Optimizing Acoustic and Perceptual Assessment of Voice Quality in Children with Vocal Nodules

by

Asako Masaki

S.B. Electrical Science and Engineering
Massachusetts Institute of Technology, 2001

Submitted to
The Harvard-MIT Division of Health Sciences and Technology
In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Speech and Hearing Biosciences and Technology

At the
Massachusetts Institute of Technology
September 2009

Signature of Author.....
Harvard-MIT Division of Health Sciences and Technology
August 10, 2009

Certified by .....
Robert E. Hillman, Ph.D.
Associate Professor of Surgery & Health Science & Technology, Harvard Medical School
Thesis Supervisor

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

Optimizing Acoustic and Perceptual Assessment of Voice Quality in Children with Vocal Nodules

by

Asako Masaki

Submitted to the Division of Health Sciences and Technology
on August 10, 2009 in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Speech and Hearing Biosciences and Technology

Abstract

Few empirically-derived guidelines exist for optimizing the assessment of vocal function in children with voice disorders. The goal of this investigation was to identify a minimal set of speech tasks and associated acoustic analysis methods that are most salient in characterizing the impact of vocal nodules on vocal function in children. Hence, a pediatric assessment protocol was developed based on the standardized Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) used to evaluate adult voices. Adult and pediatric versions of the CAPE-V protocols were used to gather recordings of vowels and sentences from adult females and children (4-6 and 8-10 year olds) with normal voices and vocal nodules, and these recordings were subjected to perceptual and acoustic analyses. Results showed that perceptual ratings for breathiness best characterized the presence of nodules in children's voices, and ratings for the production of sentences best differentiated normal voices and voices with nodules for both children and adults. Selected voice quality-related acoustic algorithms designed to quantitatively evaluate acoustic measures of vowels and sentences, were modified to be pitch-independent for use in analyzing children's voices. Synthesized vowels for children and adults were used to validate the modified algorithms by systematically assessing the effects of manipulating the periodicity and spectral characteristics of the synthesizer's voicing source. In applying the validated algorithms to the recordings of subjects with normal voices and vocal nodules, the acoustic measure tended to differentiate normal voices and voices with nodules in children and adults, and some displayed significant correlations with the perceptual attributes of overall severity of dysphonia, roughness, and/or breathiness. None of the acoustic measures correlated significantly with the perceptual attribute of strain. Limitations in the strength of the correlations between acoustic measures and perceptual attributes were attributed to factors that can be addressed in future investigations, which can now utilize the algorithms that were developed in this investigation for children's voices. Preliminary recommendations are made for the clinical assessment of pediatric voice disorders.

Thesis Supervisor: Robert E. Hillman, Ph.D.
Title: Associate Professor of Surgery and Health Science and Technology, Harvard Medical School

Acknowledgments

I don't know where I can begin in listing all the people who have helped me get to the point I am now. First of all, this work would not have been possible without the wonderful support and guidance from my wonderful advisor and mentor, Dr. Hillman, for being very very patient with me and guiding me in the right direction. I can't thank you enough for helping me get into this program. Thank you soo much for editing this thesis a million times. I would also like to whole heartedly thank my Thesis committee chair, Dr. Quatieri, for helping me implement and understand all my signal processing woes. I also wanted to thank my other committee members Dr. Stevens and Dr. Nuss for being kind enough to agree to be on my committee and taking time out of their busy lives to give me guidance. I am also very grateful for the assistance of Dr. Massaro for helping me with my statistical analysis.

Thank you too for everybody at the MGH Voice Center for all of their wonderful support and encouragement. I would especially like to massively thank Daryush for always being there to answer my many numerous MATLAB and signal processing questions. Thank you for massively helping me! I very very much appreciated all the hours we spent talking about my problems. And of course, what would life be at the lab without Cara? It would be a very sad world. Thank you thank you Cara for putting up with my emotional woes and trying to help me become a stronger confident person. You're the best! Thank you thank you Prakash too for stopping by our lab and encouraging us to continue to work harder. I very much appreciate all the insight you shared with me about life. Thank you thank you thank you Janet, for also imparting me wisdom and letting me interrupt your busy work to ask you random acoustic signal processing problems. Thank you thank you.

I would also like to thank my wonderful dear husband, Dan, for always being very supportive and encouraging me every step of the way even when I was depressed beyond belief. Without you, I am not sure how I would have survived the insanity of trying to write my thesis! I seriously cannot say or write enough about you to show you how much I appreciate you being in my life. Thank you soo much too for letting me be a part of your loving family! I would especially like to thank: Leo, Judy, Jennifer, Jaime, Peter, Grandma and Grandpa Sowatzke and Grandma and Grandpa Wehner for your constant support and love!

To my big sister Jocelyn, thank you for being the bestest girlfriend in the whole wide world and always being there to cheer me up and being more supportive than I could ever ask for. I can't count how many times you dropped everything to make sure I was okay ☺ Thank you soo soo much to for your wonderful family Ken, Lucille, Evan and Travis for adopting me into your family and being unbelievably supportive. I couldn't have asked for a better family! Thank you again to everyone for your unending emotional and academic support and guidance. I can't tell you how much I appreciate your thoughtfulness.

Thank you soo much too to my dear friends, Alison, Xiao and Jane. Thank you for believing in me and giving me the confidence to continue working hard. I would also like to thank Geri, for trying to instill some of her expertise in working with children. Thank you thank you to everybody else I didn't specifically mention for all of your love and support!

List of Figures

List of Tables

Contents

Chapter 1: Introduction and Background

1.1 Introduction

Clinical evaluation of voice production in children is more challenging than in adults. During the voice assessment process, children often have difficulty cooperating because of their limited attention span, reduced comprehension, immature speech and language skills, and a tendency to be more apprehensive about unfamiliar procedures. As a result, the type and amount of information that can be collected from children is limited, making it imperative that voice evaluation procedures be brief and focus on the most salient tasks for accurate assessment (Andrews & Summers, 1988). Currently, few empirically-derived guidelines exist for optimizing the assessment of vocal function in children with voice disorders.

As part of a typical voice evaluation procedure, children undergo invasive endoscopic imaging to obtain a definitive assessment of their laryngeal status. However, endoscopic imaging often provides an incomplete picture of the vocal structures and function due to children's tendency to have reduced cooperation during the procedure. The traumatic nature and reduced effectiveness of pediatric endoscopy provides motivation for efforts to develop non-invasive (non-traumatic) vocal function measures. Such measures could be used to reliably assess and document vocal function in order to supplement endoscopic imaging. For example, after making a diagnosis using endoscopic imaging, subsequent use of non-invasive measures to frequently monitor vocal status would prevent children from being subjected to repeated endoscopic examinations, particularly in cases in which the pathology is non-life threatening.

Auditory perceptual and acoustic measures are the most commonly used non-invasive methods for assessing voice disorders. Extensive research on the development of acoustic measures for evaluating voice disorders and vocal function in adults has shown mixed results (Eadie & Doyle, 2005; Parsa & Jamieson, 2001). Compared to the large number of studies with adults, only few studies have focused on developing acoustic measures in children (Lotz, D'Antonio, Chait, & Netsell, 1993). Furthermore, some acoustic algorithms used to assess vocal function in adults may not be suitable for use

with children's voices due to differences in the basic acoustic properties of children and adult voices. For example, the higher fundamental frequency of children's voices creates sparser spectral sampling, which negatively impacts the performance of many algorithms that depend on adequate resolution of vocal tract formants. Systematic modification and verification of existing acoustic analysis algorithms designed for use with adults is necessary to optimize the use of these acoustic algorithms for assessing vocal function in children.

At a minimum, acoustic measures of voice production should correlate with auditory perceptual characteristics of voice to be clinically useful (Eadie & Doyle, 2005). This is also in keeping with the fact that auditory perceptual judgments are already used routinely by clinicians to assess voice disorders (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993; McAllister, 1997). To be clinically useful, acoustic measures should account for more than fifty percent of the variance of the perceptual ratings (i.e., $r^2 \geq 0.5$). Statistically, the correlation coefficient, r^2, reveals the proportion of variance in the dependent variable that is predictable from the independent variable. Preferably, acoustic measures should also correlate with underlying vocal fold physiology. However, given the challenges associated with directly studying vocal fold physiology in children, a first step focuses on establishing relationships between the most salient acoustic and perceptual measures for pediatric voice evaluation.

The Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) has been developed and adopted by the American Speech-Language-Hearing Association (ASHA) as the recommended standard protocol for perceptual assessment of disordered voices (http://www.asha.org/about/membership-certification/divs/div_3.htm). A copy of the CAPE-V form is provided in Appendix A for reference. Administering the CAPE-V protocol involves having the patient read six sentences that are designed to elicit different laryngeal behaviors. Adaptation of the CAPE-V protocol for use with pre-literate children requires the development of a comparable protocol that is specifically designed to accommodate their immature speech, language, and literacy skills. In addition, determining which types of speech materials (e.g., sustained vowels, sentences, varying phonemic contexts, etc.) are best suited for eliciting the salient characteristics of voice

disorder in children is also critical for developing effective and time-efficient auditory perceptual and acoustic assessment protocols for children.

Pediatric vocal mechanisms are not simply smaller versions of adult mechanisms. Differences in the morphology and histological structure of adult (mature) and child (developing) vocal folds may produce corresponding differences in the underlying biomechanics of vocal fold vibration (Hirano, Kurita, & Nakashima, 1983). Differences in child and adult vocal mechanisms may not only influence normal phonation, but may also differentially affect the impact that vocal pathology has on voice production. Therefore, different acoustic and perceptual measures may better characterize disordered voice production in children compared to adults.

The studies in this thesis focus on voice production in subjects with vocal fold nodules because the presence of vocal fold nodules is one of the most common voice disorders affecting children. Vocal nodules are callous-like bumps that form on the medial surface of one or both vocal folds, typically at a mid-membranous location. During phonation, the presence of nodules can cause irregularities in vocal fold vibration (perturbation) and prevent normal glottal closure (turbulence noise from air escape), resulting in a dysphonic-sounding voice.

A primary goal of this thesis is to identify a minimal set of speech tasks and associated acoustic analysis methods that are most salient in characterizing the impact of vocal nodules on vocal function in children. Accordingly, a pediatric assessment protocol based on the ASHA-sanctioned CAPE-V protocol was developed. The pediatric CAPE-V protocol was specifically designed to address the challenges associated with pediatric voice assessment. Auditory-perceptual listening studies were used to determine which speech material elicited the largest perceptual differences between normal voices and voices with vocal nodules, and to test hypotheses about the differential impact of vocal nodules on voice production in children versus adults. To complement the perceptual studies, selected voice quality-related acoustic analysis approaches previously used to evaluate voice disorders in adults, were optimized for use in children. Results from the acoustical analyses were correlated with data obtained from the perceptual

studies to identify the most clinically relevant measures. The results of this work should help improve clinical strategies for assessing voice disorders in children.

The outline of the thesis is as follows. In the remainder of Chapter 1, the differences in vocal mechanisms between the developing child and mature adult are discussed. A short description of the etiology and patho-physiology of vocal fold nodules and various approaches for acoustic and perceptual assessment of voice disorders in adults and children are also presented. Chapter 1 ends by describing the project goals of this thesis. In Chapter 2, the methods and results of two perceptual experiments investigating whether the production of vowels or sentences best differentiates normal voices and voices with nodules for children and adults, and whether the sentences created for children elicit similar perceptual attributes as the sentences in the CAPE-V protocol is described. In Chapter 3, description of the acoustic measures used for analyzing the production of vowels and sentences produced by both children and adults is provided. Also the methods and results of an experiment using these acoustic measures to analyze synthesized vowels and vowels and sentences produced by children and adults is described. In Chapter 4, some of the results from Chapters 2 and 3 are compared using linear regression to determine which acoustic measures best correlate with various perceptual attributes. Finally, Chapter 5 provides a general summary and discussion of the findings presented in the earlier chapters. The results of this thesis provide clinical guidance regarding the types of speech material and acoustic measures that are clinically useful for evaluating the vocal characteristics of children with vocal nodules.

1.2 Background

1.2.1 Differences between Pediatric and Adult Vocal Mechanisms

The developing vocal mechanisms of children are not merely miniature replicas of the mature adult vocal apparatus. Figure 1.1 shows a schematic of the larynx as viewed from above with the vocal folds depicted as two masses of tissue which stretch between the front and back of the larynx. Both the length and the inner structure of the vocal folds grow and develop from early infancy to adulthood. For example, in newborns, the entire length of the vocal fold is 2.5-3 mm (Hirano, et al., 1983), whereas

the vocal fold length in adults is typically 17-21 mm for males and 11-15 mm for females (Hirano, et al., 1983). Vocal fold length differences between males and females are generally not observed under the age of 10, although the vocal folds are typically longer in males than in females by the age of 15 (Hirano, et al., 1983).

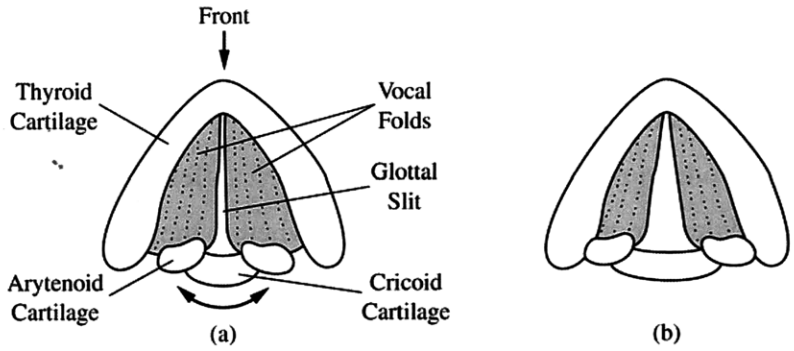
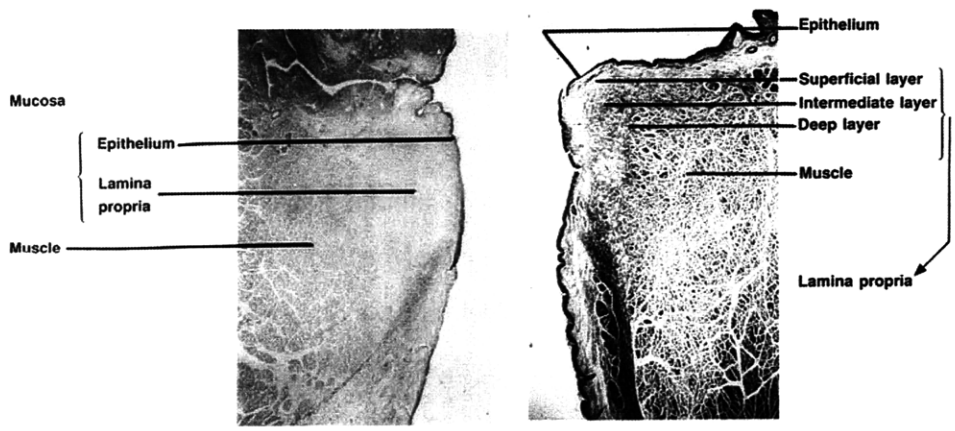


Figure 1.1: Schematic drawing of human larynx seen from above: a) voicing (vocal folds adducted); b) breathing (vocal folds abducted) (Stevens, 1998).

The internal structure of the vocal folds also changes during development. The vocal folds are composed of two clearly defined anatomical parts, the membranous and cartilaginous portions. The portion anterior to the tip of the vocal process of the arytenoid cartilage is called the membranous portion and the portion posterior to the tip of the vocal process is the cartilaginous portion. The ratio of membranous to cartilaginous tissue is 1.1-1.8 in newborns but increases to 4.7-6.2 in adult males and 3.3-4.5 in adult females (Hirano, et al., 1983). Figure 1.2 shows the inner structure of a vocal fold of a newborn and 16 year old. The membranous portion of the vocal folds consists of the vocalis muscle and mucosal layer. The mucosal layer covers the vocalis muscle and is composed of epithelium and lamina propria. Together the vocalis muscle and the mucosal layer form a distinct layered structure particularly evident around the membranous glottis (space between the vocal folds). Initially, the tissues connecting the glottal mucosa to the underlying structures are looser in children, as the lamina propria is not yet fully differentiated into layers (Hirano, Hibi, Terasawa, & Fujiu, 1986). In contrast, the lamina propria of the adult vocal folds consists of three distinct layers: the superficial layer, the intermediate layer, and the deep layers. The superficial layer has a

consistency similar to a mass of soft gelatin and is very pliant. The intermediate layer consists mainly of elastic fibers and the deep layers consist of collagenous fibers. Together, the intermediate and deep layers of the mucosa are referred to as the vocal ligament. Newborns lack a vocal ligament and the entire lamina propria of the mucosa is nearly uniform in structure. The differentiation of the vocal ligament into layers appears between the ages 6 and 12, with the distinct three-layered structure of the lamina propria consistently observed after age 15 (Gray, Hirano, & Sato, 1993). In contrast, the fibers in the vocalis muscle of a newborn are very thin and do not appear to be fully developed until around the age of 25 (Hirano, et al., 1983; Sederholm, 1996).



A B

Figure 1.2: Inner structure of vocal folds from a newborn (A) and 16 year old boy (B) (Bless & Abbs, 1983).

The above stated differences in vocal fold structure between children and adults have lead Hirano and colleagues (1983) to maintain that "the fact that the membranous portion is not only absolutely, but also relatively, longer in adults than children presents morphological evidence that adults are superior to children in the potential ability of voice control." Hence, the larger membranous part of the adult vocal fold, compared to the vocal folds of children, may provide adults with a wider range of vocal function compared to children. In other words, the relatively short membranous vocal folds of children may make adjustments more difficult, limiting the vocal folds' physiologic range

within which they can produce phonation (Lucero & Koenig, 2005). Furthermore, the length and mass of vocal folds in children (shorter and less massive than adults) creates a higher vibrational rate (i.e., fundamental frequency) than adults (Wilson, 1987).

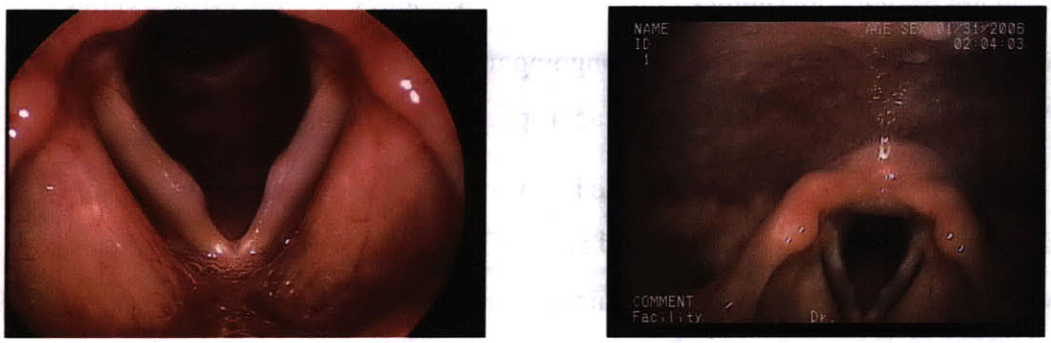
Based on differences in vocal structure and function between healthy children and adults, it is hypothesized that the presence of vocal nodules will differentially impact voice production in children compared to adults. Examination of the developmental anatomy of larynx (Hirano, et al., 1983) shows that the laryngeal size for adult females and children is more similar than the laryngeal size of children and adult males. It is expected that the differences in vocal structure between children and adults will affect performance on vocal tasks, and these differences will be reflected in perceptual and acoustic measurements of voice quality.

1.2.2 Vocal Fold Nodules

Vocal nodules are the most common voice-use related pediatric voice disorder, and their development has been associated with aggressive vocalization (i.e., excessive screaming) (Dobres, Lee, Stemple, Kummer, & Kretschmer, 1990; Shah, Woodnorth, Glynn, & Nuss, 2005). Some studies using endoscopy have found that the presence of vocal nodules in school children was twice as prevalent in males compared to females (Dobres, et al., 1990; Shah, et al., 2005). By contrast, vocal nodules are not the most common vocal pathology for adults, although nodules are far more prevalent in adult females compared to adult males (Fritzell, 1996). Indeed, patients with vocal nodules account for approximately 20% of the adult population with voice disorders (Ramig & Verdolini, 1998).

Vocal nodules, as shown in Figure 1.3, are formed by repeated collision and friction-associated trauma to the lamina propria of the vocal folds (Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989). Nodules develop over time through a sequence of histological changes, and they initially form as focal swelling, or edema, at the midpoints of the membranous portions of the vocal folds. With additional trauma, the edema is coupled with a progressive buildup of fibrous tissue. Over time, nodules are characterized by a marked increase in vascularity with well-established fibrous tumors

(Moore, 1971). Nodules often form bilaterally, and their location represents the region of greatest trauma, as this location marks the point of widest vibratory excursion and the place of maximum pressure upon contact (Hillman, et al., 1989). The presence of nodules can affect voice characteristics by increasing vocal fold mass or stiffness, interfering with the vibratory pattern of the mucosal wave, and preventing the vocal folds from closing completely during phonation (Sapienza, Ruddy, & Baker, 2004).



A B

Figure 1.3: Vocal fold nodules in an adult (A) and a 5 year old girl (B). The nodules are located on the anterior one-third of the vocal fold where contact is most forceful.

The experiments in this thesis compare acoustic and perceptual data from both children and female adults to test the hypothesis that the presence of nodules impacts pediatric and adult vocal mechanisms differently.

1.2.3 Auditory-Perceptual Assessment of Voice

For auditory-perceptual evaluation of voice disorders in adults, a standardized assessment has been created by ASHA, called the Consensus Auditory Perceptual Evaluation of Voice (CAPE-V). Clinicians use the CAPE-V to rate several perceptual attributes related to voice quality including overall severity of dysphonia (disordered voice), roughness (perceived irregularity in sound generated by vibrating vocal folds), breathiness (the audible air escape at glottis during phonation), and strain (the auditory impression of excessive vocal effort). Pitch and loudness are also rated. All perceptual attributes are rated on a continuous 100 mm visual analogue scale (VAS) ranging from normal on the left side of the scale to severely deviant on the right side of the scale. The

score for each attribute is expressed as the distance in mm from the left end of the VAS. Ratings are based on a clinician's observation of the patient's overall performance on the following tasks: three productions of sustained vowels /a/ and /i/, six standard sentences designed to elicit various laryngeal behaviors, and natural conversational speech. Clinicians also provide separate scores for any of the tasks that produce voice quality that is noticeably different from that produced by the other tasks. The tasks in the CAPE-V protocol are designed to sample a range of vocal behaviors, but the diverse subject matter included in the CAPE-V also reflects ongoing uncertainty about whether sustained vowels or continuous speech are more appropriate for assessing voice quality.

One advantage of using sustained vowels for voice quality analysis is that production can be easily controlled and standardized, provided that the characteristics of the vocal folds, vocal tract, and articulators are relatively time-invariant (de Krom, 1994). Additionally, sustained vowels are devoid of individual speech characteristics that may influence perceptual judgments of voice quality such as speaking rate, dialect, intonation, phonetic context, stress, and idiosyncratic articulatory behavior (de Krom, 1994; Wilson, 1987; Yiu, Worrall, Longland, & Mitchell, 2000). Sustained vowels provide a static characterization of the voice apparatus (Klingholtz, 1990) as they are produced when the vocal folds are adducted (closed) once before phonation (de Krom, 1994).

One advantage of using continuous speech is that it contains variations in fundamental frequency and amplitude that are important indicators of abnormal voice quality for a variety of voice pathologies (Fritzell, 1996; Hammarberg, Fritzell, Gauffin, & Sundberg, 1986; Parsa & Jamieson, 2001). More specifically, continuous speech incorporates important vocal function attributes such as rapid voice onset and termination (de Krom, 1994). In contrast to sustained vowel production, production of continuous speech involves more complex and dynamic patterns of laryngeal musculature activity characterized by a series of well-coordinated adductory and abductory vocal fold gesture activities (de Krom, 1994). Unfortunately, continuous speech also contains various perceptual cues that do not strictly relate to vocal fold characteristics including dialect, speaking rate, intonation, loudness, articulatory effects and emotional features (de Krom, 1994; Wilson, 1987; Yiu, et al., 2000). These perceptual cues may vary from speaker to speaker and often vary within one continuous speech stimulus.

A few studies have investigated the effect of voice quality ratings as a function of speech material produced. For example, deKrom (1994) examined perceptual ratings of overall severity, roughness, and breathiness, and found no differences between utterances when speakers produced continuous speech compared to when speakers produced different fragments of a vowel (e.g., vowel onset, vowel post-onset and whole vowel). However, the results of this study may be inconclusive due to several limitations. For example, only six listeners were included and the study lacked rigorous statistical analyses. In contrast to deKrom's (1994) results, Zraick and colleagues found that listeners rated overall dysphonia as more severe when they heard sustained vowels than when they heard continuous speech (Zraick, Wendel, & Smith-Olinde, 2005). Specifically, the overall severity ratings for the sustained vowel /a/ were higher ($p < 0.05$) than for a continuous speech sample (Zraick, et al., 2005). These conflicting results fail to resolve whether sustained vowels or continuous speech provides different and/or more reliable information regarding perceptual severity ratings of adult voices, and there have been few, if any, studies that have addressed this issue for children's voices. Thus, in this thesis, both sustained vowels and continuous speech were studied for adults and children.

A comparison of studies investigating vocal characteristics in children and adults supports the notion that nodules affect voice quality differently in children compared to adults. In adults, the dominant perceptual characteristics associated with nodules are roughness and breathiness (Colton & Casper, 1996). However, in children, roughness is more frequently reported than breathiness as an initial symptom of nodules. This may be because breathiness, unlike roughness, is commonly observed in many children with and without vocal dysfunction (Nagata, et al., 1983; Sederholm, McAllister, Sundberg, & Dalkvist, 1993; Wilson, 1987). In fact, preliminary data from a pilot study conducted at Boston Children's Hospital (the data collection site for this project) has suggested that the perceptual attributes that best characterize children with vocal fold nodules are roughness and strain.

1.2.4 Acoustic Assessment of Disordered Voice

The majority of studies examining acoustic measures during voice production have used sustained vowels rather than continuous speech, perhaps due to the relative simplicity of analyzing sustained vowels and because most acoustic measures assumes quasi-steady state. However, sustained vowels may not provide the best representation of everyday (typical) speech. Rather, an acoustic analysis of continuous speech may more closely approximate the perceived voice quality associated with conversational speech. The acoustic experiments in this thesis analyze and compare both sustained vowels and continuous speech utterances. As such, the following review focuses on acoustic analysis techniques that are applicable to both types of utterances. The following acoustic measures can be categorized into three groups: fundamental frequency measures, time-based perturbation measures and frequency-domain perturbation measures. The sections below describe selected measures for each of these three categories.

Fundamental Frequency Measures

Fundamental frequency measures reflect the vibratory rate of the vocal folds and provide information about the number of glottal openings per second (Colton & Casper, 1996). Fundamental frequency may be measured in a variety of ways using sustained vowels or continuous speech. For periodic sounds, fundamental frequency is defined as the lowest frequency component of a complex periodic wave. When an utterance is viewed as a spectrogram, fundamental frequency is estimated as the frequency spacing between the component frequencies. Frequently used fundamental frequency measures include mean fundamental frequency, standard deviation of fundamental frequency (long term frequency variability), and fundamental frequency range (the range of fundamental frequencies that can be produced) (Colton & Casper, 1996). Although fundamental frequency range appears to characterize vocal fold changes, many patients with voice disorders do not exhibit significant changes in mean speaking fundamental frequency (Baken & Orlikoff, 1987).

Although there are numerous methods for detecting fundamental frequency, "the best candidate for the acoustic pitch period of a sound can be found from the position of

the maximum of the autocorrelation function of the sound" (Boersma, 1993). One speech signal processing tool that incorporates this measure of fundamental frequency is the PRAAT algorithm, which has been shown to work for a wide range of pitches (16 Hz to1.8 kHz). Hence, in the acoustic experiments presented below, fundamental frequency was calculated using the PRAAT software (Boersma & Weenink, 2006).

Aperiodicity Measures

Aperiodicity measures characterize irregularities in vocal fold vibrations. Specifically, aperiodicity of the speech waveform may be due to the presence of jitter, shimmer, additive noise, glottal wave shape change, or some combination of these factors (Hillenbrand, 1987; Murphy, 1999). Such irregularities may be categorized as either time-based perturbation measures or frequency-based perturbation measures.

Time-based Acoustic Perturbation Measures

Time-based acoustic perturbation measures refer to cycle-to-cycle variations in speech which require accurate identification of cycle period boundaries (i.e., where each cycle begins and ends in the speech waveform). Some commonly used time-based aperiodicity measures include variability in amplitude (shimmer) and fundamental frequency (jitter) (Colton & Casper, 1996; Murphy, 1999).

Extracting cycle-to-cycle aperiodicity measures from continuous speech samples is challenging because the high variability in fundamental frequency during the voice-onset and voice-termination phases of continuous speech. Thus, determining the region of fundamental contour that is selected for analysis is an important consideration for reliably estimating fundamental frequency variability measures from continuous speech (Parsa & Jamieson, 2001). Attempts to compensate for the effect of long-term trends in fundamental frequency variation on acoustic perturbation measures during continuous speech have failed to produce consistent results when analyzing disordered voices (Hecker & Kreul, 1971; Schoentgen, 1989). Since determination of cycle period boundaries is often problematic for analysis of voices with moderate to severe dysphonia,

the use of such measures for analyzing these types of voices is frequently not possible (Titze, 1995).

Frequency-Based Acoustic Perturbation Measures

In contrast to time-based perturbation measures, frequency-domain perturbation measures do not require accurate identification of cycle period boundaries, which may be advantageous when analyzing disordered voice. In this study, spectral measures of HNR, spectral slope measures, and cepstral measures such as cepstral peak prominence and first rahmonic amplitude measures were examined.

First Cepstral Peak

Cepstral-based perturbation measures can be derived from the cepstrum of a speech signal. Although, in general, the cepstrum is defined as the spectrum of the log of the spectrum of the voice signal (Noll, 1967), there are several methods to calculate the cepstrum. In this thesis two different cepstral measures (cepstral peak prominence and first rahmoic amplitude) were calculated using different cepstral equations. By transforming the signal from the frequency domain to the cepstral domain, a better visual picture of the degree of harmonic organization may be obtained.

One method of calculating the cepstrum is to use a fixed window (41 ms) and apply a Fourier transform to the logarithmic power spectrum of the speech signal, where the power spectrum is the logarithm of the squared amplitude of the Fourier transform of the signal (Awan & Roy, 2005). A cepstral measure that is related to the first prominent peak is called the cepstral peak prominence (CPP). Since CPP can be calculated from both vowels and continuous speech, both voiced and unvoiced portions of the speech signal can be used. Although the location of the CPP may be associated with the period of the fundamental frequency for most normal voices, CPP often does not correspond with the period of the fundamental frequency in signals that have been severely perturbed (Hillenbrand & Houde, 1996). The magnitude of the CPP measure is obtained by normalizing the amplitude of the peak at the location of the CPP based on the overall

amplitude (Hillenbrand, Cleveland, & Erickson, 1994). For example, a linear regression line may be calculated which relates the time-measure quefrency (i.e. measure of time in the cepstrum) to the cepstral magnitude. Thus, the CPP value reflects the ratio of the CPP value to the corresponding value on the regression line that is directly below the peak, converted to dB (Awan & Roy, 2005). An example of CPP for normal voice is shown in Figure 1.4 and CPP for severely dysphonic voice is shown in Figure 1.5.

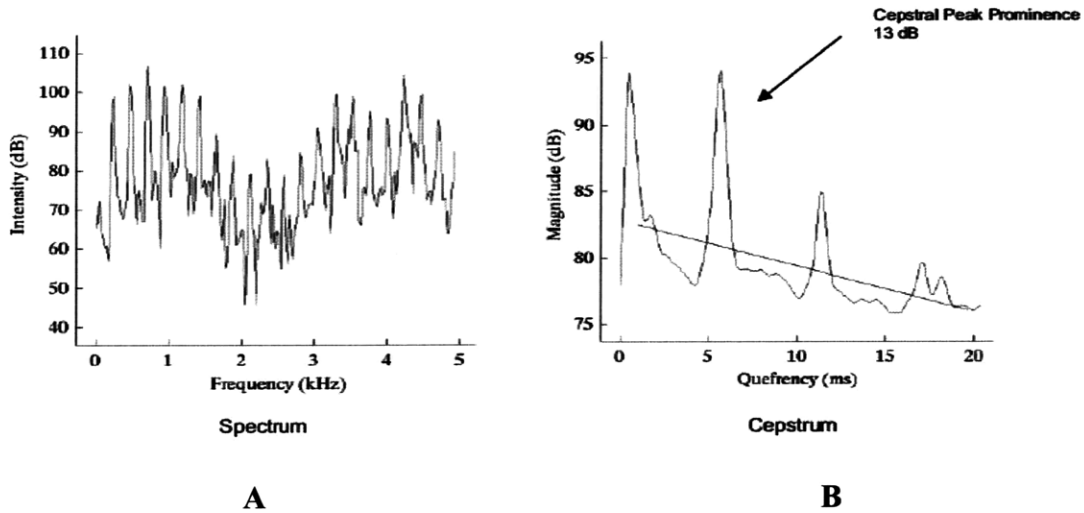


Figure 1.4: Spectrum (A) and cepstrum (B) of a normal voice (Heman-Ackah, Michael, & Goding, 2002).

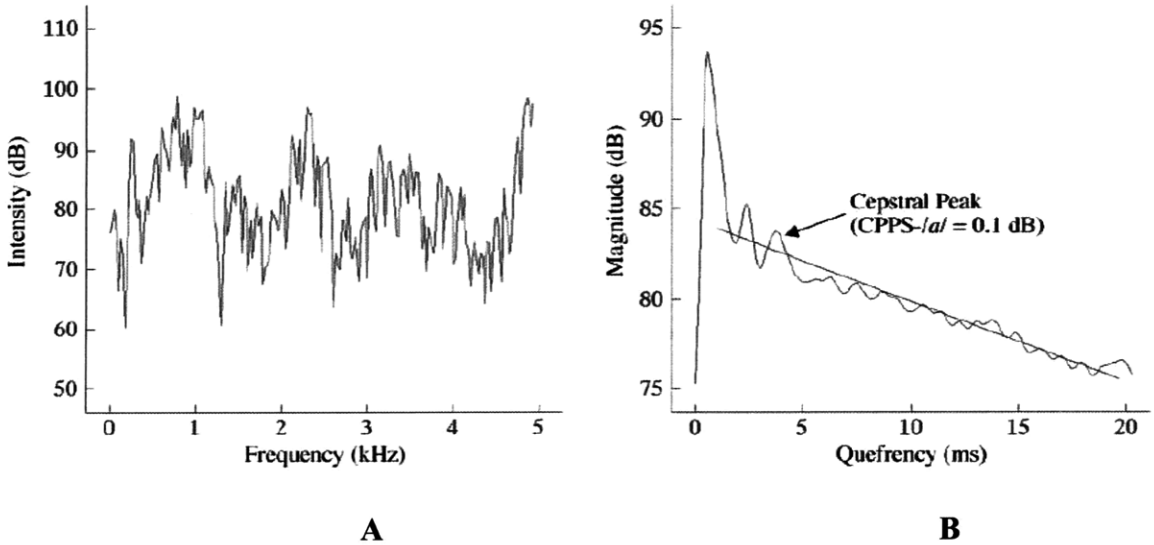


Figure 1.5: Spectrum (A) and cepstrum (B) of a severely dysphonic voice (Heman-Ackah, et al., 2002).

The disadvantage of using a fixed window length is that the CPP measure is dependent on the fundamental frequency of speech. This is problematic when analyzing speech having a wide range of fundamental frequency, as in this investigation when studying children and adults. Hence, a possible solution to the pitch dependence is another cepstral measure of the first prominent peak, termed the first rahmonic (R1), which claims to be pitch independent when analyzing synthesized vowels with known fundamental frequency. In calculation of R1, the cepstrum is calculated with a pitch period dependent window and the inverse Fourier transform of the logarithmic power spectrum (in dB) of the speech signal was computed (Murphy, 2006). In practice, a discrete Fourier transformation (DFT) is applied to an acoustic signal to create a frequency spectrum representing the intensity of each frequency component within the signal, and then an inverse DFT of the power spectrum is used to calculate the cepstrum. Hence, for periodic signals, the power spectrum represents energy at harmonically related frequencies. R1 is typically located using a peak-picking algorithm which searches for the maximum amplitude in the region of the expected fundamental frequency. The unit for R1 is dB as R1 is directly proportional to the mean of the individual spectral harmonics-to-noise ratio in dB (Murphy, 2006). Figure 1.6 illustrates the difference between the two first cepstral peak algorithms.

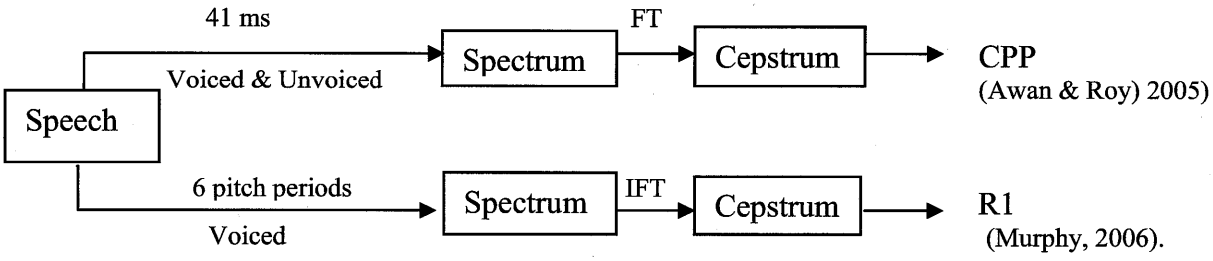


Figure 1.6: Comparison of Cepstral Peak Prominencne (CPP) and First Rahmonic Amplitude (R1) algorithm. FT represents Fourier Transform. IFT represent inverse Fourier Transform.

Regardless of how the cesptrum is calculated, the cepstrum can be separated into low and high quefrency regions. The low quefrency region encompasses all quefrencies below the first cepstral peak and consists of a broad peak that reflects the low frequency formant structure of the vocal tract (Murphy & Akande, 2007). In contrast, the high

quefrency region is characterized by rahmonic peaks, which are spaced at multiples of the fundamental period (Oppenheim, 1969; Verhelst, 1986).. Hence, the high quefrency region contains information about the source component. Since the cepstral components of source and filter (i.e., formant structure) occupy approximately disjoint quefrency regions, a cepstral lifter (a low-pass filter that operates on the cepstrum) can be applied to separate these contributions. Thus, to recover the vocal tract impulse response, a cepstral lifter having a length that is half of a pitch period may be used (Verhelst & Steenhaut, 1986).

Various research studies have suggested that the first cepstral peak correlates with overall voice quality and breathiness (Dejonckere & Wieneke, 1996; Heman-Ackah, et al., 2003; Heman-Ackah, et al., 2002; Hillenbrand & Houde, 1996; Wolfe, Martin, & Palmer, 2000). This correlation may be due to the dependence of the first cepstral peak on voice periodicity. Murphy (2006) showed that the first cepstral peak is directly proportional to the geometric mean harmonics-to-noise ratio (gmHNR), where gmHNR is the mean of the individual harmonic to between harmonic ratios in dB. The term "geometric" is used since an average of dB values is equivalent to determining the geometric mean HNR of the power spectrum (Murphy, 2006). Since, the first cepstral peak is determined by the average of the differences between the individual harmonic and between-harmonic energies in the dB spectrum, the first cepstral peak provides an overall index of the richness of the harmonic spectrum. For example, the amplitude of the first cepstral peak of a signal with a well-defined harmonic structure (i.e., a periodic signal) is large. In contrast, signals with increased levels of noise (modeled as random, mean zero, or Gaussian noise), and acoustic perturbations in the voice signal, result in a reduction in the amplitude of the first cepstral peak (de Krom, 1993). An additional characteristic of the first cepstral peak that makes it a robust indicator of voice quality, especially for disordered voices, is that the calculation of the first cepstral peak does not require an accurate determination of pitch period, although a rough estimate of the pitch period is necessary to determine the location of this peak. Also, since it has been shown that the first cepstral peak accounts for greater variance in overall severity of disordered voice for continuous speech than for sustained vowel phonation, the first cepstral peak may be a

more reliable measure of dysphonia in continuous speech than other measures of periodicity (Heman-Ackah, et al., 2002).

Using Cepstral-Based Approaches to Analyze the Voices of Women and Children

Implementing cepstral techniques to analyze the high-pitched voices of women and children poses several difficulties. For example, the vocal tract impulse response of a high-pitched voice is more likely to overlap the high-quefrency portion of the cepstrum representing the vocal source due to the location of the first cepstral peak. Hence the separation of the cepstral components (source and vocal tract) is more challenging for high-pitched speakers. This possible distortion in the cepstrum can be controlled by selection of an appropriate low pass lifter. Thus, in order to extract low quefrency components from high-pitched speech, the filter length (i.e., liftering in the quefrency domain) may be slightly increased to allow for accurate formant estimation (Rahman & Shimamura, 2005). More specifically, a lifter of length 0.6 pitch period may help facilitate analyzing speech with fundamental frequencies up to 250 Hz and a lifter of length 0.7 pitch period may facilitate analyzing speech with higher fundamental frequency values (Rahman & Shimamura, 2005).

Sampling the frequency domain signal also poses a challenge for speech signals with higher fundamental frequencies. The higher fundamental frequency makes the harmonics spaced further apart which results in sampling the spectrum at a lower rate. This poor sampling of the spectrum can possibly impede the accurate evaluation of the frequencies of the formants due to the loss in resolution.

One way to avoid the confounding factor of cepstral-based analyses being dependent on fundamental frequency is to modify the cepstral acoustic analyses to make them independent of fundamental frequency and reflect source aperiodicity. Murphy (2006) suggests two modifications to achieve fundamental frequency independence. A first modification is to estimate R1 with a pitch period dependent window, where the analysis window contains the same number of pitch periods. This modification ensures that the harmonic resolution in the spectrum stays the same across different fundamental frequencies. A second modification is to limit the number of harmonics in the spectrum

prior to taking the inverse Fourier transform. This modification alleviates the effect of having different number of harmonics associated with different fundamental frequencies and ensures that the noise level does not exceed the harmonic levels as noise increases. However, by eliminating the noise and high frequency components, this might affect the correlation between R1 and noise. Hence, to ensure that the first cepstral peak values differ due to aperiodicity or differences in voice quality, Murphy (2006) suggests applying a hamming window to limit the spectrum prior to computing the cepstrum.

Harmonics-to-Noise Ratio (HNR)

An example of a spectral-based perturbation measure which estimates the level of noise in the speech signal is the harmonics-to-noise ratio (HNR). Estimation of HNR in the frequency domain is calculated as the energy of the spectral peaks that exceed the noise level at the harmonic peak frequencies (Qi & Hillman, 1997). The harmonic components, corresponding to the fundamental frequency and its harmonics, are the spectral peaks identified from the spectrum. In Figure 1.7, the harmonic peaks in the spectrum are depicted with red circles. For calculation of the noise level in the HNR measure, since it is difficult to obtain the noise level from the spectrum, one way to get a better estimate of the average noise components (depicted as the solid line in Figure 1.7) is by calculation of the cepstrum and low pass liftering it and converting it back to the spectral domain. Hence, rather than analyzing the whole noise spectrum to calculate the noise components, only noise reference levels at the harmonic peak frequencies need to be estimated, as illustrated by green circles in Figure 1.7. Thus, HNR was computed in the acoustic experiments presented in Chapter 3 as the mean difference between the harmonic peaks and the reference levels of noise at these peak frequencies (Qi & Hillman, 1997). Figure 1.8 illustrates this cepstral-based HNR algorithm.

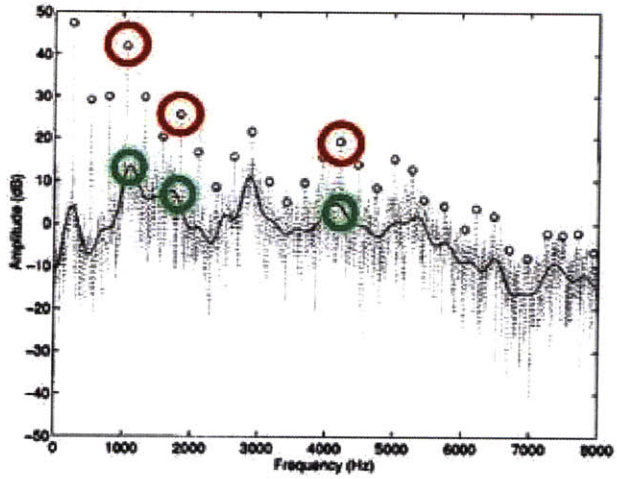


Figure 1.7: Example of low-quefrency, cepstrum-based estimation of noise floor (green circles) together with spectrum (dotted line) and its harmonic peaks (red circles). (Qi & Hillman, 1997).

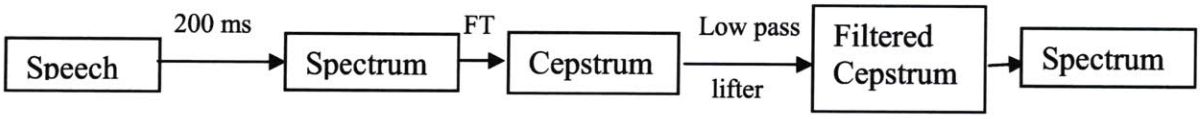


Figure 1.8: Flow chart of cepstral-based HNR algorithm, where FT represents Fourier transform.

In acoustical studies of pathological voice quality, the relative level of spectral noise is often considered an important parameter, as it correlates with a number of physiological and perceptual characteristics of voice such as insufficient vocal fold adduction and/or an irregular pattern of vocal fold vibration (Baken & Orlikoff, 1987). However, low HNR may also be related to additive noise (Baken & Orlikoff, 1987). Thus, these conflicting findings demonstrate that HNR is limited in its ability to distinguish noise contributions from different physiological mechanisms.

Although this algorithm was shown to be successful in analyzing adult speech, in this study, this algorithm needed to be applied to children to create an HNR measure that was independent of fundamental frequency. Hence, similar to the procedure used for cepstral analysis, a pitch period dependent window where the analysis window contains the same number of pitch periods was used in the acoustic experiments in Chapter 3,

thereby ensuring that the harmonic resolution in the spectrum stayed the same across different fundamental frequencies.

Spectral Slope

Another spectral measure frequently used in voice analysis is spectral slope, which represents the relative amount of energy concentrated in low versus high frequency regions of the speech spectrum (i.e. slope of speech spectrum) (de Krom, 1995; Klatt & Klatt, 1990; Linville, 2002; Mendoza, Valencia, Munoz, & Trujillo, 1996). Energy in the higher frequency regions can be attributed to a mixture of harmonic and noise energy components, with dysphonic voices typically exhibiting relatively weak harmonic components (de Krom, 1995).

Spectral slope measures have been calculated from the spectrum resulting from the production of vowels and continuous speech (Parsa & Jamieson, 2001). In the acoustic experiment presented in Chapter 3, spectral slope measures were calculated from both the individual and average spectrum of sustained vowels and continuous speech. The advantage of averaging successive spectra of the speech signal is that the resulting spectrum can be used to obtain information about the spectral distribution of speech signal over a period of time. However, there are conflicting opinions on which measures of spectral slope are preferred.

One common method to obtain spectral slope measures from the average spectrum of vowels and the average spectrum of continuous speech is to calculate the relative measure of low vs. high frequency energy concentration in the spectrum, in dB. Frokjaer-Jensen and Prytz (1976) calculated the energy below and above 1 kHz to determine the spectral distribution of the speech signal over a period of time. Decreased energy in high frequency regions corresponds to strained voice quality whereas increased energy in high frequency regions corresponds to breathy voices (Frokjaer-Jensen & Prytz, 1976). A different spectral slope measure used by Awan and Roy (2005) was to calculate the ratio of the energies below to above 4 kHz. This measure was shown to be a strong predictor of overall severity of dysphonia (Awan & Roy, 2005). Since the higher frequency spectrum is normalized relative to the lower frequency spectrum, the ratio of

the two portions of the spectrum is independent of recording variables such as microphone distance, amplification level, etc (Frokjaer-Jensen & Prytz, 1976).

One limitation of using the average spectrum of continuous speech to characterize voice is that the vocal tract can influence the spectrum and mask laryngeal features (Klingholtz, 1990). For example, the average spectrum of continuous speech fails to detect fine temporal details of the speech signal, and therefore does not characterize cycle-to-cycle aperiodicity of either period or amplitude (Schoentgen, 1989).

Spectral slope measures provide insight into physiological mechanisms of the larynx. For example, Klatt and Klatt (1990) suggests that the breathier quality of female voices can be explained by a longer opening and the presence of a posterior opening between the vocal folds, which generates aspiration noise, particularly at higher frequencies where the noise may actually replace harmonic excitation of the third and higher formants. Hence, non-simultaneous closure along the length of the vocal folds can result in reduced high-frequency content which lowers the spectral slope measure (Hanson & Chuang, 1999).

Several studies have also shown that spectral slope may be an important index of vocal severity. For example, spectral slope measures have been shown to predict vocal severity better for the continuous speech compared to sustained vowels (Hillenbrand & Houde, 1996). Specifically, spectral slope measures accounted for approximately 40% of the variance in breathiness ratings for sustained vowels compared to approximately 70% of the variance in breathiness ratings for continuous speech. Awan and Roy (2006) also found that spectral slope measures, calculated as a ratio of low- to high-frequency spectral energy contributes to an accurate prediction of dysphonia severity.

Using Spectral-based measures of overall slope to Analyze the Voice of Women and Children

The fundamental frequency of adult female and children's voices is typically higher than that for adult males. As mentioned above, this higher fundamental frequency may cause under-sampling of the formants in the spectrum. Assuming that spectral slope measures are not dependent on an exact estimation of the formant frequencies, the

spectral under-sampling observed in adult women and children's voice should not affect accurate calculations of spectral slope measures in these populations. Also contrary to the procedures used for first cepstral peak and HNR measures, a pitch period dependent window was not used to achieve independence with respect to fundamental frequency when calculating the spectral slope measures. Such measures are energy ratios which should not be dependent on the pitch period.

1.2.5 Relationships between Perceptual and Acoustic Measures of Disordered Voice Quality

Several studies have correlated perceptual voice qualities with various acoustic measures, although their findings have been inconsistent. These conflicting results may be due, in part, to the adoption of different definitions for perceptual qualities, a lack of one-to-one correspondence between a perceptual quality and a single acoustic measure, and the non-homogeneity of algorithms used to calculate acoustic measures (Eadie & Doyle, 2005). In this thesis, both sustained vowels and continuous speech data were analyzed and compared. Thus, the following sections focus on acoustic correlates of roughness, strain, breathiness, and overall severity as they have been applied to both vowels and continuous speech.

Roughness

Since roughness may originate from irregular vocal fold vibration, pathologies that affect the vibratory behavior of the vocal folds often result in some degree of perceived roughness (Fukazawa, el-Assuooty, & Honjo, 1988). Aperiodic vocal fold vibrations resulting from vocal pathologies can be observed as cycle-to-cycle variability in amplitude and fundamental frequency (Heman-Ackah, et al., 2002). Roughness of dysphonic voice quality has been associated primarily with measures of spectral noise suggesting a high level of irregularity across the time domain of acoustic signals (Deal & Emanuel, 1978). In investigating the acoustic correlates of roughness for both sustained vowels and continuous speech, it was found the highest correlation ($r = -0.9$) was between the first cepstral peak and roughness for continuous speech, and that the

perception of roughness in continuous speech was more detectable than in sustained vowels (Halberstam, 2004).

Strain

Perception of vocal strain is associated with increased and poorly regulated laryngeal muscle tension (Colton & Casper, 1996; Titze, 1995). Furthermore, imbalances and increases in muscle tension produce abnormally stiff vocal folds and/or incomplete closure of the membranous glottis (Holmberg, Hillman, Hammarberg, Sodersten, & Doyle, 2001). Unfortunately, the auditory perception of strain has not been well characterized by objective measurements (Colton & Casper, 1996). For example, although strained voice quality in continuous speech was shown to be positively correlated with spectral slope measures, (Hammarberg, Fritzell, Gauffin, Sundberg, & Wedin, 1980) the absence of consistent within-rater and between-rater ratings associated with this study makes the results difficult to interpret with confidence.

Breathiness

The perception of breathiness results from an incomplete closure of the vocal folds and from posterior glottal opening, causing turbulent airflow (i.e., noise) in the area of the glottis (Colton & Casper, 1996). This turbulent airflow produces significant high-frequency acoustic energy (i.e., above 2-3 kHz), which reduces the inherently weak periodic component of the vocal source at the vocal folds in the mid- and high-frequencies (Hammarberg, Fritzell, & Schiratzki, 1984; Klatt & Klatt, 1990; Titze, 1995). An increased perception of breathiness has been associated with many acoustic measures including increased noise energy in high frequency region, low periodicity, and low HNR in the higher frequency bands (Hammarberg, et al., 1980; Heman-Ackah, et al., 2002; Klatt & Klatt, 1990).

Spectral slope measures may be important in predicting the severity of breathy voice. Specifically, the significant increase in noise energy in the higher-frequency region should reduce the strength of higher harmonics, which in turn reduces both the

HNR in the higher frequency bands and the spectral slope value (Frokjaer-Jensen & Prytz, 1976; Hammarberg, et al., 1980; Klatt & Klatt, 1990; Morrison & Rammage, 1994). Additionally, additive noise produced at the glottis also causes breathy voices to be less periodic than a non-breathy voice due to the unpredictable cycle-to-cycle variation in intensity (Heman-Ackah, et al., 2002). The reduction in periodicity observed for breathy voices is also reflected in the first cepstral peak measure. In fact, Hillenbrand and Houde (1996) observed an inverse relationship between the perception of increased severity of breathiness and decreased relative amplitude of the first cepstral peak. More specifically, the first cepstral peak measure has been shown to account for 92% of the variance in perceptual estimates of breathiness in vowels (Hillenbrand & Houde, 1996). For continuous speech, both the first cepstral peak and spectral slope accounted for 70-85% of the variance for breathiness ratings (Hillenbrand & Houde, 1996). This high correspondence between breathiness and the first cepstral peak values may be due to the flat shape of the cepstrum for breathy voices.

Overall Severity

Several studies have suggested the usefulness of cepstral measures in predicting the overall severity of dysphonia. In a study of male and female dysphonic voice samples, Wolfe and colleagues used multiple regression analysis and found a multiple correlation of 0.84, accounting for 71% of variance, between measures derived from the first cepstral peak and degree of abnormality (i.e., deviation from normal) in female voices (Wolfe, et al., 2000). No other examined acoustic variables contributed significantly to the amount of perceptual variance. Specifically, increased abnormality was associated with a decrease in overall first cepstral peak, due to less harmonic energy. For male voices, a multiple correlation of 0.98, accounting for 96% of variance, was observed between combinations of measures which included the first cepstral peak (Wolfe, et al., 2000). In another study, Dejonckere and Wieneke (1994) showed that the magnitude of the first cepstral peak "is fairly correlated with the perceptual scaling of voice deviance." Heman-Ackah and colleagues (2002, 2003) also reported that the first cepstral peak index provides an indication of the harmonic structure for voiced speech,

and that the first cepstral peak strongly correlates with perceptual measures of overall dysphonia. Compared to other acoustic indices (e.g., jitter, shimmer and various spectral slope and noise measures), first cepstral peak may provide a superior correlate of overall voice quality (Awan & Roy, 2006; Dejonckere & Wieneke, 1994; Heman-Ackah, et al., 2002).

1.2.6 Acoustic-perceptual comparisons

As the above review of the literature indicates, numerous studies have suggested various relationships between selected auditory-perceptual characteristics of disordered voices and specific acoustic measures. This thesis focuses on determining which of these acoustic measures (i.e., mean fundamental frequency, first cepstral peak, cepstrum-based HNR, and spectral measure of overall spectral slope) correlate best with each of the perceptual attributes of roughness, strain, breathiness and overall severity. Both sustained vowel and continuous speech utterances were assessed for children and women with vocal nodules. The acoustic algorithms used to calculate the aforementioned acoustic measures were modified, as detailed above, to optimize their performance in assessing the pathological voices of children.

1.3 Statement of Problem and Hypotheses

A primary goal of this thesis was to identify the minimal set of speech tasks and associated acoustic analysis methods that are most salient in delineating the impact of vocal nodules on voice production in children. The following set of specific hypotheses was tested to address this goal:

1. Perceptual judgments of voice quality will correlate better with acoustic measures when both are extracted from CAPE-V sentences than from sustained vowels.

2. The pediatric and adult forms of the CAPE-V protocol will elicit similar vocal behaviors, reflected as a lack of significant differences in perceptual judgments of voice quality when both versions are produced by older children who can read.

3. Larger auditory-perceptual differences between the voices of adult women with and

without vocal nodules will be observed for roughness and breathiness compared to strain.

4. Larger auditory-perceptual differences between the voices of children with and without vocal nodules will be observed for roughness and strain compared to breathiness.

5. Different combinations of acoustic measures (first cepstral peak, spectral slope, fundamental frequency and HNR) will show the highest correlations with each of the perceptual parameters of overall severity, roughness, breathiness and strain.

Chapter 2: Perceptual Experiments

Perceptual evaluation of voice is considered to be an essential aspect of the conventional voice diagnostic, and is the primary method by which dysphonias are identified and therapy progress is tracked.

2.1 Methods

A pediatric version of the CAPE-V protocol was created to provide a standardized set of sentences designed to evoke a variety of vocal behaviors. Subjects with vocal nodules and normal voices from different age ranges were selected to study how the relationships between acoustic measures (described in chapter 3) and perceptual measures change with age (i.e., from young children to adults). Recordings of the production of the vowel /a/ and sentences (adult version of the CAPE-V and/or pediatric version of the CAPE-V) for each of the subjects were collected and analyzed using perceptual measures. In addition, listeners rated the production of vowels and sentences by children and female adults to determine which perceptual attribute(s) show the largest perceptual differences between normal voices and voices with nodules.

2.1.1 Development of a Pediatric Version of the CAPE-V

A pediatric version of the CAPE-V was developed to elicit a set of laryngeal behaviors similar to those targeted in the adult version of the CAPE-V sentences (the other CAPE-V tasks are sustained vowels and conversational speech which can be produced by children). The pediatric CAPE-V sentences were designed to account for normal speech and language development in three-year-old English-speaking children. By this age, children typically have acquired the consonant sounds /m, n, h, w, p, b, t, d, k, g/ (Smit, Hand, Freilinger, Bernthal, & Bird, 1990), and their mean phrase length for an utterance is typically 3-3.75 morphemes (Mclaughlin, 1998). Three-year-olds have also typically developed the phonological processes of gliding liquids and reducing clusters (Haelsig & Madison, 1986), and they are able to produce two-syllable length structures (Shriberg, 1993).

The pediatric CAPE-V sentences were comprised of age-appropriate noun words presented as easily identifiable pictures. Pictures were chosen as they have been used with children as young as three years of age to elicit voice samples (Sapienza, et al., 2004). All sentences in the pediatric version of the CAPE-V protocol have the form "A (1st picture) and (2nd picture)."

Three pediatric CAPE-V sentences were created for each adult CAPE-V sentence so that the production of three short pediatric CAPE-V sentences better approximates the duration for the production of one adult CAPE-V sentence. The pediatric CAPE-V sentences, which were designed to target similar laryngeal behaviors as in the adult CAPE-V, are listed below:

1) Sentences with different vowels that require divergent tongue positions
    a) Adult sentence: "The blue spot is on the key again."
    b) Pediatric sentences:
        i) A boot and pen.
        ii) A pot and key.
        iii) A cup and pig.
2) Sentences having easy onset with /h/
    a) Adult sentence: "How hard did he hit him?"
    b) Pediatric sentences:
        i) A honey and hay.
        ii) A head and hook.
        iii) A hotdog and hat.
3) Sentences that are all voiced
    a) Adult sentence: "We were away a year ago."
    b) Pediatric sentences:
        i) A dog and bed.

ii) A baby and bee.

iii) A boy and money.

4) Sentences that elicit hard glottal attack

a) Adult sentence: "We eat eggs every Easter."

b) Pediatric sentences:

i) A egg and apple.

ii) A ear and onion.

iii) A ant and eye.

5) Sentences that incorporate nasal sounds

a) Adult sentence: "My mama makes lemon muffins."

b) Pediatric sentences:

i) A mitten and bone.

ii) A moon and donut.

iii) A banana and comb.

6) Sentences that are weighted with voiceless plosive sounds

a) Adult sentence: "Peter will keep at the peak."

b) Pediatric sentences:

i) A boat and cake.

ii) A pie and cat.

iii) A kite and cookie.

2.1.2 Subjects

Two groups of children with nodules (4-6 year-olds and 8-10 year-olds) were studied. Recordings from several children were collected until eight "usable" recordings for each age group are collected. Usable recordings were required to be high quality (i.e.,

having minimal background noise) with the picture naming and/or sentence production considered to be fluent (e.g., pairs of pictures not produced with long pauses). Although the number of males and females in each age group was not equivalent, previous research has shown that a child's vocal fold anatomy does not appreciably differ between males and females younger than 10 years of age (Hirano, et al., 1983).

In addition to the two groups of children with nodules, two control groups of children (4-6 years of age and 8-10 years of age) were selected. Children in the control groups had no history of a voice disorder and had normal hearing, speech, and language development. Children in the control groups were judged to have normal voice by a speech-language pathologist.

Two groups of adult females (18 years of age or older, having normal hearing, speech, and language development) were recruited for the study. One group consisted of eight subjects with nodules and the other group consisted of eight subjects without a history of a voice disorder. Adults in the control group were judged to have normal voice by a speech-language pathologist.

Recordings were collected from 153 3-12 year old children with disordered voice having a diagnosis of vocal nodules and with normal speech, hearing, and resonance from the Pediatric Voice Clinic at Children's Hospital Boston. Recordings from 53 3-12 year old children with normal voice, speech and hearing were also collected. The age distribution of children who were recorded is shown in Figure 2.1.

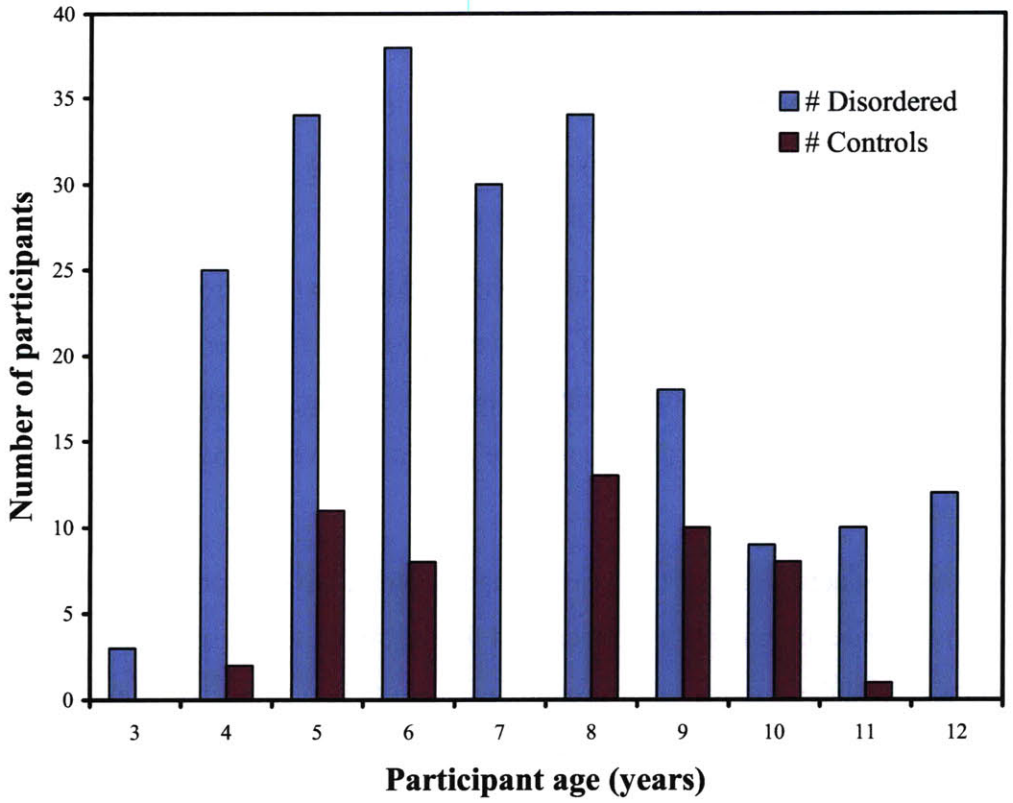


Figure 2.1: Age distribution of children recorded with the Pediatric CAPE-V sentences.

2.1.3 Recorded speech material

After initial testing and modification of procedures, the pediatric version of the CAPE-V sentences was used to collect the data for this study. In addition to producing the pediatric CAPE-V sentences, all children (both age groups, normal and disordered voice) produced three repetitions of the sustained (3-5 seconds) vowel /a/, and a short segment of conversational speech. Recordings of the 8-10 year-old children producing the sentences in the adult CAPE-V protocol were also collected.

Although many recordings were collected, not all recordings were of good enough quality to be used for acoustic analysis. For example, some recordings of poor quality had excessive extraneous background noise (e.g., resulting from body movement such as rustling clothes). Refinements were also made in the recording protocol to better ensure high quality audio recordings. For example, the clinician reminded the children to keep their hands and feet quiet during the recording session. Also a highly unidirectional

microphone was used to decrease environmental noise. Another refinement that was made to the protocol was driven by the fact that young children (less than 9 years old), despite practicing the sentences prior to the recordings, occasionally produced each word (i.e., individually) with equal stress, rather than as a sentence. This type of production lacked natural fluency and intonation. Hence, in order to obtain better recordings, children practiced the sentences numerous times until they were able to produce continuous speech, prior to the recordings.

Figure 2.2 illustrates the difficulties with having young children produce the pediatric CAPE-V sentences across different ages by showing the percentage of children that were successful in producing all of the sentences. Figure 2.2 shows that children in the 3-5 year-old age range had particular difficulties with producing high quality recordings of the pediatric CAPE-V sentences. Children in this age range often lacked cooperation, failed to sit quietly, and had difficulty understanding the task or sentence structure. In fact, the youngest child to have successfully recorded all of the pediatric CAPE-V sentences was 3 years and 9 months old.

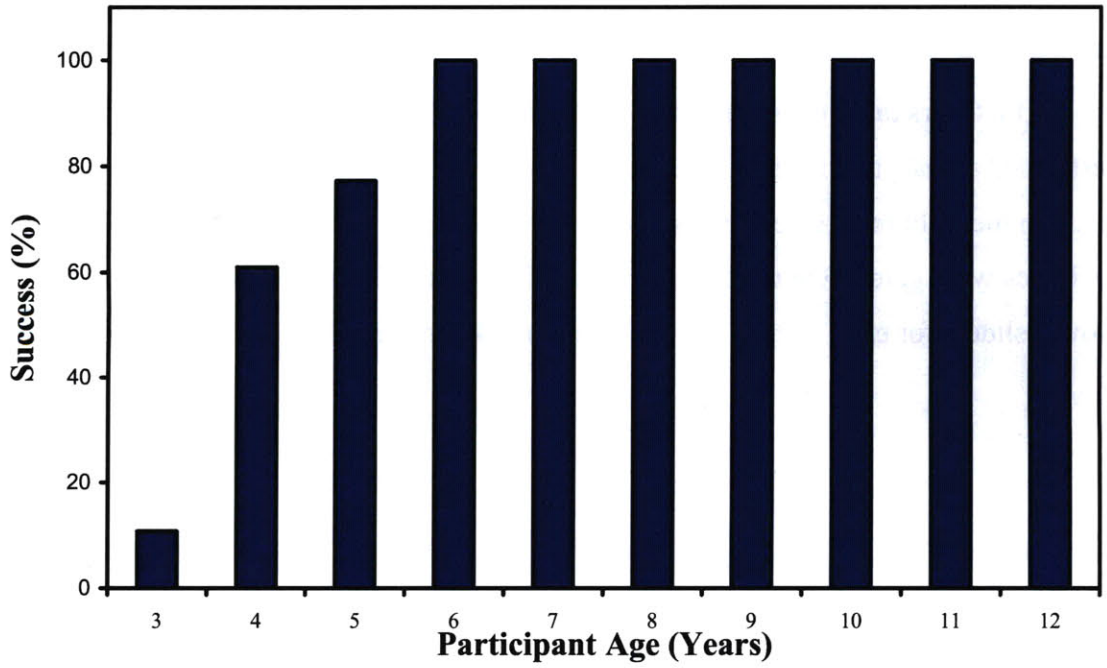


Figure 2.2: Age distribution of successful implementation of the Pediatric CAPE-V sentences.

All adult subjects were instructed to produce three repetitions of the sustained (3-5 seconds) vowel /a/, produce a short segment of conversational speech, and read the sentences in the adult CAPE-V protocol.

2.1.4 Recording Procedures

All speech material was produced at a comfortable pitch and loudness level and was recorded in a quiet environment with a high quality condenser microphone. The microphone was kept at a constant distance of 10 cm from the lips using an adjustable head device (head-pointer for non-vocal communication). The microphone was connected to an amplifier, the output of which was connected to the input of the computer interface for the KayPentax Computerized Speech Laboratory (CSL model 4400). This setup enabled direct A/D conversion and digital recording of sentences (32 kHz sampling rate) and vowels (50 kHz sampling rate). Sentences were recorded with a lower sampling rate than for vowels due to memory limitations.

2.2 Experimental Design

Listeners rated the speech stimuli during two 30-45 minute listening sessions, performed at least twenty-four hours apart using a computer interface shown in Figure 2.3. On the right hand side of the screen, instructions and definitions of the perceptual attributes were given. On the left hand side of the screen, listeners clicked on buttons and moved sliders for each perceptual attribute to mark the perceived deviance from normal.

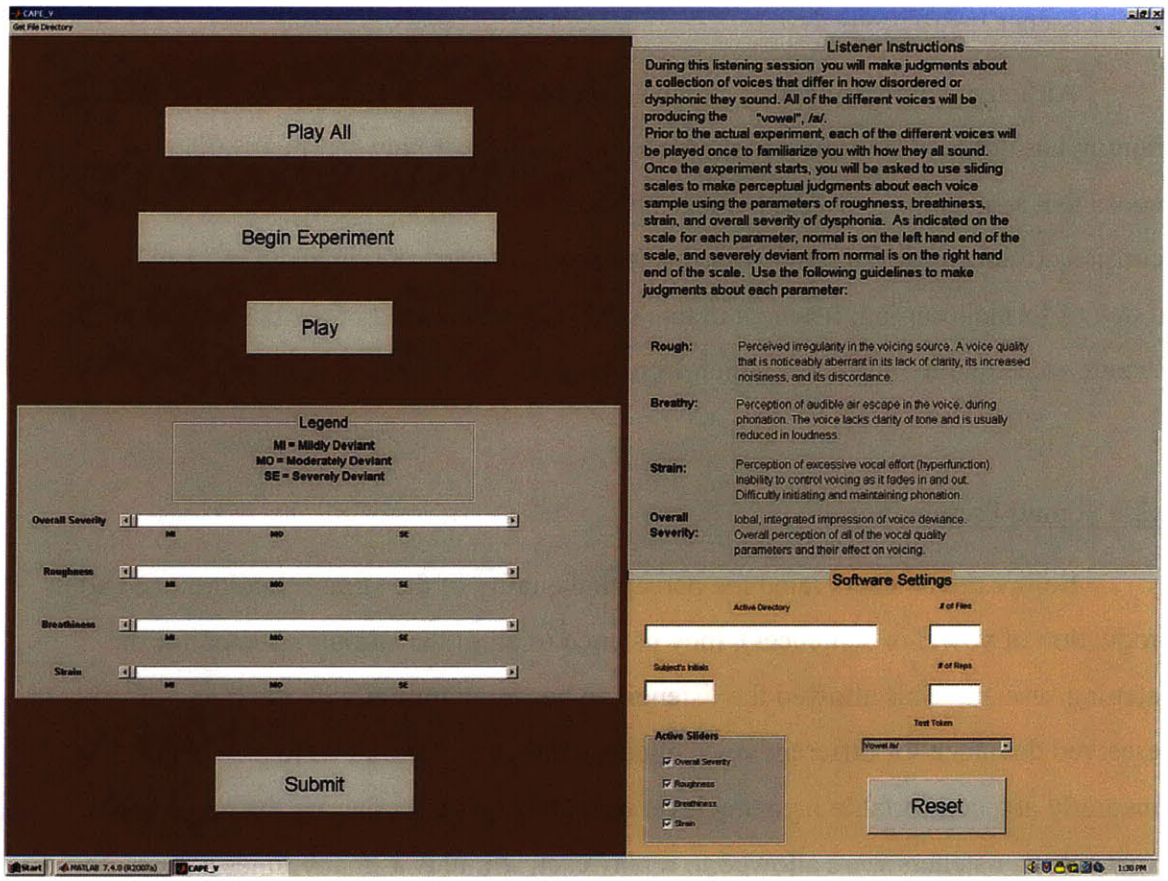


Figure 2.3: Matlab-based computer interface for perceptual experiments.

The general structures for two perceptual experiments are described in the following sections.

2.2.1 Listeners

Seventeen graduate students in speech-language pathology who had taken a class in voice disorders were recruited as listeners, which ensured that all listeners had a similar degree of exposure and training in making voice quality judgments (Shrivastav, Sapienza, & Nandur, 2005). All listeners were native speakers of American English and were screened to ensure normal bilateral hearing (air conduction pure-tone threshold below 25dB HL at 500 Hz, 1 kHz, 2 kHz, and 4 kHz). Listeners were paid for participating in the experiment.

2.2.2 Listener Instructions

Although some perceptual studies have employed a training phase for listeners, training has not been definitively shown to increase either between-rater reliability (consistent ratings for individual voices between listeners) or within-rater agreement (consistent ratings for the same voice by the same listener) (Kreiman, et al., 1993). Thus, in lieu of formal training, listeners in this study were reminded of the definitions of the various voice quality attributes which were judged.

2.2.3 Stimuli Presentation

Before the listeners rated the perceptual quality of the stimuli (children and adult production of vowel or sentences), they listened to all of the stimuli selected for the listening session. This allowed the listeners to become familiar with the range of perceived deviance for different voice quality attributes. Stimuli were presented binaurally at a comfortable listening level using headphones. During the perceptual experiments, listeners were allowed to replay each stimulus as many times as necessary before making judgments.

2.2.4 Rating Stimuli

During the rating phase of the experiment, each of the stimuli was randomly presented twice to allow for an assessment of within-rater reliability. The presentation order of the stimuli was randomized to help control for perceptual effects, such as the range of stimuli and the order of presentation of stimuli (Shrivastav, et al., 2005). After listening to each stimulus, each of the listeners rated the stimulus on each of four voice quality attributes: overall severity, breathiness, roughness, and strain. Listeners indicated their rating for each attribute using a mouse click on a computer screen, shown in Figure 2.3, to indicate their degree of perceived deviance from normal using a separate visual-analog scale (VAS) for each attribute.

2.2.5 Analysis

All statistical analyses were conducted using a statistical analysis program, PASW Statistics 17.0 (SPSS Inc,. Chicago, IL, USA).

To assess within-rater reliability, Pearson's correlation coefficients (i.e., r-value) were calculated for all speech samples which were rated twice by each listener. A Pearson's correlation coefficient measures the strength and direction of the linear relationship between two variables while implicitly assuming that the two variables are jointly normally distributed.

Between-rater reliability for dysphonic severity ratings was determined by calculating an intra-class correlation coefficient (ICC). The ICC is used to assess the reliability of measurements made by multiple raters measuring the same quantity. A goal of the present study was to generate a single average rating for each voice to provide a basis for studying the correlation between perceptual ratings of voice and objective measures of voice (discussed in Chapter 3). Hence, the average measure gives the reliability of the mean of the ratings of all the raters. ICCs and 95% confidence intervals for all ratings were determined to compare between-rater reliability across attributes. ICC was calculated using a two-way fully random ANOVA (analysis of variance) model, since a random sample of raters was selected from a larger population, with each rater rating each target. ANOVA is a general technique that can be used to test the hypothesis that the means among two or more groups are equal. Specifically, ANOVA methodology is quite effective in determining if two or more group means differ due to chance, or if observed differences are indeed the result of true difference between phenomena. Two-way ANOVA is a multiple regression which measures the effects of two independent variables and assesses whether there is an interaction between the factors. Additionally, fully random ANOVA allows the results to be generalized to other raters and for other voices, thereby providing a more clinically relevant assessment.

The following sections describe two perceptual experiments along with their methods and results.

2.3 Perceptual Experiment 1

One goal of this thesis was to determine a minimal combination of speech tasks that provided the most salient assessment of vocal function in children with vocal fold nodules. More specifically, speech material that consistently elicited the largest perceptual differences between normal and disordered voices was considered to provide the most salient information about vocal function. The effect of vocal pathology (vocal nodules) on which perceptual characteristics was most salient in children versus adult females was examined by systemically comparing perceptual data for children and adult females with nodules. Based on previous literature, it was hypothesized that roughness and breathiness, rather than strain, would show the largest difference in perceptual quality between normal adult female speakers and adult females with vocal nodules. In contrast, based on a pilot study, it was hypothesized that roughness and strain, rather than breathiness, would show the greatest difference in perceptual quality between children with normal vocal pathology and children with vocal nodules.

Due to the rapid change in laryngeal structure over the course of development, a narrow age range of children 4-6 year-old was chosen to provide a more homogenous children's group. Thus, the stimuli which were recorded from 4-6 year-old children and adult females with normal voices or voices with vocal nodules were rated by each of the listeners. Each group contained eight subjects (i.e., 8 normal children, 8 children with nodules, 8 normal adult females, and 8 adult females with nodules).

Three types of stimuli were used: 1) a one second sustained vowel /a/ (the production having fundamental frequency closest to each subject's average fundamental frequency in continuous speech was chosen for perceptual analysis) produced by both children and adults, 2) the sentence weighted with voiceless plosives taken from the adult version of the CAPE-V produced by adults, and 3) the sentences weighted with voiceless plosives taken from the pediatric version of the CAPE-V produced by children. The CAPE-V sentences weighted with voiceless plosives were selected because the laryngeal adjustments required to produce these sentences were presumed to be the most dissimilar to those required for elicitation of the sustained vowel /a/. More specifically, to produce a sentence weighted with voiceless plosives, the vocal folds are adducted (closed) and

abducted (opened) numerous times, whereas for a sustained vowel, the vocal folds are adducted only once.

Experiment 1 was divided into two separate listening sessions separated by at least one day. In the first session, listeners ascribed a rating for each perceptual attribute to the production of the sustained vowel /a/ produced by all subjects, for a total of 64 vowel stimuli (32 subjects x 2 repetitions). In the second session, listeners ascribed a rating for each perceptual attributed to the production of adult and pediatric versions of the CAPE-V sentences weighted with voiceless plosives produced by all subjects, for a total of 64 sentence stimuli (32 subjects x 2 presentations). Since there were three sentences in the pediatric version of the CAPE-V sentence for every one sentence in the adult version of the CAPE-V sentence, the first two pediatric CAPE-V sentences produced by the children were concatenated to form one sentence for the purposes of rating. At the midpoint of both sessions of the experiment, listeners were provided a 10-minute break in order to optimize listener's attention and minimize fatigue.

The speech material that provided the largest perceptual difference between subjects with normal voices and voices with vocal nodules was determined using a four-factor analysis of variance (ANOVA) where there are four independent variables and one dependent variable. In an ANOVA, the main effects are the simple effects of an independent variable on a dependent variable. Hence, it is the effect of the independent variable alone averaged across the levels of the other independent variables. The four main effects for each perceptual attribute (overall severity, breathiness, roughness and strain) were: age (young children, adults), voice (normal, disordered), speech material (vowel, continuous speech) and rater. Material, age, and voice were fixed factors, and rater was a random variable. There is an interaction effect between two independent variables in the ANOVA if the effect of one independent variable depends on the levels of the second independent variable. The main effects and two-way interactions (age vs. voice, voice vs. speech material, age vs. speech material) were tested for significance using a p-value of 0.05.

In order to determine which perceptual parameters were the best predictors of differences in voice type (normal, disordered), binary logistic regression was performed

for each independent perceptual parameter (overall severity, roughness, breathiness, and strain). The dichotomous dependent variable was voice type (normal, disordered), and the covariate variables were rater and material.

As described above, it was hypothesized that the presence of nodules would differentially impact phonation in young children compared to adults due to children having a relatively immature vocal mechanism, and that these differences in phonation would be reflected as differences between salient perceptual attributes and/or speech material that elicits these perceptual attributes.

2.3.1 Results: Vowels

Within-rater reliability

Within-rater reliability for all of the rated stimuli was assessed by calculating the Pearson's correlation between the first and second ratings of the same stimuli for each listener. Tables 2.1 and 2.2 summarize the mean Pearson's r-value (the correlation between the first and second ratings) for each perceptual parameter for the production of the vowel /a/ by children and adults.

Listener	Overall Severity	Roughness	Breathiness	Strain
1	0.660	0.614	0.766	0.724
2	0.696	0.388	0.887	0.557
3	0.635	0.434	0.831	0.818
4	0.201	0.191	0.955	0.613
5	0.858	0.93	0.762	0.603
6	0.823	0.851	0.776	0.619
7	0.914	0.862	0.933	0.845
8	0.811	0.757	0.738	0.555
9	0.873	0.825	0.684	0.802
10	0.684	0.443	0.768	0.649
11	0.935	0.391	0.857	0.433
12	0.627	0.15	0.762	0.804
13	0.284	0.26	0.731	0.357
14	0.905	0.8	0.832	0.927
15	0.575	0.467	0.651	0.484
16	0.552	0.356	0.634	0.662
17	0.535	0.723	0.633	0.760

Table 2.1: Pearson's r-values for production of the vowel /a/ by children.

Listener	Overall Severity	Roughness	Breathiness	Strain
1	0.618	0.484	0.696	0.480
2	0.738	0.688	0.887	0.685
3	0.838	0.896	0.695	0.278
4	0.262	0.342	0.493	0.273
5	0.894	0.535	0.809	0.800
6	0.836	0.630	0.871	0.393
7	0.541	0.476	0.580	0.403
8	0.721	0.687	0.722	0.480
9	0.532	0.761	0.711	0.492
10	0.517	0.921	0.229	0.371
11	0.602	0.625	0.684	0.628
12	0.672	0.489	0.854	0.665
13	0.642	0.726	0.654	0.504
14	0.643	0.698	0.548	0.061
15	0.539	0.100	0.412	0.558
16	0.386	0.237	0.692	0.465
17	0.658	0.800	0.590	0.133

Table 2.2: Pearson's r-values for the production of the vowel /a/ by adults.

As seen in Tables 2.1 and 2.2, individual listeners varied in their ratings of vowels produced by children with correlations ranging from 0.150 to 0.955, with a mean correlation of 0.668. For the production of vowels by adults, correlations ranged from 0.061 to 0.921 with a mean Pearson's r value of 0.581. By squaring the Pearson's r values (i.e., r^2, which measures the amount of variance common to the two sets of ratings) it was determined that approximately 45% of variance was consistent across ratings for the production of vowels by children and 34% of variance was explained for the production of vowels by adults.

The mean Pearson correlation coefficients were highest for breathiness ratings, regardless of age group (children: mean r =0.776, adults: mean r = 0.655). The perceptual attribute with the lowest mean r value was roughness for the production of vowels by children (mean r = 0.555), and strain for the production of vowels by adults (mean r = 0.451).

Due to the large range of Pearson's correlation coefficients, further analysis of the perceptual data focused on ratings in which the Pearson's correlation coefficients were greater than 0.7. Overall, for children's production of vowels, 50% of the ratings had a Pearson's correlation coefficient greater than 0.7. When separating the ratings based on the individual vocal attributes, breathiness ratings had the highest percentage of ratings with Pearson's correlation coefficients greater than 0.7 (76.4%). By contrast, only 25% of the ratings for adults' production of vowels had Pearson's correlation coefficients greater than 0.7. Similar to children, among the perceptual attributes that were rated, breathiness ratings for adults had the greatest percentage of ratings with Pearson's correlation coefficient greater than 0.7 (35.3%). For the production of vowels by children, the perceptual attributes of overall severity, roughness and strain each had 41.2% of ratings with Pearson's correlation coefficients greater than 0.7, and for the adults' production of vowels, the perceptual attribute of strain had the lowest percentage with only 5.9% of the ratings having Pearson's correlation coefficient greater than 0.7.

Between-rater reliability

In order to investigate between-rater reliability, only listeners having ratings with a within-rater reliability (i.e., Pearson's correlation coefficient) greater than 0.7 were analyzed for the remaining measures. Intra-class correlation coefficients (ICCs) and 95% confidence intervals (CI) of all ratings were calculated to compare between-rater reliability across the different perceptual attributes.

Average ICCs were calculated using a two-way random ANOVA model. The average ICCs and 95% CI for the perceptual attributes of overall severity, breathiness, roughness and strain are shown in Table 2.3. Note that ICCs were not calculated for the perceptual attribute of strain for the production of vowels by adults because only one listener had a within-rater reliability with a correlation value greater than 0.7. As seen from Table 2.3, all of the perceptual attributes showed a moderately strong between-rater reliability for the production of vowels by children and adults except for the parameter strain for children, and roughness for adults. The between-rater reliability was strongest for perceptual ratings of breathiness for the production of vowels by children (mean ICC = 0.94), whereas, for adults, the measure with the strongest between-rater reliability was overall severity (mean ICC value =0.83).

Attributes	**Children vowel: ICC% (95% CI)**	**Adult vowel: ICC% (95% CI)**
Overall Severity	0.72 (0.40, 0.89)	0.83 (0.66, 0.93)
Roughness	0.71 (0.42, 0.88)	0.60 (0.21, 0.84)
Breathiness	0.94 (0.88, 0.98)	0.77 (0.49, 0.91)
Strain	0.64 (0.32, 0.85)	

Table 2.3: Intra-class correlation coefficient (ICC) for children and adult's production of /a/. Mean ICC and 95% Confidence Intervals (CI) for each perceptual attribute taken from listeners with Pearson's r-values > 0.7 for each perceptual attribute. Note that ICC was not calculated for the perceptual attribute of strain for the production of vowels by adults.

2.3.2 Results: Sentences

Within-rater reliability

Mean within-rater reliability values for each of the perceptual ratings of the production of the pediatric and adult versions of the CAPE-V sentences are summarized in Tables 2.4 and 2.5.

Listener	Overall Severity	Roughness	Breathiness	Strain
1	0.836	0.402	0.818	0.750
2	0.714	0.642	0.565	0.623
3	0.739	0.578	0.433	0.728
4	0.660	0.586	0.753	0.719
5	0.882	0.910	0.816	0.802
6	0.934	0.621	0.905	0.901
7	0.748	0.320	0.558	0.726
8	0.703	0.659	0.706	0.654
9	0.726	0.671	0.759	0.639
10	0.459	0.457	0.410	0.646
11	0.764	0.915	0.709	0.835
12	0.923	0.939	0.907	0.863
13	0.888	0.863	0.773	0.790
14	0.951	0.869	0.608	0.905
15	0.955	0.889	0.746	0.799
16	0.846	0.705	0.730	0.814
17	0.852	0.845	0.852	0.810

Table 2.4: Pearson's r values for production of the pediatric version of the CAPE-V sentences by children.

Listener	Overall Severity	Roughness	Breathiness	Strain
1	0.699	0.862	0.353	0.897
2	0.759	0.888	0.145	0.675
3	0.855	0.887	0.465	0.83
4	0.952	0.944	0.964	0.937
5	0.89	0.914	0.861	0.827
6	0.557	0.68	0.548	0.649
7	0.312	0.347	0.186	0.374
8	0.788	0.765	0.508	0.651
9	0.675	0.716	0.661	0.847
10	0.639	0.703	0.481	0.384
11	0.917	0.855	0.76	0.775
12	0.953	0.995	0.87	0.935
13	0.936	0.969	0.87	0.933
14	0.889	0.859	0.851	0.699
15	0.797	0.861	0.615	0.652
16	0.665	0.63	0.576	0.745
17	0.934	0.922	0.905	0.921

Table 2.5: Pearson's r values for production of the adult version of the CAPE-V sentences by adults.

As seen from Tables 2.4 and 2.5, the listeners in this study had correlation coefficient values between 0.320 and 0.955, with a mean r-value of 0.743 for rating the production of the pediatric version of the CAPE-V sentences by children (Table 2.4), and correlation coefficient values between 0.145 and 0.995, with a mean r-value of 0.741 for the production of the adult version of the CAPE-V sentences by adults (Table 2.5). The mean Pearson correlation coefficients were highest for overall severity ratings (mean r-value= 0.799) for children's production of the pediatric version of the CAPE-V sentences, whereas, for the production of the adult version of the CAPE-V sentences by adults, the mean Pearson's correlation coefficient was highest for roughness (mean r-value= 0.812). The perceptual attribute with the lowest Pearson's r-value was roughness for the production of pediatric CAPE-V sentences by children (mean r = 0.698) and breathiness for the production of the adult version of the CAPE-V sentences by adults (mean r = 0.625).

Due to the large range of Pearson's correlation coefficients among the listeners, further analysis of the perceptual data was focused on listeners with the highest within-rater reliability values (i.e., mean r > 0.7). For the production of the pediatric version of the CAPE-V sentences by children, 70.6% of the ratings were judged with Pearson's correlation coefficient greater than 0.7, with overall severity having the greatest percentage of ratings with Pearson's correlation coefficient greater than 0.7 (88.2%). For production of the adult version of the CAPE-V sentence by adults, 61.8 % of the ratings had Pearson's greater than 0.7, with roughness having the greatest percentage of ratings with Pearson's correlation coefficients greater than 0.7 (82.3%). By contrast, roughness had the lowest percentage of ratings with Pearson's correlation coefficient greater than 0.7 (47%) for production of sentences by children, whereas breathiness ratings had the lowest percentage of ratings with Pearson's correlation coefficient greater than 0.7 (41.2%) for production of sentences by adults.

Between-rater reliability

To investigate between-rater reliability, only listeners with a Pearson's correlation coefficient greater than 0.7 were used in the remaining analysis. The average intra-class

correlation coefficients (ICCs) and 95% CI for the perceptual attributes of overall severity, breathiness, roughness, and strain are shown in Table 2.6. Production of the pediatric and adult versions of the CAPE-V sentences showed strong associations for each of the perceptual attributes. Between-rater reliability was highest for breathiness ratings for sentence production by children (ICC = 0.97), whereas roughness ratings showed the highest between-rater reliability for sentence production by adults (ICC=0.96).

Attributes	Children sentence: ICC% (95% CI)	Adult sentence: ICC% (95% CI)
Overall Severity	0.94 (0.86, 0.98)	0.93 (0.86, 0.97)
Roughness	0.93 (0.82, 0.97)	0.96 (0.92, 0.98)
Breathiness	0.97 (0.94, 0.99)	0.89 (0.77, 0.96)
Strain	0.91 (0.82, 0.97)	0.85 (0.70, 0.94)

Table 2.6: Intra-class correlation coefficient (ICC) for children and adult's production of CAPE-V. Average ICC and 95% Confidence Intervals (CI) for each perceptual attribute from listeners having Pearson's r values > 0.7 for each perceptual attribute.

2.3.3 Results: Comparing vowels and sentences

A summary of the percentages of reliable raters with Pearson's correlation coefficient greater than 0.7 and associated mean Intra-class correlation coefficients (ICCs) for judgments of vowels and sentences from children and adults are shown in Table 2.7.

	Reliable Raters (%)		ICC (%)	
	Vowel	Sentence	Vowel	Sentence
Children	0.5	0.71	0.75	0.94
Adult	0.25	0.62	0.73	0.91

Table 2.7: Percentage of listeners having Pearson's r values > 0.7 and mean intra-class correlation coefficients (ICCs) for reliable raters for production of vowel and sentences from both children and adult females.

Prior to performing a statistical analysis to determine which speech material provided the largest perceptual difference between normal voices and voices with vocal nodules for both children and adults, box plots were created to visually examine trends in

the data. Figures 2.4-2.7 show groups of numerical data for the perceptual ratings of overall severity (Figure 2.4), roughness (Figure 2.5), breathiness (Figure 2.6), and strain (Figure 2.7) for the production of vowels and sentences by 4-6 year old children and adult females.

The box plots indicate the median, lower and upper quartiles, and minimum and maximum data values for each of the perceptual attributes which were rated. Each box in the box plot represents values in the inter-quartile range (25th to 75th percentile) of the data set. The ends of the vertical lines in the box plot, known as whiskers, indicate the minimum and maximum data values, unless outliers are present, in which the whiskers extend to a maximum of 1.5 times the inter-quartile range. Data points greater than 1.5 times the inter-quartile range were considered outliers.

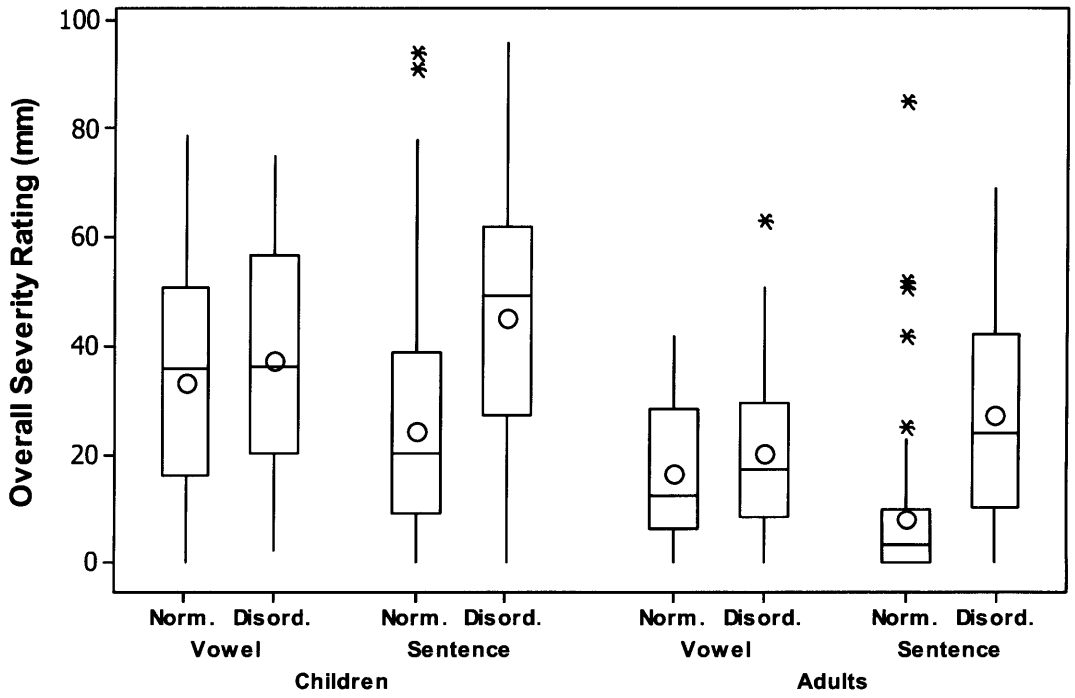


Figure 2.4: Children and adult's overall severity ratings for /a/ and CAPE-V. Box plot of overall severity ratings for the production of vowels and sentences by 4-6 year old children and adult females, with normal (Norm.) or disordered (Disord.) voices. Circles indicate the mean value, and stars (*) indicate outlier values.

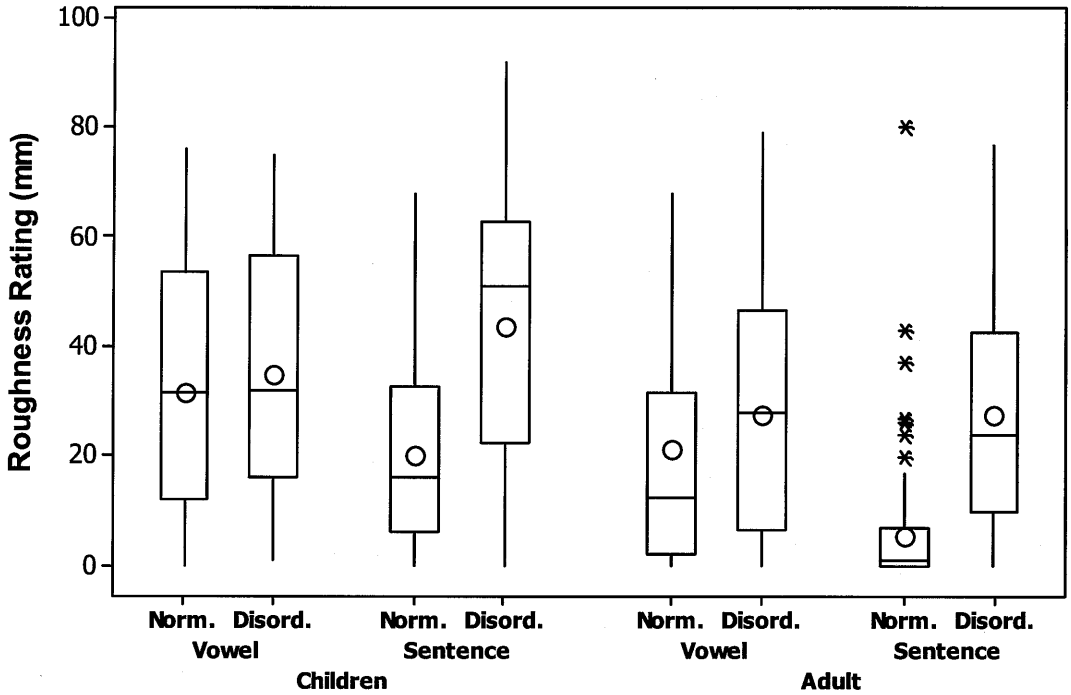


Figure 2.5: Children and adult's roughness ratings for /a/ and CAPE-V. Box plot of roughness ratings for the production of vowels and sentences by 4-6 year old children and adult females, with normal (Norm.) or disordered (Disord.) voices. Circles indicate the mean value and stars (*) indicate outlier values.

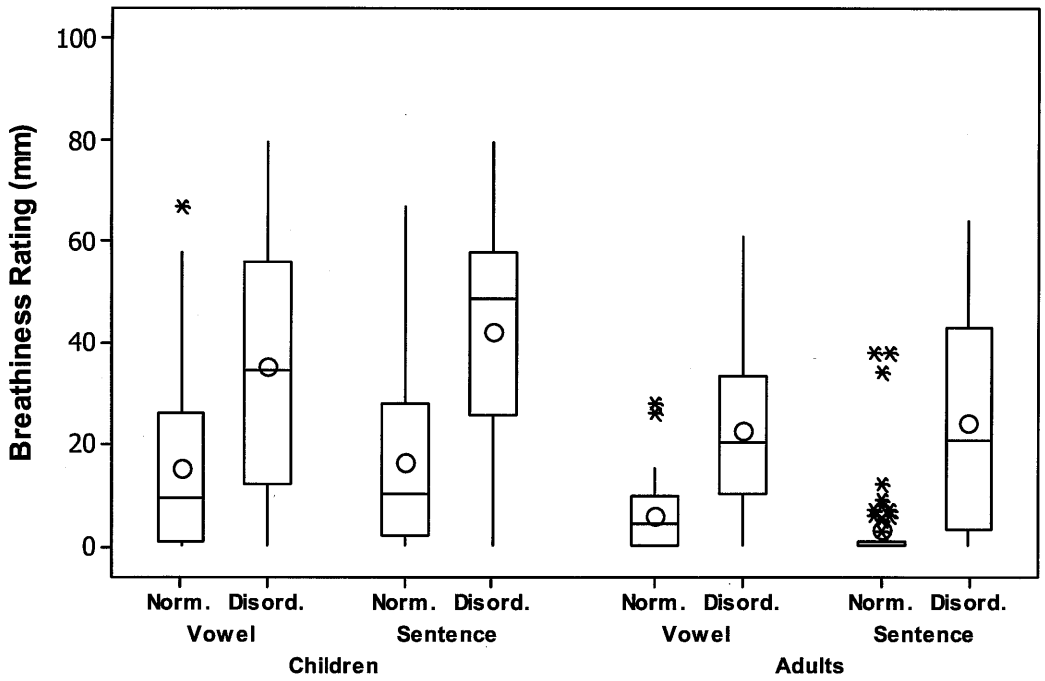


Figure 2.6: Children and adult's breathiness ratings for /a/ and CAPE-V. Box plot of breathiness ratings for the production of vowels and sentences by 4-6 year old children and adult females, with normal (Norm.) or disordered (Disord.) voices. Circles indicate the mean value and stars (*) indicate outlier values.

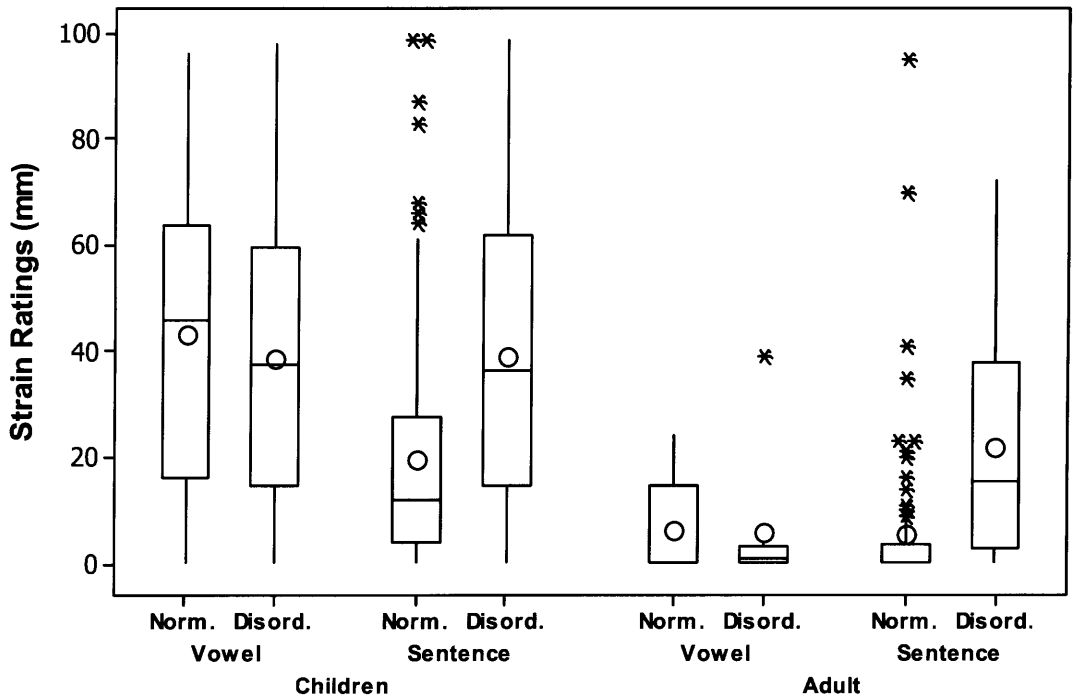


Figure 2.7: Children and adult's strain ratings for /a/ and CAPE-V. Box plot of strain ratings for the production of vowels and sentences by 4-6 year old children and adult females, with normal (Norm.) or disordered (Disord.) voices. Circles indicate the mean value and stars (*) indicate outlier values.

Figures 2.4-2.7 show that there was a greater difference between the means of normal voices and voices with vocal nodules (i.e. disordered voices) for the production of sentences compared to the production of vowels irrespective of age or perceptual quality. Additionally, the ranges and variances of all the perceptual ratings (overall severity, roughness, breathiness and strain) were smaller for adult females and skewed more toward the normal end compared to 4-6 year-old children. Since fifty percent of the perceptual ratings of adult sentences occupy a very small perceptual range in Figures 2.4-2.7, ratings for the production of adult sentences had the most outliers.

Table 2.8 shows the results of a four-way ANOVA calculated to determine the statistical significance of the effects observed in Figures 2.4-2.7. A brief description of the ANOVA parameters are described next, but for further clarification please refer to Cohen (1996). In an ANOVA, the F-ratio is the statistic used to test the hypothesis that the means are significantly different from one another. It is defined as the ratio of the

between group variation divided by the within group variation to the mean square within groups. The number of degrees of freedom for the model (df) is equal to the number of values in the final calculation of a statistic that are free to vary. In statistical hypothesis testing, the p-value is a measure of how much evidence we have against the hypothesis that the means of the two groups are the same.

As observed in Table 2.8, each of the factors of rater, age and voice were significant for all of the perceptual attributes. However, the effect of speech material (vowels vs. sentences) was not significantly different for any of the perceptual attributes. The interaction between speech material (vowels vs. sentences) and voice (normal vs. disordered) was significant for each of the perceptual parameters except breathiness.

Variable	Overall Severity	Roughness	Breathiness	Strain
Rater				
F (df)	13.458 (15,586)	10.151 (16,521)	8.501 (16,585)	14.215 (13,476)
p	< 0.001	< 0.001	< 0.001	< 0.001
Material				
F (df)	2.156 (1,586)	3.333 (1,521)	1.208 (1,585)	3.633 (1,476)
p	0.143	0.068	0.272	0.057
Age				
F (df)	136.202 (1,586)	84.809 (1,521)	94.308 (1,585)	43.778 (1,476)
p	< 0.001	< 0.001	< 0.001	< 0.001
Voice				
F (df)	67.756 (1,586)	79.535 (1,521)	205.566 (1,585)	8.992 (1,476)
p	< 0.001	< 0.001	< 0.001	0.003
Material * Age				
F (df)	0.875 (1,586)	0.108 (1,521)	0.34 (1,585)	0 (1,476)
p	0.35	0.742	0.56	0.998
Material * Voice				
F (df)	33.009 (1,586)	33.142 (1,521)	3.801 (1,585)	29.135 (1,476)
p	< 0.001	< 0.001	0.052	< 0.001
Age * Voice				
F (df)	0.13 (1,586)	< 0.001 (1,521)	2.305 (1,585)	0.381 (1,476)
p	0.718	0.978	0.129	0.537

Table 2.8: Analysis of Variance (ANOVA) table for children and adult's production of /a/ and CAPE-V. Results of 4-way ANOVA to determine the differences in perceptual ratings for the production of vowels and sentences for children versus adults. Star (*) indicates interaction of two main effects.

The inclusion of non-significant interactions (i.e. p> 0.05) has been shown to affect the significance of the main effects in an ANOVA (Cohen, 1996). Thus, the ANOVA was recalculated after excluding the non-significant factors, with the results being shown in Table 2.9. The results in Table 2.9 confirm that the interaction of speech material and voice (i.e., material * voice) was not significant for the perceptual attribute of breathiness.

Variable	Overall Severity	Roughness	Breathiness	Strain
Rater				
F (df)	13.423 (15,588)	10.474 (16,523)	8.622 (16,587)	15.175 (13,478)
p	< 0.001	< 0.001	< 0.001	< 0.001
Age				
F (df)	150.082 (1,588)	98.732(1,523)	93.869 (1,587)	116.133 (1,478)
p	< 0.001	< 0.001	< 0.001	< 0.001
Voice				
F (df)	70.327(1,588)	80.066 (1,523)	243.693 (1,587)	12.372 (1,478)
p	< 0.001	< 0.001	< 0.001	< 0.001
Material * Voice				
F (df)	17.904 (2,588)	19.203 (2,523)	2.207 (2,587)	18.412 (2,478)
p	< 0.001	< 0.001	0.111	< 0.001

Table 2.9: Recalculated ANOVA with non-significant interactions extracted from Table 2.8. Star (*) indicates interaction of two main effects.

The lack of any significant interactions for age and material suggests that the significant main effect of age ($p < 0.001$) does not differ between the productions of vowels versus sentences.

The main effects of age and voice were statistically significant, and these main effects were further investigated by examining the differences in the means and standard deviations for these two factors as shown in Tables 2.10 and 2.11. Table 2.10 shows that in general, perceptual attributes were rated more severely for children compared to adults. Similarly, Table 2.11 shows that listeners rated normal voices as being less severe than voices with vocal nodules (i.e., disordered voices).

Age	Overall Severity	Roughness	Breathiness	Strain
Child	36.4 (1.0)	33.0 (1.3)	26.0 (0.9)	33.7 (1.2)
Adult	17.8 (1.2)	15.0 (1.2)	10.8 (1.4)	10.6 (2.0)

Table 2.10: Perceptual ratings of /a/ and CAPE-V for children and adults. Mean and standard error for each perceptual attribute (rated using a 0-100 mm visual analog scale) for 4-6 year-old children and adult females.

Voice	Overall Severity	Roughness	Breathiness	Strain
Normal	21.1 (1.1)	17.2 (1.1)	7.7 (1.1)	18.6 (1.6)
Disorderd	33.1 (1.1)	30.8 (1.1)	29.2 (1.1)	25.7(1.6)

Table 2.11: Perceptual ratings of /a/ and CAPE-V for normal voices and voices with vocal nodules. Mean and standard error for each perceptual attribute (rated using a 0-100 mm visual analog scale) for subjects with normal and disordered voices (i.e., voices with vocal nodules).

The mean and standard error for the main effects of speech material and voice were also calculated to examine their interactions. As shown in Figures 2.4-2.7, Table 2.9 confirms a statistical significant difference between speech material and voice for all of the perceptual parameters except breathiness. Table 2.12 further shows that the largest perceptual difference between normal voices and voices with vocal nodules for each of the perceptual parameters except breathiness was for the production of sentences. These findings hold for both children and adults, since a separate analysis showed no significant age-by-material-by-voice interactions ($p > 0.4$).

Material	Voice	Overall Severity	Roughness	Breathiness	Strain
Vowel	Normal	26.6 (1.9)	23.5 (1.9)	8.2 (1.5)	27.2 (2.8)
	Disordered	30.4 (1.9)	28.1 (1.9)	27.1 (1.5)	23.3 (2.8)
Sentence	Normal	15.6 (1.2)	11.1 (1.4)	7.2 (1.5)	9.9 (1.5)
	Disordered	35.9 (1.2)	33.4 (1.4)	31.2 (1.5)	28.1 (1.5)

Table 2.12: Mean and standard error for each perceptual attribute (rated using a 0-100 mm visual analog scale) for the interaction of material and voice.

A binary logistic regression was performed to determine which perceptual quality best distinguished normal voices and voices with vocal nodules for 4-6 year-old children

and adult females. Tables 2.13 and 2.14 show that ratings for all of the perceptual attributes significantly distinguished between normal voices and voices with vocal nodules. Additionally, the ratio of the regression coefficient (β) to its standard error (SE) was calculated, and the resulting values were ordered from highest to lowest to determine the order of significance for the perceptual attributes. Thus, although each of the perceptual attributes significantly distinguished between normal voices and voices with vocal nodules for both children and adults, breathiness was the perceptual attribute that best distinguished between normal voices and voices with vocal nodules for children (Table 2.13), whereas roughness best differentiated between normal voices and voices with vocal nodules for adults (Table 2.14).

	β (SE)	p	β/SE
Breathiness	5.114 (0.572)	< 0.001	8.941
Overall Severity	3.371 (0.543)	< 0.001	6.208
Roughness	2.92 (0.628)	< 0.001	4.650
Strain	1.695 (0.449)	< 0.001	3.775

Table 2.13: Perceptual quality differentiating children with normal voices and voices with vocal nodules. Binary logistic regression coefficient (β), standard error (SE), p-value and binary logistic regression coefficients divided by standard error for 4-6 year old children.

	β (SE)	p	β/SE
Roughness	5.949 (0.845)	< 0.001	7.040
Overall Severity	5.946 (0.987)	< 0.001	6.024
Breathiness	11.472 (1.949)	< 0.001	5.886
Strain	5.418 (1.216)	< 0.001	4.456

Table 2.14: Perceptual quality differentiating adult females with normal voices and voices with vocal nodules. Binary logistic regression coefficient (β), standard error (SE), p-value and binary logistic regression coefficients divided by standard error for adult females.

2.4 Perceptual Experiment 2

The goal of Perceptual Experiment 2 was to determine whether the picture stimuli in the pediatric version of the CAPE-V sentences elicited speech that was perceptually similar to the speech elicited by the adult version of the CAPE-V sentences. This had to be done with recordings from the 8-10 year-old group because younger children could

not reliability read the adults CAPE-V sentences. Hence, in this experiment, listeners rated speech stimuli from sixteen 8-10 year-old children (8 children with normal voices and 8 children with nodules). Each of the children produced both the adult and pediatric versions of the CAPE-V sentences weighted with voiceless plosives. A total of 64 sentence stimuli were recorded (16 subjects x 2 repetitions x 2 sentences).

A three-factor ANOVA was used to determine significant differences between the voice quality elicited by the two versions of the CAPE-V sentences for each perceptual attribute (overall severity, breathiness, roughness and strain). The three-factor ANOVA was used to calculate each of the three main effects: voice (normal, disordered), speech material (pediatric, adult CAPE-V sentences), and rater, and the two-way interaction (voice vs. speech material). It was hypothesized that differences between the voice samples elicited from the two versions of the CAPE-V sentences would not be observed for any of the perceptual attribute ratings.

2.4.1 Results: Pediatric vs Adult version of the CAPE-V sentence

Within-rater reliability

Mean within-rater reliability values for the ratings of the production of pediatric and adult versions of the CAPE-V sentences are summarized for each perceptual attribute in Tables 2.15 and 2.16.

Listener	Overall Severity	Roughness	Breathiness	Strain
1	0.762	0.732	0.232	0.899
2	0.764	0.86	0.438	0.807
3	0.774	0.811	0.797	0.587
4	0.844	0.831	0.726	0.8
5	0.887	0.852	0.768	0.763
6	0.862	0.662	0.625	0.654
7	0.594	0.637	0.256	0.283
8	0.809	0.713	0.466	0.634
9	0.719	0.748	0.758	0.63
10	0.956	0.913	0.645	0.701
11	0.837	0.83	0.524	0.612
12	0.822	0.901	0.859	0.668
13	0.847	0.897	0.817	0.783
14	0.919	0.947	0.642	0.788
15	0.847	0.856	0.747	0.895
16	0.778	0.859	0.532	0.76
17	0.846	0.892	0.79	0.781

Table 2.15: Pearson's r values for production of pediatric version of the CAPE-V sentences by 8-10 year-old children.

Listener	Overall Severity	Roughness	Breathiness	Strain
1	0.605	0.429	0.867	0.607
2	0.763	0.342	0.732	0.869
3	0.772	0.561	0.88	0.743
4	0.696	0.733	0.569	0.667
5	0.896	0.841	0.879	0.945
6	0.558	0.491	0.834	0.842
7	0.278	0.079	0.985	0.642
8	0.627	0.588	0.414	0.494
9	0.77	0.71	0.76	0.902
10	0.448	0.522	0.611	0.308
11	0.727	0.643	0.883	0.803
12	0.838	0.914	0.772	0.867
13	0.901	0.877	0.874	0.649
14	0.838	0.851	0.833	0.47
15	0.67	0.677	0.701	0.958
16	0.841	0.344	0.834	0.723
17	0.729	0.767	0.803	0.81

Table 2.16: Pearson's r values for production of the adult version of the CAPE-V sentences by 8-10 year-old children.

The range of Pearson's correlation coefficients for listeners in this study was between 0.232 and 0.956, with a mean value of 0.742 for rating the production of the pediatric version of the CAPE-V sentences. By contrast, the range of Pearson's correlation coefficients for rating the production of the adult version of the CAPE-V sentences was between 0.079 and 0.985, with a mean r-value of 0.70. The highest within-rater reliability was for roughness ratings from judging the production of the pediatric version of the CAPE-V sentences (mean r = 0.820), whereas breathiness ratings had the highest within-rater reliability for rating the production of the adult version of the CAPE-V sentences (mean r = 0.778). The perceptual attribute with the lowest Pearson's correlation coefficient was breathiness for rating the production of pediatric CAPE-V sentences (mean r = 0.625) and roughness for rating the production of the adult version of the CAPE-V sentences (mean r = 0.610).

Due to the large range in Pearson's correlation coefficients, further analysis of the perceptual data focused on listeners having ratings with Pearson's correlation coefficients greater than 0.7. For the production of the pediatric CAPE-V sentences, 72.1% of the ratings were judged with Pearson's correlation coefficients greater than 0.7. When examining the individual perceptual attributes, overall severity ratings had the greatest percentage of ratings with Pearson's correlation coefficient greater than 0.7 (94.1%). For the production of the adult CAPE-V sentences, 60.3 % of the ratings had Pearson's correlation coefficients greater than 0.7. Breathiness had the greatest percentage of ratings (82.3%) with Pearson's r greater than 0.7 for the production of adult CAPE-V sentences. For the production of pediatric CAPE-V sentences, breathiness had the lowest percentage of ratings with Pearson's r greater than 0.7 (47%), whereas for the production of the adult CAPE-V sentences, roughness had the lowest percentage of ratings with Pearson's r greater than 0.7 (41.2%).

Between-rater reliability

The average ICCs and 95% CI for the perceptual attributes of overall severity, breathiness, roughness, and strain for all the raters are shown in Table 2.17. Production of the pediatric and the adult versions of the CAPE-V sentences showed strong

associations for all of the perceptual attributes. Between-rater reliability was highest for overall severity and breathiness for the production of the pediatric version of the CAPE-V sentences (ICC = 0.94) whereas, between-rater reliability was highest for roughness and breathiness ratings for the production of the adult version of the CAPE-V sentences (ICC = 0.92). Average ICC for the pediatric CAPE-V was 0.93, whereas the average ICC for the adult CAPE-V was 0.89.

Attributes	Pediatric CAPE-V: ICC% (95% CI)	Adult CAPE-V: ICC% (95% CI)
Overall Severity	0.94 (0.86, 0.97)	0.89 (0.79, 0.96)
Roughness	0.92 (0.84, 0.97)	0.92 (0.84, 0.97)
Breathiness	0.94 (0.86, 0.98)	0.92 (0.85, 0.97)
Strain	0.92 (0.82, 0.97)	0.84 (0.69, 0.94)

Table 2.17: Average ICC and 95% Confidence Intervals for each perceptual attribute for listeners having Pearson's r values > 0.7 for each CAPE-V parameter.

Prior to performing a statistical analysis to determine whether the perceptual characteristics elicited from the pediatric and adult version of the CAPE-V sentence were similar, box plots representing the data were created as shown in Figures 2.8-2.11. Groups of numerical data for the perceptual ratings of overall severity (Figure 2.8), roughness (Figure 2.9), breathiness (Figure 2.10), and strain (Figure 2.11) for production of the pediatric and adult versions of the CAPE-V sentences by 8-10 year-old children are shown.

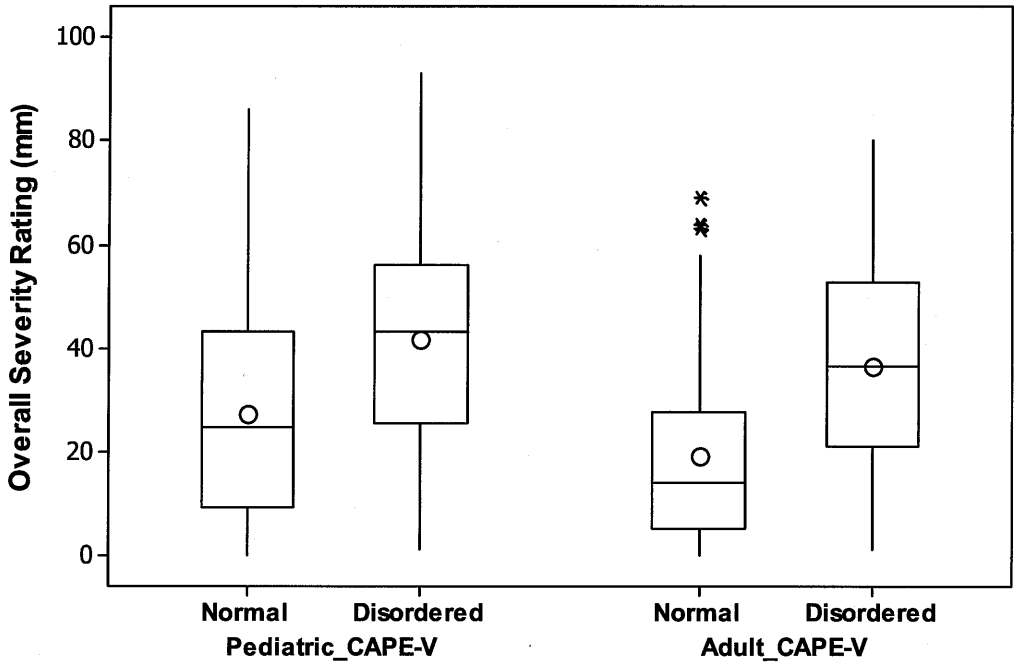


Figure 2.8: Box plot of overall severity ratings for production of the pediatric and adult versions of the CAPE-V sentences by 8-10 year-old children with normal voices or voices with vocal nodules. Circles indicate the mean value and stars (*) indicate outlier values.

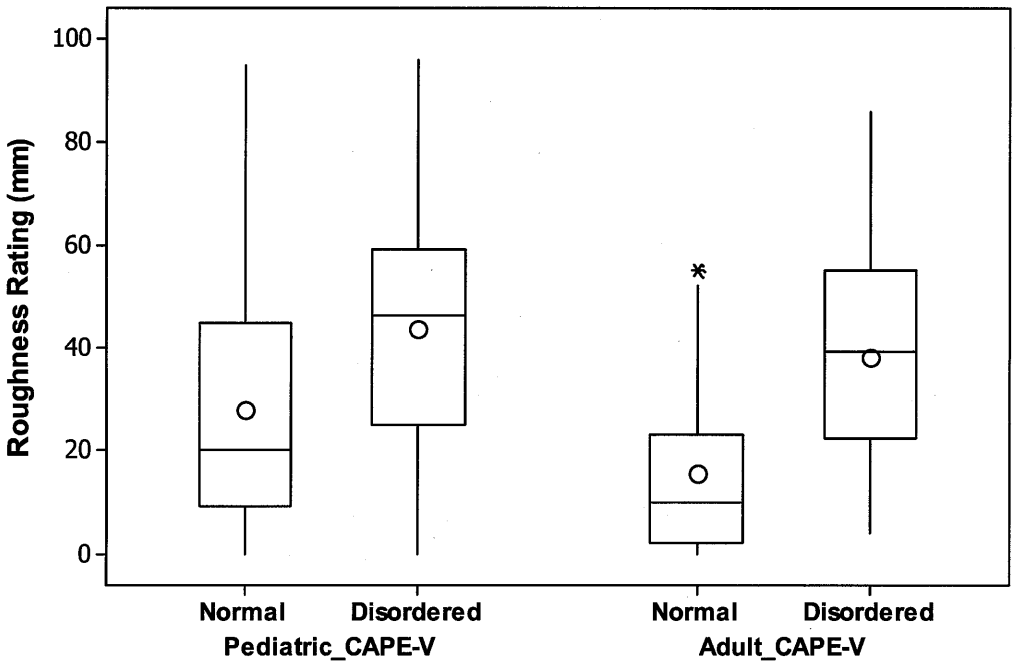


Figure 2.9: Box plot of roughness ratings for the production of the pediatric and adult versions of the CAPE-V sentences by 8-10 year-old children with normal voices or voices with vocal nodules. Circles indicate the mean value and stars (*) indicate outlier values.

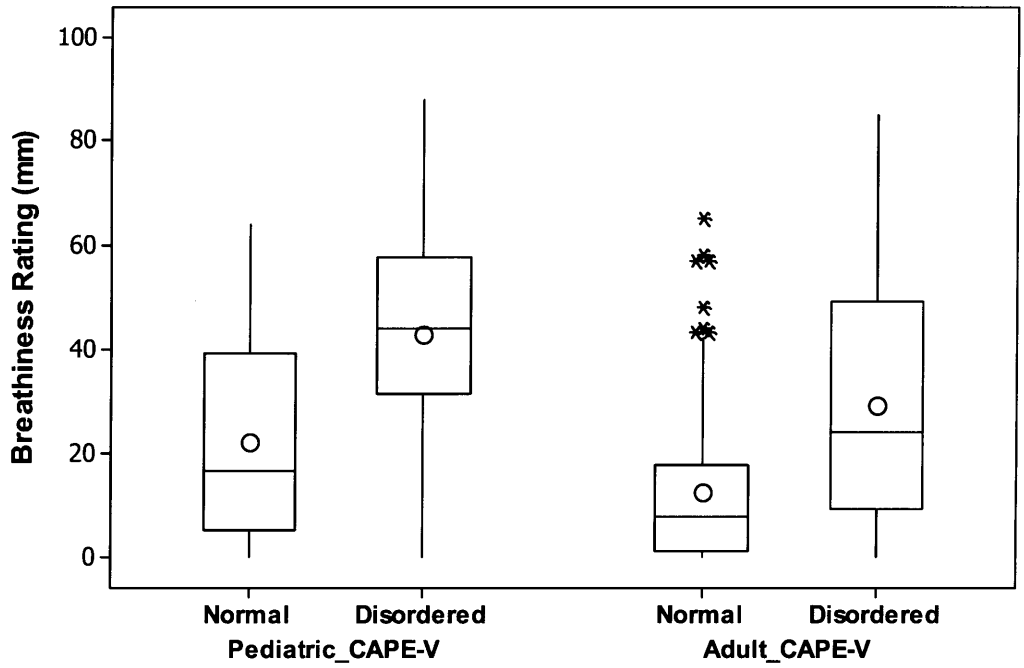


Figure 2.10: Box plot showing breathiness ratings for 8-10 year old children, with normal voices or voices with vocal nodules, producing both pediatric and adult version of the CAPE-V sentences. Circle indicates mean value and stars (*) indicates outliers.

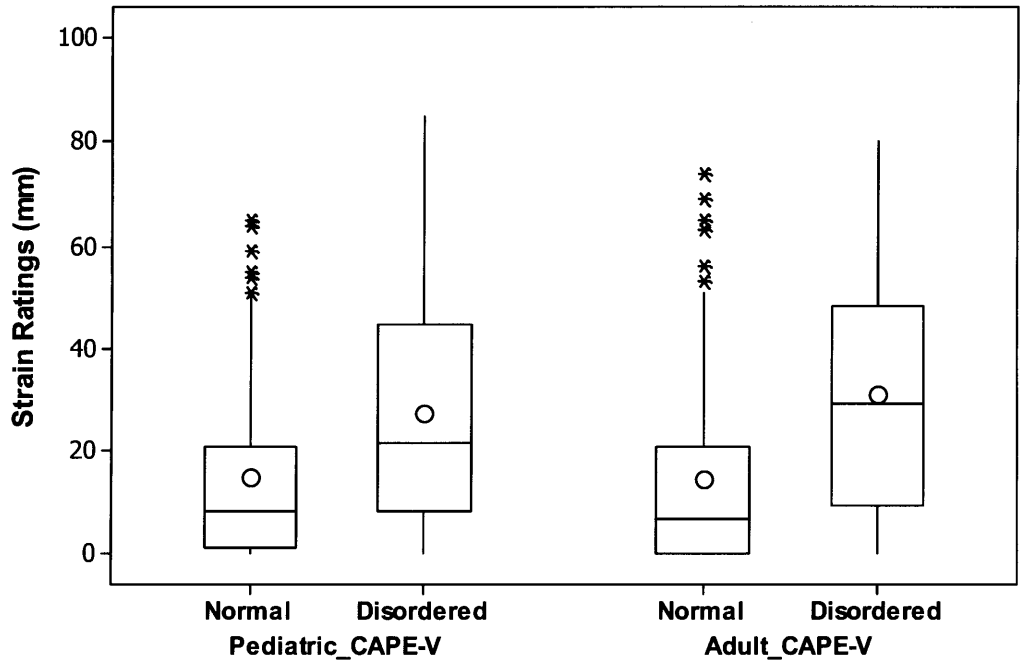


Figure 2.11: Box plot of strain ratings for the production of the pediatric and adult versions of the CAPE-V sentences by 8-10 year-old children with normal voices or voices with vocal nodules. Circles indicate the mean value and stars (*) indicate outlier values.

As seen from Figures 2.8-2.11, the speech samples produced by the pediatric and adult versions of the CAPE-V both resulted in the group with nodules being rated as more disordered for all of the perceptual attributes. However, the pediatric version tended to produce higher average CAPE-V scores (more disordered) and larger ranges and variances as compared to the adult sentences.

A three-factor ANOVA was used to examine the statistical significance of the above findings. In the ANOVA, material (pediatric version of CAPE-V sentences, adult version of CAPE-V sentences) and voice (normal, disordered) were used as fixed factors and rater was used as a random factor. Table 2.18 shows that the main effects of rater, material and voice were significant for each of the perceptual parameters, but the interaction between material and voice (i.e., material * voice) was not significant (i.e., $p > 0.05$).

Variable	**Overall Severity**	**Roughness**	**Breathiness**	**Strain**
Rater				
F (df)	11.664 (15,397)	9.362 (14,334)	8.98 (14,334)	12.112 (14,302)
p	< 0.001	< 0.001	< 0.001	< 0.001
Material				
F (df)	32.688 (1,397)	34.161 (1,334)	8.952 (1,334)	4.696 (1,302)
p	< 0.001	< 0.001	0.003	0.031
Voice				
F (df)	88.952 (1,397)	79.562 (1,334)	95.27 (1,334)	59.182 (1,302)
p	< 0.001	< 0.001	< 0.001	< 0.001
Material * Voice				
F (df)	0.791 (1,397)	2.208 (1,334)	1.19 (1,334)	1.375 (1,302)
p	0.374	0.138	0.276	0.242

Table 2.18: Results of 3-way ANOVA to determine the differences in perceptual ratings for the production of pediatric and adult version of the CAPE-V sentences for 8-10 year old children. Star (*) indicates interaction of two main effects.

The inclusion of non-significant interactions has been shown to affect the significance of the main effects (Cohen, 1996). Thus, the ANOVA was recalculated after extracting the non-significant factors from Table 2.18.

Variable	Overall Severity	Roughness	Breathiness	Strain
Rater				
F (df)	11.670 (15,398)	9.328 (14,335)	8.975 (14,335)	12.097 (14,303)
p	< 0.001	< 0.001	< 0.001	< 0.001
Material				
F (df)	32.705 (1,398)	34.038 (1,335)	8.947 (1,335)	4.690 (1,303)
p	< 0.001	< 0.001	0.003	0.031
Voice				
F (df)	89.959 (1,398)	80.622 (1,335)	96.692 (1,335)	59.109 (1,303)
p	< 0.001	< 0.001	< 0.001	< 0.001

Table 2.19: Recalculated ANOVA with non-significant interactions extracted from Table 2.18. Star (*) indicates interaction of two main effects.

Mean values for the material (pediatric/adult version of CAPE-V) and voice (normal/disordered) are shown in Tables 2.20 and 2.21. In general, the ratings for the production of the pediatric version of the CAPE-V sentences were rated significantly more severe than the ratings for the production of the adult version of the CAPE-V sentences. Additionally, normal voices were rated as being less severe compared to voices with vocal nodules.

Material	Overall Severity	Roughness	Breathiness	Strain
Pediatric version of CAPE-V	34.5 (1.1)	35.6 (1.2)	26.6 (1.9)	25.2 (1.7)
Adult version of CAPE-V	23.7 (1.6)	21.1 (2.2)	19.8 (1.2)	19.4 (1.7)

Table 2.20: Perceptual ratings of pediatric and adult CAPE-V sentence. Mean ratings and standard error for each perceptual attribute (rated using a 0-100 mm visual analog scale) for the production of the pediatric versus adult version of the CAPE-V sentences from 8-10 year old children.

Voice	Overall Severity	Roughness	Breathiness	Strain
Normal	21.1 (1.3)	18.9 (1.6)	14.0 (1.4)	15.0 (1.4)
Disorderd	37.1 (1.3)	37.9 (1.6)	32.4 (1.4)	29.5 (1.4)

Table 2.21: Perceptual ratings of CAPE-V sentence for normal voices and voices with vocal nodules. Mean ratings and standard error for each perceptual attribute (rated using a 0-100 mm visual analog scale) comparing normal voices and voices with vocal nodules of 8-10 year old children.

2.5 Discussion

In Perceptual Experiment 1 it was found that the voices of children with vocal nodules were rated as more disordered (severe) than the voices of adults with nodules, regardless of the type of speech material produced. One possible implication of this finding is that the presence of nodules had a greater impact on vocal function in children than in adults. Physiologically, nodules may have a greater impact on vocal fold function in children because they have a relatively smaller portion of the glottis that is comprised of vibratory membrane (approximately half) as compared to the adult (approximately two-thirds). The smaller structural size of the pediatric vocal fold could make its vibratory function more sensitive/susceptible to the presence of a growth like a nodule, thus creating a higher degree of dysphonia.

In Perceptual Experiment 1 it was also found that ratings for the production of sentences (overall severity, roughness and strain) were better at differentiating normal voices and voices with vocal nodules than ratings for the production of vowels. Indeed, this finding appears to hold for both children and adults as evidenced by the lack of a significant interaction between speech material and age. This result is generally supportive of our view that sentence production would be a better platform for more accurately/realistically delineating disordered voice quality. Results did not show a statistically significant difference between vowels and sentences in terms of how severe each perceptual attribute was rated. This finding is at odds with a previous study which reported that listeners rated dysphonia as being more severe when they listened to a sustained vowel compared to when they listened to a continuous speech sample (Zraick, et al., 2005).

Perceptual Experiment 1 has demonstrated that differences between children and adults with normal voices and voices with vocal nodules may best be captured by evaluating their production of continuous speech via reading sentences or picture naming. Thus, it seems that the procedure created in this study for eliciting continuous speech in a standard fashion from children as young as 3 years old is an important contribution to the pediatric voice evaluation process.

The results for Perceptual Experiment 2 did not support the prediction that there would be no differences between the perceptual ratings elicited by the pediatric and adult versions of the CAPE-V sentences. Although ratings for the production of both the pediatric and adult version of the CAPE-V sentences were successful in distinguishing between subjects with normal voices and voices with vocal nodules, in general, the pediatric CAPE-V sentences were rated as more severe than the production of the sentences in the adult version of the CAPE-V. The reason for these differences in severity ratings between the two versions of the CAPE-V is not readily apparent. Perhaps differences in the mode of speech elicited by the two types of tasks had an unexpected impact on voice quality ratings. The adult version entailed more fluent/natural reading of a sentence, whereas the pediatric version involved more staccato production of monosyllables to name pictures in a repeating carried phrase, which also produced a more rhythmic cadence. Also, the higher reliability of ratings of the pediatric CAPE-V sentences could be due to the fact that there could be less time variation in the production of pediatric CAPE-V sentences than for the adult CAPE-V sentences. The higher ratings (more disordered) and larger ranges elicited by the pediatric version of the CAPE-V may actually be clinically advantageous if it provides better differentiation from normal and better separation between different levels of dysphonia severity. Perhaps an inadvertent outcome of designing the pediatric version is that it is elicits vocal function that is more sensitive to the presence of nodules than normal read speech. These conjectures would need to be explored in studies involving larger numbers of subjects.

Also supporting clinical use of the pediatric CAPE-V were the findings that it was associated with higher within-rater and between-rater reliability than the adult version. This may have also been a by-product of the larger overall range (spread of scores) in listener ratings than were elicited by the adult version.

Hypotheses about which perceptual features would predominate in differentiating normal speakers from those with nodules in children versus adults were not entirely supported by the results of the perceptual experiments. It was predicted that roughness and strain would be the main features in differentiating normal children's voices from those with nodules, but results instead showed that breathiness was more dominant than either of these. For adults, the prediction that roughness would be one of the top two features in differentiating normal voices from those with nodules was borne out by the results, but breathiness did not end up being as dominant as expected. Even though all of the perceptual parameters showed significant differences between normal and nodules groups in children and adults, strain was always the least dominant, which is consistent with the finding that it was also the least reliability rated parameter.

The listeners in this study showed a lower within-rater reliability for vowels ($r = 0.62$) compared to sentences ($r = 0.74$). This may be due to the fact that even though a commonly used duration of 1-second vowel samples was employed (Awan & Roy, 2006), listeners in the present study indicated that they had difficulty rating the vowels due to the unnaturalness associated with such a short production.

Chapter 3: Implementation, Validation and Application of Acoustic Analysis Algorithms

In view of the aforementioned limitations inherent in the perceptual analysis of voice, the experiments in this chapter focus on determining reliable objective and quantitative methods of voice analysis. Acoustic methods of evaluating voice are advantageous because the acoustic algorithms may be applied to a variety of voice samples in a consistent manner, with the results being provided in a numerical format that can more easily be compared with "normal" values (Awan & Roy, 2006). Another benefit of investigating the acoustic characteristics of voice is that such measures can provide non-invasive estimates of vocal fold function (Colton & Casper, 1996).

In the second portion of this chapter, the following acoustic measures were calculated for both sustained vowels and continuous speech produced by children and female adults: mean fundamental frequency, first cepstral peak (R1, CPP), cepstrum-based HNR, and overall spectral slope. The acoustic algorithms were first validated and then adapted for use with children. Synthesized samples for the normal voice production of the vowel /a/ with representative parameters (e.g., formant frequencies) for adult females, 9 year-old children, and 5 year-old children were used for validation of the acoustic algorithms. Various parameters (e.g., fundamental frequency, jitter, shimmer, noise, source spectral slope) of the synthesized vowels were changed to determine the extent to which the output of the acoustic algorithms was sensitive to changes in these parameters.

3.1 Acoustic Measures

An overview of the various acoustic measures was provided in Chapter 1. The following sections describe details regarding how each of these algorithms was implemented in this thesis.

3.1.1 Fundamental Frequency

Since there is widespread agreement that pitch is an important factor in characterizing both the etiology and symptoms of vocal disorders (Baken & Orlikoff, 1987), this thesis included measurements of fundamental frequency to help characterize the voices of normal speakers and speakers with vocal nodules. The fundamental frequency of the voice stimuli was determined using Pratt software (Boersma & Weenink, 2006). As described in Chapter 1, this software calculates the fundamental frequency of speech stimuli using an autocorrelation-based method.

3.1.2 First Cepstral Peak

Most of the acoustic algorithms discussed below require the calculation of a cepstrum. Cepstral analysis provides information about the extent to which the spectral harmonics (e.g., fundamental frequency) are distinguishable from the background noise level (Hillenbrand & Houde, 1996). A general description of cepstral analysis was provided in the section 1.1.4 of Chapter 1.

The amplitude of the first rahmonic peak in the cepstrum is representative of the periodicity of the signal (Murphy, 2006). Four cepstral measures were calculated: 1) a non-pitch period dependent method for calculating the normalized first cepstral peak prominence (CPP) from each cepstral frame and then averaged, 2) a non-pitch period dependent method for calculating the normalized CPP from the cepstrum of the average spectrum, 3) a pitch-period dependent method of calculating the amplitude of the first rahmonic peak (R1) in the cepstrum from each frame and then averaged, and 4) a pitch-period dependent method of calculating R1 from the cepstrum of the average spectrum.

When calculating both of the non-pitch period dependent CPP measures (measures 1 and 2 above), the input speech was resampled to 25 kHz to facilitate a comparison of the results with studies that have used a 25 kHz sampling rate. Since, a fixed window length has generally been used to calculate the cepstrum (Awan & Roy, 2005; Hillenbrand & Houde, 1996), speech samples were segmented into 41 ms frames

(Awan & Roy, 2005). Consecutive frames had a 50% overlap. A Hamming window was applied to each frame, which was then transformed into the spectral and cepstral domain by using Discrete Fourier Transforms (DFT) having a length of 1024 points.

Normalized CPP was calculated as the ratio of the amplitude of the CPP to the expected amplitude of the cepstral peak derived using linear regression. Only quefrencies greater than 2 ms were used in the linear regression to avoid including the low quefrency components which are sensitive to vocal tract resonances. For the analysis of continuous speech, a voicing detection method was not used when calculating CPP, since removal of the low amplitude signals (due to unvoiced sounds) has been shown to not improve correlations of normalized CPP measures with breathiness ratings (Hillenbrand & Houde, 1996).

Previous cepstral measures (e.g., CPP) have been shown to exhibit a dependence on fundamental frequency. However, the subjects (e.g., children and adults) in the experiments presented in this thesis produced speech with a wide range of fundamental frequencies. To minimize the dependence of cepstral methods on fundamental frequency, an attempt was made to reduce the sensitivity of the cepstral algorithms to differences in pitch (i.e., pitch independence), thereby enabling a comparison of different first cepstral peak values across speakers with different fundamental frequencies.

To this end, a pitch-period dependent method using a six pitch period window length was developed to reduce the pitch dependence of R1 (Murphy, 2006). In contrast to the rectangular window used by Murphy, in this experiment a Blackman window was used since tapering of the Blackman window smoothly to zero, reduces the side lobe amplitudes (Oppenheim & Schafer, 1989). Hence, the window length was multiplied by a Blackman window and successive windows having a 50% overlap was used to obtain the spectrum. Figures 3.1 and 3.2 show the average spectrum and corresponding cepstrum for the synthesized vowel /a/ with fundamental frequency of 165 Hz using a Hamming window versus a Blackman window respectively.

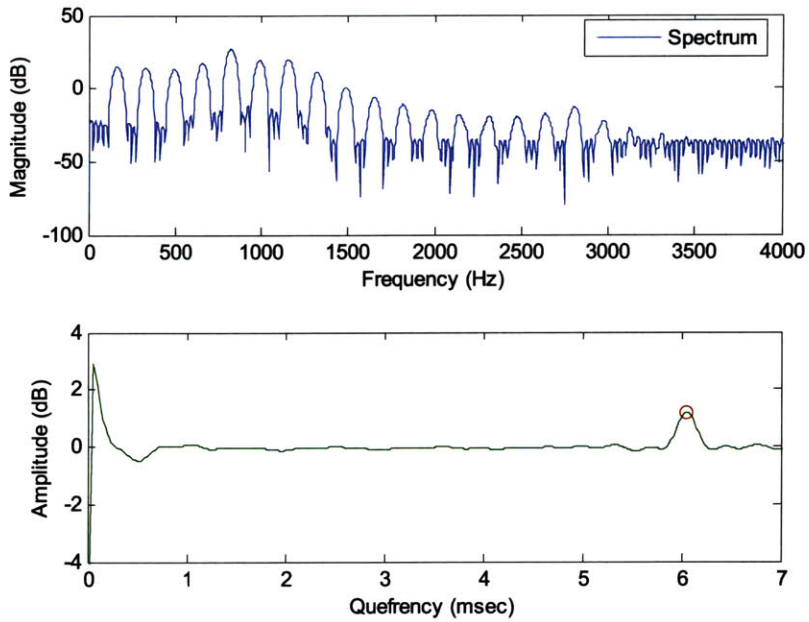


Figure 3.1. Top: spectrum of vowel /a/ (F0 = 165 Hz) produced by an adult female using a Hamming window. Bottom: corresponding cepstrum with red circle indicating the amplitude of R1.

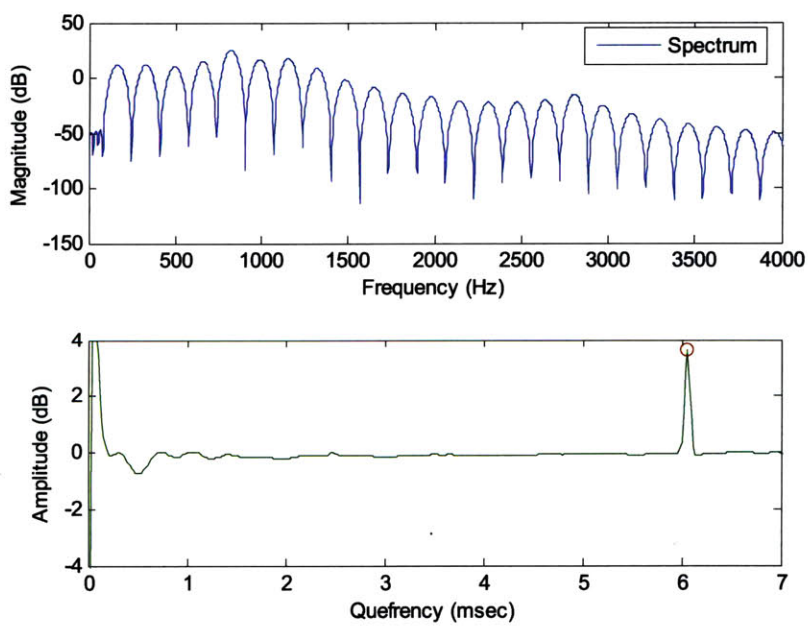


Figure 3.2. Top: spectrum of vowel /a/ (F0 = 165 Hz) produced by an adult female using a Blackman window. Bottom: corresponding cepstrum with red circle indicating the amplitude of R1.

A comparison of Figures 3.1 and 3.2 demonstrates that when a Blackman window is used to obtain the spectrum, R1 in the corresponding cepstrum is more clearly defined than when the spectrum is calculated using a Hamming window. Thus, for R1 measures, a Blackman window was used to calculate the spectrum in the experiments described below.

For each of the pitch-period dependent R1 methods (measures 3 and 4 above), the pitch period was calculated for each window. As a reminder, in calculation of R1, Murphy used a synthesized vowel with known fundamental frequency (i.e., pitch period), but in real speech pitch period is unknown so preprocessing needs to be performed to determine the length of the pitch period. The flow chart illustrating the pitch-period dependent R1 algorithm is depicted in Figure 3.3. In this pitch-independent algorithm, pitch period was determined using a fixed window length (41 ms) to calculate the cepstrum to find the pitch period and corresponding cepstrum (Awan & Roy, 2005). Once, the pitch period is known, R1 was calculated in the conventional way. Specifically, the pitch period was identified in the quefrency range between 3 ms and 8 ms for adult females and 2 ms and 6 ms for 4-6 and 8-10 year-old children, at the location of the first rahmonic peak. After the pitch period was determined, a six pitch period window length having a period corresponding to the pitch period of R1 was used to recalculate the cepstrum and derive the R1 amplitude.

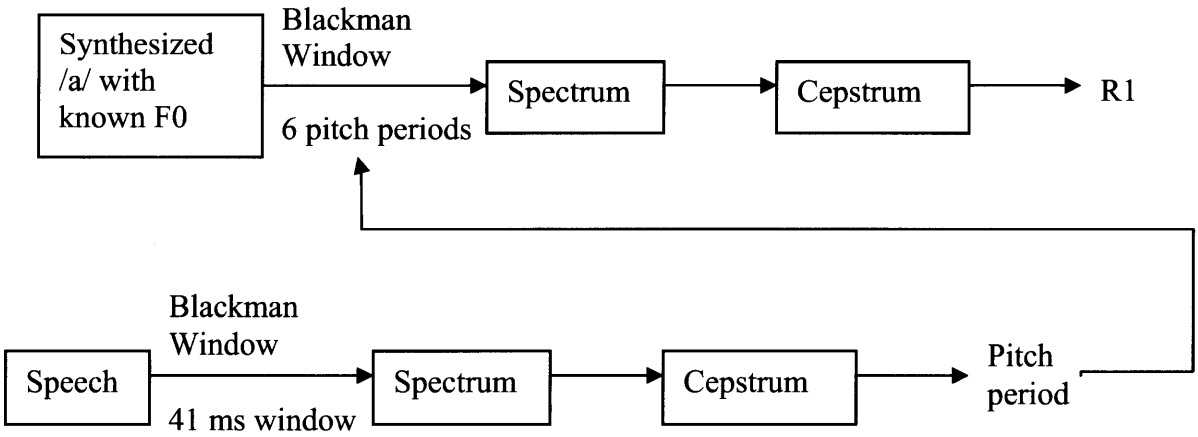


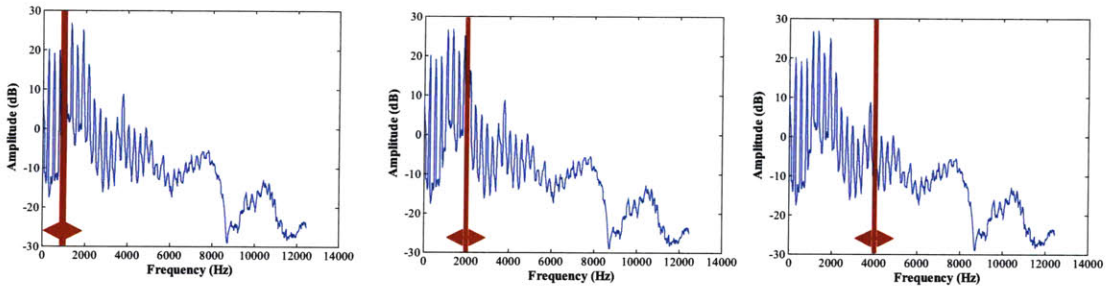
Figure 3.3: Flow chart of fundamental frequency independent R1 algorithm.

When calculating the spectrum with the pitch period dependent window length, each spectral frame was low-pass filtered using a Blackman window having a bandwidth of 12.5 kHz. The length of the DFT used for the analysis of vowels was 20480 points. R1 amplitude was calculated using two methods: 1) from the cepstrum obtained from the average spectrum of the signal (hereafter referred to as the "average spectrum" method) and 2) by averaging R1 amplitudes calculated for each cepstral window of the signal (hereafter referred to as the "individual spectra" method).

Although voicing detection was not used in calculating CPP values from continuous speech, voicing detection was used for calculation of R1, since R1 is only valid for segments of voiced speech. As described in Chapter 1, CPP is defined as the first cepstral peak prominence value in the cepstrum derived from a spectrum containing both voiced and voiceless regions. On the contrary, R1 is defined as the first rahmonic amplitude in the cepstrum derived from a spectrum only having voiced segments. The voicing detection algorithm selected the voiced segments on the basis of an autocorrelation method (Boersma & Weenink, 2006) . Furthermore, since sentences were sampled at 32 kHz as opposed to 50 kHz for vowels, the DFT length used for the analysis of sentences was 13180 points.

3.1.3 Spectral Slope Measures

While the characterization of the vocal source is important in describing voice quality, there is disagreement about which measure of source spectral slope best characterizes vocal quality. Thus, in the experiments in this thesis, three measures of spectral slope were used, with each measure being defined as an energy ratio with different cutoff frequencies (1 kHz, 2 kHz, and 4 kHz). These spectral slope measures are illustrated in Figure 3.4. Each spectral slope measure was calculated for both vowels and continuous speech. The energy ratio, in dB was calculated in two ways: 1) average of the energy ratios calculated for each individual spectrum (i.e., individual spectra) and, 2) energy ratio calculated from the average spectrum (i.e., average spectrum).



(a) (b) (c)

Figure 3.4: Average spectrum (shown in blue) of /a/ produced by child with vocal nodules. Red line shows cut-off frequency of (a) 1 kHz, (b) 2 kHz, and (c) 4 kHz for calculation of spectral slope measures.

For the analysis of sustained vowels, the input speech was segmented into 41 ms frames with 50% overlap between consecutive frames (Awan & Roy, 2005). Each frame was multiplied with a Hamming window and was transformed into the spectral domain using a DFT. Awan and Roy (2005) used a DFT length of 1024 points and a sampling rate of 25 kHz for vowels. However, since the sampling rate for vowels in the experiments of this thesis was 50 kHz, a DFT length of 2048 was used. For the analysis of sentences, a smaller window length of 20 ms window was used, with a DFT length of 1024, to ensure that the spectral representation of each sound did not get blurred.

3.1.4 Cepstral-based HNR Algorithm

The experiments in this thesis also examined HNR measures, as it has been shown that the determination of the relative amount of additive noise in the voice signal is useful for assessing patients with voice disorders. Conventional methods of quantifying aperiodicity in voice utterances (e.g., jitter, shimmer, and HNR) are limited by their need to accurately identify pitch period boundaries. As an alternative to conventional aperiodicity measures which derive measures from the spectrum of the signal, cepstral analysis was used in the experiments of this thesis. Cepstral analysis enables a quantification of the amount of additive noise in the speech signal because cepstral-based

HNR values (in dB) calculated at each harmonic location have been shown to provide information regarding the richness of the harmonic structure in the spectrum of voiced speech (Qi & Hillman, 1997).

Spectral HNR measures for vowels and continuous speech were calculated using cepstral analysis based on an algorithm proposed by Qi and Hillman (1997). In order to compare HNR measures across speakers with different fundamental frequencies (e.g., children and adults), a pitch-independent cepstral-based HNR algorithm, illustrated in Figure 3.5, was derived and used.

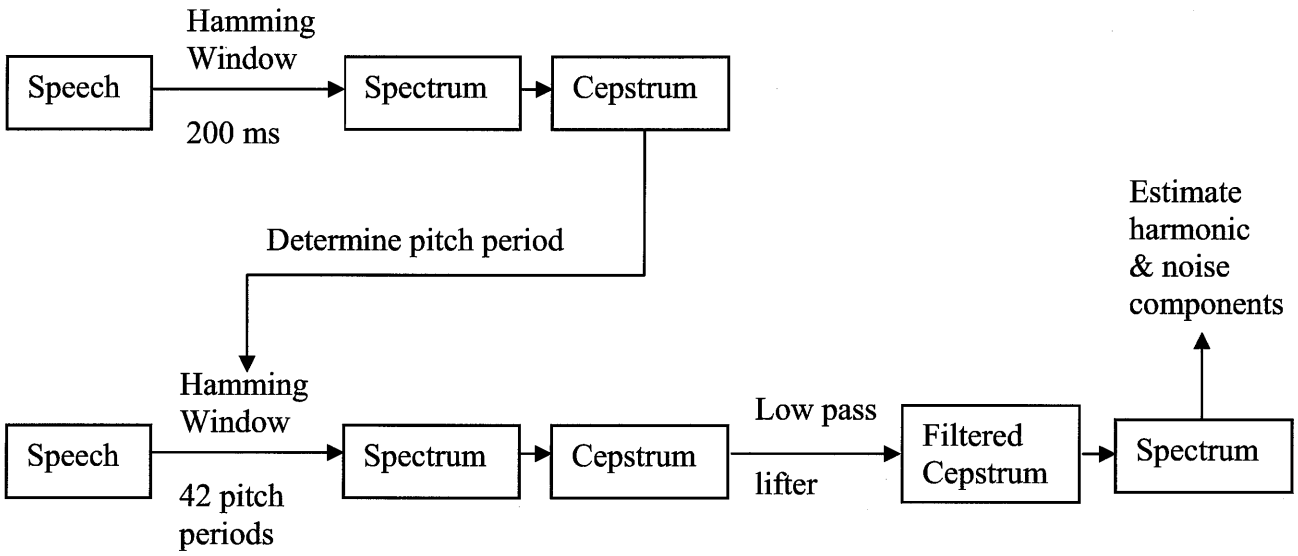


Figure 3.5: Block diagram of fundamental frequency independent cepstral-based HNR algorithm.

To calculate a fundamental frequency independent HNR measure, only the front-end of the existing HNR algorithm depicted in Figure 1.8 was used to determine the pitch period. Specifically, the pitch period was calculated by dividing the input speech into 200 ms frames (Qi & Hillman, 1997). Consecutive frames had a 50% overlap. Each frame was multiplied with a Hamming window, and was transformed into the spectral and cepstral domain using a DFT as described above. The location of the first rahmonic (R1) in the cepstrum was calculated to determine the pitch period of each frame.

After determining the pitch period, HNR was calculated using a forty-two pitch period window length. This window length was chosen because a window length of 200 ms contains approximately forty-two pitch periods of a typical adult female voice. Each window was multiplied with a Hamming window, and was used to compute both the spectrum and cepstrum. A DFT length of 20480 points was used for vowels sampled at 50 kHz, whereas a window length of 13108 points was used for sentences sampled at 32 kHz. These DFT lengths were chosen to be longer than the length of forty-two pitch periods for a voice sample having a fundamental frequency of 165 Hz. Since HNR measures are only valid for voiced segments, only voiced frames were considered (Parsa & Jamieson, 2001). Hence, similar to the calculation of R1 discussed above, a voicing detection algorithm was used to detect the voiced segments when analyzing continuous speech (Boersma & Weenink, 2006).

The average noise level, used to calculate the noise for HNR, was computed by low-pass liftering (i.e., filtering in the cepstral domain) the cepstrum using a pitch dependent window. The cut-off quefrency of the low-pass lifter was determined by considering various factors including whether the cut-off quefrency adequately modeled the variations of the spectral envelope and whether the cut-off quefrency was low enough to reduce possibility of removal of the first cepstral peak.

When extracting low quefrency components, a lifter length of 0.6 pitch periods was used for speech having fundamental frequencies less than 250 Hz, whereas a lifter length of 0.7 pitch periods was used for speech having fundamental frequencies greater than 250 Hz (Rahman & Shimamura, 2005). The liftered cepstrum was then transformed back into the frequency domain, resulting in a spectrum that represented the smoothed noise level of the original signal.

The HNR was calculated using the harmonics (i.e., peaks) in the spectrum, which correspond to multiples of the fundamental frequency. HNR for each frame was calculated as the average of the ratios between the harmonic amplitudes in the spectrum up to 8 kHz and the corresponding noise levels at these harmonic locations. This procedure was repeated for each spectral window of the signal and then averaged.

3.2 Validation of Acoustic Algorithms

Vowels having speech characteristics (e.g., formant frequencies) of a female adult, a 5 year-old child, and a 9 year-old child with normal voices were synthesized. These synthesized vowels were used to not only validate the modifications made to the non-pitch period period dependent acoustic algorithms to make the algorithms fundamental frequency independent but also to assess whether the modified algorithms worked well for analyzing the voices of children.

3.2.1 Creating Synthetic Vowels

A one-second production of the vowel /a/ was synthesized using speech parameters for two children (5 year-old and 9 year-old) and an adult female with normal voices. The parameters used for the speech synthesis are shown in Table 3.1. Formant frequencies (Lee, Potamianos, & Narayanan, 1999; Peterson & Barney, 1952), and bandwidths of the formants (Mannell, 1998) were estimated.

	F0 (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)	B1 (Hz)	B2 (Hz)	B3 (Hz)
Adult female	212	850	1220	2810	100	109	147
9 year old	261	1037	1639	3265	105	119	158
5 year old	269	1195	1796	3424	109	123	162

Table 3.1: Spectral characteristics of synthesized /a/ for adult and children. Fundamental frequency (F0), formant frequencies (F1, F2, F3) and bandwidths of the formants (B1, B2, B3) used to synthesize the vowel /a/ for an adult female, 9 year-old, and 5 year-old.

A Matlab-implemented formant synthesizer, based on the source-filter theory of speech production (Fant, Liljencrants, & Lin, 1985) was used to create the synthesized vowels (Mehta, 2006). This synthesizer allows for independent modelling of the voicing source and vocal tract parameters.

The voicing source of the synthesizer combines a periodic source and an aspiration noise source. Hence, the speech waveform consists of a linear combination of

a periodic component, which is the output of a periodic glottal pulse linearly filtered by the vocal tract (Rosenberg, 1971), and a noise component, which is the output of the vocal tract filter with an unmodulated aspiration noise source. For the synthesis of the vowel /a/, the vocal tract was modelled as an all-pole filter, with pairs of poles representing a resonance frequency (formant) and its bandwidth.

The following parameters were selected to generate the synthesized vowels: fundamental frequency, sampling frequency, duration of vowel, DC flow (fraction of maximum noise amplitude for DC airflow), open quotient (ratio of open time of glottal cycle to total duration of period, in %), jitter (deviation from nominal pitch period value, in %), shimmer (cycle-to-cycle variation in the amplitude of the source excitation, in %), first three formant frequencies, first three formant bandwidths, spectral tilt (slope of voicing source spectrum), speed quotient (ratio of opening of glottis to duration of closing), and HNR (ratio of power in the harmonic and noise components computed on the signals at the output, in dB) (Mehta, 2006).

3.2.2 Manipulating Speech Synthesizer Parameters

Source characteristics of normal voices for the vowel /a/ synthesized using speech parameters for each of three ages (female adults, 5 year old, and 9 year olds) were manipulated to investigate the sensitivity of the acoustic algorithms to changes in these source characteristics. The source characteristics that were manipulated were fundamental frequency, HNR, jitter, shimmer, and spectral tilt. Fundamental frequency was varied between 165 Hz to 325 Hz in 5 Hz increments to span the fundamental frequency range typically evident in the voices of adult females and children as young as four years old. HNR was varied between 0 dB and 100 dB in 10 dB increments, jitter was varied between 0% and 5% in 0.5% increments, shimmer was varied between 0% and 5% in 0.5% increments, and spectral tilt was varied between -12dB/octave and -17dB/octave in -1dB/octave increments.

Systematic manipulation of each of these source characteristics enabled an examination of an individual characteristic's effect on each of the following acoustic algorithms for the analysis of the vowel /a/: a) normalized CPP measure, b) pitch period

dependent R1 measure, c) spectral slope measure calculated using the 2 kHz energy ratio (i.e., below to above 2 kHz energy), and d) cepstral-based HNR. Each of the acoustic algorithms were calculated using both the individual spectra method and the average spectrum method, except for the cepstral-based HNR algorithm which was only calculated using the individual spectra method. Source characteristics that were not being systematically manipulated were set to their "default" values (jitter = 0, shimmer = 0, spectral tilt = 0dB, DC flow = 0.1, Open Quotient = 60, Speed quotient = 270).

3.2.3 Results of Manipulating Speech Synthesizer Parameters

Fundamental Frequency

Fundamental frequency independence for each of the acoustic algorithms was assessed using the above-described synthesized vowels sampled at 50 kHz and 32 kHz. Hence, the effect of changing fundamental frequency for vowels synthesized using the three formant frequency parameter sets corresponding to the speech characteristics for adult females, a 9 year-old and a 5 year-old shown in Table 3.1, was examined. It was expected that the algorithms for calculating R1, 2 kHz energy ratio, and HNR would be relatively fundamental frequency independent, whereas the conventional CPP algorithm was expected to exhibit fundamental frequency dependence.

Figure 3.6 shows that the normalized CPP value calculated using the individual spectra method for synthesized vowels (32 kHz and 50 kHz sampling rates) having adult formant frequencies, maximally varied by 9.46 dB and 3.57 dB, respectively, over the frequency range from 165 Hz to 325 Hz. By contrast, R1 calculated using the individual spectra method only maximally varied by 0.26 dB and 0.34 dB, for the synthesized vowel created using a 50 kHz and a 32 kHz sampling rate, respectively.

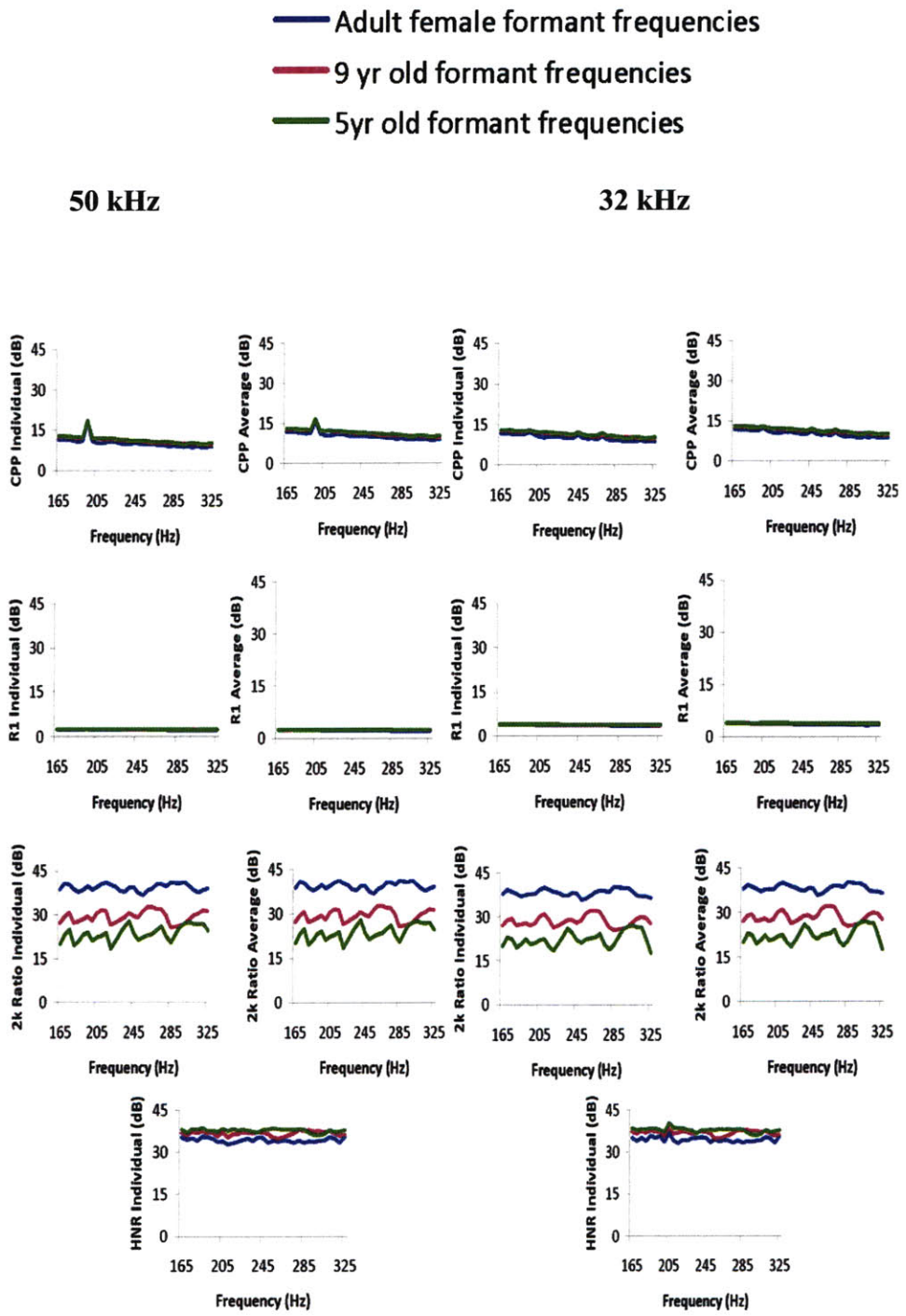


Figure 3.6: Manipulation of fundamental frequency for each of the acoustic algorithms using the individual spectra method (Individual) or the average spectrum method (Average). The synthesized vowel /a/ created using sampling rates of 50 kHz (left two columns) and 32 kHz (right two columns) was analyzed for each of three different formant frequency sets corresponding to an adult female, a 5 year-old, and a 9 year-old.

Similarly, the normalized CPP value calculated using the average spectrum method for synthesized vowels (32 kHz and 50 kHz sampling rates) having adult formant frequencies, maximally varied by 7.28 dB and 3.65 dB, respectively, over the frequency range from 165 Hz to 325 Hz. By contrast, R1 calculated using the average spectrum method only maximally varied by 0.41 dB and 0.59 dB, for the synthesized vowel created using a 50 kHz and a 32 kHz sampling rate, respectively. Hence, a comparison of the CPP and R1 values calculated using the two different methods and two different sampling rates, shows that the normalized CPP values are more sensitive to fundamental frequencies than the R1 values. Furthermore, as seen in Figure 3.6, peaks in the normalized CPP values occur at 195 Hz, 245 Hz and 270 Hz, which are frequencies that correspond to window frames containing an integer number of periods. The slight differences in the presence of the peaks observed between the synthesized vowels created using a 50 kHz sampling rate versus a 32 kHz sampling rate may be due to the interaction between pitch period and sampling rate.

The spectral slope measure of 2 kHz energy ratio was relatively insensitive to variations in fundamental frequency regardless of whether the individual spectra method or the average spectrum method was used. Specifically, Figure 3.6 shows that the 2 kHz energy ratio calculated for the synthesized vowel having adult formant frequencies, maximally varied by 4.40 dB for both sampling rates over the frequency range from 165 Hz to 325 Hz when calculated using the individual spectra method and maximally varied by 4.41 dB (50 kHz) and 4.46 dB (32 kHz), when calculated using the average spectrum method.

HNR was relatively insensitive to variations in fundamental frequency. Specifically, Figure 3.6 shows that HNR values calculated from a synthesized vowel having adult formant frequencies, maximally varied by 2.82 dB (50 kHz) and 3.58 dB (32 kHz) over the frequency range from 165 Hz to 325 Hz.

CPP, R1 and HNR measures were similar for a given fundamental frequency regardless of the formant frequency values used to synthesize the vowel. However, the 2 kHz energy ratio values differed as a function of formant frequency values used to create the vowels. Specifically, the 2 kHz energy ratios decreased as the formant frequency

values increased (with decreases in age). This systematic difference may be due to the differences in spectral shape for the three different formant frequency values. Figures 3.7 and 3.8 demonstrate this concept for the vowel /a/ synthesized with a constant fundamental frequency of 325 Hz for formant frequencies for an adult female and a 5 year-old child.

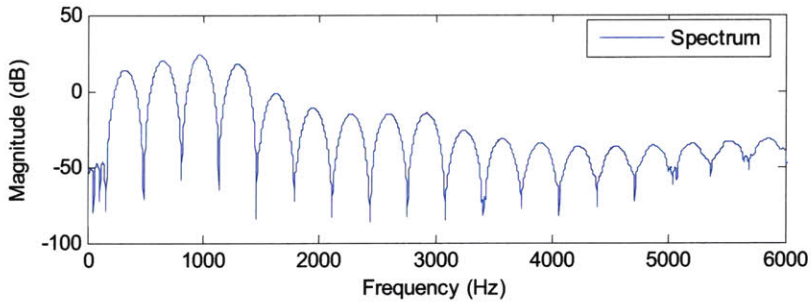


Figure 3.7: Spectrum of synthesized vowel /a/ (F0 = 325 Hz, adult formant frequency values).

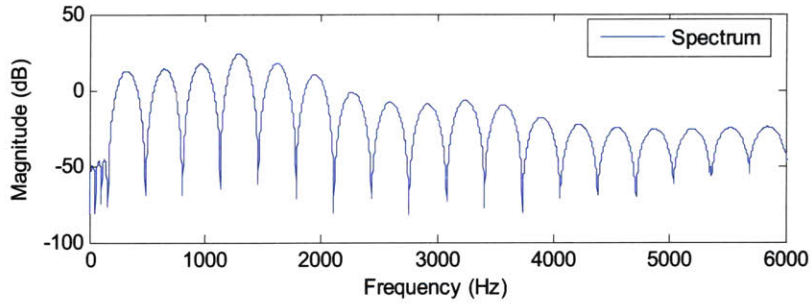


Figure 3.8: Spectrum of synthesized vowel /a/ (F0 = 325 Hz, 5 year-old formant frequency values).

A comparison of Figures 3.7 and 3.8 reveals that the downward slope of the spectral shape for the synthesized vowel with 5 year-old formant frequencies (Figure 3.8) is less steep than the spectral shape of the synthesized vowel with adult formant frequencies (Figure 3.7). This difference in spectral shape results in higher energy in the lower frequency region of the spectrum and correspondingly lower energy in the high frequency region of the spectrum for adult formant frequencies.

Spectral Tilt

Spectral tilt is related to the slope of the harmonics of the source spectrum, and can be varied by changing the roll-off of the spectrum, with greater spectral tilt values reflecting a steeper decline of the higher frequency amplitudes (Klatt & Klatt, 1990). It was expected that the algorithms for calculating CPP and R1 measures would be relatively unaffected by changes in spectral tilt, although it was expected that the HNR measure and 2 kHz energy ratio measures would depend on spectral tilt.

As seen in Figure 3.9, CPP and R1 measures did not change as a result of varying spectral tilt for the synthesized vowel /a/ created using sampling rates of 50 kHz and 32 kHz with fundamental and formant frequency values for an adult female, a 9 year-old child, and a 5 year-old child. CPP and R1 measures were also not affected by differences in formant frequencies associated with the different age groups. Additionally, the 2 kHz energy ratio increased as the spectral tilt value was increased (i.e., became more negative). The nonlinear change between 2 kHz energy ratio and spectral tilt may be caused by the properties of the synthesizer, in which the spectral tilt value changes with different fundamental frequencies. The 2 kHz energy ratio also appears to decrease with when higher formant frequencies are used to synthesize the vowel (i.e., for younger children). This may be explained by the spectral characteristic of young children versus adults as illustrated in Figures 3.7 and 3.8, which show higher energy in the lower frequency region of the spectrum and lower energy in the higher frequency region of the spectrum for adult formant frequencies. A slight change in HNR was also observed with changes in spectral tilt. Also, for a given spectral tilt value, HNR was also slightly different for the different spectral characteristics of adults and children.

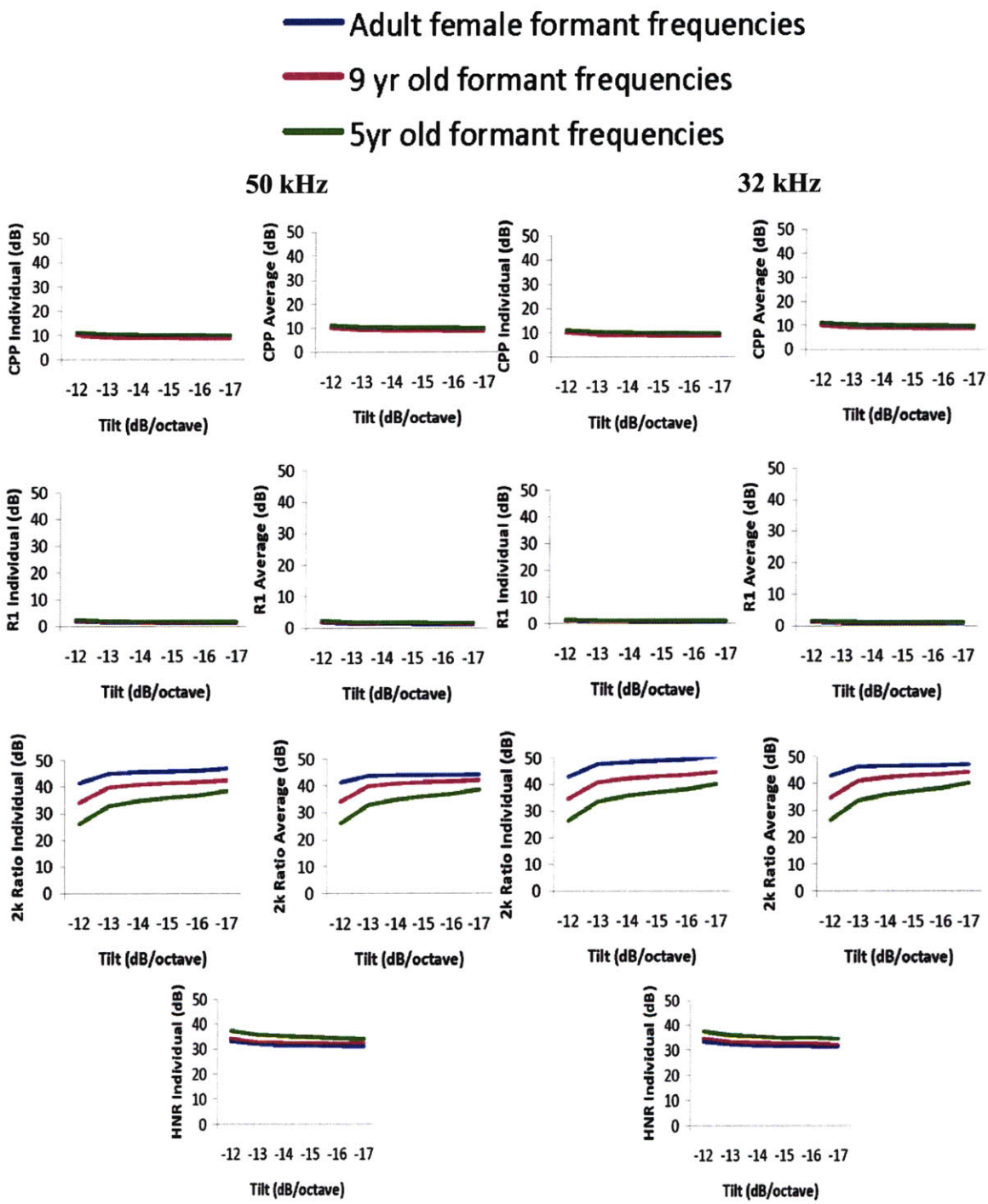


Figure 3.9: Manipulation of spectral tilt for each of the acoustic algorithms calculated using the individual spectra method (Individual) or the average spectrum method (Average). The synthesized vowel /a/ created using sampling rates of 50 kHz (left two columns) and 32 kHz (right two columns) was analyzed for each of three different fundamental frequency and formant frequency sets corresponding to an adult female, a 5 year-old, and a 9 year-old.

HNR

The effect of adding noise to the speech signal is reflected as additional energy between the harmonics, while the harmonic amplitudes remain unchanged (on averaging) provided that the harmonic amplitudes are greater than the noise level (Murphy, 2006). It was expected that the algorithms calculating CPP, R1, HNR, and 2 kHz energy ratio would be affected by changes in HNR since these algorithms are known to be sensitive to additive noise.

CPP, R1, HNR, and the 2 kHz energy ratio measure were affected by variations in HNR for the synthesized vowel /a/ created using sampling rates of 50 kHz and 32 kHz with fundamental and formant frequencies for an adult female, a 9 year-old child, and a 5 year-old child. Figure 3.10 shows that all of the acoustic measures increase with increases in HNR up to an HNR value of approximately 40 dB, and then changes little or plateaus beyond that. This may be due to the fact that the acoustic measures cannot resolve the noise floor when the input signal has high HNR values (e.g., greater than 50 dB). However, although this may be a concern for synthesized speech, resolving the noise floor is not an issue for actual speech, since normal HNR values for actual speech are usually not greater than 20 dB (Qi & Hillman, 1997).

For each HNR value, R1 and CPP were similar regardless of which set of formant frequencies were used to synthesize the vowel. However, due to the sensitivity of formant frequency locations when calculating the 2 kHz energy ratios different energy ratio values were obtained for the different sets of formant frequencies as shown in Figure 3.10. Specifically, the 2 kHz energy ratio decreased as the formant frequencies increased (e.g., for younger children). This decrease in 2 kHz energy ratio for younger children may be explained by the spectral characteristic of young children versus adults as shown in Figures 3.7 and 3.8, which show that for vowels with adult formant frequencies, there is higher energy in the lower frequency region of the spectrum and lower energy in the higher frequency region of the spectrum.

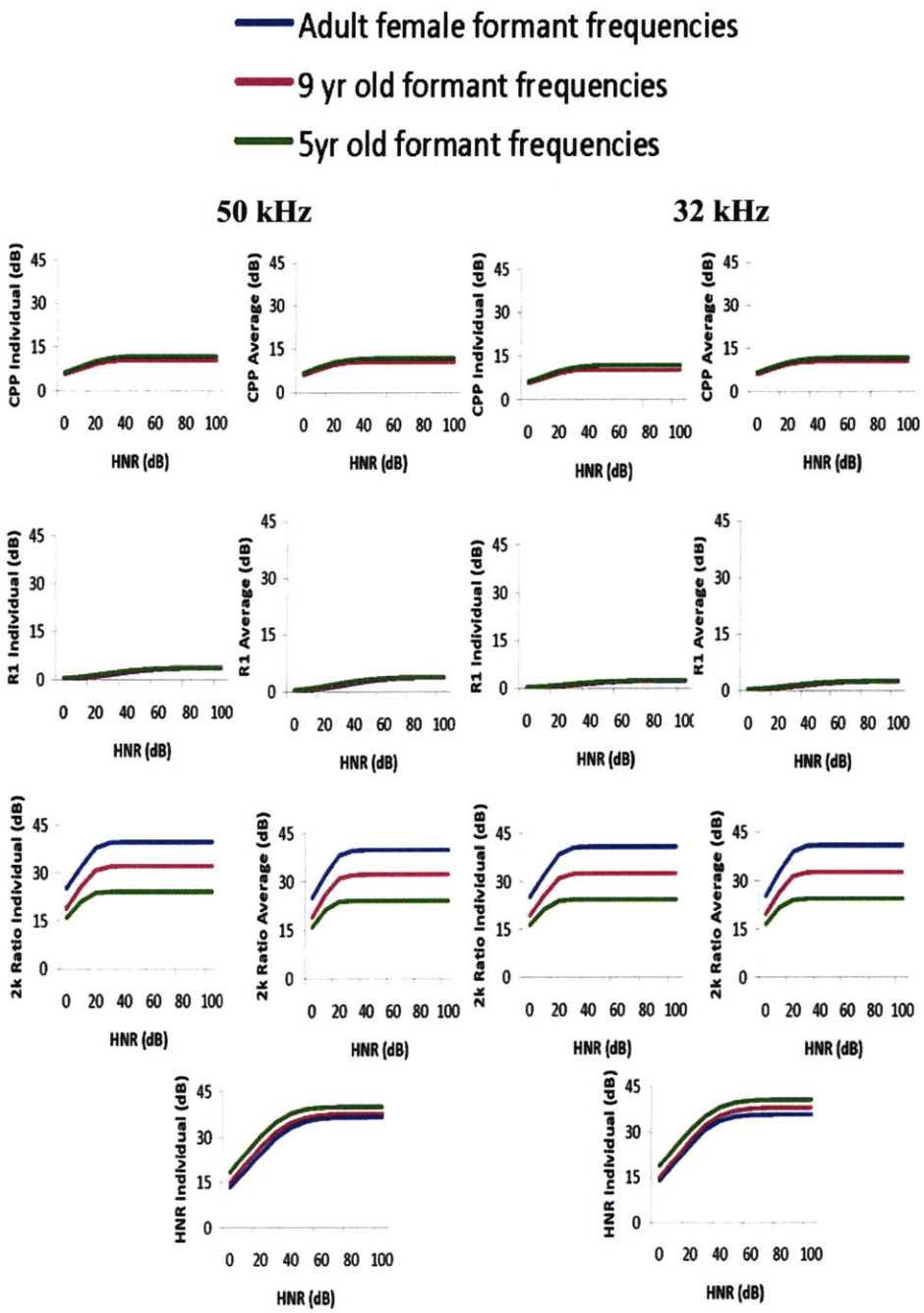


Figure 3.10: Manipulation of HNR for each of the acoustic algorithms calculated using the individual spectra method (Individual) or the average spectrum method (Average). The synthesized vowel /a/ created using sampling rates of 50 kHz (left two columns) and 32 kHz (right two columns) was analyzed for each of three different fundamental frequency and formant frequency sets corresponding to an adult female, a 5 year-old, and a 9 year-old.

Jitter

Adding jitter to a speech signal is reflected as a decrease in the amplitude of the harmonic components and an increase in the amplitude of the between-harmonic components of the speech signal (Dejonckere & Wieneke, 1994; Murphy, 2000). It was expected that the CPP, R1, and HNR algorithms would be affected by changes in jitter because these algorithms are known to be sensitive to aperiodicities in the speech signal. By contrast, the 2 kHz energy ratio was not expected to be affected by jitter because an increase in jitter should decrease the harmonic components while increasing the noise components of the entire frequency range resulting in an energy ratio that remains relatively unchanged. This is contrary to the effect of adding noise, discussed above, which was expected to affect the spectral energy mostly in the higher frequency region of the spectrum.

Slight variations in CPP and R1 measures were observed with changes in jitter for the synthesized vowel /a/ created using sampling rates of 50 kHz and 32 kHz with fundamental and formant frequencies for an adult female, a 9 year-old child, and a 5 year-old child. Specifically, CPP and R1 decreased as jitter was increased. By contrast, HNR varied considerably with changes in jitter, with HNR decreasing as jitter was increased as seen in Figure 3.11. The 2 kHz energy ratio was not affected by changes in jitter, although 2 kHz energy ratio values were again observed to be different for vowels synthesized using different sets of formant frequencies. However, unlike the 2 kHz energy measure, R1, CPP and HNR were not impacted by the choice of formant frequencies used to synthesize the vowels.

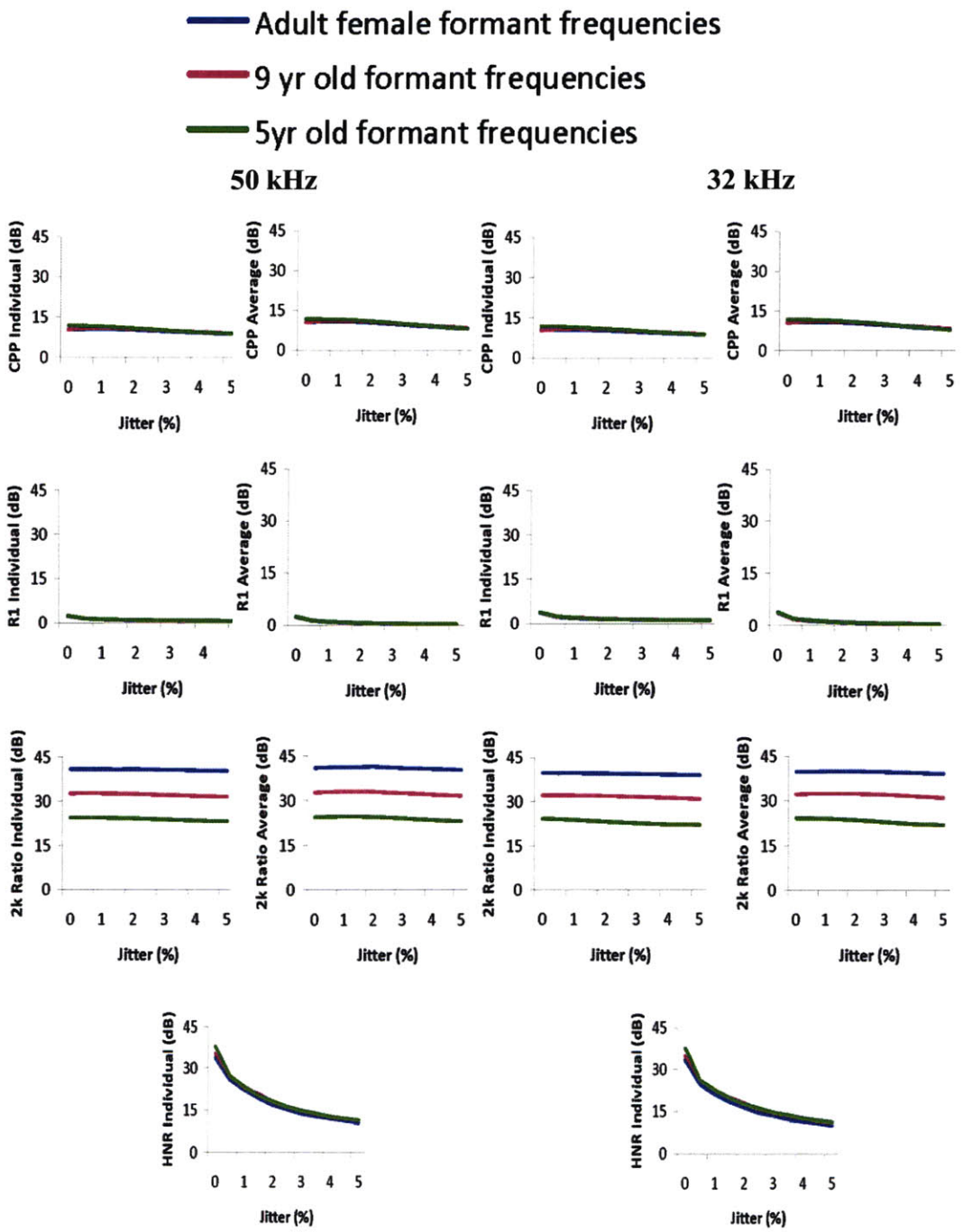


Figure 3.11: Manipulations of jitter for each of the acoustic algorithms calculated using the individual spectra method (Individual) or the average spectrum method (Average). The synthesized vowel /a/ created using sampling rates of 50 kHz (left two columns) and 32 kHz (right two columns) was analyzed for each of three different fundamental frequency and formant frequency sets corresponding to an adult female, a 5 year-old, and a 9 year-old.

Shimmer

Speech signals containing shimmer retain a strong harmonic structure throughout the spectrum, although increases in the between-harmonic amplitudes is often observed (Murphy, 2000). The algorithms for calculating CPP, R1, and HNR measures were expected to be affected by changes in shimmer because these algorithms are known to be sensitive to aperiodicities in the speech signal. However, the sensitivity of these algorithms to changes in shimmer was expected to be less than their sensitivity to changes in jitter because variations in shimmer do not change to the same extent the harmonic structure of the speech signal, unlike variations in jitter. It was expected that the 2 kHz energy ratio, would not be affected by variations in shimmer because increases in shimmer should decrease the harmonic components in proportion to increasing the noise components of the entire frequency range, resulting in an energy ratio that remains relatively unchanged. This prediction is in contrast to the effect of adding noise, as discussed above for HNR, in which the additional noise was expected to affect the spectral energy mostly in the higher frequency region of the spectrum.

Figure 3.12 shows that variations in shimmer did not affect CPP, R1, and the 2 kHz energy ratio measures in the same way as variations in shimmer affected HNR. Specifically, HNR decreased as shimmer was increased. As observed previously, the 2 kHz energy ratios differed for different sets of fundamental and formant frequencies. Contrary to spectral slope (2 kHz energy ratio) and HNR measures, R1 and CPP values were not affected by the different formant frequencies associated with the different age groups.

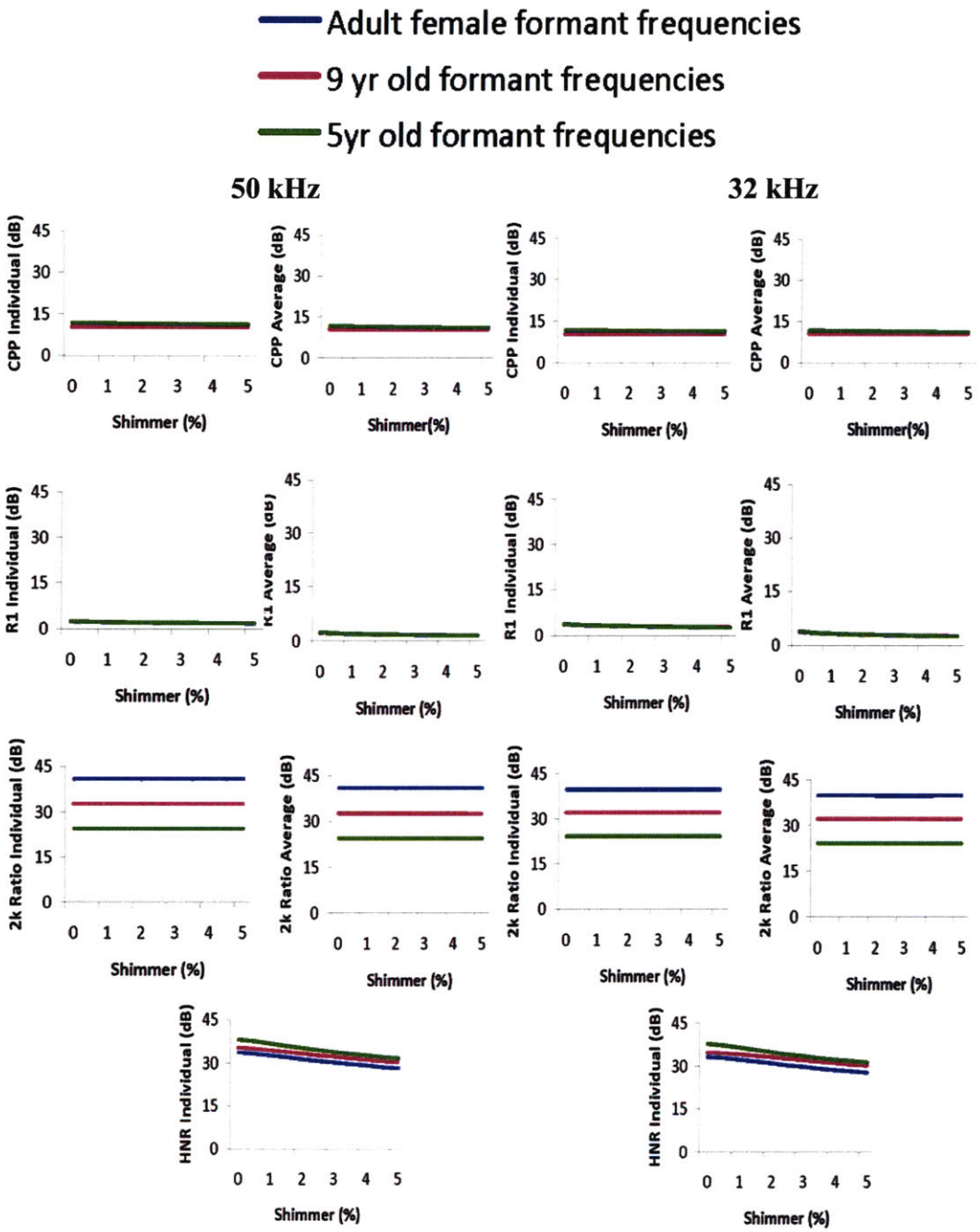


Figure 3.12: Manipulations of shimmer for each of the acoustic algorithms calculated using the individual spectra method (Individual) or the average spectrum method (Average). The synthesized vowel /a/ created using sampling rates of 50 kHz (left two columns) and 32 kHz (right two columns) was analyzed for each of three different fundamental frequency and formant frequency sets corresponding to an adult female, a 5 year-old, and a 9 year-old.

3.2.4 Discussion of Manipulating Speech Synthesizer Parameters

By manipulating various speech synthesizer parameters, it was found that the conventional CPP algorithm was most sensitive to variations in fundamental frequency compared to the other acoustic algorithms. Thus, our attempt to create fundamental freqeucny independent measures was warranted.

Furthermore, the algorithms which were sensitive to the spectral characteristics of the speech signal were the spectral slope measure (2 kHz energy ratio) and HNR. Although a close examination of the R1 and CPP cepstral measures showed sensitivity to variations in spectral tilt reflected as decreases in CPP and R1 with increased (more negative) spectral tilt, as predicted by Murphy (2006), the sensitivity of R1 and CPP was not as prominent as the sensitivity of the spectral slope and HNR algorithms to spectral characteristics of the speech signal. By systematically varying HNR, it was found that decreases in HNR resulted in decreases in R1 and CPP, which is consistent with the findings of Murphy (2006). Similarly, increases in jitter and shimmer also resulted in decreases in R1 and CPP. The finding that R1 and CPP values decrease more with jitter than shimmer was expected because it has been shown that the prominence of the harmonics in the spectrum decrease more for changes in jitter than shimmer, resulting in a corresponding decrease in the amplitude of the first cepstral peak (Dejonckere & Wieneke, 1994; Murphy, 2006). With regards to sensitivity to variations in additive noise, all of the acoustic algorithms used in this thesis were found to be sensitive to variations in additive noise.

3.3 Results and Discussion of Acoustic Analysis of Children and Adult Females with Normal Voices and Vocal Nodules

The validated acoustic analysis algorithms on sustained vowels were used to analyze the same voice samples recorded for Perceptual Experiment 1 discussed in Chapter 2. By way of review, to reduce the variability of laryngeal structure differences in children, a narrow age range was chosen to create a more homogeneous group. Hence, the stimuli were recorded from 4-6 year-old children with normal voices (n = 8) or with vocal nodules (n = 8), and adult females with normal voices (n = 8) or vocal nodules (n =

8). Each subject produced both the sustained vowel /a/ and a CAPE-V sentence (pediatric CAPE-V produced by children, adult CAPE-V produced by adults). Hence, although the acoustic algorithms were not validated on running speech, the algorithms were used to analyze real continuous speech.

Table 3.2 shows values for each of the acoustic measures for sustained vowels and continuous speech utterances (CAPE-V sentences) produced by 4-6 year-old children and adult females.

	4-6 year old				Adult female			
	Vowel		Sentence		Vowel		Sentence	
Measures	**Normal**	**Disordered**	**Normal**	**Disordered**	**Normal**	**Disordered**	**Normal**	**Disordered**
F0 (Hz)	256.3 (19)	243.4 (51.4)	243.9 (23.3)	254.1 (33.9)	217.7 (17.1)	218.5 (28.3)	212 (25)	206.9 (27.2)
HNR Indiv (dB)	24.2 (4.6)	21.3 (2.4)	16.8 (1.9)	15.9 (3)	25.4 (2.9)	24.2 (3.3)	13.7 (2.4)	14.4 (2.1)
CPP Indiv (dB)	9.7 (2.1)	8.6 (1.4)	7.5 (0.9)	6.8 (0.9)	10.2 (1.6)	9.3 (1.2)	7.5 (0.6)	7.5 (0.2)
CPP Avg (dB)	10.9 (1.1)	10.1 (1.2)	8.3 (1.5)	7.1 (0.4)	11.6 (1.5)	11.1 (1.2)	7.1(1.1)	7.8 (1.4)
R1 Indiv (dB)	0.9 (0.4)	0.7 (0.2)	1.1 (0.3)	0.9 (0.3)	0.9 (0.3)	0.8 (0.2)	0.9 (0.3)	0.9 (0.1)
R1 Avg (dB)	0.7 (0.2)	0.5 (0.1)	0.4 (0.2)	0.3 (0)	0.8 (0.2)	0.7 (0.2)	0.3 (0.1)	0.4 (0.1)
1k Indiv (dB)	-1 (7.4)	-4.4 (4.4)	13.8 (2.8)	15.3 (4.7)	6.6 (4)	4.7 (3.6)	22.6 (4.8)	25.3 (1.5)
1k Avg (dB)	-0.5 (7.4)	-3.6 (4.4)	6.6 (3.7)	6.3 (3.8)	6.7 (4)	4.9 (3.6)	14.1 (4)	12.7 (1.9)
2k Indiv (dB)	15.9 (5.1)	12.6 (4.7)	17.9 (2.6)	17.9 (4.6)	23.3 (2.2)	20.4 (2.4)	27.1 (3.3)	28.3 (1.2)
2k Avg (dB)	16.9 (5.5)	13.1 (4.8)	11.6 (3.2)	9.4 (2.9)	24.1 (2.3)	21.1 (2.5)	16.2 (3.7)	15.2 (2.1)
4k Indiv (dB)	19.9 (5.9)	17 (5.2)	20.6 (2.9)	20 (4.8)	30.1 (4.5)	25.3 (2.1)	31.5 (3.1)	32.3 (1.3)
4k Avg (dB)	20.9 (6.1)	17.4 (5.2)	14.7 (4)	13.3 (4.4)	31 (4.6)	26 (2.3)	22.4 (3.5)	21.4 (2.7)

Table 3.2: Average acoustic measures for the vowel /a/ and CAPE-V sentence produced by 4-6 year-old children and adult females with normal voices or voices with vocal nodules. Some acoustic measures were calculated using the individual spectra method (Indiv) or the average spectrum method (Avg) as discussed above. Mean and standard deviation is shown.

The results in Table 3.2 confirm that most of the average acoustic measures displayed the expected differences between normal voices and voices with vocal nodules for vowel production in children and adults, although the magnitudes of the differences tended to be smaller between the normal voices and voices with vocal nodules for adults. Since the acoustic measures calculated in this study measures periodicity, the measures obtained from normal voices was expected to be greater than measures obtained from voices with vocal nodules. Specifically, the acoustic measures of HNR, CPP, R1, and spectral energy ratio (spectral slope) measures were expected to be reduced for voices with vocal nodules relative to normal voices. All of the acoustic measures analyzing sentence production for sentences, except the 1 kHz spectral energy ratio, also displayed

the same expected differences between normal and disordered voices. However, for sentence production by adults, most of the measures showed either no difference, or a small difference in the opposite direction from what was expected. These unexpected results were seen for measures of HNR, CPP obtained from average spectrum, R1 obtained from average spectrum and all the spectral energy ratio measures based on the individual spectra method. This could be due to the fact that it is not clear that certain acoustic measures have meaning for continuous speech.

Table 3.2 also shows the expected higher mean fundamental frequency values for children as compared to adults, but the differences between normal voices and voices with vocal nodules did not appear to be significant given the associated standard deviations. These findings are consistent with studies which have found that patients with benign vocal fold lesions/vocal nodules do not exhibit significant changes in mean speaking fundamental frequency (Kilic, Okur, Yildirim, & Guzelsoy, 2004).

The results in Table 3.2 show that for the most part, the acoustic algorithms which were developed and validated using synthetic vowels for analyzing the voices of adult females and children can be applied to real speech, both vowels and continuous speech. In general, most of the measures showed the potential to be sensitive to the types of dysphonia that are associated with vocal nodules in adult females and children. These expected differences between normal and disordered voices are based on previous studies which have mostly described the development and application of these measures in synthetic speech and real speech in adults, as reviewed in Chapter 1. By way of summary, dysphonia is associated with the addition of noise, jitter and shimmer to the speech signal, causing a degradation of harmonic structure which is reflected in reductions in measures of HNR, CPP and R1 (e.g., Murphy, 2006). A reduction in spectral ratio (spectral slope) values is also expected because dysphonic voices (e.g., breathy voice) are typically characterized as having noticeable spectral noise above 2-3 kHz (e.g., Klatt & Klatt, 1990).

The seemingly unexpected results for adult sentence production are probably mainly attributable to observation from the perceptual investigation that showed adults were significantly less dysphonic than children with scores skewed towards the lower end

of the perceptual scales. This created more overlap in voice quality parameters among the adult subjects as compared to the pediatric subjects, which would contribute to the diminished number and magnitude of expected differences between adults with normal voices and voices with vocal nodules. It is somewhat understandable that the reduced separation in voice quality among adults with normal voices and voices with vocal nodules seems to have been further exacerbated by the influence of additional challenges/factors associated with the extraction of the acoustic measures from continuous speech (sentences). Another potential challenging issue is that these acoustic measures were not tested on synthetic continuous speech.

As already stated, these initial results are to simply establish proof-of-concept that the developed measures can be applied to the real speech of women and children. The apparent overlap in acoustic measurement results for sentence production in adults with normal voices and voices with vocal nodules in no way precludes moving to the final phase of this investigation, the examination of relationships between perceptual judgments and acoustic parameters.

Chapter 4: Relationship between Auditory-Perceptual Ratings and Acoustic Measures

Acoustic measures are clinically useful to the extent that they reflect the underlying voice production system or explain perceptual ratings. However, a comprehensive theory that describes the correspondences between vocal physiology, acoustics, and perceived voice quality has not been forthcoming (Kreiman, et al., 1993). To date, most studies have attempted to validate an acoustic measure by examining the correlations between the acoustic measure and ratings of specific voice qualities (Kreiman, et al., 1993). Hence, determining an objective voice measure that correlates well with listener ratings of the severity of a voice disorder would be clinically useful. Consequently, a goal of this chapter is to identify a subset of spectral and/or cepstral-based acoustic measures that predicts the degree of dysphonia associated with vocal nodules in women and children. Although several attempts have been made to investigate the relationship between acoustic measures and perceptual parameters of voice quality, most of these studies have focused on adult speech. As observed from the perceptual results in Chapter 2, the voices of adult women and children with nodules differ perceptually. Thus, it was expected that different acoustic measures would correlate with different perceptual attributes for women versus children.

4.1 Analysis

The relationship between auditory-perceptual and acoustic measures was investigated using the data previously described for the perceptual (Chapter 2) and acoustic (Chapter 3) analysis portions of this investigation. Specifically, the stimuli were produced by 4-6 year-old children with normal voices (n = 8) or vocal nodules (n = 8) and adult females with normal voices (n = 8) or vocal nodules (n = 8). Data from the perceptual and acoustic analyses of the vowel /a/ and a CAPE-V sentence were used.

Linear regression was performed for each acoustic measure to determine which acoustic measures significantly accounted for each of the perceptual ratings for the production of the vowel /a/ and a CAPE-V sentence. Specifically, each acoustic measure

was used as an independent predictor variable in a linear regression analysis to separately predict each of the perceptual parameters of overall severity, breathiness, and roughness for both vowel and sentence stimuli. The arithmetic means of the ratings for overall severity, roughness, breathiness and strain for each listener were entered as dependent variables in the regression analysis. Hence, the perceptual ratings for each sample were averaged across all reliable listeners (defined as listeners with a Pearson's correlation coefficient greater than 0.7), and the averaged ratings were used to explore the relationship between perceived severity for each of the perceptual parameters and each of the acoustic variables. The ability of the acoustic measures to predict the auditory-perceptual ratings for the production of vowels and sentences was examined by determining the acoustic measure(s) in the regression analysis which had significant correlations ($p < 0.05$).

4.2 Results

The results of the linear regression analyses showed that different acoustic measures were significantly correlated with different perceptual qualities for vowels and sentences.

4.2.1 Relationship between Acoustic and Perceptual Measures for Vowels and Sentences

Table 4.1 illustrates the significant correlations ($p< 0.05$) between acoustic measures and individual perceptual attributes. It also shows the few acoustic measures which exceeded a moderate correlation level as defined by the square of the pair-wise correlations ($r^2 \geq 0.5$) with perceptual attributes. Squaring the r-value provides an estimate of how much of the variance in one parameter can be accounted for by the variance in the other parameter so that $r^2 \geq 0.5$ means that at least 50% of the variance can be accounted for.

		Overall severity							Roughness							Breathiness							
		R1	CPP	HNR	F0	1k	2k	4k	R1	CPP	HNR	F0	1k	2k	4k	R1	CPP	HNR	F0	1k	2k	4k	Total
Child	Vowel																						18
	Sentence																						
Adult	Vowel																						9
	Sentence																						

Table 4.1: Significant correlations ($p<0.05$) between each of the acoustic measures and the perceptual rating of overall severity, roughness and breathiness for the vowel /a/ and a CAPE-V sentence produced by children and adult are shown in yellow. Acoustic measures which showed significant correlations and exceed moderate level of $r^2 \geq 0.5$ are shown in red.

The following sections describe the acoustic measures which were correlated with individual perceptual qualities for the vowel /a/ and CAPE-V sentences produced by 4-6 year old children and adult females. In presenting the results only the magnitudes of the correlations are shown because all of those that attained a meaningful level (in this case all that were ≥ 0.285) were negative. The negative correlations reflect the expected relationships in which judgments of increased severities for perceptual attributes were associated with reductions in acoustic measures.

Overall Severity

Figure 4.1 shows correlation coefficients from the regression analysis comparing perceptual ratings of overall severity with each of the examined acoustic measures for vowels and sentences produced by 4-6 year old children and adult females with normal voices and voices with nodules.

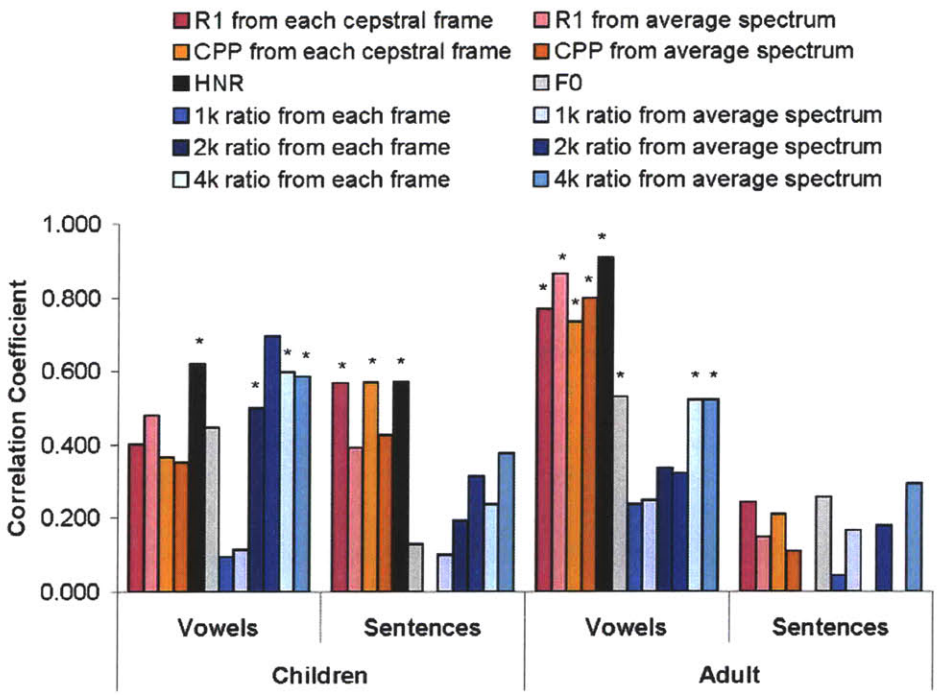


Figure 4.1: Correlation coefficients demonstrating the relationship between each of the acoustic measures and the perceptual rating of overall severity for the vowel /a/ and a CAPE-V sentence produced by 4-6 year old children and adult females, with normal voices and vocal nodules. Stars (*) indicate significant correlations ($p<0.05$).

As seen from Figure 4.1, the acoustic measures that significantly correlated with overall severity ratings for the vowel /a/ produced by 4-6 year old children were: HNR and the spectral slope measures of 2 kHz and 4 kHz energy ratio. For sentences, the acoustic measures that significantly correlated with overall severity ratings were: HNR, R1, and CPP.

Furthermore, the acoustic measures that significantly correlated with overall severity ratings for the vowel /a/ produced by adult females were: HNR, R1, CPP, F0, and the spectral slope measure of 4 kHz energy ratio. However, none of the examined acoustic measures were significantly correlated with the overall severity ratings for sentences produced by adult females.

Roughness

The correlation coefficients from the regression analysis comparing the perceptual ratings of roughness to each of the acoustic measures for vowels and CAPE-V sentences produced by 4-6 year old children and adult females with normal voices and voices with nodules are shown in Figure 4.2.

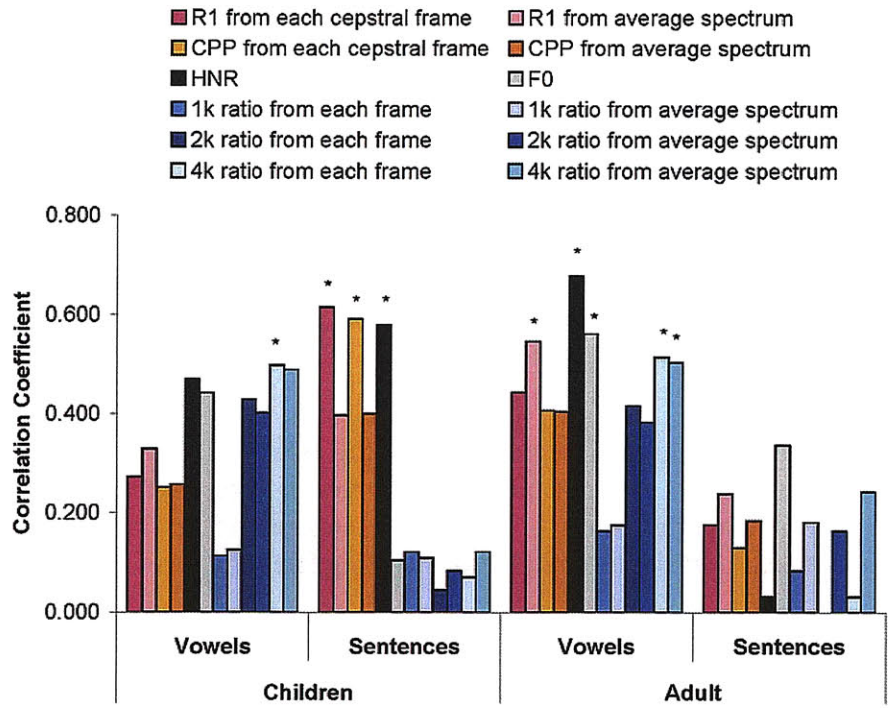


Figure 4.2: Correlation coefficients for each of the acoustic measures compared to roughness ratings for the vowel /a/ and a CAPE-V sentence produced by 4-6 year old children and adult females, with normal voices and vocal nodules. Stars (*) indicate significant correlations ($p<0.05$).

Figure 4.2 shows that the spectral slope measure of 4 kHz energy ratio correlated with roughness ratings for the production of the vowel /a/ for 4-6 year old children, whereas the acoustic measures of HNR, R1, and CPP significantly correlated with roughness ratings of sentences.

For vowels produced by adult females, the acoustic measures of HNR, R1, F0, and the spectral slope measure of 4 kHz energy ratio significantly correlated with the roughness ratings, whereas for CAPE-V sentences produced by adult females, none of the examined acoustic measures significantly correlated with the roughness ratings.

Breathiness

Figure 4.3 shows correlation coefficients from the results of the linear regression comparing the perceptual ratings for breathiness to each of the acoustic measures for vowels and CAPE-V sentences produced from 4-6 year old children and adult females with normal voices and voices with nodules.

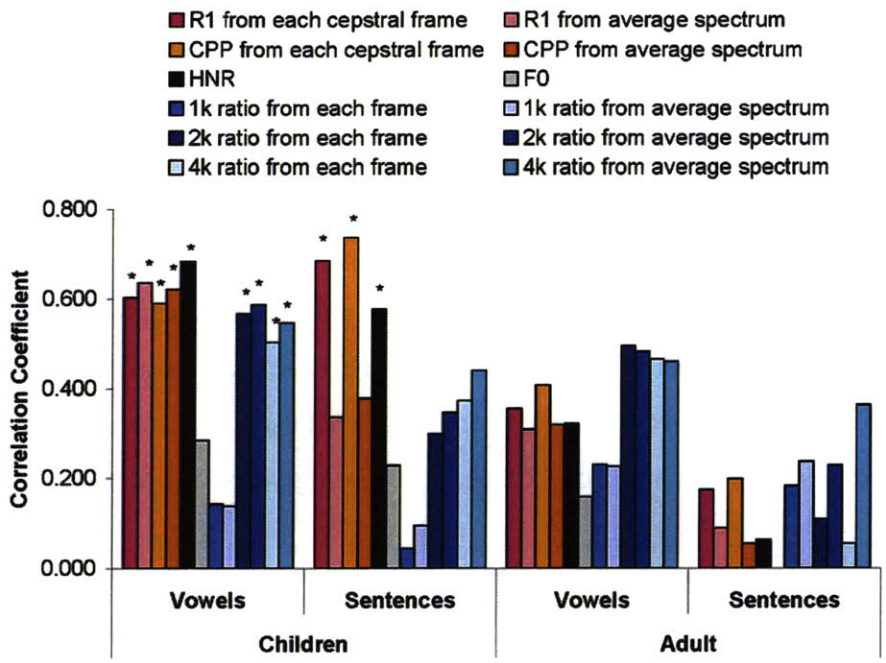


Figure 4.3: Correlation coefficients for each of the acoustic measures compared with the perceptual rating of breathiness for the vowel /a/ and a CAPE-V sentence produced by 4-6 year old children and adult females, with normal voices and vocal nodules. Stars (*) indicate significant correlations ($p<0.05$).

Figure 4.3 shows that the HNR, R1, CPP, and the spectral slope measures of 2 kHz and 4 kHz energy ratio significantly correlated with breathiness ratings for vowels, whereas for breathiness ratings of CAPE-V sentences, HNR, R1, and CPP significantly correlated for 4-6 year-old children.

For vowels and sentences produced by adult females, none of the examined acoustic measures significantly correlated with the breathiness ratings.

Strain

None of the examined acoustic measures significantly correlated with the perceptual ratings of strain for vowels or sentences produced by adult females or 4-6 year-old children.

4.3 Discussion

Since voice disorders may simultaneously affect multiple voice attributes, each perceptual rating is likely to be influenced by multiple acoustic properties rather than being influenced by any one particular parameter (de Krom, 1995). This view is supported by the results of the linear regression analyses which show that multiple acoustic measures can be significantly correlated with individual perceptual attributes. More specifically, these results suggest that a comprehensive analysis of the voice signal should consider both spectral-based and cepstral-based acoustic measures, which agrees with recent findings by Awan and Roy (2006) for assessing voice disorders in adults.

None of the acoustic measures were significantly correlated with the perceptual attribute of strain for either children or adults. This finding is consistent with previous observations that the auditory perception of strain has not been well characterized by objective measurements (Colton & Casper, 1996).

Although the significant correlations that were found between acoustic measures and perceptual attributes of overall severity, roughness and breathiness clearly indicate the existence of potentially useful relationships for quantifying voice quality, only a few of these attained or exceeded a moderate level as defined by the square of the pair-wise correlations ($r^2 \geq 0.5$). Moderate to strong relationships were found between overall severity of dysphonia in adult vowel production and measures of HNR ($r^2 = 0.828$), R1 from average spectrum ($r^2 = 0.750$), CPP from average spectrum ($r^2 = 0.638$), R1 from each frame ($r^2 = 0.591$), CPP from each frame ($r^2 = 0.540$). A moderate relationship was also found between breathiness and CPP from each frame ($r^2 = 0.543$) in vowel production for children. Moderate to strong correlations with perceptual attributes are required before an acoustic measure would be considered to have potential as a clinical

tool for quantifying voice quality. Thus, possible ways to increase the strength of relationships between the acoustic measures implemented in this investigation and perceptual attributes are discussed at the end of this section.

The fact that more of the correlations did not attain or exceed moderate levels is somewhat at odds with previous studies that have applied versions of some of these acoustic measures to adults with voice disorders and reported higher correlations with perceptual attributes (e.g., Awan & Roy, 2006; Heman-Ackah, et al., 2002; Hillenbrand, et al., 1994; Qi, Hillman, & Milstein, 1999). In addition to the obvious point that differences in actual implementation of acoustic algorithms reported in the literature could explain some of this variance in results, there are other factors that are probably more responsible. The biggest difference between the present study and previous investigations that have examined relationships between acoustic measures and perception is the current study's limitation of only using subjects with vocal nodules in the disordered groups. This was done to produce more homogeneity in the disordered groups and to limit comparisons between children and adult females to a disorder that occurs commonly in both cohorts. However, restricting the study to vocal nodules limited the nature and severity of the dysphonia that was manifested, which was clearly reflected across all perceptual attributes by ratings that fell mostly in the mild to moderate range (most scores were between 20 and 60 on the 100 point scale). Most previous studies have used combinations of normal and disordered voice samples that span all (or most) of the range for perceptual attributes (often using several types of voice disorders); from normal to severely disordered. This wider spread of perceptual responses better facilitates correlation analyses; as compared to using data that is restricted to only a part of the scale, as was the case in this investigation. It seems quite probable that the restricted ranges of the perceptual data in this study had a negative impact on the strength of the correlations that could be attained with acoustic measures. The fact that the adult group was rated as significantly less dysphonic as compared to the group of children seems to have further exacerbated this issue as evidenced by the lack of significant correlations across all perceptual attributes for the adult sentence productions, and for the breathiness parameter in adult vowel production. The lack of significant correlations between acoustic measures and ratings of breathiness in adult vowel production seems particularly

surprising given previous reports (e.g., Awan & Roy, 2006; Heman-Ackah, et al., 2002; Hillenbrand, et al., 1994; Qi & Hillman, 1997), until it is recalled that breathiness was not one of prominent perceptual features in adult females with nodules (see Chapter 2).

Since the focus of this work was on developing voice measures for children, it is somewhat encouraging that, as shown in Table 4.1, a larger number of significant correlations between perceptual attributes and acoustic measures were found for children (18 significant correlations) than for adults (9 significant correlations). In the very least this is further evidence that this investigation was successful in developing algorithms that can be applied with some confidence to analyzing the voices of children, i.e., the special challenges in dealing with the higher pitches and formants of children's speech have been met.

The results of the correlation analyses lend further general support to the view that the algorithms developed in this investigation have the potential to provide quantitative assessment of voice quality in children; a capability that has not previously been made readily available. The significant correlations that were found between acoustic measures and perceptual attributes of overall severity, roughness and breathiness clearly indicate the existence of potentially useful relationships that need to be explored further. Future attempts to further develop these algorithms for the clinical assessment of children's voices should include: 1) application to a wider range of voice disorders and severities of dysphonias in children to produce perceptual data that spans the entire ranges of perceptual attributes, 2) collection of a larger set of reliable perceptual data, perhaps using more recently described psychoacoustic-based methods (Kreiman, et al., 1993; Shrivastav, et al., 2005), and 3) use of multivariate statistical approaches (e.g., multiple regression) to relate sets of acoustic measures to perceptual attributes since it is clear that more than one acoustic measure will be needed to capture the variance in perceptual attributes (this will require larger sets of reliable perceptual data), and 4) test acoustic algorithms on synthesized continuous speech samples.

Chapter 5: Summary of Main Findings and Conclusions

A primary goal of this thesis was to identify speech tasks and acoustic analysis measures that were most salient in characterizing the impact of vocal nodules on vocal function in children. Two types of speech materials were compared: production of the sustained vowel /a/ and production of a CAPE-V sentence (pediatric version of the CAPE-V produced by 4-6 year-old and 8-10 year-old children and adult version of the CAPE-V produced by 8-10 year old children and adult females). The acoustic measures extracted from both the vowel and the continuous speech samples were mean fundamental frequency, first cepstral peak, cepstrum-based HNR, and spectral measure of overall spectral slope. The stimuli were recorded from 4-6 year-old children with normal voices (n = 8) or disordered voices with vocal nodules (n = 8), 8-10 year-old children with normal voices (n = 8) or disordered voices with vocal nodules (n = 8), and adult females with normal voices (n = 8) or disordered voices with vocal nodules (n = 8).

In an effort to elicit a set of standardized speech stimuli from children that mimicked the vocal behaviors elicited by the production of the standardized CAPE-V sentences read by adults, pediatric CAPE-V sentences were created using picture stimuli that enabled children as young as three years old to "read" the sentences. It was hypothesized that the pediatric and adult versions of the CAPE-V would produce equivalent perceptual results. Despite carefully controlling the parameters used to create the pediatric CAPE-V sentences, perceptual ratings of voice quality for the pediatric and adult versions of the CAPE-V protocol differed. More specifically, perceptual ratings for the production of the sentences from the pediatric version of the CAPE-V protocol were consistently indicated as being more severe than the ratings for production of sentences from the adult version of the CAPE-V. Even though the original hypothesis was not supported, the higher ratings (more disordered) and larger ranges elicited by the pediatric version of the CAPE-V may actually be clinically advantageous if it provides better differentiation from normal and better separation between different levels of dysphonia severity. Perhaps an inadvertent outcome of designing the pediatric version is that it is elicits vocal function that is more sensitive to the presence of nodules than normal read speech. These conjectures would need to be explored in studies involving larger numbers

of subjects. Also supporting clinical use of the pediatric CAPE-V were the findings that it was associated with higher within-rater and between-rater reliability than the adult version. The findings in this thesis supported the hypothesis that perceptual judgments of voice quality would correlate better with acoustic measures when both were obtained from the production of CAPE-V sentences compared to the production of sustained vowels. Specifically, production of sentences rather than vowels differentiated normal voices and voices with vocal nodules on the perceptual qualities of overall severity, roughness and strain for both children and adults.

Even though each of the perceptual parameters significantly distinguished between normal voices and voices with vocal nodules for both children and adults, the hypothesis that roughness and strain would most differentiate normal voices and voices with vocal nodules for children was not confirmed. Rather, breathiness was found to be the perceptual parameter that best differentiated normal voices and voices with vocal nodules for children. For adults, the hypothesis that roughness would be among the top perceptual parameters in differentiating normal voices and voices with vocal nodules was confirmed, although, contrary to predictions, ratings for breathiness were not as good at differentiating normal voices and voices with vocal nodules for adults.

Conventional acoustic algorithms, typically used to analyze adult voices, were modified to enable an analysis of vowels and sentences produced by children. Unlike the conventional acoustic algorithms, the modified acoustic algorithms used a pitch-period dependent window length in order to maintain fundamental frequency independence, which was important for comparisons between children and adult voices having a wide range of fundamental frequencies. Validation of the modified acoustic algorithms with synthesized vowels created with speech characteristics (e.g., formant frequencies) of children and adults indicated that the modified acoustic algorithms were insensitive to variations in fundamental frequency. Furthermore, the potential sensitivity of the modified acoustic algorithms for evaluating a variety of dysphonic voices was demonstrated.

Testing the hypothesis that different combinations of acoustic measures would show the highest correlations with the perceptual parameters was difficult to assess given

the limited number of voice stimuli tested in this thesis. However, numerous individual acoustic measures showed moderate correlations with perceptual ratings for the same speech stimuli. This preliminary evidence suggests that the pitch-independent acoustic measures developed in this thesis correlate with various perceptual characteristics of children and adults. However, because only a few strong correlations between acoustic measures and perceptual ratings were observed, additional studies are warranted to give better statistical significance and extend the results of this thesis. One possible improvement may be to use a larger number of subjects to explore questions raised by the perceptual experiment such as why the pediatric CAPE-V sentences were rated as being more severe than the production of the sentences in the adult version of the CAPE-V protocol and why the pediatric version of the CAPE-V produced higher within-rater and between-rater reliability than the adult version. In summary, the results of this thesis indicate that eliciting the sentences in the pediatric version of the CAPE-V proved useful in the analysis of children's voices, and such measures may be clinically useful as the results of the perceptual experiments showed that sentences as opposed to vowels most differentiated normal voices and voices with vocal nodules. These results also suggest that a comprehensive analysis of the voice signal should consider both spectral-based and cepstral-based acoustic measures. Findings for the correlation analyses generally support the view that the algorithms developed in this investigation have the potential to provide quantitative assessment of voice quality in children; a capability that has not previously been made readily available. The significant correlations that were found between acoustic measures and perceptual attributes of overall severity, roughness and breathiness clearly indicate the existence of potentially useful relationships that need to be explored further.

Future attempts to further develop these algorithms for the clinical assessment of children's voices should include: 1) application to a wider range of voice disorders and severities of dysphonias in children to produce perceptual data that spans the entire ranges of perceptual attributes, 2) collection of a larger set of reliable perceptual data, perhaps using more recently described psychoacoustic-based methods (Kreiman, et al., 1993; Shrivastav, et al., 2005), and 3) use of multivariate statistical approaches (e.g., multiple regression) to relate sets of acoustic measures to perceptual attributes since it is

clear that more than one acoustic measure will be needed to capture the variance in perceptual attributes (this will require larger sets of reliable perceptual data).

Appendix A: CAPE-V Protocol Form

Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)

Name:_____ **Date:**_____

The following parameters ofvoice quality will be rated upon completion of the following tasks:
1.Sustained vowels, /a/ and /i/for 3-5 seconds duration each.
2.Sentence production:

a.The blue spot ison the key again.
b.How hard did he hit him?
c.We were awaya year ago.
d. We eat eggs every Easter.
e. Mymama makes lemon muffins.
f.Peter will keep at the peak.

3.Spontaneous speech in response to: "Tell me about your voice problem." or "Tell me how your voice is functioning."

Legend: C = Consistent I= Inconsistent
MI= Mildly Deviant
MO= Moderately Deviant
SE= Severely Deviant

				SCORE
Overall Severity	_____			C I ___/100
	MI	MO	SE	
Roughness	_____			C I ___/100
	MI	MO	SE	
Breathiness	_____			C I ___/100
	MI	MO	SE	
Strain	_____			C I ___/100
	MI	MO	SE	
Pitch	(Indicate the nature of the abnormality):_____			
	_____			C I ___/100
	MI	MO	SE	
Loudness	(Indicate the nature of the abnormality):_____			
	_____			C I ___/100
	MI	MO	SE	

COMMENTS ABOUT RESONANCE: NORMAL OTHER (Provide description):_____

ADDITIONAL FEATURES (for example, diplophonia,fry,falsetto,asthenia,aphonia,pitch instability,tremor,wet/gurgly, or other relevant terms):

Bibliography

Andrews, M., & Summers, A. (1988). *From voice therapy for adolescents*. Boston, MA: College-Hill Press.

Awan, S. N., & Roy, M. (2005). Acoustic Prediction of Voice Type in Women with Functional Dysphonia. *J Voice, 19*(2), 268-282.

Awan, S. N., & Roy, N. (2006). Toward the development of an objective index of dysphonia severity: a four-factor acoustic model. *Clin Linguist Phon, 20*(1), 35-49.

Baken, R., & Orlikoff, R. (1987). The effect of articulation on fundamental frequency in speakers and singers. *J Voice, 1*, 68-76.

Bless, D. M., & Abbs, J. H. (1983). *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. San Diego: College Hill Press.

Boersma, P. (1993). Accurage short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Inst Phon Sci, 17*, 97-110.

Boersma, P., & Weenink, D. (2006). Praat version 4.4.04, from Available at http://www.fon.hum.uva.nl/praat

Cohen, B. H. (1996). *Explaining psychological statistics*. Pacific Grove, CA: Brooks/Cole Publishing Co.

Colton, R. H., & Casper, J. K. (1996). *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment* (2nd ed.). Baltimore, MD: Lippincott Williams & Wilkins.

de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J Speech Hear Res, 36*(2), 254-266.

de Krom, G. (1994). Consistency and reliability of voice quality ratings for different types of speech fragments. *J Speech Hear Res, 37*(5), 985-1000.

de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J Speech Hear Res, 38*(4), 794-811.

Deal, R. E., & Emanuel, F. W. (1978). Some waveform and spectral features of vowel roughness. *J Speech Hear Res, 21*(2), 250-264.

Dejonckere, P. H., & Wieneke, G. H. (1994). Spectral, cepstral and aperiodicity characteristics of pathological voices before and after phonosurgical treatment. *Clin Linguist Phon, 8*(2), 161-169.

Dejonckere, P. H., & Wieneke, G. H. (1996). *Cepstra of normal and pathological voices: Correlation with acoustic, aerodynamic and perceptual data*. Amsterdam: John Benjamins Publishing Co.

Dobres, R., Lee, L., Stemple, J. C., Kummer, A. W., & Kretschmer, L. W. (1990). Description of laryngeal pathologies in children evaluated by otolaryngologists. *J Speech Hear Disord, 55*(3), 526-532.

Eadie, T. L., & Doyle, P. C. (2005). Classification of dysphonic voice: acoustic and auditory-perceptual measures. *J Voice, 19*(1), 1-14.

Fant, G., Liljencrants, J., & Lin, Q. (1985). A four-parameter model of glottal flow. *Speech Transmission Laboratory Quarterly Progress and Status Report, 4*, 1-13.

Fritzell, B. (1996). Voice disorders and occupations. *Logoped Phoniatr Vocol, 21*, 7-12.

Frokjaer-Jensen, B., & Prytz, S. (1976). Registration of voice quality. *Bruel and Kjaer Technical Review, 3*, 3-17.

Fukazawa, T., el-Assuooty, A., & Honjo, I. (1988). A new index for evaluation of the turbulent noise in pathological voice. *J Acoust Soc Am, 83*(3), 1189-1193.

Gray, S. D., Hirano, M., & Sato, K. (1993). *Vocal fold physiology: Frontiers in basic science*. San Diego: Singular.

Haelsig, P. C., & Madison, C. L. (1986). A study on phonological processes exhibited by 3-, 4-, and 5-year old children. *Lang Speech Hear Serv Sch, 17*, 107-114.

Halberstam, B. (2004). Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL J Otorhinolaryngol Relat Spec, 66*(2), 70-73.

Hammarberg, B., Fritzell, B., Gauffin, J., & Sundberg, J. (1986). Acoustic and perceptual analysis of vocal dysfunction. *J Phonetics, 14*, 533-547.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol, 90*(5-6), 441-451.

Hammarberg, B., Fritzell, B., & Schiratzki, H. (1984). Teflon injection in 16 patients with paralytic dysphonia: perceptual and acoustic evaluations. *J Speech Hear Disord, 49*(1), 72-82.

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

Hecker, M. H., & Kreul, E. J. (1971). Descriptions of the speech of patients with cancer of the vocal folds. I. Measures of fundamental frequency. *J Acoust Soc Am, 49*(4), Suppl 2:1275+.

Heman-Ackah, Y. D., Heuer, R. J., Michael, D. D., Ostrowski, R., Horman, M., Baroody, M. M., et al. (2003). Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol, 112*(4), 324-333.

Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S., Jr. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice, 16*(1), 20-27.

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *J Speech Hear Res, 30*(4), 448-461.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *J Speech Hear Res, 37*(4), 769-778.

Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Hear Res, 39*(2), 311-321.

Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: an experimental framework and initial results. *J Speech Hear Res, 32*(2), 373-392.

Hirano, M., Hibi, S., Terasawa, R., & Fujiu, M. (1986). Relationship between aerodynamic, vibratory, acoustic and psychoacoustic correlates in dysphonia. *J. Phonetics, 14*, 445-456.

Hirano, M., Kurita, S., & Nakashima, T. (1983). Growth, development, and aging of human vocal folds. In D. M. Bless & J. H. Abbs (Eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. San Diego: College Hill Press.

Holmberg, E. B., Hillman, R. E., Hammarberg, B., Sodersten, M., & Doyle, P. (2001). Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *J Voice, 15*(3), 395-412.

Kilic, M. A., Okur, E., Yildirim, I., & Guzelsoy, S. (2004). The prevalence of vocal fold nodules in school age children. *International Journal of Pediatric Otorhinolaryngology, 68*, 409-412.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am, 87*(2), 820-857.

Klingholtz, F. (1990). Acoustic recognition of voice disorders: a comparative study of running speech versus sustained vowels. *J Acoust Soc Am, 87*(5), 2218-2224.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res, 36*(1), 21-40.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: developmental changes of temporal and spectral parameters. *J Acoust Soc Am, 105*(3), 1455-1468.

Linville, S. E. (2002). Source characteristics of aged voice assessed from long-term average spectra. *J Voice, 16*(4), 472-479.

Lotz, W. K., D'Antonio, L. L., Chait, D. H., & Netsell, R. W. (1993). Successful nasoendoscopic and aerodynamic examinations of children with speech/voice disorders. *Int J Pediatr Otorhinolaryngol, 26*(2), 165-172.

Lucero, J. C., & Koenig, L. L. (2005). Phonation thresholds as a function of laryngeal size in a two-mass model of the vocal folds. *J Acoust Soc Am, 118*(5), 2798-2801.

Mannell, R. H. (1998). *Formant diphone parameter extraction utilising a labelled single-speaker database*. Paper presented at the ICSLP.

McAllister, A. (1997). *Acoustic, perceptual and physiological studies of ten-year-old children's voices.* KTH.

Mclaughlin, S. F. (1998). *Introduction to Language Development*. San Diego: Singular Publishing Group, Inc.

Mehta, D. D. (2006). *Aspiration noise during phonation: Synthesis, analysis and pitch-scale modification.* Massachusetts Institute of Technology, Cambridge, MA.

Mendoza, E., Valencia, N., Munoz, J., & Trujillo, H. (1996). Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). *J Voice, 10*(1), 59-66.

Moore, G. P. (1971). *Organic voice disorders*. Englewood Cliffs: Prentice Hall.

Morrison, M., & Rammage, L. A. (1994). *The management of voice disorders*. San Diego: Singular Publishing Group, Inc.

Murphy, P. J. (1999). Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis. *J Acoust Soc Am, 105*(5), 2866-2881.

Murphy, P. J. (2000). Spectral characterization of jitter, shimmer, and additive noise in synthetically generated voice signals. *J Acoust Soc Am, 107*(2), 978-988.

Murphy, P. J. (2006). On first rahmonic amplitude in the analysis of synthesized aperiodic voice signals. *J Acoust Soc Am, 120*(5 Pt 1), 2896-2907.

Murphy, P. J., & Akande, O. O. (2007). Noise estimation in voice signals using short-term cepstral analysis. *J Acoust Soc Am, 121*(3), 1679-1690.

Nagata, K., Kurita, S., Yasumoto, S., Maeda, T., Kawasaki, H., & Hirano, M. (1983). Vocal fold polyps and nodules. A 10-year review of 1,156 patients. *Auris Nasus Larynx, 10 Suppl*, S27-35.

Noll, A. M. (1967). Cepstrum pitch determination. *J Acoust Soc Am, 41*(2), 293-309.

Oppenheim, A. V. (1969). Speech analysis-synthesis system based on homomorphic filtering. *J Acoust Soc Am, 45*(2), 458-465.

Parsa, V., & Jamieson, D. G. (2001). Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res, 44*(2), 327-339.

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of Vowels. *J Acoust Soc Am, 24*(2), 175-184.

Qi, Y., & Hillman, R. E. (1997). Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *J Acoust Soc Am, 102*(1), 537-543.

Qi, Y., Hillman, R. E., & Milstein, C. (1999). The estimation of signal-to-noise ratio in continuous speech for disordered voices. *J Acoust Soc Am, 105*(4), 2532-2535.

Rahman, M. S., & Shimamura, T. (2005). Forman frequency estimation of high-pitched speech by homomorphic prediction. *Acoust Sci Tech, 26*(6), 502-510.

Ramig, L. O., & Verdolini, K. (1998). Treatment efficacy: voice disorders. *J Speech Lang Hear Res, 41*(1), S101-116.

Rosenberg, A. E. (1971). Effect of glottal pulse shape on the quality of natural vowels. *Journal of the Acoustical Society of America, 49*(2).

Sapienza, C. M., Ruddy, B. H., & Baker, S. (2004). Laryngeal structure and function in the pediatric larynx: clinical applications. *Lang Speech Hear Serv Sch, 35*(4), 299-307.

Schoentgen, J. (1989). Jitter in sustained vowels and isolated sentences produced by dysphonic speakers. *Speech Comm, 8*, 61-79.

Sederholm, E. (1996). *Hoarseness in ten-year-old children: perceptual characteristic, prevalence and etiology.* Karolinska Institute and KTH, Stockholm.

Sederholm, E., McAllister, A., Sundberg, J., & Dalkvist, J. (1993). Perceptual analysis of child hoarseness using continuous scales. *Scand J Log Phon.*

Shah, R. K., Woodnorth, G. H., Glynn, A., & Nuss, R. C. (2005). Pediatric vocal nodules: correlation with perceptual voice analysis. *Int J Pediatr Otorhinolaryngol, 69*(7), 903-909.

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *J Speech Hear Res, 36*(1), 105-140.

Shrivastav, R., Sapienza, C. M., & Nandur, V. (2005). Application of psychometric theory to the measurement of voice quality using rating scales. *J Speech Lang Hear Res, 48*(2), 323-335.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *J Speech Hear Disord, 55*(4), 779-798.

Stevens, K. N. (1998). Acoustic Phonetics. Cambridge, MA: MIT Press.

Titze, I. (1995). *Summary statement: workshop on acoustic voice analysis.* Paper presented at the National Center for Voice and Speech.

Verhelst, W., & Steenhaut, O. (1986). A new model for the short-time complex cepstrum of voiced speech. *IEEE Trans Acoust, Speech Sig Proc, ASSP-34*, 43-51.

Wilson, D. K. (1987). *Voice Problems in Children* (3rd ed.). Baltimore: Williams and Wilkins.

Wolfe, V. I., Martin, D. P., & Palmer, C. I. (2000). Perception of dysphonic voice quality by naive listeners. *J Speech Lang Hear Res, 43*(3), 697-705.

Yiu, E. M., Worrall, L. E., Longland, J., & Mitchell, C. (2000). Analyzing vocal quality of connected using Kay's Computerized Speech Lab: A preliminary finding. *Clin Linguist Phon, 13*, 295-305.

Zraick, R. I., Wendel, K., & Smith-Olinde, L. (2005). The effect of speaking task on perceptual judgment of the severity of dysphonic voice. *J Voice, 19*(4), 574-581.