequence, so long as some change, whether positive or negative (small or large), was apparent in the posttreatment sample collected. Therefore, for each participant there were two sets of voice/speech recording samples, one pretreatment and the other posttreatment. The participants were recorded reading (1) "The Rainbow Passage," (Fairbanks, 1960) which served as the connected speech sample, and (2) a vowel /ɑ/ production sustained for approximately 5 seconds at comfortable loudness and pitch. The samples were digitized at a sampling rate of 25 kHz using Multi-Speech (Model 3700). The samples were then edited for ADSV analysis and for the listening task. For connected speech, the second and third sentences were extracted from the Rainbow Passage, and for the sustained vowel, the middle 3 seconds of the sustained vowel were isolated for analysis. A description of participant characteristics for each diagnostic category is presented in Table 1. In addition, definitions of each diagnostic category are provided in the Appendix.

#### Acoustic Analyses

All of the connected speech and sustained vowel samples were analyzed using the ADSV. The samples of the participants were deidentified, coded, and digitized. Once a sample was uploaded to the program, the ADSV allowed the user to manually or

Table 1  
Participant Characteristics

Diagnostic Category	Number of Participants	Age (years)		Gender Ratio (Males: Females)	Duration of Symptoms	
		Mean (years)	Range (years)		Mean (months)	Range
Unilateral Vocal Fold Paralysis	20	49.1 (15.5)	17-81	7:13	33.3 (59.1)	3 weeks - 10 years
Adductor Spasmodic Dysphonia	20	49.2 (13.9)	23-71	10:10	78.7 (83.9)	6 months - 30 years
Primary Muscle Tension Dysphonia	20	44.1 (13.6)	22-72	1:19	80.3 (110.7)	2 weeks - 25 years
Benign Vocal Fold Lesions	20	48.5 (12.2)	27-66	9:11	20.9 (39.2)	3 weeks - 15 years
Presbylaryngis	20	78.9 (5.7)	67-89	15:5	46.1 (56)	6 weeks - 20 years
Mutational Falsetto	12	18 (3.1)	14-25	12:0	39.9 (33.8)	2 months - 10 years

Summary of participant characteristics within each diagnostic category with standard deviations displayed parenthetically

automatically select the portion of the sample to be analyzed by placing cursors around the desired sample. The program requires several frames of information for each data point that is generated, and as such, automatic selection in the program results in the first 0.05 seconds not being selected for analysis. To ensure that a sample is analyzed in its entirety, the cursors must encompass the sample in its totality. For the purpose of this study, when analyzing sustained vowels, the user selected the entire 3 seconds. As the 3 second sustained vowel was selected from the middle of the participant's /a/, there was no silence or white noise due to breathing prior to the sample. As such, the whole sample could be selected and analyzed. In the connected speech samples, however, there was no waveform at the very beginning of the sample, as the participant had paused prior to beginning a new sentence. In an effort to not include any silence or white noise in the analysis while ensuring that the voice sample in its entirety was analyzed, cursors were placed 0.05 seconds before the waveform demonstrated any evidence of the participant initiating speech. This allowed for the connected speech sample to be analyzed in its entirety without including any unnecessary nonspeech or voice related information in the analysis.

Next, the software analyzed the voice/speech samples and the CSID measurement, an estimate of dysphonia severity, was recorded. The CSID estimate is theoretically a number between 0 and 100, with 100 being rated the most severe. However, at times, the CSID can generate a number below 0 or above 100; such an estimate represents an extremely normal and periodic voice, or a profoundly abnormal and aperiodic voice, respectively. The ADSV applies a number of steps during the

acoustic analysis of the voice/speech samples, as reported by Awan et al. in 2010. The steps employed within the ADSV are as follows:

1. A total of 1024 overlapping frames are created by separating the voice/speech samples into a sequence of frames. The total overlap between frames is 75%. For each frame, a 1024 point discrete Fourier transformation (DFT) is created, which is then converted into the log power spectrum (Baken, 1987). Next, a second DFT is computed, resulting in cepstrum. A very periodic signal results in a cepstrum with a clear cepstral peak prominence (CPP). The CPP is simply the most prominent peak in a cepstrum, and is also called a signal's dominant rahmonic. A dominant rahmonic is the fundamental period of a signal, and the word "rahmonic" was derived from the word harmonic, with the first syllable spelled backwards. A very aperiodic signal, on the other hand, results in a weak dominant rahmonic or CPP (Hillenbrand & Houde, 1996; Hillenbrand et al., 1994). Smoothing is then applied to the cepstrum to allow for further identification of the CPP (Awan et al., 2010; Hillenbrand & Houde, 1996).
2. In each cepstral frame, both smoothed and unsmoothed, a variety of measures are calculated. From the smoothed frames, the CPP is identified, as is the ratio between the observed and expected CPP, which is estimated using a linear regression analysis (Awan et al., 2010). From the unsmoothed frames, only the low/high frequency (L/H) is calculated for spectral energy. Upon computing these measures, the means and standard deviations of the L/H spectral energy ratio and CPP are calculated (Awan et al., 2010). The standard deviations of the CPP and the L/H spectral ratio are included within the analysis, as variability was deemed

to be an important measure in identifying dysphonia severity (Awan & Roy, 2006, 2009; Callan, Kent, Roy, & Tasko, 1999; Wolfe & Steinfatt, 1987). Each of these predictor variables has reason for being included within the ADSV analysis, and the measurements included were the same used by Awan et al. (2010).

3. The CSID for the Rainbow Passage was then calculated manually using three factors: the CPP, the L/H spectral ratio, and the L/H spectral ratio standard deviation. The manual computation utilized the formula provided in the Analysis of Dysphonia and Voice (ADSV): An Application Guide (Awan, 2011), as the current version of the ADSV does not contain an automatic computation of the Rainbow Passage CSID. The formula provided in the application guide is based upon the previous work conducted by Awan and colleagues. The sustained vowel CSID was calculated automatically by using five factors: gender of the participant, the CPP, the L/H spectral ratio, and their respective standard deviations.

In order to evaluate the remeasurement reliability of the CSID, 20% of the samples ( $n = 22$  samples) were randomly selected and reanalyzed by the original user. These CSID estimates were extracted and recorded in the same manner as the original estimates. As expected, mean Pearson correlation coefficients ( $r$ ) revealed excellent remeasurement reliability for connected speech ( $r = 0.99, p < 0.0001$ ) and sustained vowels ( $r = 0.99, p < 0.0001$ ).

### Auditory-Perceptual Ratings

Eight graduate speech-language pathology students at the University of Utah served as listeners to establish auditory-perceptual ratings of dysphonia severity for all 446 samples. All listeners had completed coursework in the assessment and management of voice disorders, though none had extensive research or clinical experience with these disorders. The samples were presented to the listeners at a comfortable loudness level in a quiet environment using research-quality speakers. To ensure that all samples were presented at a comparable loudness, the samples were first normalized within each diagnostic category using Adobe Audition CS5.5. The samples were normalized to ensure that any perceived changes in loudness between samples were minimized, thus limiting this factor as a possible influence in listener ratings.

During the rating session, each diagnostic category was presented as its own listening experiment, and the listeners were blinded to the diagnostic category. At the beginning of a rating session, 10 samples that were representative of the experimental samples in the diagnostic category (but not included in the listening task) were presented to the listeners to orient them to the task. The pre- and posttreatment samples were presented as pairs, and each sample was rated using a computer program with a 100-millimeter visual analog scale (VAS). Listeners were instructed to place a vertical marker using the cursor on a VAS for each presented sample, which was labeled “Normal Voice” on the far left side, and “Profoundly Abnormal Voice” on the far right side. Thus, a rating of zero reflected a normal voice, whereas 100 reflected a severely dysphonic voice. Within each diagnostic category, the order of the samples was randomized, as was the presentation order of the paired samples. This resulted in each listener being exposed to

the samples in different sequences. The mean scores for each sample were then calculated by averaging the eight listeners' ratings.

In order to determine intrajudge reliability, 20% of the samples ( $n = 22$  samples) were randomly selected and rated a second time. Once all ratings were completed, intrarater and interrater reliability were calculated. Intrarater reliability, estimated using the Pearson correlation coefficients ( $r$ ), was acceptable for both sustained vowels ( $r = 0.83, p < 0.0001$ ) and connected speech ( $r = 0.92, p < 0.0001$ ). Table 2 provides intrarater reliability estimates for each diagnostic category, and reveals acceptable reliability across all categories. Similarly, intraclass correlation coefficients (ICCs) revealed acceptable interrater reliability for sustained vowels,  $ICC = 0.95 (p < 0.0001)$  and connected speech,  $ICC = 0.96 (p < 0.0001)$ . Table 3 provides interrater reliability by diagnostic category, where similar to intrarater reliability, the listeners were found to be reliable across all categories.

Table 2. Intrarater Reliability

Intrarater ( $r$ )	Connected Speech	Sustained Vowel
UVFP	0.848	0.963
ADSD	0.876	0.834
PMTD	0.901	0.964
BVFL	0.794	0.590
Presbylaryngis	0.856	0.738
Mutational Falsetto	0.728	0.873

All Pearson's  $r$  correlations are significant at  $p < 0.0001$

Table 3. Interrater Reliability with all 95% confidence intervals expressed parenthetically

Interrater (ICC)	Connected Speech	Sustained Vowel
UVFP	0.976 (0.964 - 0.985)	0.948 (0.923 - 0.968)
ADSD	0.971 (0.956 - 0.982)	0.956 (0.935 - 0.973)
PMTD	0.987 (0.980 - 0.992)	0.984 (0.976 - 0.990)
BVFL	0.954 (0.931 - 0.971)	0.943 (0.915 - 0.965)
Presbylaryngis	0.952 (0.929 - 0.970)	0.921 (0.883 - 0.951)
Mutational Falsetto	0.947 (0.911 - 0.972)	0.928 (0.877 - 0.963)

All ICCs are significant at  $p < 0.0001$



## RESULTS

### Data Analysis

Pearson's Product Moment Correlations were used to determine the degree of association between auditory-perceptual ratings and acoustic estimates of severity across three different conditions. The first measure examined the degree of association between the pretreatment listener severity ratings and the acoustic estimates, the second examined the degree of association between posttreatment listener severity ratings and the acoustic estimates, and the third examined the degree of association between the change observed in listener severity ratings and acoustic estimates from pre- to posttreatment. Change in dysphonia severity was calculated by subtracting mean posttreatment severity ratings from pretreatment severity ratings. This last correlation permitted an assessment of the CSID's sensitivity to the change in severity, and its potential as a treatment outcomes measure. These correlation coefficients were calculated separately for each of the six diagnostic categories, and for sustained vowels and connected speech. Finally, these three correlation coefficients were calculated for all 112 participants aggregated across the six diagnostic categories (for sustained vowels and connected speech separately). This allowed for the degree of association between auditory-perceptual ratings of severity and acoustic estimates of severity to be examined across a large, heterogeneous group. Furthermore, to examine the CSID's performance across a spectrum of dysphonia severity, and to assess possible spectrum effects, voice samples were also stratified into

tertiles, or categories (mild, moderate, and severe), based upon the listener severity ratings of dysphonia. A Generalized linear model (GLM) was then computed to examine the performance of the CSID across the severity spectrum.

### Predicted Severity: Connected Speech

Within the connected speech context, estimates of dysphonia severity were generated for all pretreatment and posttreatment samples ( $n = 112$ ) using the CSID. A Pearson's Product Moment correlation was calculated to determine the degree of association between the estimates of dysphonia severity and the mean perceived ratings of severity. Overall results combining pre- and posttreatment severity ratings for connected speech samples indicated a moderately strong, significant relationship between perceived ratings and acoustic estimates of dysphonia severity ( $r = 0.72, p < 0.05$ ). Aggregated correlations across all diagnostic categories revealed a moderately strong, significant relationship for pretreatment samples ( $r = 0.673, p < 0.05$ ) and posttreatment samples ( $r = 0.66, p < 0.05$ ), and a strong, significant relationship in the change in dysphonia severity ( $r = 0.767, p < 0.05$ ). Within each diagnostic category, correlations were also calculated for pre- and posttreatment samples, as well as for change in dysphonia severity (Table 4). While all correlations for the change in dysphonia severity were significant, the strength of the relationship between mean perceived ratings and acoustic estimates varied. Primary muscle tension dysphonia had the strongest relationship ( $r = 0.866, p < 0.05$ ), while presbylaryngis had the weakest ( $r = 0.468, p < 0.05$ ). None of the change in dysphonia severity correlation coefficients differed significantly (i.e., all  $p$  values  $> 0.05$ ) with the exception of a significant difference

Table 4. Correlation coefficients between mean perceived and mean predicted ratings in connected speech within each diagnostic category for pre- and posttreatment samples and the change in dysphonia (pretreatment – posttreatment)

Diagnosis	Pretreatment		Posttreatment		Change	
	Pearson's		Pearson's		Pearson's	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
UVFP	0.771	<0.0001	0.826	<0.0001	0.691	0.001
ADSD	0.749	0.0001	0.774	<0.0001	0.752	0.0001
PMTD	0.845	<0.0001	-0.386	0.092	0.866	<0.0001
BVFL	0.512	0.021	0.681	0.001	0.659	0.002
Presbylaryngis	0.536	0.015	0.635	0.003	0.468	0.037
Mutational Falsetto	0.585	0.046	0.538	0.071	0.681	0.015

observed between primary muscle tension dysphonia and presbylaryngis ( $p < 0.05$ ). In addition, the CSID appeared to consistently underestimate change in dysphonia severity in connected speech, as can be seen in Figure 1, which illustrates the perceived versus acoustically-estimated change in severity by diagnostic category. However, based upon pairwise *t* test comparisons, no significant differences (i.e.,  $p < 0.05$ ) were identified between the listener perceived and acoustically estimated change scores, regardless of diagnostic category. Figure 2 provides a scatter plot of mean perceived and predicted severity ratings for all connected speech samples.

Within each diagnostic category, the mean perceived severity ratings of all voice samples were averaged and compared to the averaged estimates of dysphonia severity. This was completed for pretreatment and posttreatment samples separately. With the exception of mutational falsetto, results indicated that when perceived ratings were aggregated within a category, they were very closely related to the aggregated predicted dysphonia ratings (Figure 3), particularly for PMTD. Inspection of Figure 3 reveals that the dysphonia rating predicted by the CSID was noticeably lower than the perceived

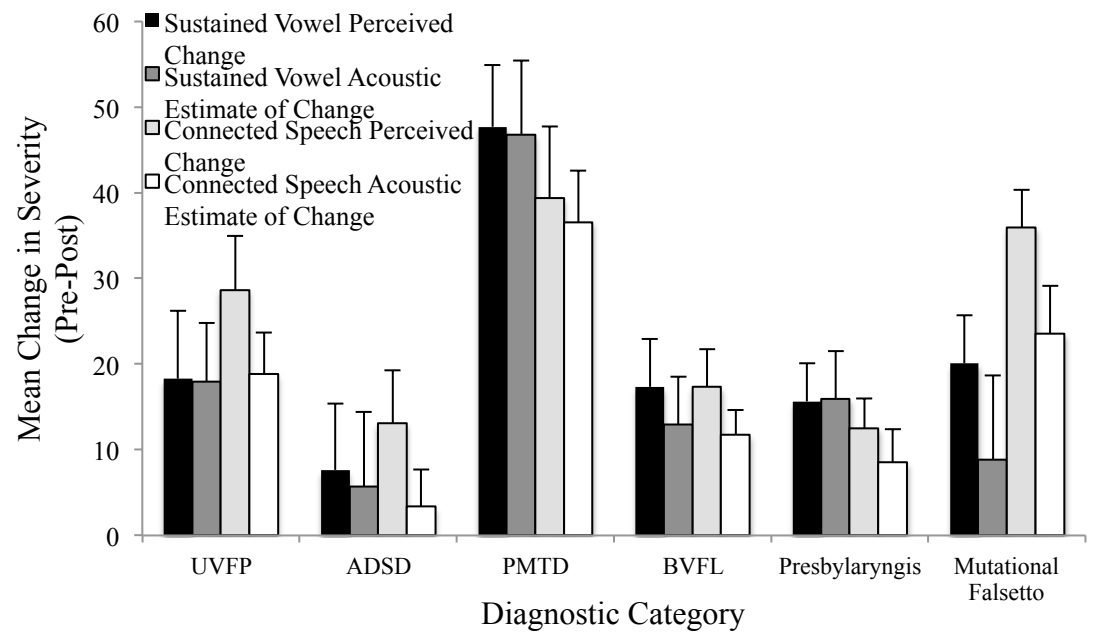


Figure 1. Mean perceived versus predicted change in severity by diagnostic category

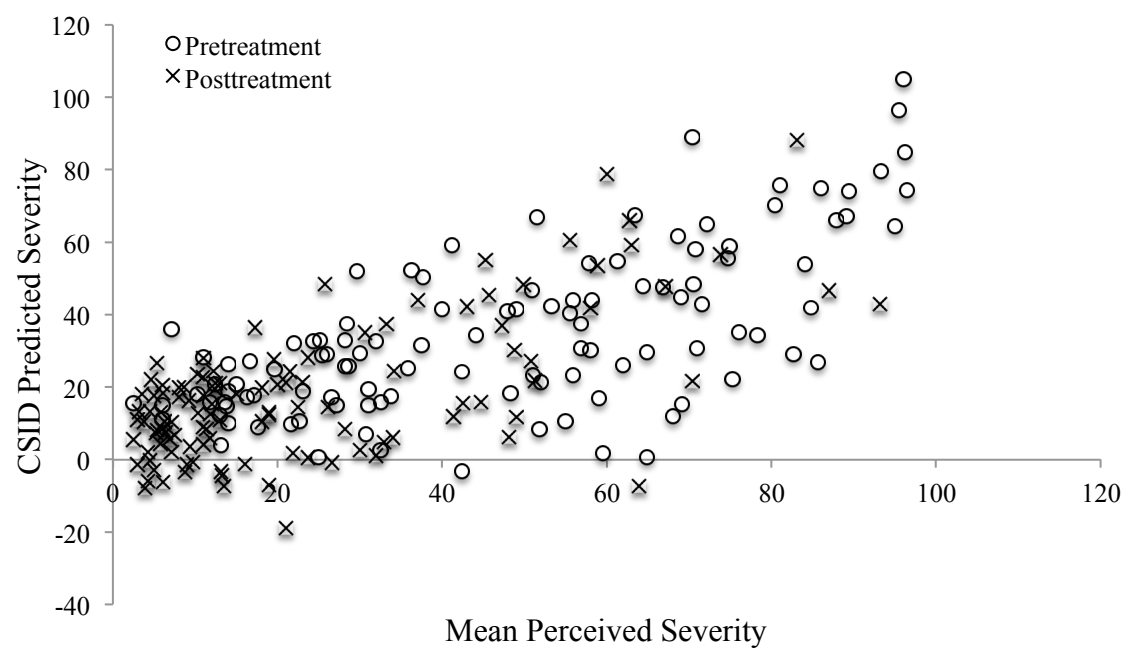


Figure 2. Scatter plot of mean perceived and predicted severity ratings for all connected speech samples

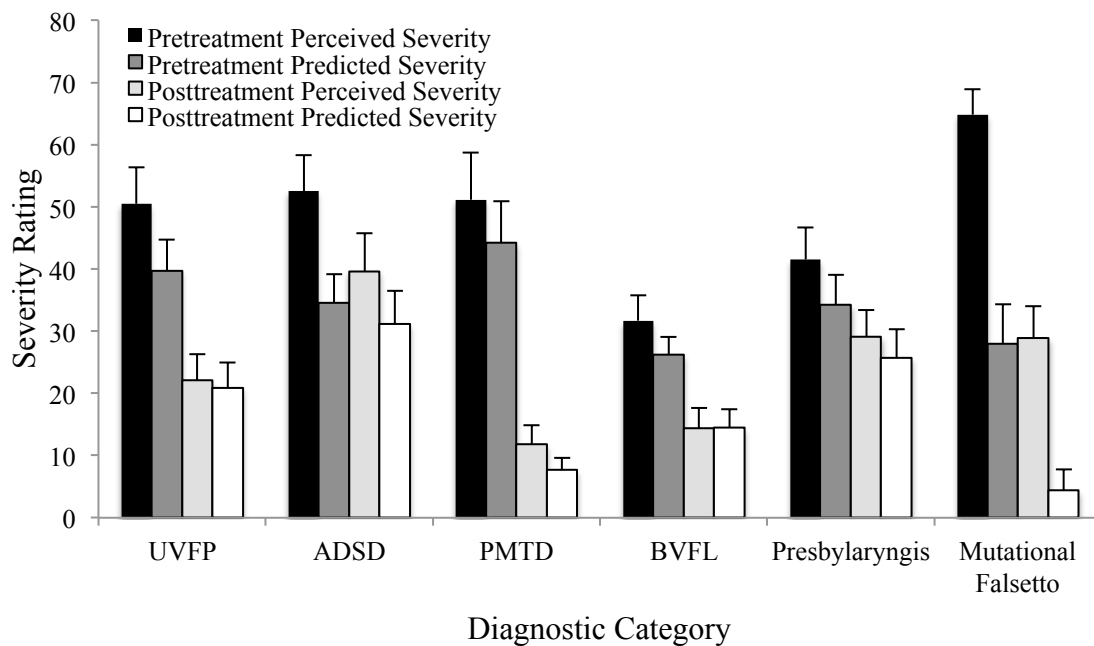


Figure 3. Mean perceived versus predicted severity ratings for connected speech by diagnostic category

dysphonia ratings in both pre- and posttreatment context for mutational falsetto; yet, the relationship between the mean perceived and acoustically predicted change in dysphonia severity was still moderately strong in this diagnostic category ( $r = 0.681, p < 0.05$ ). Furthermore, a nonsignificant GLM test confirmed that the relationship between the mean perceived and predicted ratings of dysphonia in connected speech was not dependent upon diagnostic category ( $F[1, 5] = 1.39, p = 0.23$ ). That is, the diagnostic category did not influence the reported relationships.

#### Predicted Severity: Sustained Vowel

Estimates of dysphonia severity were generated for all pre- and posttreatment samples within the sustained vowel context ( $n = 111$ ) using the CSID. As with connected

speech, correlations were computed to determine the strength of the association between mean listener ratings and acoustically estimated severity. The overall results, which combined all pre- and posttreatment severity ratings, indicated a strong relationship between perceived and predicted ratings of dysphonia severity in sustained vowels ( $r = 0.836, p < 0.05$ ). Correlations aggregated across all diagnostic categories indicated significant relationships in pretreatment samples ( $r = 0.809, p < 0.05$ ), posttreatment samples ( $r = 0.810, p < 0.05$ ), and the change in dysphonia severity ( $r = 0.814, p < 0.05$ ). (Figure 4 provides a scatter plot of mean perceived and predicted severity ratings for all sustained vowel samples). In addition, correlations were computed within each diagnostic category. Similar to connected speech, all correlations for the change in dysphonia severity across the diagnostic categories were significant; however, the strength of the

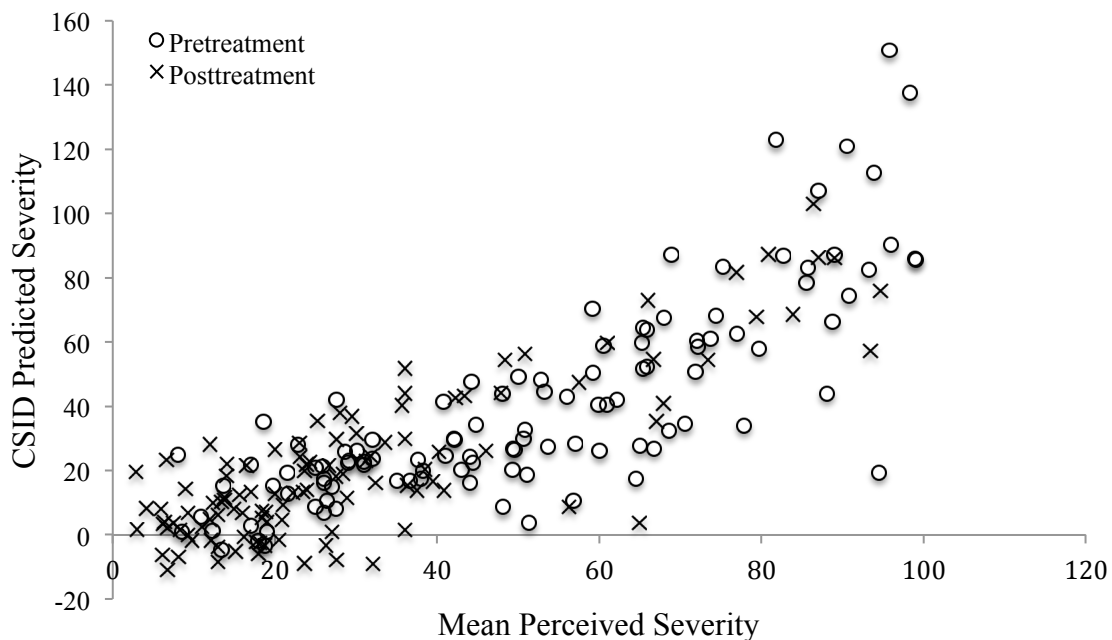


Figure 4. Scatter plot of mean perceived and predicted severity ratings for all sustained vowel samples

association varied. Mutational falsetto had the strongest relationship between predicted and estimated change in dysphonia severity ( $r = 0.844, p < 0.05$ ), while the weakest relationship was again found in presbylaryngis ( $r = 0.74, p < 0.05$ ). Table 5 provides the correlations between the mean perceived severity ratings and estimates of dysphonia severity by diagnostic category. None of the change in dysphonia severity correlation coefficients differed significantly (i.e.,  $p > 0.05$ ). Within the context of sustained vowels, the CSID seemed to consistently underestimate change in dysphonia as compared to mean perceived ratings, particularly for mutational falsetto (Figure 1). However, none of the differences between perceived versus estimated change scores reached statistical significance for the sustained vowel context, based upon pairwise  $t$  test comparisons (i.e., all  $p$  values  $> 0.05$ ).

As in the connected speech analyses, mean perceived ratings of severity were compared to the mean CSID estimates of sustained vowels within each diagnostic category. This was completed for pretreatment and posttreatment samples separately (Figure 5). The results showed that the means of perceived versus predicted dysphonia severity were closely related, particularly for primary muscle tension dysphonia.

Table 5. Correlation coefficients between mean perceived and mean predicted ratings in sustained vowel within each diagnostic category for pre- and posttreatment samples and the change in dysphonia (pretreatment – posttreatment)

Diagnosis	Pretreatment		Posttreatment		Change	
	Pearson's		Pearson's		Pearson's	
	$r$	$p$	$r$	$p$	$r$	$p$
UVFP	0.859	<0.0001	0.859	<0.0001	0.806	<0.0001
ADSD	0.773	<0.0001	0.903	<0.0001	0.821	<0.0001
PMTD	0.853	<0.0001	0.032	0.895	0.785	<0.0001
BVFL	0.601	0.005	0.781	<0.0001	0.755	0.0001
Presbylaryngis	0.899	<0.0001	0.686	0.001	0.740	0.0002
Mutational Falsetto	0.833	0.001	0.937	<0.0001	0.844	0.001

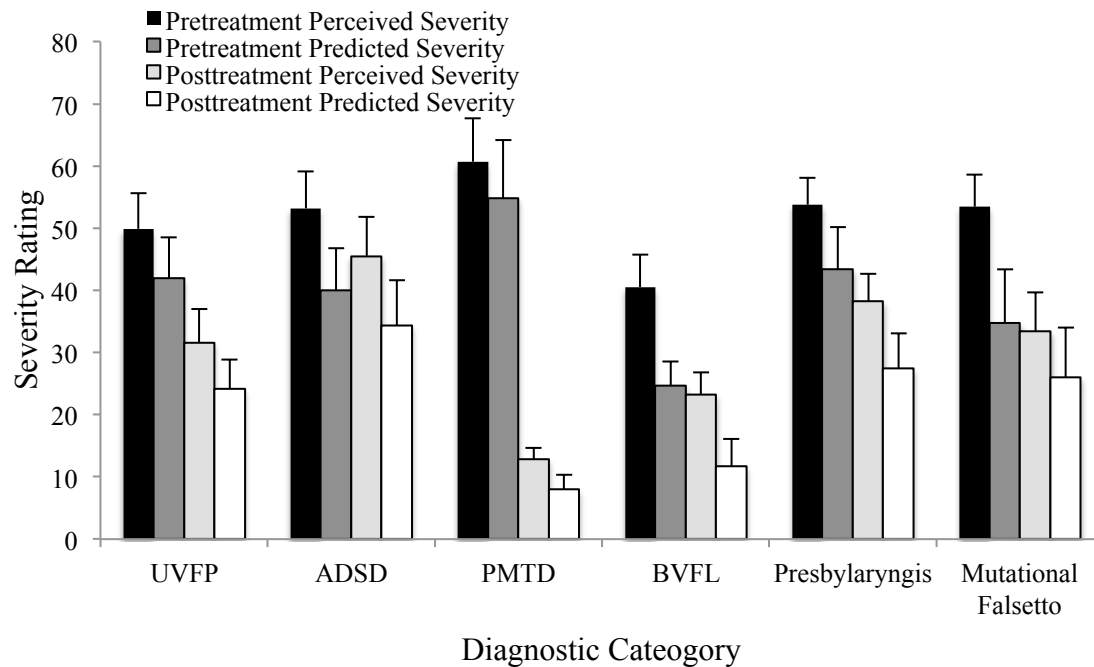


Figure 5. Mean perceived versus mean predicted severity ratings for sustained vowel by diagnostic category

Inspection of Figure 5 reveals that the CSID underestimated dysphonia severity across the diagnostic categories in pre- and posttreatment, resulting in an underestimation of the change in dysphonia severity (Figure 1). A nonsignificant GLM interaction test, however, showed that the relationship between the mean perceived and predicted ratings of dysphonia in sustained vowels was not dependent upon diagnostic category ( $F[1, 5] = 0.55, p = 0.74$ ).



### Effect of Dysphonia Severity

To assess the influence of possible spectrum effects on the performance of the CSID, a Generalized linear model (GLM) was computed. First, within each voice context, all pre- and posttreatment samples were separated into tertiles based upon the mean perceived ratings of severity. The interaction effect from this model tested the possible influence of varying levels of dysphonia severity on the performance of the CSID. A nonsignificant F-test associated with the interaction effect indicated that overall severity does not significantly influence the relationship between perceived severity ratings and predicted severity ratings in connected speech ( $F[1, 2] = 0.58, p = 0.56$ ). Thus, the CSID appears to perform similarly across the dysphonia severity spectrum in connected speech. Furthermore, in connected speech, there was also a nonsignificant F-test associated with the interaction effect that indicated the diagnostic category does not significantly influence the relationship between mean perceived severity and predicted severity ( $F[1, 5] = 0.62, p = 0.68$ ). However, the results indicated that there was a significant interaction effect associated with overall severity that influenced the relationship between the perceived severity and predicted severity ratings in sustained vowels ( $F[1, 2] = 6.22, p < 0.05$ ). That is to say, the performance of the CSID appeared to vary depending upon the perceived severity of the voice disorder. Furthermore, a significant  $F$  test associated with the interaction effect indicated that the diagnostic category significantly influenced the relationship between mean perceived severity and predicted severity in the sustained vowel context ( $F[1, 5] = 2.23, p = 0.05$ ). Within the diagnostic categories for the sustained vowel context, a significant interaction effect was observed for the diagnoses PMTD ( $F[1, 2] = 5.62, p < 0.05$ ) and mutational falsetto ( $F[1,$

2] = 6.35,  $p < 0.05$ ), indicating that the performance of the CSID varied depending upon the perceived severities of these voice disorders. Table 6 provides the  $F$  Values within each diagnostic category for connected speech and sustained vowel contexts, and Figures 6 and 7 display the estimated versus mean perceived dysphonia severity for both contexts across the tertiles. For nearly all severity groups within the two contexts, the acoustic estimates of severity underestimated mean listener perceived ratings. This finding was consistent within the diagnostic categories as well (Figures 3 and 5).

Due to these findings, an analysis of variance was computed for each tertile examining the relationship between the estimated and listener perceived severity in sustained vowels. The results indicated that there was a significant relationship at all levels of severity (i.e., tertiles), with the coefficient of determination becoming increasingly stronger with increased severity ( $F[1, 2] = 6.22, p < 0.05$ ). Within the “mild” tertile ( $F[1, 68] = 5.15, p < 0.05$ ), approximately seven percent of the variance was accounted for, whereas within the “moderate” tertile ( $F[1, 73] = 13.76, p < 0.05$ ), approximately sixteen percent of the variability was accounted for. The most variance

Table 6. Interaction effect of severity within each diagnostic category for connected speech and sustained vowel contexts

Diagnosis	Connected Speech		Sustained Vowel	
	$F$ Value	$p$	$F$ Value	$p$
UVFP	1.87	0.1695	0.96	0.393
ADSD	0.79	0.464	0.08	0.919
PMTD	0.41	0.664	5.62	0.008
BVFL	0.45	0.644	1.25	0.298
Presbylaryngis	0.34	0.713	0.71	0.501
Mutational Falsetto	0.74	0.491	6.35	0.009
Aggregate	0.58	0.559	6.22	0.002

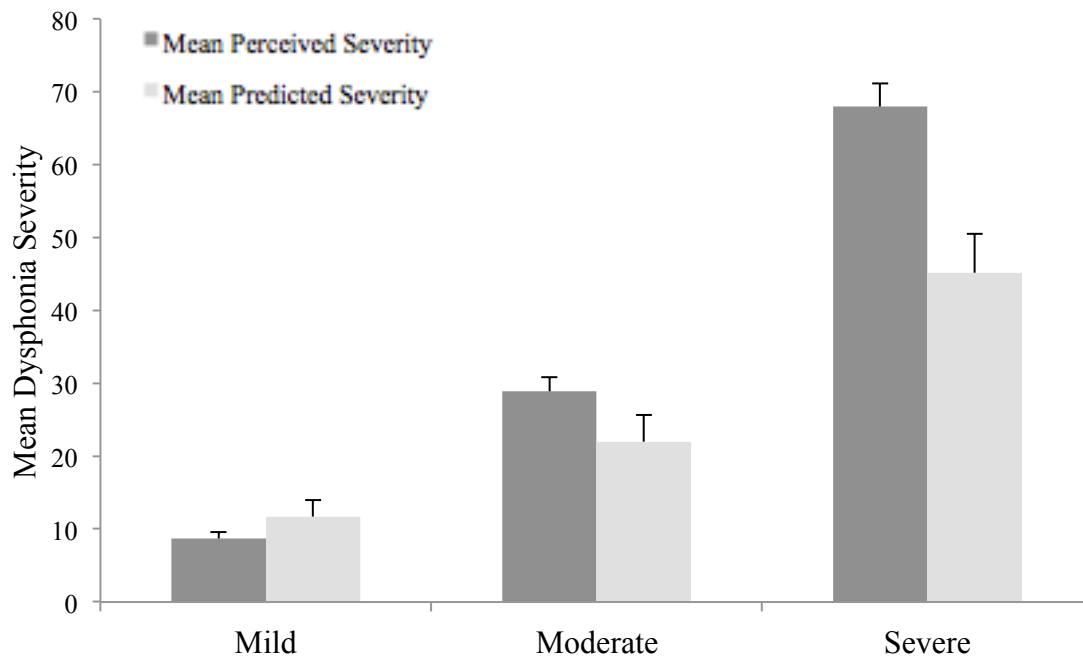


Figure 6. Mean acoustic severity estimates versus mean listener ratings per sample for all pre- and posttreatment samples combined for connected speech

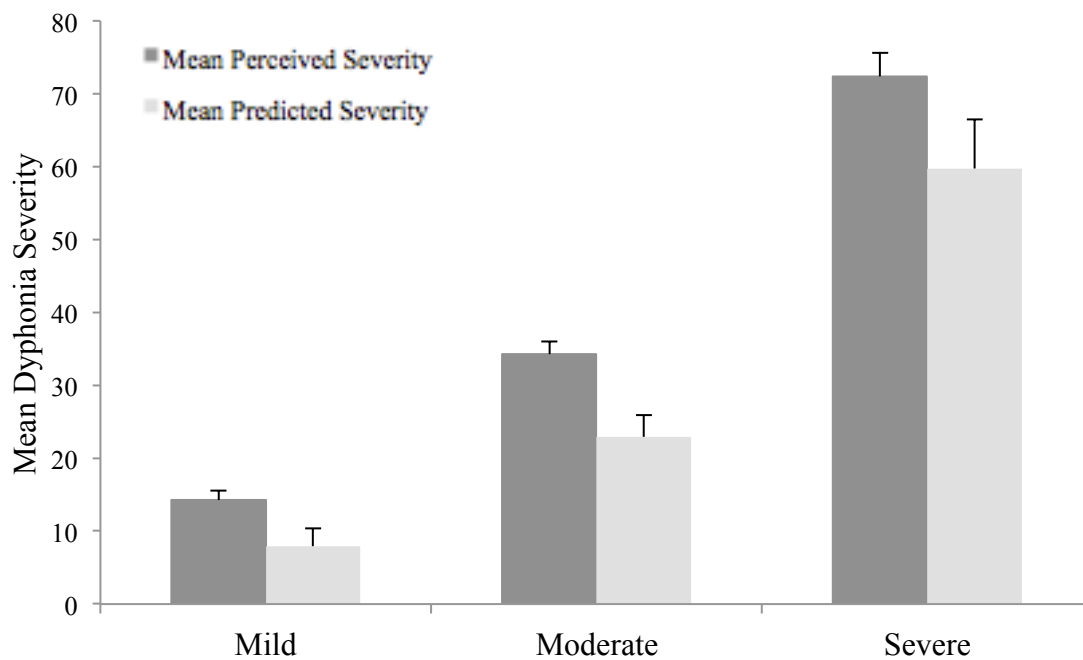


Figure 7. Mean acoustic severity estimates versus mean listener ratings per sample for all pre- and posttreatment samples combined for sustained vowel

was accounted for in the severe category, with a coefficient of determination of 0.49 ( $F[1, 75] = 71.63, p < 0.05$ ). Mean predicted ratings versus the mean perceived ratings for the entire sample set were then computed within the three tertiles separately for connected speech and sustained vowels. Pairwise  $t$  tests were computed between mean predicted and mean listener differences between mean listener ratings and mean acoustically predicted ratings across all tertiles ( $p < 0.05$ ); specifically, the CSID significantly underestimated mean listener ratings across all severities in connected speech and sustained vowels, with the exception of mild connected speech samples, where the CSID overestimated severity (Figures 6 and 7).

Further analyses examining the size of change in dysphonia severity was investigated in connected speech and sustained vowels separately to determine the sensitivity of the CSID across small, medium, and large changes in dysphonia severity (Figures 8 and 9). This was achieved by calculating mean change in dysphonia severity as judged by the listeners, and creating tertiles based upon the size of change. Pairwise  $t$  tests were then computed between mean predicted and mean listener ratings in the tertiles (small, medium, large) to determine the accuracy of the CSID in capturing varying sizes of change scores. While the CSID consistently underestimated mean change in dysphonia severity, differences between mean listener ratings and mean predicted change ratings across all tertiles in sustained vowels were not significant. This suggests that the CSID did not significantly underestimate small, medium, and large changes in dysphonia severity in the sustained vowel context. Results also indicated nonsignificant differences between mean listener ratings and mean perceived ratings in the tertiles representing small and medium changes in dysphonia severity in connected speech; however, results

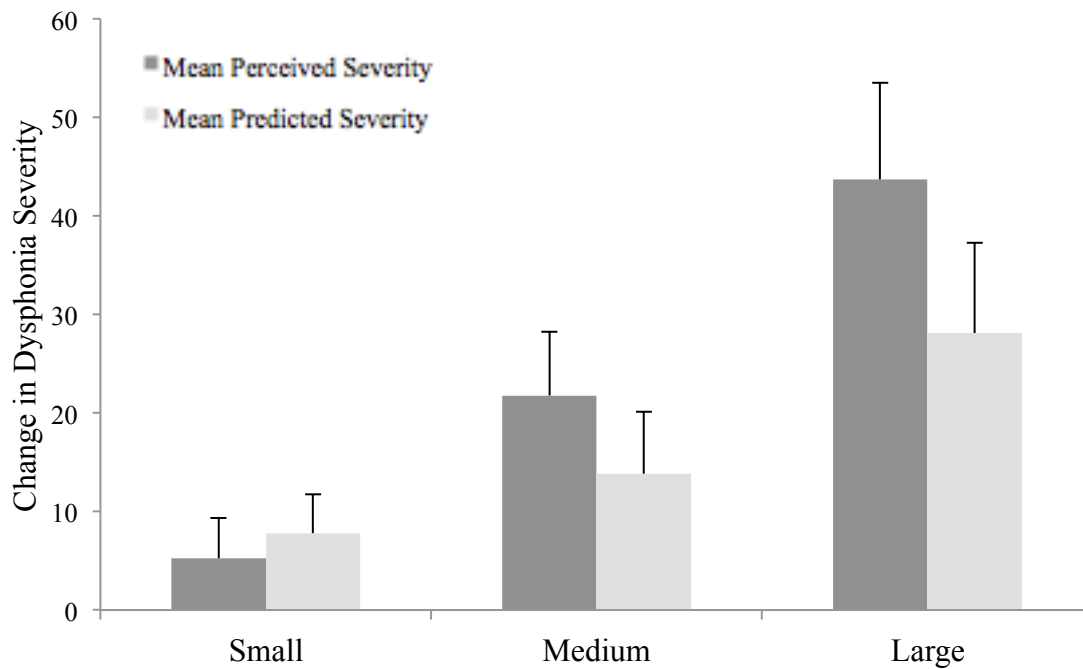


Figure 8. Mean perceived versus predicted size of change in dysphonia severity for connected speech

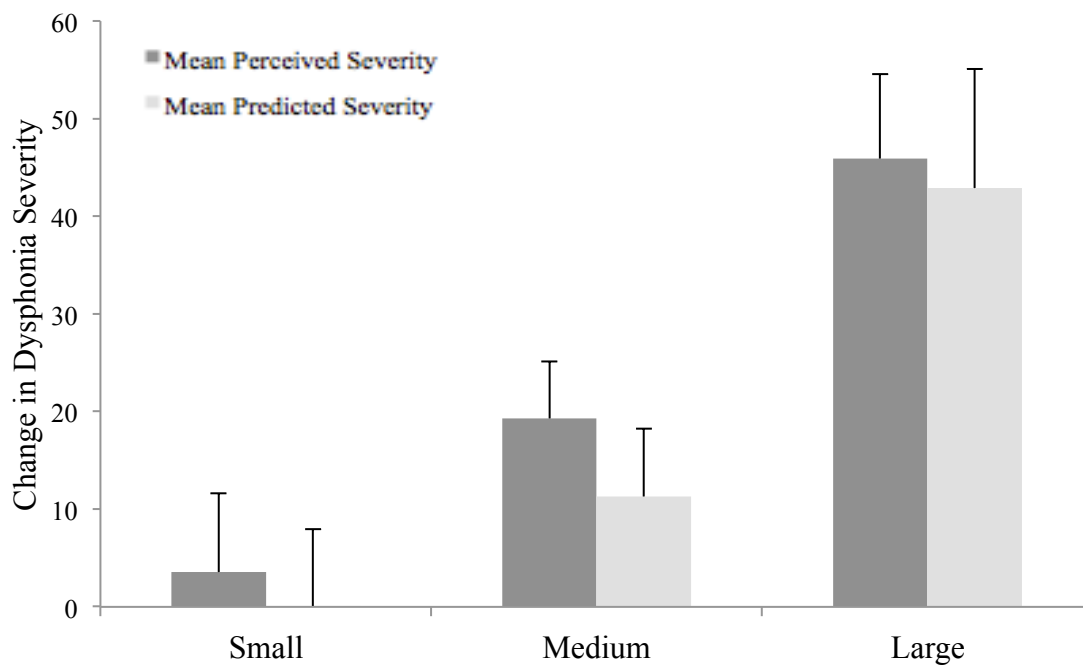


Figure 9. Mean perceived versus predicted size of change in dysphonia severity for sustained vowel

indicated significant differences between mean predicted and mean perceived change in the tertile representing large change ( $t[37] = 2.28, p < 0.05$ ). Specifically, the acoustically predicted change significantly underestimated the mean perceived change in large changes of dysphonia severity pre- to posttreatment (Figure 8).

## DISCUSSION

The primary purpose of this study was to assess the validity of the CSID, an acoustic analysis algorithm comprised of spectral and cepstral-based measures, as a possible objective treatment outcomes measure across a range of diagnostic categories. Although some variations existed, the results of this study confirmed that the CSID generally provides valid estimates of dysphonia severity. Furthermore, the CSID appears to be sensitive to changes in dysphonia severity following intervention, as was seen when comparing the mean perceived and acoustically estimated change scores, wherein the difference was often less than 10 millimeters (Figure 1). On the visual analog scale that was used by the listeners to rate the voice samples, this would translate into less than one centimeter of difference, indicating a very high level of correspondence between the change perceived by the listeners as compared to the CSID. In fact, there was no significant difference between the magnitude of change following treatment perceived by listeners as compared to the magnitude of change estimated by the CSID, regardless of diagnostic category or voice context (sustained vowel versus connected speech). Thus, the CSID's sensitivity to change before and after management supports its potential utility as an objective treatment outcomes measure. The performance of the CSID varied somewhat on the basis of voice context (sustained vowel versus connected speech), diagnostic category, and severity spectrum. In the following section, these factors will be discussed in more detail.

## Performance of the CSID: Connected Speech

### Versus Sustained Vowels

The results indicated that when all samples were aggregated across pre- and posttreatment, predicted severity ratings were more highly correlated with listener ratings within a sustained vowel context ( $r = 0.84, p < 0.05$ ) versus connected speech ( $r = 0.72; p < 0.05$ ). Thus, the results of the current study corroborate previous findings regarding the overall strength of the relationships between the perceived and predicted dysphonia severity ratings in sustained vowels versus connected speech (Awan et al., 2010). Specifically, Awan et al. (2010) also found that the correlation observed within the sustained vowel context was superior to connected speech. However, the results of this study also broaden the utility of spectral/cepstral measures, and the CSID, as it was a strong predictor of change in dysphonia severity in connected speech across a range of diagnostic categories.

Although the relationships between pretreatment, posttreatment, and change in dysphonia severity were generally strong across the diagnostic categories in connected speech and sustained vowels (Tables 4 and 5), the CSID consistently had a lower dysphonia rating when compared to listener ratings (Figures 3 and 5). This underestimation may be due to a number of reasons. First, predicted underestimation of the mean listener ratings of severity may be due to “end effects” in listener ratings. That is, listeners had a tendency to rate severely dysphonic samples as profoundly abnormal, or 100 on the VAS (Figures 2 and 4), whereas the CSID estimates have been reported to more closely approximate a more normal distribution (Awan et al., 2009). Another possible reason for underestimation of mean listeners ratings may be attributed to a



difference in scaling (Awan et al., 2009). Listeners were bound to rate dysphonia samples between 0 (“Normal”) and 100 (“Profoundly Abnormal”) on the VAS, while the acoustic estimates of severity could be rated negatively (i.e., less than 0) for highly periodic voices or greater than 100 for extremely dysphonic samples. As was seen in Figures 2 and 4, many posttreatment samples had negative predicted severities, however, listeners could only rate samples they found to be close to normal as “0”. This scaling difference may have also impacted the strength of the correlation between mean perceived and estimated ratings of severity. Although the scaling difference may have negatively influenced the strength of the relationship between perceived and predicted ratings of severity, the fact that the CSID does not have absolute boundaries may be effective in identifying subtle changes in mild or severe dysphonia that may not be perceptible to some trained listeners. While the predicted ratings underestimate the mean listener ratings of severity, the fact that the estimates are based upon a normal distribution and without absolute boundaries may actually be desirable.

While the CSID underestimated severity as compared to listener ratings within the contexts of connected speech and sustained vowels, it particularly underestimated severity within the context of connected speech. This underestimation resulted in the overall relationship between perceived and predicted severity to be weaker in connected speech than in sustained vowels. Connected speech is more complicated to analyze than sustained vowels, and as such, contributes more error variance to regressions derived from predictions of perceived dysphonia severity. In dysphonic samples, it is challenging to successfully separate vowels from consonants due to the aperiodicity that may be present. The CSID has been automatically programmed to use zero decibels (dB) as a cut

off for removal for all signals' normalized CPP values to assist in eliminating any aperiodic signals of low amplitude (i.e., a CPP would not be included if it had a value that was lower than expected based upon subsequent linear regression analyses). A 0dB cutoff may have caused various CPPs to be excluded in a severely dysphonic sample when they should have been included in the analysis (Awan et al., 2010). Although the 0dB remains a cutoff during the analyses of sustained vowels, there is no need to separate vowels from consonants, and as such, all of the sustained vowel samples were acoustically analyzed and were not affected by the 0dB threshold. Future research examining the adjustment of the 0dB threshold and its effect on the relationship between predicted and perceived severities of dysphonia in connected speech is warranted, as continuous speech samples are more ecologically valid measure of change in dysphonia severity than sustained vowels.

#### Performance of CSID Across Diagnostic Category

In addition to the CSID being sensitive to changes in PMTD as established previously by Awan & Roy (2009), the results indicated that the acoustic index is sensitive to change across the other five diagnostic categories tested (Tables 4 and 5). However, the strength of these relationships varied. Of the six diagnostic categories, the diagnosis that resulted in the weakest correlations with perceived severity changes pre- and posttreatment samples in connected speech and sustained vowel was presbylaryngis. These moderately strong correlations may possibly be attributed to two causes. The first is that there may be idiosyncratic characteristics of voices associated with presbylaryngis that are not adequately captured by the CSID measure, such as pitch breaks combined

with strain or effort due to decreased respiratory strength (Sauder, Roy, Tanner, Houtz, & Smith, 2010). Two voice disorders commonly associated with strain, presbylaryngis and ADSD, were both predicted to be more mildly disordered by the CSID than listeners perceived (Figures 3 and 5). While strain is perceptually salient and a change in dysphonia was perceived pre- to posttreatment for these two diagnoses, it may not be acoustically salient, as the CSID predicted modest change in dysphonia severity (Figure 1). A second reason for the moderately strong correlations seen in presbylaryngis samples may be the previously discussed “end effect” in listener ratings (Awan & Roy, 2009; Awan et al., 2009; Awan et al., 2010). Specifically, listeners perceived severity ratings of presbylaryngis connected speech samples pre- and posttreatment to be more mild, which resulted in limited variability in the ratings, and thereby reduced correlation coefficients. These findings support the need for clinical observations in addition to acoustic measures. When combined with other clinical observations, the CSID may be able to provide an objective measure of dysphonia severity that will assist in quantifying a dysphonic voice and its change following treatment.

This idea of an “end effect” or a restricted range of ratings resulting in a reduced, and at times nonsignificant, correlation coefficient can be seen in the posttreatment correlations for PMTD (Tables 4 and 5). Many of the PMTD posttreatment samples for connected speech and sustained vowel were rated close to normal by the listeners (i.e., close to “Normal Voice” on the VAS). The absence of variability in posttreatment ratings likely attenuated the ability to detect associations, and contributed to the non-significant correlation coefficients (Figures 3 and 5). Although the posttreatment PMTD findings were nonsignificant, the correlations observed for the change in dysphonia severity were

strong for both connected speech and sustained vowel. This indicates that the CSID is sensitive to change in dysphonia severity for PMTD, which truly measures the effect of intervention.

In addition to weaker associations observed with change in dysphonia severity in presbylaryngis patients, a moderately strong relationship in the change between pre- and posttreatment connected speech samples of patients with mutational falsetto was also observed. While the relationship reflecting the change in dysphonia severity between estimated and perceived ratings of dysphonia severity in sustained vowels were markedly stronger for mutational falsetto ( $r = 0.84, p = 0.001$ ), the CSID underestimated the change in dysphonia severity for connected speech (Figure 1). The CSID also noticeably underestimated dysphonia severity in pre- and posttreatment connected speech samples (Figure 3). These findings may be attributed to the fact that the CSID connected speech algorithm does not include gender as a factor, and mutational falsetto is a disorder found in male adolescents characterized by a high pitch voice. Although gender was not found to be a significant contributor to the model for connected speech in previous studies, none of the samples in the study had a diagnosis of mutational falsetto (Awan et al., 2010). Allowing for an additional, optional, gender variable to be included in quantifying mutational falsetto dysphonia may allow for a stronger relationship between pre- and posttreatment sample change. Another factor that possibly influenced the correlation coefficients may have been the small sample size, as only 12 patients were available for analyses. Future studies with larger sample sizes must be completed to investigate the significance of gender as a contributor to this rather idiosyncratic diagnostic category.

### Performance of the CSID Across the Severity Spectrum

Although the relationships between perceived and predicted severity in sustained vowels and connected speech were strong, awareness of the influence of severity and diagnosis upon the CSID's ability to estimate dysphonia is essential. The results of this study indicated that there is an effect of severity on the CSID's ability to estimate dysphonia severity. Specifically, the results indicated that there was a significant interaction effect associated with overall severity that influenced the relationship between the perceived severity and predicted severity ratings in sustained vowels, while there was no significant interaction within the context of connected speech. This difference in significance may be due in part to the addition of acoustic variables in the analysis of sustained vowels, as the CPP standard deviation and gender are included in estimates of sustained vowel severity. While these variables were found to be significant contributors to sustained vowel estimates, they may affect the performance of the CSID across the dysphonia severity spectrum. Further research is necessary to determine if the inclusion of these variables in the sustained vowel model results in the dysphonia estimation being dependent upon the severity rating.

Posthoc analyses were conducted to ascertain which tertile was influencing the CSID's ability to predict dysphonia severity in sustained vowels. The results indicated that for sustained vowels, there was a significant relationship between the perceived and predicted ratings within all three tertiles, but within the mild tertile, the coefficient of determination was the lowest, with listener severity accounting for seven percent of the variation in estimated sustained vowels. Further posthoc analyses demonstrated that across all the tertiles for connected speech and sustained vowels, the CSID

underestimated listener ratings of severity significantly (Figures 6 and 7). As previously discussed, this underestimation may be due to a scaling difference, as the CSID is not bound to rating a sample between 0 and 100 (Awan et al., 2009). While there was a significant relationship between the estimated and perceived ratings in sustained vowels across all categories, the spectrum effects may warrant a reexamination of the regression model used to calculate estimated severity in sustained vowels. This reexamination should include a larger and more diverse set of dysphonic samples, specifically mild dysphonia, as this level of severity had the lowest coefficient of determination. If the future model continues to have a strong relationship with listener ratings but is not influenced by the severity of the sample, then it will be more sensitive to changes in dysphonia severity than the current model contained in the ADSV. Nevertheless, the results of this study show a strong overall relationship between changes in dysphonia severity in sustained vowels ( $r = .836, p < 0.0001$ ).

The results also indicated that in sustained vowels, the diagnosis influenced the placement of a sample into the tertiles, which in turn influenced the CSID's ability to estimate severity. This may be due to the relative distribution of severity within the diagnostic categories in sustained vowels. The disorders that were perceived as having the most change (Figure 1), PMTD and mutational falsetto, were also the diagnoses whose estimated severity for sustained vowels was dependent upon the diagnostic category. These diagnoses' mean pretreatment samples were rated as severe dysphonia, and as these diagnoses were particularly responsive to treatment, mean posttreatment ratings averaged as being mild (Figure 5). As such, these diagnoses represented 37% (26/70) of the total samples in the sustained vowel "mild" category, over representing the

expected one third of the samples. However, in connected speech, where the diagnosis did not influence the placement of a sample into the tertiles, these diagnoses represented only 29% (21/73) of the total samples in the “mild” category. This over-representation in sustained vowels may be the reason that diagnostic categories have an influence on the estimated severity of the sustained vowels, particularly in the “mild” tertile. It is believed that if the performance of the CSID in estimating severity of sustained vowels was no longer dependent upon the perceived severity ratings (i.e., tertiles), there would cease to be an influence of diagnostic category. However, further research must be undertaken to determine this relationship. These issues notwithstanding, there was a strong and significant relationship in the change in severity between perceived and predicted ratings in sustained vowels in mutational falsetto and in PMTD.

Additional posthoc analyses also revealed that while the CSID underestimated changes in dysphonia severity pre- to posttreatment, it did not do so significantly in sustained vowels, and it is sensitive to small, medium, and large changes in dysphonia severity in this context (Figure 9). Furthermore, while the CSID continued to underestimate change in dysphonia severity in connected speech, results indicated non-significant differences between mean listener ratings and mean perceived change ratings in the tertiles representing small and medium changes in dysphonia severity (Figure 8). However, within the context of connected speech there was a significant difference between the mean change scores in voice samples that experienced large changes in dysphonia severity. This may be additive, as the CSID nonsignificantly underestimated medium changes in dysphonia severity compared to listener ratings; however, it more dramatically and significantly underestimated large changes in dysphonia severity. The

acoustic tool's sensitivity to small and medium changes in dysphonia severity in connected speech and sustained vowel contexts speaks to its utility and validity, as small-to-medium changes in dysphonia are more difficult to characterize perceptually, particularly for novice clinicians. In addition, changes in dysphonia severity are not often large (Figures 3 and 5), and as such, an objective assessment of dysphonia severity must be sensitive to more conservative changes in dysphonia. Thus, while the acoustic index significantly underestimates large changes in dysphonia severity in connected speech, its performance in small and medium changes in dysphonia severity is essential in an objective treatment outcomes measure, and further validates the CSID.



## CONCLUSION

The effects of treatment have traditionally been determined through auditory-perceptual ratings; however, an objective measure that accurately quantifies change in dysphonia severity would be particularly useful in demonstrating the effects of intervention. The CSID, a spectral/cepstral-based acoustic measure contained within the ADSV, possesses a strong relationship with mean perceived ratings of severity within the contexts of connected speech and sustained vowels across a range of diagnostic categories and varying degrees of dysphonia severity. As such, it offers great promise as a means of objectively quantifying dysphonia severity and serving as an objective treatment outcomes measure. Although the influence of severity may at first appear to limit the clinical utility of the CSID, it rather supports the utility of the acoustic index as a means of reinforcing, quantifying, and supplementing trained listener ratings of dysphonia severity. The CSID's sensitivity to small, medium, and large changes in dysphonia severity further validates its utility. Future studies with larger samples of dysphonia severity are needed to further address and resolve the spectrum effects in sustained vowel samples.

## APPENDIX

The diagnostic categories chosen for the purposes of this study were selected for three reasons. First, these pathologies are frequently encountered in multidisciplinary voice clinics. Thus, if the ADSV proves to be sensitive to treatment outcomes, the ADSV will be useful in assessing dysphonia severity in the most commonly encountered diagnostic categories. Second, these diagnostic categories are characterized by heterogeneous and possibly idiosyncratic voice qualities and severities. Finally, the diagnostic categories included in this study are those that respond to intervention, whether it is medical, surgical, or behavioral.

### Unilateral Vocal Fold Paralysis

Unilateral vocal fold paralysis (UVFP) is generally caused by nerve damage or lesions to the recurrent laryngeal nerve. In UVFP, one vocal fold functions normally, whereas the other is paralyzed in either an adducted and abducted position. Perceptually, the most common signs and symptoms are breathiness and hoarseness. Intervention for UVFP may be surgical or behavioral; both forms of intervention have been found to be effective (Colton, Casper, & Leonard, 2006).

### Adductor Spasmodic Dysphonia

There are two types of spasmodic dysphonia, abductor spasmodic dysphonia (ABSD) and adductor spasmodic dysphonia (ADSD). Speech samples in this study will

only include ADSA, the most common variant. The etiology of ADSA is unknown, however, it is believed to have a neurological basis. The onset has been associated with traumatic emotional events, major upper respiratory infections, and other unknown causes (Colton et al., 2006). The perceptual signs and symptoms associated with ADSA include strain and struggle, voice stoppages, harshness, hoarseness, and tremors. Adductor Spasmodic Dysphonia is typically treated medically with intrafold Botox injections. Recently, surgical procedures such as selective denervation-reinnervation procedures have shown promise (Colton et al., 2006).

### Primary Muscle Tension Dysphonia

The etiology of muscle tension dysphonia (MTD) is unknown, however, there are many known sources of excessive and dysregulated laryngeal muscle activity. Technical misuse of the vocal mechanism, learned adaptations following an upper respiratory infection, extreme compensation for an underlying vocal fold pathology, increased laryngeal tone, psychological and/or personality factors all may lead to dysregulated laryngeal activity. Perceptually, MTD is characterized by assorted voice qualities and severities. Behavioral intervention is the primary form of treatment for MTD, and within one therapy session, many patients may experience normal voice (Colton et al., 2006).

### Benign Vocal Fold Lesions

Benign vocal fold lesions (BVFL) is an umbrella term for many different laryngeal lesions, each with their own effect on the vibratory characteristics of the vocal folds. Benign vocal fold lesions encompass the following: vocal fold nodules, vocal fold

polyps, vocal fold cysts, and Reinke's Edema. Vocal fold nodules are caused by loud voice, and are often found in women due to vocal fold size, fundamental frequency, and lower levels of hyaluronic acid. Perceptual characteristics of vocal fold nodules include hoarseness, breathiness, and vocal strain. Intervention for nodules is typically behavioral. Vocal fold polyps are typically caused by a single phonotraumatic event, although chemical exposure, such as cleaning solvents, may also be a cause. Perceptually, polyps are characterized by breathiness, hoarseness, and vocal strain. Polyps may resolve on their own; however, if they do not, surgical and/or behavioral intervention may be necessary. Vocal fold cysts have no definitive known cause, and are perceptually characterized by hoarseness and breathiness. Intervention for vocal fold cysts is typically surgical, and behavioral intervention may play a role postsurgery. Reinke's Edema is characterized by a collection of water or swelling in Reinke's space. The primary risk factor is smoking, although there may be other contributing factors. Voices of individuals with Reinke's Edema are characterized by hoarseness and effortful speech, and intervention is typically surgical combined with a smoking cessation program (Colton et al., 2006).

#### Mutational Falsetto

Mutational falsetto is believed to be a functional and/or psychogenic voice disorder typically occurring in male adolescents, and is thought to be caused by a variety of factors, such as resisting the changes of puberty. The disorder is characterized by a high pitch voice (often breathy), which may reflect the individual's preadolescent pitch

(Berry, 1991). Intervention for mutational falsetto is most often behavioral, and occasionally surgical.

### Presbylaryngis

Presbylaryngis may be found in individuals with good health who are over 65 years of age. Presbylaryngis is caused by aging, and may be attributed to thinning of muscle and tissue in the larynx. This thinning results in bowed vocal folds. Perceptually, presbylaryngis is characterized by a breathy voice, with pitch breaks, and reduced loudness. Intervention for presbylaryngis is behavioral and/or surgical (Colton et al., 2006).

## REFERENCES

- Awan, S. N. (2011). *Analysis of dysphonia and voice (ADSV): An application guide*. Montvale, NJ: KayPENETAX.
- Awan, S. N., Helou, L. B., Stojadinovic, A., & Solomon, N. P. (2011, April). Tracking voice change after thyroidectomy: application of spectral//cepstral analyses. *Clinical Linguistics & Phonetics*, 25(4), 302-320.
- Awan, S.N., & Roy, N. (2005). Acoustic prediction of voice type in adult females with functional dysphonia. *Journal of Voice*, 19, 268-282.
- Awan S.N., & Roy, N. (2006). Toward the development of an objective index of dysphonia severity: A four-factor model. *Clinical Linguistics & Phonetics*, 20(1), 35-49.
- Awan, S. N., & Roy, N. (2009). Outcome measures in voice disorders: Application of an acoustic index of dysphonia severity. *Journal of Speech, Language, and Hearing Research*, 52, 482-499.
- Awan, S. N., Roy, N., & Dromey, C. (2009). Estimating dysphonia severity in continuous speech: Application of a multiparameter spectral/cepstral model. *Clinical Linguistics & Phonetics*, 23(11), 825-841.
- Awan, S. N., Roy, N., Jette, M., Meltzner, G. S., & Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgments from the CAPE-V. *Clinical Linguistics & Phonetics*, 24(9), 742-758.
- Baken, R.J. (1987). *Clinical measurements of speech and voice*. Boston, MA: Little, Brown and Co.
- Berry, Q. C. (1991, July). Psychosocial aspects of adolescent dysphonia: An approach to treatment. *Language, Speech, and Hearing Services in Schools*, 22, 163-167.
- Callan, D.E., Kent, R.D., Roy, N., & Tasko, S.M. (1999). Self-organizing maps for the classification of normal and disordered female voices. *Journal of Speech and Hearing Research*, 42, 355-366.

- Colton, R. H., Casper, J. K., & Leonard, R. (2006). *Understanding voice problems: A physiological perspective for diagnosis and treatment* (3rd ed.). Baltimore, MD: Lippincott Williams & Wilkins.
- Fairbanks, G. (1960). *Voice and articulation drillbook* (2<sup>nd</sup> Ed.) New York, NY: Harper & Row.
- Frattali, C. M. (1998). *Measuring outcomes in speech-language pathology*. New York, NY: Thieme.
- Hillenbrand, J., & Houde, R.A. (1996). Acoustic correlates of breathy voice quality: dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing, 39*, 298-310.
- Hillenbrand, J., Cleveland, R.A., & Erickson, R.L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*, 769-778.
- Kempster, G.B., Gerratt, B.R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R.E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*, 124-132.
- Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *Journal of the Acoustical Society of America, 126*(5), 2619-2634.
- Sauder, C., Roy, N., Tanner, K., Houtz, D. R., & Smith, M. E. (2010, July). Vocal function exercises for presbylaryngis: A multidimensional assessment of treatment outcomes. *Annals of Otolaryngology, Rhinology & Laryngology, 119*(7), 460-467.
- Watts, C. R., & Awan, S. N. (2011, December). Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech, Language, and Hearing Research, 54*, 1525-1537.
- Wolfe, V., & Steinfatt, T.M. (1987). Prediction of vocal severity within and across vocal types. *Journal of Speech and Hearing Research, 30*, 230-240.