# Cepstral Analysis of Voice in Phononormic Adults in the Age Range of 20-40 Years

Sujitha P S

**Register Number: 13SLP027**

**A Dissertation Submitted in Part Fulfillment of Final Year**

**Master of Science (Speech Language Pathology)**

**University of Mysore, Mysore**

**ALL INDIA INSTITUTE OF SPEECH AND HEARING,**

**MANASAGANGOTHRI, MYSORE- 570 006**

**MAY, 2015**

# CERTIFICATE

This is to certify that this dissertation entitled **"Cepstral Analysis of Voice in Phononormic Adults in the Age Range of 20-40 Years"** is a bonafide work submitted in part fulfillment for the Degree of Master of Science (Speech Language Pathology) of the student (Registration No.: 13SLP027). This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any of the University for the award of any Diploma or Degree.

**Mysore**
**May 2015**

**Prof. S. R. Savithri**
**Director**
**All India Institute of Speech and Hearing**
**Manasagangothri, Mysore – 570006**

# CERTIFICATE

This is to certify that this dissertation entitled **"Cepstral Analysis of Voice in Phononormic Adults in the Age Range of 20-40 Years"** is a bonafide work in part fulfillment for the degree of Master of Science (Speech-Language Pathology) of the student (Registration No. 13SLP027). This has been carried out under my guidance and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

P.G.Kishore 14/5/15

**Mysore**
**May, 2015**

**Mr. Gopikishore Pebbili**
**Guide**
Lecturer in Speech Pathology
Department of Speech-Language Pathology
All India Institute of Speech and Hearing
Manasagangothri, Mysore – 570006

# DECLARATION

This is to certify that this dissertation entitled “Cepstral Analysis of Voice in Phononormic Adults in the Age Range of 20-40 Years” is the result of my own study under the guidance of Mr. Gopi Kishore Pebbili., Lecturer in Speech Pathology, Department of Speech-Language Pathology, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier in other University for the award of any Diploma or Degree.

**Mysore** **Register No.: 13SLP027**

**May 2015**

Dedicated to,
Amma & Acha

# ACKNOWLEDGEMENT

**"Be thankful for what you have; you'll end up having more. If you concentrate on what you don't have, you will never, ever have enough." –** ***Oprah Winfrey***

Every happy moments, difficult moments and quiet moments have been accompanied by God's blessings, support and grace.  Thank you Lord, for granting me wisdom and strength throughout these years.

This journey would not have been possible without the support of my family, teachers and friends. To my family, thank you for encouraging me in all of my pursuits and inspiring me to follow my dreams. I am especially grateful to my parents (Mr P K Sreekumar & Mrs Geetha K), who loved me unconditionally supported me, encouraged me, and. I always knew that you believed in me and wanted the best for me. To my dynamic brother Appu (Sudhy P S), you have been a silent support for me through all the ups and downs in my life. To my Grandma, I express my thanks to Lakshmikutty ammamma who always remembered me in her prayers.

I would like to express my sincere thanks to our director (Dr. S.R. Savithri) for giving me an opportunity to carry out this dissertation. To my Sir (Mr Gopi Kishore Pebbili), thank you for your patience, motivation, enthusiasm, and immense knowledge Teachers like you are hard to find, and I'm extremely grateful for everything you've taught me and your guidance enabled me to grow in my career. I gratefully acknowledge Navya ma'am for her valuable advises and crucial contribution to this study. Your patience and willingness to guide me would always be appreciated and remembered. Thank you ma'am for everything.

I would like to express my sincere gratitude to Dr Vasanthalakshmi and Mr Santhosha for your help in Statistical Analysis. I would also like to thank Dr Pushpavathi for your support and guidance. I also extend my gratitude to Reuben Sir, Irfana ma'am and Preethi ma'am for your insightful comments, valuable suggestions and your constant support.

To my friends (Merlin, Salome, Malavika, Swathi, Prathiba & Arun) and to all my classmates, thank you for listening, offering me advice, and supporting me through this entire process. Special thanks to my Malu, Sallu & Merlu for being there whenever I needed a friend.

I would like to thank Kavya, Sumanth, Gopika, Nadeer, Sachin, Gowtham & Prajwal for their timely help during my data collection and also thanks to all participants of my study for their cooperation. There are loads more people who directly or indirectly helped me for making this dissertation a success.

………Thanks to all………

# LIST OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

A cepstrum is a log power spectrum of a log power spectrum (Hillenbrand, Ronald, Cleveland, & Erickson, 1994). It is well-defined as a discrete Fourier transform of the logarithm power spectrum (Hillenbrand et al, 1994, & Hillenbrand and Houde, 1996). Cepstrum was first introduced by Bogert, Healy and Tukey (1963) to explain seismic signal processing. In 1966, Noll, designed cepstral analysis for measuring fundamental frequency. Later, Hillenbrand et al (1994) and Hillenbrand et al (1996) applied this cepstral analysis for determining breathiness in dysphonic voices.

The cepstral operations are labeled as ‘quefrency’ for frequency, ‘rhamonics’ for harmonics and ‘liftering’ for filtering. Though Cepstral analysis gives various measures like Cepstral Peak Prominence (CPP), smoothened Cepstral peak prominence (sCPP); the literature reveals that these are the strong measures to indicate dysphonia severity in comparison to noise to harmonic ration (NHR), amplitude perturbation quotient (APQ), relative average perturbation (RAP), and smoothened pitch perturbation quotient (sPPQ) (Kumar, Bhat & Prasad, 2010). Heman-Ackah (2004) enlightened about a robust algorithm that is used for voice analysis, it is the CPP; this algorithm examine the amount of harmonic structure prevailing in a selected voice signal.

The peak with highest amplitude in the cepstrum is the cepstral peak. The CPP is determined by measuring the amplitude difference from the highest peak of the cepstrum to the corresponding regression line which is drawn directly below to the cepstral peaks. Thus, CPP is considered to a measure of the degree of harmonics. This also explains how prominently the peaks emanate out from the cepstral “background noise” (Hillenbrand et al, 1994). A voice signal which is considered as normal will have a well-defined harmonic structure, and such as voice signal would have a strong cepstral peaks too. A dysphonic voice signal which is aperiodic in nature or with an increased spectral noise exhibits decrease in amplitude of the cepstral peak. When CPP is measured rather than considering absolute amplitude, its prominence needs to be measured. This is because the degree of periodicity, overall energy and window size can affect the amplitude of cepstral peak (Hillenbrand et al, 1994).

It is well established in literature that traditional voice analysis methods, majorly the jitter measures, similarly shimmer measures which are the time-based measures, are extensively biased by extraneous variables, and those variables can be the intensity of voice signal, the vowel being selected for analysis, and the frequency of phonation (Awan, Giovinco, & Owens, 2012). As these time-based measures require cycle to cycle boundary identification on the estimation of fundamental frequency for analysis, they cannot be applied to severe dysphonic voices with high aperiodicity (Moers, Bernd, Rosanowski, Elmar, Ulrich, & Tino, 2012). Further, the time-based measures of voice quality cannot be applicable to connected speech. Moreover, in many studies NHR, jitter, and shimmer measures, which are the traditional measures, are reported to be the unreliable predictors of dysphonia (Balasubramanium et al 2010; Heman-Ackah, Michael, Baroody, Ostrowski, Hillenbrand, Heuer, Horman, & Sataloff, 2003). Cepstral analysis has been reported to be an effective alternate measure to overcome these limitations of time based measures.

Cepstral analysis methods estimate aperiodicity or additive noises in a voice signal without identification of cycle boundaries; this is one of its principal advantage (Awan et al 2012). Thus, by measuring cepstral peak prominence using cepstral analysis the dysphonia severity can be identified on sustained vowel productions. Also, cepstral method can be used to evaluate dysphonia severity in connected speech. It is an easy acoustic analysis method which can done even using freely downloadable software.

Thus, growing pressure on technology has encouraged discovering more objective method which is further reliable and also valid to measure fundamental frequency in the field of voice evaluation. Beginning of relatively developed use of personal computers, acoustic analysis software of low cost and greater availability of digital audio recorders, leads the acoustic analysis of voice to be more popular and easier task.

**Need for the study**: Cepstral analysis of voice has been reported to be a more reliable and valid parameter to evaluate voices even with a high level of aperiodicity. It permits use of speech as a stimulus, thus making the analysis more naturalistic. Among the other parameters obtained from the cepstral analysis of voice, CPP has been reported to be a strong correlate of perceptually normal and dysphonic voices by a number of studies. Studies also reported successful use of CPP as a parameter to differentiate normal and dysphonic voices and as a parameter to track therapy/surgical outcomes. The vast applications and the ease involved in measuring, makes CPP an essential part of the routine voice evaluation. However, as CPP is reported to be influenced by the factors such as age and gender of the participant, stimuli and algorithm used for analysis etc, it is essential to have age and gender specific norms to a particular geographic region to serve as reference. Therefore, considering the relevance of cepstral measures in evaluating quality of voice and the dearth of normative data for CPP and sCPP, the present study investigates the CPP and sCPP in phononormic adults in the age range of 20-40 years.

**Aim of the study:** To establish reference data for CPP and sCPP in phononormic adults in the age range of 20-40 years.

**Objective:** The objectives of the study are 1. To establish reference values for CPP and sCPP in phonation and connected speech in phononormic adults between 20 to 40 years of age. 2. To investigate the effect of age on CPP and sCPP. 3. To investigate the effect of gender on CPP and sCPP. 4. To investigate the effect of stimulus on CPP and sCPP.

# CHAPTER II

# REVIEW OF LITERATURE

Voice is a multidimensional entity and it reveals the speaker's physical and emotional health, personality and identity. The production of voice depends on the coordination among the respiratory, phonatory, and resonatory systems. Voice production is an aerodynamic process in which the acoustical waves are created by laryngeal modulations of respiratory airflow. These waves are amplified and filtered accordingly by vocal tract resonance. For an economical or optimum vocal output, stability across these respiratory, phonatory, and resonatory subsystems is essential. A disturbance at any of these subsystems may lead to compensatory changes and resulting in voice problem. Therefore, for a voice evaluation to be comprehensive, it should consider assessment of functioning of each of these subsystems of voice production (Awan, 2011). According to Dejonckere, Bradley, Clemente, et al., (2001) clinical assessment of voice disorders should consist of perceptual voice assessment, videolaryngostroboscopy, objective measurements (acoustic analysis and aerodynamic measurements), and subjective self-evaluation of voice.

Auditory perceptual assessment plays a vital role in voice evaluation despite its inherent subjectivity and the lingering debate regarding its reliability and validity (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993; Kent, 1996; De Bodt, Van de Heyning, Wuyts, & Lambrechts, 1996; Webb, Carding, Deary, MacKenzie, Steen, & Wilson, 2004; Behrman, 2005; Ma & Yiu, 2005; Eadie, Kapsner, Rosenzweig, Waugh, Hillel, & Merati, 2010). The most evident advantages of using the perceptual voice assessment are accessibility of the test materials and simplicity in implementation procedures. Auditory perceptual measures are often considered as gold standard and used as a reference for other objective voice assessment tools such as acoustic analysis, aerodynamic analysis etc. However, perceptually rating voice quality is universally acknowledged as a difficult task and the clinician requires intensive training in the vocal dimensions that identify pathology most effectively. Eadie and Baylor (2006) have shown that the reliability of ratings can be affected by rater's exposure to different types of voices, as well as their level of training, and fatigue. In addition, the type of speech task (continuous speech versus sustained vowel) and the type of scale used have been shown to affect perceptual judgment.

The instrumental analysis provides quantitative and objective data on a wide range of different speech parameters far beyond the scope of an auditory-based judgment (Baken, 1987). The aerodynamic component of the voice evaluation provides information related to the valving efficiency of the glottis during phonation, as well as to the respiratory capacity, which yields measurements for air pressure, airflow, and air volume (Sapienza, 1996). Aerodynamic analysis of voice includes static measures of respiration and dynamic measures of laryngeal valving. The static measures help in understanding the volumes of air that can be inhales/exhaled in a breath and maximum capacities of an individual's respiratory system. The dynamic measures such as subglottic pressure and laryngeal resistance provide information about the efficiency of laryngeal valving in converting the expiratory airstream to acoustic energy.

The vocal imaging techniques such as videolaryngoscopy, high speed digital videoendoscopy, or videokymography are used for the visualizing the glottic and supraglottic structures of larynx. The vocal fold vibratory characteristics observed through these imaging techniques are useful in understanding the etiology of the voice disorder and in documenting the therapeutic/surgical outcomes. Each of these imaging techniques provides information about different physiological aspects of vocal function that is complementary to each other. Additionally, electroglottographic (EGG) measurements can provide objective data about the fundamental frequency and generally about the opening and closing phase within the oscillation cycles.

The use of ‘‘quality-of-life’’ measures in the area of voice disorders has gained momentum in the last decade as part of the routine clinical voice evaluation battery. Quality of life measurement is one way to assess the overall outcome of the physical, mental, and social well-being of a patient after a health-related problem. The available self-assessment rating scales are the Voice Handicap Index (VHI) (Jacobson, Johnson, Grywalski, Silbergeit, Jacobson, Benninger, et al, 1997), the Voice Outcome Survey (Glicklich, Glovsky, & Montgomery, 1999), Voice Handicap Index-10 (VHI-10) (Rosen, Lee, Osborne, Zullo & Murphy, 2004), Voice-Related Quality of Life (Hogikyan, & Sethuraman, 1999), Voice Activity and Participation Profile (VAPP) (Ma & Yiu, 2001) and the Outcome Scale (Casperi, 2001) are currently being used throughout the world to assess dysphonia as an outcome measure in adult populations (Benninger, Ahuja, Gardner, & Grywalski, 1998; Murry, & Rosen, 2000).

Among the different types of instrumental analysis that could be used in speech disorders (e.g., acoustic, aerodynamic, electromyographic), acoustic analysis is reported to be highly advantageous (Perkell, Guenthe, Lane, et al. 2000). The computerized multidimensional acoustic voice analysis enables visual and numeric information on the analyzed voice. The aim of such analysis program is to provide objective data and to support perceptual voice evaluation (Hammarberg, Fritzel, Gauffin, & Sundberg, 1986; Johns, Garrett, Hwang, Ossoff, & Courey, 2004). Also, acoustic analysis is noninvasive and does not require semi-invasive tools like endoscopes. Therefore participants may be tested easily for different voice tasks, such as sustained phonation of different vowels or running speech. Such an analysis of different voice tasks is important to assess the everyday performance and range of the voice. Acoustic measurements are usually conducted by means of dedicated computer software.

Traditionally the clinical acoustic voice analysis and majority of research studies examined sets of acoustic parameters such as fundamental frequency (F0), measures of frequency- perturbation (e.g., jitter), measures of amplitude-perturbation (e.g., shimmer) and various noise-indices (Llorente, Ruiz, Lechon, et al., 2008). Fundamental frequency is often reported in voice research and this measure corresponds directly to the vibratory rate of the vocal folds (Colton, & Casper, 1996) and reflects their biomechanical characteristics (Baken, & Orlikoff, 2002). Perturbation can be considered as a generic term used to describe some form of variation in the voiced speech waveform from cycle to cycle. This is usually simply stated or defined as the variation in the measurement of the assumed underlying period from cycle to cycle (jitter), or the variation in the amplitude of the waveform from period to period (shimmer), although many variations exist.

Perturbations are obtained by analysis of prolonged vowel phonation samples and quantify unintentional irregularity in the acoustic waves generated by the larynx. The definition of these parameters indicates that they require some level of periodicity, and recent studies have shown that, in order to calculate these perturbation parameters, the vocal signal must be nearly periodic or else the ability of these parameters to describe the vocal signal greatly degrades (Jiang, Zhang, & McGilligan, 2006; Titze, 1995; Zhang & Jiang, 2005; Zhang, Tao, & Jiang, 2006). Applications of jitter and shimmer analysis have been dissatisfied by the measurement of reliability, sensitivity, and specificity. This has been particularly true in the analysis of hoarse voices with potentially severe aperiodic sound signal structure (Ludlow, Bassich, Connor, et al., 1987; Zyski, Bull, McDonald, & Johns, 1984). When fundamental frequencies and peak amplitudes vary for consecutive phonatory cycles for irregular phonations, jitter and shimmer estimations become unstable and unreliable (Jiang, Zhang, & Ford, 2003).

Professionals interested in voice analysis have several methods for measuring the voice signals. But their most commonly available acoustic analysis measures have not met success in their abilities to consistently and reliably quantify the voice. Validity and reliability of acoustic analysis performed with different tools was shown to be affected by several factors such as microphone type, placement and angle, noise levels, data acquisition system, sampling rate and software used for analysis (Deliyski, Shaw, & Evans, 2005, 2006). To overcome these weaknesses, a researcher needs to look more into the complex signal of 'voice' and its components.

This complex signal is produced by combinations of several sine waves with diverse frequencies and amplitudes. Each sine wave has amplitude and frequency. The amplitudes of each component waves can be added together at any given moment of time, and then the amplitude of the complex wave at that moment of time can be derived. If the amplitudes of the component waves are graphed in terms of function of frequency, the amplitude (or power) spectrum is produced. The graphical representation of the complex wave is amplitude graphically presented as a function of time (time domain) and graphical representation of spectrum is amplitude graphed as a function of frequency (frequency domain).

The manner of signal transformation from the time domain to frequency domain is called Fourier transformation. It is not the absolute amplitude; rather, the amplitude present in the spectral representation it is a logarithm of the amplitude. The greater resolution difference in smaller and larger amplitudes is obtained using logarithm. Fundamental frequency with the largest amplitude of all of the frequencies in the voice spectrum will correspond to pitch. There are also other frequencies that are amplified by the resonators in the vocal tract and that are usually multiples of the fundamental frequency. These also produce characteristic amplitude peaks in the spectrum and are referred to as the harmonic frequencies.

Human voice signal is not periodic in nature, and consists of energy concentrations with low-amplitude, along with the frequency components in the spectrum. In an aperiodic voice signal the amplitude will be equally distributed across the frequencies and there will not be a definable fundamental frequency. Measures such as jitter (frequency perturbation), shimmer (amplitude perturbation), and noise-to-harmonic ratio (NHR) have the ability to accurately identify and track changes in fundamental frequency. But the relevance, validity, and clinical practicality of particular perturbation measures are still unsure, particularly considering moderate or severely disordered voices. This issue is because these perturbation measurements are influenced by the accurate identification of cycle boundaries (i.e. where the cycle of vibration begins and ends in the signal). So it is increasingly noticeable that the existence of significant noise in the signal makes it less periodic and become difficult to accurately track these cycle boundaries. Thus, these measures are possible only in mildly dysphonic voice signal, which is reasonably periodic.

Moreover, the jitter and shimmer measures, which are time-based measures, are suitable in analysis of sustained vowels; these acoustic analysis methods have limitations when considered with analysis for connected speech. As, continuous speech contains rapid onsets and offsets, fundamental frequency variations, amplitude variations, voiced and voiceless phonemes, variations related to prosody, rate of speech, phonetic contexts in which the speech is elicited, vocal pauses and stress makes the measurement practically inaccurate. Meltzner, and Hillman (2010) stated that the jitter and shimmer values in connected speech from time-based analysis, may incorrectly inflate the acoustically predicted ratings of dysphonia severity, because of the presence of voiceless phonemes and variations in prosody of connected speech. The authors also reported that shorter vowel duration in connected speech will negatively affect the ability of time-based measures to perfectly track aperiodicity and also the dysphonia severity. Thus, in order to estimate dysphonia severity in a connected speech sample, the measures other than time-based are needed. Also, there are few extraneous variables that may influence the traditional time-based analysis for example the intensity of the voice signal, the vowel being elicited, and the frequency of phonation. Therefore, in order to get more accurate measures of voice signal, an analysis method that is not a time-dependent measure would be preferable.

A best acoustic analysis method should be able to quantify the voice signal without relying on frequency or its intensity or in any other variables that can affect the accuracy of the measurement; also it should be reliable and reproducible. One such measure is cepstral measure, which is based on peak to peak calculation, not accurate calculation of fundamental frequency.

If a Fourier transformation of a spectrum is executed, a cepstrum is formed. Thus, the spectral representation of the spectrum is called as cepstrum. The term “cepstrum” is created by reversing the first syllable of “spectrum”. Initially, cepstrum was well-defined as the power spectrum of the logarithmic power spectrum, but it was later redefined as the inverse Fourier transform of the log spectrum, as it is reversible from a function of frequency to a function of time by an inverse transform. Here, the resultant Fourier transformation considers the information present in the spectrum of the signal which is in frequency domain and converts this into a time domain cepstrum. Thus, cepstrum is achieved by two Fourier transforms.

In cepstrum, the "frequency" of each wave component of the spectrum is called as "quefrency." Quefrency refers to frequency of the occurrence of the frequency in the power spectrum. The unit of quefrency (1/frequency) is cycles per frequency, measured in seconds. The amplitude components of the waves of the spectrum when marked as a function of quefrency, gives cepstrum. The uniformly spaced quefrency with smaller-amplitude peaks in the cepstrum are called rhamonics which is similar to “harmonics” in spectrum. The peak with the highest amplitude is cepstral peak of cepstrum. The predominant peak in the cepstrum is the fundamental period of the spectrum. The fundamental period is the quefrency of the dominant sine wave of the complex wave termed the spectrum, just similar to fundamental frequency which is the frequency of the dominant sine wave of voice signal. Overall, from the background noise the display of level of spectral harmonics and the vocal fundamental frequency are graphically displayed in cepstrum. The spectral and cepstral representation are given in figure 1.

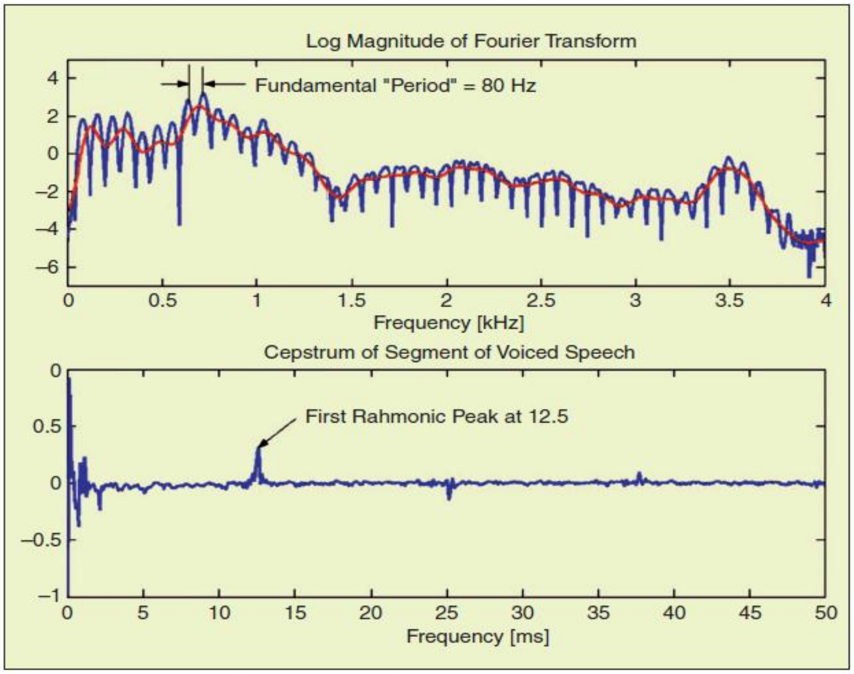


*Figure 1.*: Spectral and Cepstral Graphical representation

A strong peak present in a highly periodic voice signal represents the fundamental frequency and the small peaks indicate the multiples of fundamental frequency. These strong peaks will occur at regular intervals, and in the cepstrum these intervals denote the fundamental period. The figure 2 is the illustration for cepstral peak and rhamonics. Hence, a large-amplitude cepstral peak is recognized at the fundamental period. A multiple similar-amplitude cepstral peaks at many quefrencies will be realized in a signal with indistinct pattern or intervals which is nothing but an aperiodic voice signal. Whereas in a weakly periodic signal will result in a very low-amplitude cepstral peak. When a line called linear regression line that denotes the average sound energy, is drawn through the cepstrum, which is obtained from the voice signal, then the distance measured from the peak of the cepstrum to this linear regression line is termed as CPP. The regression line is nothing but the line is figured between 1 millisecond to the maximum quefrency. This regression line is constructed to normalize the variability in amplitude of phonation that can be from one person to another, plus from one testing situation to another within the same person. Thus, the magnitude of the cepstral peak in relation to the amplitude of phonation can be decided by an experimenter and also objective comparison from one testing situation to another without having to account for differences in loudness of phonation, microphone distance, or recording level can be well decided by adding a linear regression line by the experimenter.

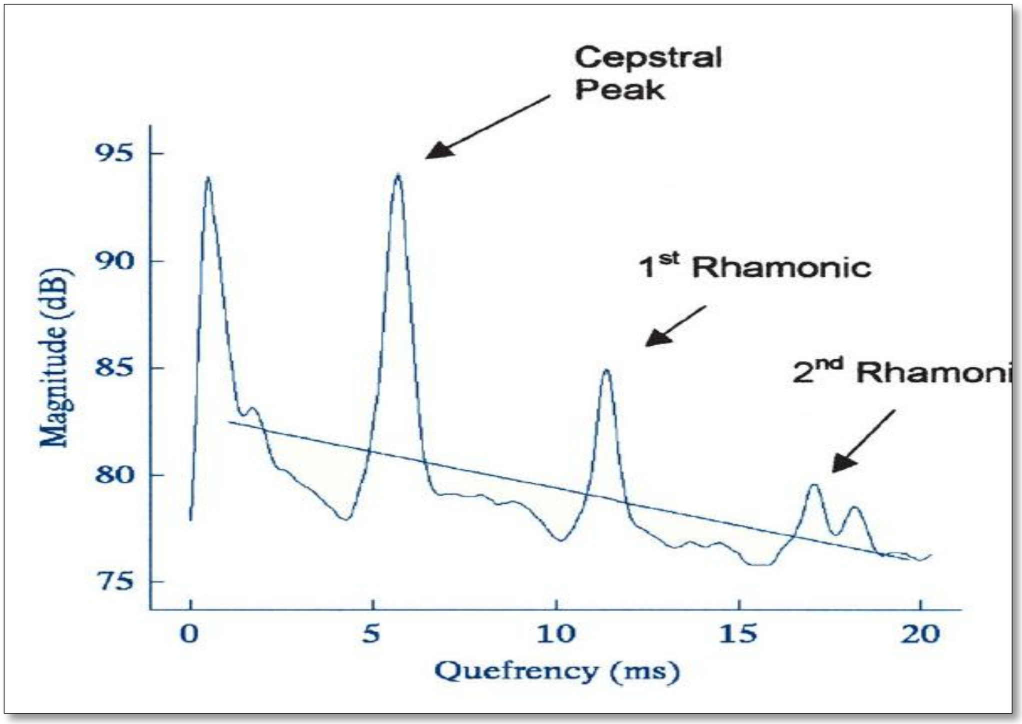
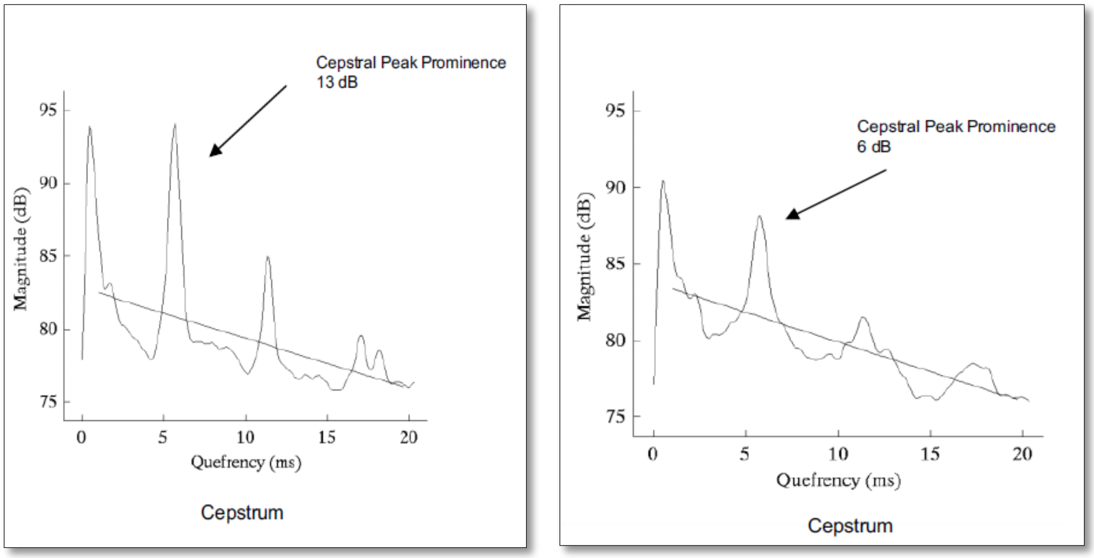


*Figure 2.*: Illustration of Cepstral peak and rhamonics

(Source: Heman Ackah, 2004)

A voice signal which is periodic will have a high-amplitude CPP (in dB), and a weakly periodic or an aperiodic voice signal will have low amplitude CPP. The CPP is an ideal acoustic measure which is be able to quantify independently a voice signal. It has a robust voice analysis algorithm that measures the degree of harmonic structure in a voice signal. Thus this measure is reliable; it correlates well with the dysphonia severity and can be reproducible. The cepstral representation of a normal voice and a moderately dysphonic voice is given in figure 3. The method of analysis ensures that this measure is not influencing any variables such as recording technique, recording volume, or aperiodicity. In machine diagnostics, the main applications of cepstrum are detection of periodic events in the log spectrum (harmonics/sidebands), detection of echoes, and separation of source and transmission path effects.



(A) (B)

*Figure 3:* Cepstral representation of (A) normal voice (B) moderately dysphonic voice

(Source: Heman Ackah et al 2003)

The idea of the cepstrum was first introduced to voice signal by Noll in 1964, but at that period the lack of high-speed computers made calculating the cepstrum cumbersome and time-consuming. Later, Hillenbrand et al (1984) developed an automated method of calculating the cepstrum by incorporating high-speed capabilities of modern computers. In addition, the concept of the linear regression line was added as a means of normalizing the measure for purposes of comparison, and the use of the CPP was introduced. Further advancement in cepstral analysis was the introduction of smoothening feature produces the sCPP. This feature enables the averaging of individual cepstra which are averaged over a given number of frames before and after the frame of interest. This smoothing algorithm is added to average the signal data over a given number of frames in the cepstrum, thus it will reduce the artifacts. In sCPP, the cepstral measurement occurs in every 2 milliseconds instead of every 10 milliseconds and the smoothening is carried out in two steps.

The CPP and the smoothed CPP are two cepstral measures which are found to be the best predictors of dysphonia severity as related to the listener ratings measures. These two measures can be incorporated to evaluate the dysphonia severity in a continuous speech and sustained vowel voice signal, thus measures overall severity. The rationale behind these measures is periodic voice signals display well-defined harmonic configuration in the spectrum and thus obtains a prominent cepstral peak of the selected signal. Therefore, a decrease in overall CPP is an indication of dysphonic voice. Besides, a sensitive and valued tool is essential to evaluate through extensive range of severity, using the CPP and CPPS, which represent spectral noise, is also applicable in the example of strongly dysphonic voices. A strongly distorted dysphonic voice has a flat cepstrum representation and a low CPP value since it consist of unharmonic structure.

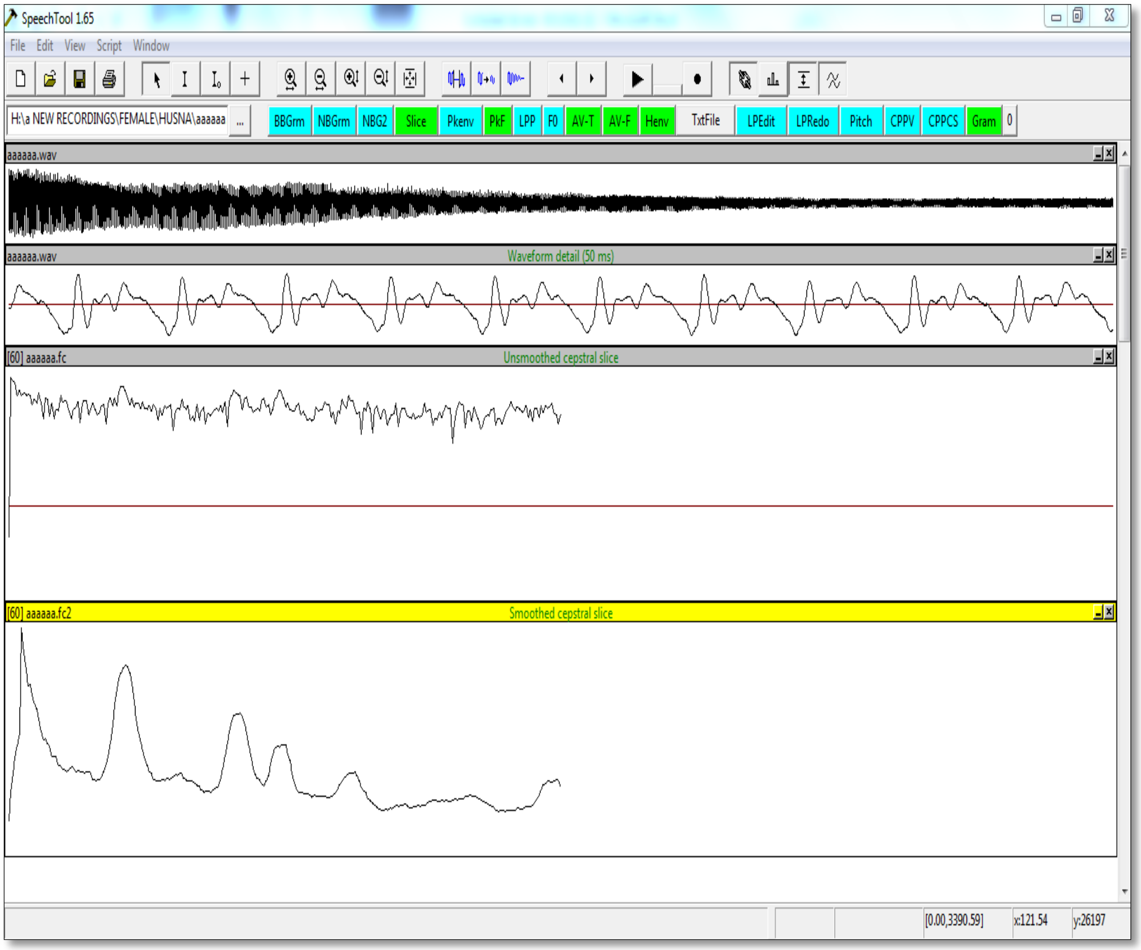


*Figure 4.:* Illustration of a phonation sample /a/ analyzed using the Speech Tool Programme (Version 1.65), from a 23years normal female participant taken as a screenshot. (a) First window indicates the waveform of the signal, (b) the second indicates the unsmoothed cepstral slice and (c) the third indicates the smoothed cepstral slice of the signal.

Cepstral analyses can be conducted in the SpeechTool program (figure 4) which is available in http://homepages.wmich.edu/_hillenbr/ (Hillenbrand, Western Michigan University, Kalamazoo, MI, USA) and also in Computerized Speech Lab (CSL) 4500 (KayPENTAX) system. The difference is CSL program does not separate voiced portion and unvoiced portion of the voice signal and thus consist of both computations and for analysis the selections are completely manual based. Also, analysis can be carried by using an automated voice detection algorithm (Analysis of Dysphonia in Speech and Voice (ADSV); KayPENTAX, Montvale, New Jersey). The ADSV model was established for the spectral - cepstral standardized analysis methods of both sustained vowels and continuous speech.

## Applications of cepstral measures

The principal advantage of Cepstral analysis methods is its facility to estimate aperiodicity or additive noises in a voice signal without identification of cycle boundaries (Awan et al 2012). Thus, by measuring cepstral peak prominence using cepstral analysis, the dysphonia severity can be identified on sustained vowel productions. Also, cepstral method can be used to evaluate dysphonia severity in connected speech. It is an easy acoustic analysis method which can done even using freely downloadable software**.**

Hillenbrand et al., 1994 measured the efficacy of 12 acoustic measures for sustained vowels in predicting breathiness ratings. Recordings were taken from eight men and seven women with no evidence or history of voice disorders. They were asked to produce three modes of phonation which includes nonbreathy, moderately breathy, and very breathy sustained vowels (/a/, /i/, /e/, and /o/). 20 listeners rated the degree of breathiness using a direct magnitude estimation procedure. They estimated the acoustic measures first harmonic (H1) amplitude, signal periodicity, and spectral tilt, using methods like CPP and Pearson ‘r’ at autocorrelation Peak (RRK). The results showed that measures of signal periodicity, based on either the time waveform or its spectrum, provided the most accurate predictions of perceived breathiness, accounting for approximately 80% of the variance in breathiness ratings. The relative amplitude of the first harmonic correlated moderately with breathiness ratings, and two measures of spectral tilt correlated weakly with perceived breathiness. They reported that the cepstral peak prominence (CPP) correlates well with perceptual ratings of breathiness. This result, cue to the amplitude of the first cepstral peak with regard to a linear regression line that is fitted to the cepstrum for normalization. Authors concluded that the relative amplitude of the dominant cepstral peak was among the strongest correlates of the severity of breathy voice quality, and that an inverse relationship existed, whereby increased severity of perceived breathiness was related to decrease relative amplitude of the CPP.

In continuation with the previous study, Hillenbrand et al (1996) extended the cepstral analyses considering connected speech samples along with sustained vowels taken from speakers with laryngeal pathologies. They obtained breathiness ratings using a sustained vowel and a 12-word sentence spoken by 20 pathological and five non-pathological talkers. The voice samples were taken from the Voice Disorders Database recorded at the Massachusetts Eye and Ear Infirmary and distributed by Kay Elemetrics Corporation. The samples from the database were auditioned by a graduate student, who was instructed to select voice samples that represented a range of breathiness percepts from mild to severe, while avoiding samples that were largely or exclusively aphonic.

Acoustic analysis was made for signal periodicity, first harmonic amplitude, and spectral tilt. The results revealed that, for the sustained vowels, a frequency domain measure of periodicity provided the most accurate predictions of perceived breathiness, accounting for 92% of the variance in breathiness ratings. The relative amplitude of the first harmonic and two measures of spectral tilt correlated moderately with breathiness ratings. For the sentences, both signal periodicity and spectral tilt provided accurate predictions of breathiness ratings, accounting for 70% to 85% of the variance.

Heman-Ackah, Michael, Goding (2002) compared the ability of smoothed cepstral peak prominence (sCPP) and other traditional acoustic analysis measures to predict overall dysphonia, breathiness, and roughness in pathologic voice conditions. They considered 38 preoperative and postoperative samples of connected speech and sustained phonation of vowel /a/ which is taken from 9 male and 10 female participants with unilateral recurrent laryngeal nerve paralysis between the age ranges of 25 to 87 years. Postoperative voice samples of seven participants who had a type I thyroplasty were obtained one month after surgery and by eight months after surgery for 12 participants who underwent reinnervation of the recurrent laryngeal nerve. The samples were analyzed perceptually for grade (overall dysphonia), roughness, breathiness, asthenia, and strain by a mark on a 120-mm line from least abnormal (0) to most abnormal (120), time domain analysis to obtain RAP, sPPQ, APQ, and NHR and Cepstral analysis to obtain CPPS for running speech (CPPS-s) and CPPS for sustained vowel /a/ phonation (CPPS-/a/).The results indicates that CPPS-s and CPPS-/a/ were the best predictors of overall dysphonia and breathiness, and the other acoustic measures NHR, APQ, RAP, and sPPQ were the less accurate predictors of overall dysphonia and breathiness. One limitation of this study was that only two listeners had perceptually rated the samples which questions the validity of the perceptual results, based on which the other parameters were compared.

A study done by Awan, Roy, Jette, Meltzner, and Hillman (2010) examined the relationship between Cepstral measures of dysphonia severity and listener rating using Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) in four male and female normal participants between 25 to 32 years of age and 12 male and female participants between 21 to 78 years of age with varying dysphonia severity (mild, moderate, and severe). They have considered sustained vowel and CAPE-V sentence samples recorder and analyzed using KayPentax CSL model 4400. Results indicated a strong correlation and a high degree of agreement with listener perceived severity ratings and acoustically estimated severity ratings across a variety of samples. The strongest correlation was present for the third CAPE-V sentence (the presence of possible voiced stoppages or spasms and the ability to maintain consistent voicing - ‘We were away a year ago’) and to the sustained vowel /ɑ/.

Balasubramanium, Bhat and Prasad (2010) measured CPP in participants with vocal nodules between the age range of 20-40 years which includes 25 males and 25 female. The participants were diagnosed by experienced laryngologist using endoscopy and in any of the participants medical treatment was not started during the period of study. A control group was also considered for the study with age and gender matched participants who had perceptually normal voice. The participants were given task to phonate vowel /a/ where they had to maintain their habitual pitch and loudness. The samples were straightly recorded to CSL 4150 hardware via a dynamic microphone with a constant speaker's mouth to microphone distance of 10cm. Results revealed the clinical group had lower values of CPP when compared to control group, this attributes to the existence of flat harmonic structure in the clinical group in turn supporting the presence of glottic chink. The chink resulted in more noise level which lowered the CPP. More the noise more lowered the value of CPP, which indicates voice as abnormal. This suggests the applications of CPP in day to day clinical training and practice. Also it can track the efficacy of treatment procedure in voice disorders. Study recommends more upcoming researches need to be directed towards obtaining normative values for cepstral measures in the Indian population.

Balasubramanium, Bhat, Fahim and Raju (2011) measured CPP in 30 individuals with Unilateral Adductor Vocal Fold Palsy using CSL (CSL model 4150; Kay Pentax, Lincoln Park, New Jersey) for phonation sample of /a/. The clinical group comprised of 30 males and females in the age range of 20 to 40 years. At the time of recording none of the participants was receiving medical treatments nor was it initiated. The clinical group was compared with age and gender matched control groups who were perceptually evaluated by three trained SLP's. Results showed lower values of CPP in the clinical group and this can be because of the presence of phonatory gap which indicates presence of high background noise resulting in a flat harmonic structure. Also the result reveals significant differences in male and female group of participants, which indicating lower measures of cepstral peak in the female group. This can be because of 80% of females groups have a posterior phonatory gap, inturn resulting softer and less-intense voices. This study lacks instrumentation for quantifying the size of the phonatory gap in female groups, which would have facilitated correlating the size of phonatory gap with the cepstral peak values obtained.

Watts and Awan (2011) performed cepstral measurements on both continuous speech and vowel prolongations. Participants includes 11 female and 5 male hypofunctional speakers in mean age of 52 years and 11 female and 5 male normal speakers in mean age of 53 years, they were asked to sustain /a/ and read the Rainbow Passage which is perceptually evaluated by two speech-language pathology graduate students. The examiners identified the speaker’s voice quality category as normal, breathy, rough, or hoarse and rated the severity on a 100-point visual analog scale that had labels for mild, moderate, and severe. The middle one second steady-state portion of the sustained vowel was isolated for spectral/cepstral analyses. Acoustic measures for continuous speech were centered on the second sentence of the passage. Cepstral analysis provided the acoustic measures CPP and CPP standard deviation. Spectral analyses utilized an algorithm that gives the L/H spectral ratio and L/H spectral ratio SD measures. The L/H spectral energy ratio compares the average energy in the entire speech signal below 4 kHz to the average energy above 4 kHz in a long-term spectral analysis. Among the measures used in this study, CPP and L/H spectral ratio showed significant differences between groups in both speaking conditions. Thus, this study demonstrates CPP and L/H spectral ratio as effective discriminatory measures of normal versus abnormal voice, and provides further evidence of the clinical value of cepstral/spectral-based measures.

Moers et al (2012) reported a retrospective study on Vowel- and Text-based Cepstral Analysis of German men (24) and women (49) with chronic hoarseness between

19 and 85 years of age. The participants were asked to produce vowel /e/ and then to read a German text *(“The North Wind and the Sun”*), which is a standard text consisting 108 words. Also, automatic text based evaluations were done using the first sentence consisting of 27 words only. This study compares text-based automatic evaluation with the German Roughness-Breathiness-Hoarseness (RBH) evaluation scheme. Perceptual evaluation was done by five trained speech therapists and physicians according to the German (RBH) scheme and the “overall quality” of voice was measured on a 4-point scale (1= “very good” to 4 = “very bad”) and also a 10 cm visual analog scale (VAS, 0.0 = “very good” to 10.0 = “very bad”). PRAAT software was used for noise based measures and Speech tool software was used for cepstral analysis. The results indicated cepstral analysis correlates well with the German RBH scheme on chronic hoarse voice participants. This study further recommends using text based recording for voice analysis and also suggests that CPPS only is not a suitable measure to provide a full hoarseness index; rather CPPS with combination of other analysis methods will provide a meaningful objective measure.

A thesis by Garrett (2013) was aimed at providing normative data of cepstral measures. The study included 60 participants with 15 males and females, aged 20-30 years, and fifteen males and fifteen females, aged 40-50 years. They were asked to produce sustained vowels /a/ and /i/ for about 3 sec at a 75 dB (± 2 dB) intensity level and also asked to read out loud four CAPE-V stimuli and the 2nd and 3rd sentence of the Rainbow Passage (Fairbanks, 1960), which was consistent with the stimuli developed by the ADSV program. The samples were analyzed in a computer installed with the Kay-PENTAX Multi-Speech (Model 3700) software running the subprogram ADSV; Model

5109, version 3.4.1). Also, the intensity of the participant's productions was monitored by a RadioShack Sound Level Meter (Catalogue Number 33-2055). The CPP, L/H spectral ratio and the fundamental frequency of CPP (CPP F0) were determined. Results indicated a significant gender effect for both vowels and connected speech segments, where the male participants had significantly better voice quality as measured by CPP and L/H spectral ratio for both the vowels /a/ and /i/. In connected speech, women had higher CPP values, denoting better voice quality in females; while men had higher L/H spectral ratio values, denoting better voice quality in males. The age aspect did not show a significant effect on vowels /a/ and /i/; however, for connected speech, age appeared to have a significant effect on CPP for all five connected speech segments. Specifically, CPP was significantly better for younger speakers compared to older speakers, indicating better voice quality in the younger age group.

Gillespie, Dastolfo, Magid & Gartner-Schmidt (2014) conducted the first study to analyze time and frequency based acoustic analyses, independent of an algorithm, after single known treatments at identical follow-up time points for patients with four carefully selected and mutually exclusive voice disorders and to determine if outcome sensitivity of certain acoustic voice laboratory measures varies with disorder type. Data were collected retrospectively from patients (older than 18 years) records from January 2009 to July 2013 were included and diagnoses were made via a team consisting laryngologist and a SLP. Also, the data were specifically chosen as pre and post intervention measures. The task given for participants was to phonate a sustained /a/ and then to read the sentence ‘‘we were away a year ago’’ from the CAPE-V protocol at their best comfortable pitch and loudness level. Recordings of these tasks were completed using the

ADSV and Multi-Dimensional Voice Profile (MDVP) software from the CSL. The ADSV program is responsible to measure CPP in the sentence (CPP speech) and vowel (CPP vowel) and respective SDs, L/H ratio in the sentence (L/H ratio speech) and L/H ratio vowel and respective SDs. In addition, a multifactorial estimate of dysphonia severity, stated as CSID was also used to calculate measures for the all-voiced sentence. Auditory-perceptual evaluations were made using CAPE-V sentences using the zero to three GRBAS scale. The perceptual analyses of the participants revealed the most prominent characteristic pre intervention was hoarseness, then roughness, followed by breathiness and strain, which were rated as equally severe. After intervention, the hoarseness rating improved the most, followed by equal decreases in roughness and strain. On acoustic measures subjects with lesions, MTD-1, and UVFP demonstrated statistically significant improvement before and after intervention. All groups, except vocal fold atrophy, also showed an improvement in VHI-10 after treatment, indicating that most subject groups experienced a reduction in voice handicap after treatment. All the results are collapsed across all groups, statistically significant changes were observed in CPP speech, CPP vowel, their SDs, CSID, and VHI-10. As hypothesized, no measure revealed significant change for all disorders. When the findings were analyzed by disorder, with the exception of the atrophy group, which did not demonstrate significant change in any parameter, CSID and CPP speech were the most consistent indicators of change in response to treatment. The robust and significant changes in CSID provide an example where a multifactor formula appears to be more sensitive to change than individual measures. In addition, both measures were taken from connected speech, which indicates that ecologically valid measures, such as those taken during speech and not a single phoneme, may be the most appropriate for phonatory analysis. Improvement in CPP speech in the patients with lesions, and UVFP after surgical treatment, indicates greater harmonic energy, a decrease in spectral noise, and corresponds with less severe vocal quality as a result of the interventions. Participants who improved in CSID also improved in VHI-10. The authors concluded that the measure that most reflected change after treatment was CSID, and therefore CSID may be a cornerstone measurement for acoustic analyses for all voice problems. This obtained finding is in agreement with previously done studies on CSID as an outcome measure. In addition, the measures that used connected speech demonstrated greater response to treatment than sustained vowels. Further, L/H ratio in speech or its SD showed large effect sizes in all groups and may be an important outcome measure to track response to surgical and behavioral intervention. The lesion and UVFP groups demonstrated change in more measures compared with the other two disorder groups. Analysis of CPP speech, CPP vowel, L/H ratio SD in speech, and CSID is recommended as part of the outcomes battery for patients with lesions. The authors further opined that assessment of both vowels and connected speech in frequency-based measures and vowels in time-based measures may be appropriate.

Yang, Fan, Tian, Liu, Gan, Chen, and Yin (2014) investigated smoothed CPP of vowel sounds of individuals with Velopharyngeal insufficiency (VPI) after surgery to fix a cleft palate with the control group having phononormic individuals. The control group and clinical group comprised of 20 boys and girls in the age range of six to nine years. The clinical group included participants who were experienced VPI after cleft palate surgery. They all were aged younger than three years when they underwent a two-flap palatoplasty and no fistula present in all the participants. Phonation of vowel /a/ of an each participant in both the groups was recorded using Praat software version 4.1.2 (developed by Boersma and Weenink of the University of Amsterdam, the Netherlands). In the clinical group, recordings were done before surgery, before speech therapy (three to four months after surgery), and after speech therapy immediately. The samples were acoustically analyzed by applying Fourier transformation to the acoustic signal to obtain cepstral values. The results revealed lower CPPs in clinical group before surgery and speech therapy than those in the control group. This may be because of the occurrence of VPI during speech development which results in hoarse voice. The results also revealed no significant difference across the control group and the clinical group and across boys and girls after speech therapy. This study recommends clinicians to assess voice characteristics in individuals with VPI after cleft palate surgery. Thus, cepstral analysis is helpful to track the efficacy of treatment not only in voice disorder populations, but also in other disorder conditions. Further, it can be used as a good complementary tool in the case of laryngeal evaluation. This study used only single vowel phonation that may not be representative of speech or any other sounds with higher intraoral air pressure demands. Another potential threat to the validity of this study is that the VPI group was getting older with the clinical treatment, but the control group was not. This maturation factor might have affected the result of the study to a certain extent. This indicates importance of incorporating more sounds in the future studies and also to form a better design in order to record the control group over time and compare findings between control and clinical groups at each time point.

## Factors affecting cepstral analysis

Lowell, Colton, Kelley and Hahn (2011) conducted a study aimed to investigate the cepstral measures and the Long Term Average Spectrum (LTAS) measures can discriminate a set of dysphonic versus normal speakers using continuous speech sample and to decide the correlation of the same acoustic measures with the perceptual voice rating scales. Also, authors have studied the relations of cepstral and LTAS measures between same speaker's samples by comparison of one sentence to a second sentence and one sentence to a constituent phrase of that sentence. The voice samples included are 27 dysphonic voice samples of speakers between 19–86 years and 27 normal between age range of 26 - 55 years were selected from a database documented by Massachusetts Ear and Eye Institute (MEEI). The recordings were done in a sound-treated room with the placement of speaker's mouth-to-microphone distance 15 cm, by means of a condenser microphone and digital voice recording device. In the first set of analysis they edited pauses and also the unvoiced segments of each sample in order to create a concatenated signal which represented the voiced portions speech samples. In the second set of analysis they included edited parts to check the effects of edited versus unedited samples. The sample rating was rated by three judges who have broad skill in voice analysis and they rated the voice samples using a 100-mm VAS. The three features of voice as given in CAPE-V (roughness, breathiness, and strain) were considered as dimensions for rating. Results indicated both the cepstral and LTAS of edited voice samples strongly separate the dysphonic from normal speaker voices. For the dysphonic group, spectral mean, spectral SD, CPP, and CPPS were lower; although skewness and kurtosis were higher and also both CPP and CPPS were lower for the dysphonic groups relative to the normal groups. This study considered four spectral moments assessments, spectral mean, kurtosis, and skewness which revealed moderate or greater relationships (0.64, 0.71, and 0.67, respectively) with CPP and CPPS (-0.78 and -0.72, respectively) of overall voice severity. The within speaker consistency of obtained measures was checked, by comparing the sentence 1 versus sentence 2 and sentence 2 versus the constituent phrase. The comparisons in each combinations involved changes in both phonetic content and length, with sentences 1 sentence and 2 sentence and comparison of sentence 2 and the constituent phrase. Generally mean values of LTAS and cepstral measures were significantly not the same among the sentences 1 and 2 but not in the case of sentence 2 and the constituent phrase. Likewise, the differences across sentence 1 and 2 tasks present in both groups, but that were found frequent in the normal speakers as well. This is because; normal speakers have greater flexibility in vocal vibratory forms. Likeness concerning the means of sentence 2 and the constituent phrase directs that a fairly short utterance of six words is enough to reveal group spectral pattern variances. Although, there is a difference in means of some tasks, the correlations between all the LTAS measures and cepstral measures for sentences 1 and 2 and also among sentence 2 and the constituent phrase were high. Thus, the length of the connected speech or the different phonemic content affects the pattern of the resultant measures; the utterances with high or low in one context were respectively high or low in another context. The absolute mean values were different in unedited versus edited sentences. In LTAS, values were weaker enough to differentiate between dysphonic and normal speakers when the unvoiced segments where included. Whereas, the Spectral SD measure did not show significant difference between groups when the unvoiced segments were edited. In contrast to that the cepstral measures showed significant differences across the groups when unvoiced segments are retained in the samples. This indicates the speech tool by Hillenbrand and colleagues gives lesser impact on unvoiced speech segments. The future recommendation by this study is to compare the variety of spectral and cepstral measures by automated and non-automated voicing detection methods and to study whether the consistency over the cepstral and LTAS based measures is maintained during post intervention.

Awan, Giovinco and Owens (2012), reported the effects of vowel intensity and type of vowel in cepstral measures. They consider i/, /ɑ/, /u/, and /æ/ vowels from 92 healthy male and female participants who are in the middle of 18 and 30 years of age. The stimulus is elicited at three diverse vocal loudness levels. The first one is at comfortable pitch and loudness level, then the second one is as softly as possible without any whispering and finally as loudly similar to screaming/straining the voice or tensing of the neck region. Hillenbrand's cepstral analysis program was used and the results revealed factors like vocal loudness/intensity and vowel type have a significant effect in measured values. The obtained CPP values are i/- 6.53, /u/- 6.78, /æ/-7.57, /ɑ/- 7.56 respectively. There is a decrease in CPP values for high vowels whereas increase in CPP values for low vowels and this is because of the large separation of first formant frequency and second formant frequency during production of vowels which are high. This results in an overall reduction in the signal amplitude. Whereas, there will be a high first formant frequency and a low second formant frequency in low vowels, this results in a broadband signal resonance that emphasizes the overall energy of low vowels. Also, the high vowels are elicited with a decreased oral cavity opening when compared to low vowels production.  Louder and more intense the voice it tends to produce decreased perturbation and also loud voice which is produced with increased vocal pitch reflects in increased tension of vocal folds. This increase in vocal F0 can alter CPP measures because when vocal F0 is increased it results in a more constant F0 with increased motor-unit firing rates and with decreased jitter components, which results in increased CPP. Also, they have indicated that it is not appropriate to compare pretreatment CPP measures with vowel /i/ versus post treatment CPP measure with vowel /a/ or vise versa. Therefore this study reveals the importance of a normative study in the field of CPP measures which need to be conducted using different vowel types.

Brinca, Batista, Tavares, Goncalves and Moreno (2013) investigated the use of CPP and CPPs to differentiate dysphonic and non-dysphonic voices using phonation and connected speech tasks in 30 participants in the age range of 19-66 years. The auditory perceptual analysis and acoustic analysis was done on phonation of vowel /a/ for one second and two sentences extracted from a Portuguese reading passage. The auditory-perceptual ratings were done using GRBAS scale (G represents the degree of overall voice abnormality, R represents roughness, B represents breathiness, A represents asthenia (weakness), and S represents strain). The GRBAS scale uses a Likert scale of 4-point value from 0 (normal) to 3 (extreme) for all the five parameters and Cepstral measures are obtained using Hillenbrand's speech tool available from http://homepages.wmich.edu/_hillenbr/. Their findings revealed the measured values for CPP in both sustained vowel and connected speech consistently occurred as a significant factor in predicting dysphonia. That is, the cepstral peak prominence is significantly lower for both vowels and connected speech in dysphonic group than the non-dysphonic group. The best correlation coefficients ($0.6 < r < 0.7$) were obtained between CPP and grade B of the GRBAS scale. There are few limitations present in this study, one of that is this study only assessed female voices. Future studies comparing females versus males may provide further information about the importance of these acoustic measures. Another limitation of the study is related to the lack of homogeneity of voice disorders in the clinical group. In future studies, to minimize inter-speaker variability, the dysphonic group should be as homogenous as possible.

The above studies provided insight into different applications of cepstral measures**.** However, considering the ethnic and racial variations, these results could not be directly applied to Indian population. In the view of clinical and research application of cepstral measures the present study is taken up with the aim of developing cepstral values for phononormic adults in the age range of 20-40 years.

# CHAPTER III

# METHOD

**Participants:** The current study aimed at establishing reference data of CPP and sCPP for adults in the age range of 20-40 years. To accomplish this, 100 phononormic individuals in the age range of 20-40 years were considered for the study. The participants were subdivided into two groups with age interval of ten years as 20-30 years and 30-40 years (with the upper limit excluded from the class interval selected). Each group consists of 50 individuals with equal number of males and females. Table 1 indicates the details of average age and distribution of the participants across the groups.

Table 1

*Distribution of participants by age and gender*

| Groups | Males | Average age | Females | Average age | Total |
|---|---|---|---|---|---|
| Group I<br>(20-30 years) | 25 | 24.13 | 25 | 23.94 | 50 |
| Group II<br>(30-40 years) | 25 | 30.01 | 25 | 33.62 | 50 |

The participants were randomly selected from in and around the Mysore city. The participants for the current study were accessed through announcements in undergraduate and graduate level classes, personal contacts, and phone and email correspondence with local community churches, businesses, agencies and schools. The selection of an individual for the study was based on the following predetermined inclusion criteria.

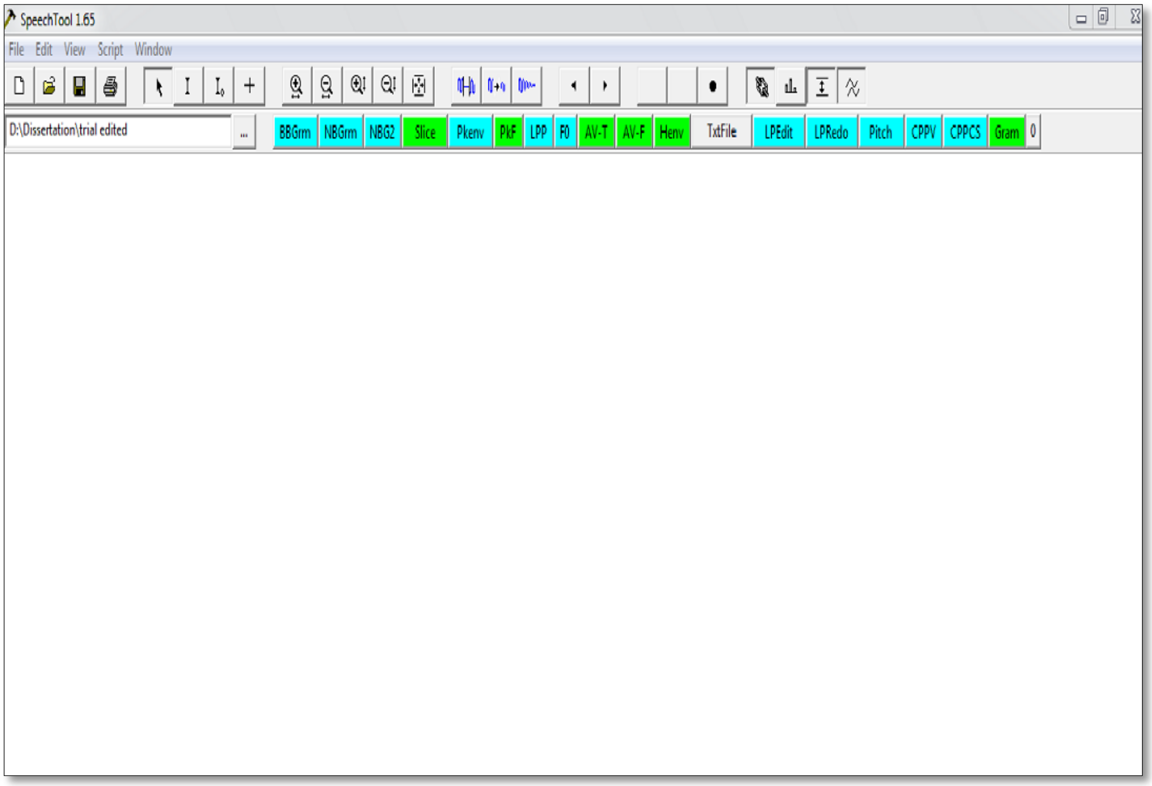
*Inclusion criteria:* Individuals with no complaints of voice problems and identified to have perceptually normal voice on the day of the recording were included in the study. It was ensured that the participants were free from sensory problems such as hearing loss, motor problems. It was also ensured that none of the participants were actively involved in vocal loading within a day prior to the recording and none of them were current smokers. Further, to be included in the study it was ensured that the participants were free from upper respiratory tract infections or allergies on the day of testing.

**Procedure:** The participants were informed regarding the purpose and procedures involved in the study and an informed consent was obtained from them. All the recordings were performed in a quiet room in a solo sitting for all the participants included in the study. The participants were made to sit on a comfortable chair with their back straight and instructed to phonate vowels /a/, /i/ and /u/ for five seconds each. Following this the participants were instructed to read the first sentences of '300 word Kannada reading passage' (Savithri & Jayaram, 2007) and the Bengaluru Kannada passage at their habitual pitch and loudness. Sentences from two different passages were considered so as to investigate the variations in cepstral measures across stimuli as in '300 word Kannada reading passage' which consist of all phonemes in Kannada language and as in Bengaluru passage with specifically voiced phonemes. Instructions were provided to participants through verbal mode in their native language.

All the phonation and reading samples were recorded using HP headphone with microphone. This headphone has microphone frequency response of 100-18000Hz and Headphone frequency response of 20-20Khz with impedance of 32 Ohm and maximum power input of 50mW, also with 108 dB sensitivity. The recordings were obtained by maintaining a constant mouth to microphone distance of 5 cm. The samples were recorded directly to Speech Tool (z-tool) software using the above specified microphone and the samples were saved for further analysis. The participants were asked to repeat the task, if participant feels his/her loudness is altered compared to the daily conversation. To check the test retest reliability, recordings were repeated for 10% of participants following a week of initial recording.

The CPP and sCPP can be measured using the instruments Computerized Speech Lab (CSL model 4150; Kay Pentax, Lincoln Park, New Jersey) as well as the Hillenbrand's Speech Tool (Version 1.65) (Hillenbrand et al (1994). Figure 5 is a screen shot of Speech Tool window (Version 1.65). However, Speech Tool is superior to CSL for measuring CPP and sCPP as this directly provides the mean CPP and mean sCPP measure, making the task easy and less time consuming. Further, Speech Tool is available in the http://homepages.wmich.edu/~hillenbr/, which is freely downloadable hence facilitates easy availability and wider utility. Therefore, in the present study Speech Tool version 1.65 by Hillenbrand et al (1994) was used for analyzing the recorded samples. The Speech Tool was used in the current study to analyze CPP and sCPP using vowels /a/, /i/, /u/ as well as the two sentence stimuli.

**Data Analysis Procedure:** The speech tool software does not allow editing of the recorded samples. Therefore the samples recorded in speech tool were opened in the Praat software (Version 5.3.53) given by Boersma and Weenink (2001) to select the stable, middle three second segment of phonation of vowels and to select the target portion of both the sentences. Thus edited samples were further saved; reopened in z-tool for cepstral analysis to obtain CPP and sCPP. Thus two dependent variables (CPP and sCPP) were obtained in five different contexts (three vowels and two sentences) for each participant.



D:\Dissertation\trial edited

*Figure 5:* Illustration of speech tool (ztool) (Version 1.65)

**Statistical analysis:** The obtained CPP and sCPP values for both phonation as well as speech samples were documented in excel sheets. This data was further subjected to statistical analysis  using Statistical Package for Social Sciences software (SPSS) version 17.0 to obtain the descriptive statistic measures mean, standard deviation, and confidence intervals for each dependent variable (CPP and sCPP) for each level of independent variables and for each of the three vowel and two sentences. Further descriptive statistics was done to obtain the mean and standard deviation for CPP and sCPP measures in phonation and reading tasks and to find the gender, age and stimulus effect in CPP and sCPP, the parametric tests such as one way MANOVA, two way MANOVA and mixed ANOVA and non parametric tests like Mann-Whitney Test and Wilcoxon Signed Rank tests were done on the data using Statistical Package for the Social Sciences (SPSS) 17.0. Independent variables of the study were gender (2 levels: male and female) and age (2 levels: 20-30 years and 30-40 years) and stimulus (2 levels: Bengaluru Kannada passage and Kannada Reading Passage).

# CHAPTER IV

# RESULTS AND DISCUSSION

Cepstral analysis of voice computes fundamental frequency and harmonic amplitude without being dependent on cycle boundary identification and has been frequently reported to be a robust measure for objective analysis of quality of voice. Several studies indicated that cepstral based measures such as cepstral peak prominence and smoothened cepstral peak prominence as reliable parameters for differentiating normal from dysphonic voice, discriminating across the levels of severity of dysphonia, and a valid outcome measure in efficacy studies. Considering the research and clinical applications of these measures, it is essential to develop reference data so as to differentiate normal from dysphonic voices and to verify whether post-intervention findings have reached the normalcy. In this context the current study is a preliminary attempt to establish the reference data of CPP and sCPP measures for individuals in the age range of 20-40 years.

To achieve this, the phonation and speech recording samples were obtained from 100 phonormic individuals including 50 males and 50 females in the age range of 20 to 40 years (20-30 years and 30-40 years age intervals with equal number of males and females). The obtained samples were analyzed using cepstral analysis to obtain the CPP and sCPP values. Further, the data was subjected to statistical analysis to find the reference values and also to verify whether there is a significant effect of age, gender or stimulus on the obtained cepstral measures. Results of the study are presented and discussed under the following headings.

1. Test retest reliability

2. Reference data for CPP and sCPP across the stimuli.
3. Effect of age on CPP and sCPP.
4. Effect of gender on CPP and sCPP.
5. Effect of stimulus on CPP and sCPP (voiced versus balanced sentence).

During the statistical analysis, data from six participants (two from 20 to 30 year males, two from 20 to 30 years females, two from 30 to 40 years males) was removed as outliers based on box plots. The data from the remaining 94 participants was considered for further statistical analysis. In order to verify the normal distribution of the data, the test of normality was performed using Shapiro Wilk test.

### Test retest reliability

In order to check the test retest reliability, recording was repeated on 10% (five males and five females) participants following a week of initial recording. The test retest reliability was performed using Cronbach's alpha test. The results indicated a good reliability across the parameters (CPPa - 0.90, CPPi-0.90, CPPu-0.94, CPPs1-0.91, CPPs2-0.64, sCPPa-0.82, sCPPi-0.88, sCPPu- 0.82, sCPPs1-0.88, sCPPs-2 0.71). The overall reliability was 0.85 which is suggestive of high test-retest reliability. The reliability of speech tool has also been reported in earlier study by Leong, Hawkshaw, Dentchev, Gupta, Hurie and Sataloff (2013) and they reported the CPP measure is moderately reliable for females and sCPP is reliable for males.

### Cepstral peak prominence and smoothened Cepstral peak prominence across the stimuli.

The CPP and sCPP were documented across the vowels /a/, /i/, /u/, voiced sentence from first line of Rainbow passage (s1) and a sentence with voiced and voiceless consonants in equal ratio (balanced sentence) (s2) from first line of 300 word Kannada reading passage. The mean and standard deviation values of CPP and sCPP for vowels /a/, /i/, /u/ and sentences s1, s2 are given in table 2 and figure 6. The reference data for cepstral based measures have not been previously established in Indian population. However, when compared similar studies reported in Western population (Awan et al, 2011; Watts et al, 2011; and Garret et al, 2013), the values obtained in the current study are comparatively higher. Awan et al (2011) reported for the vowel /a/ in females averaging a CPP of 10.74 dB, and males averaging a CPP of 13.03 dB. In addition to these studies descriptive statistics of control groups were provided in a few previous studies concerning cepstral measures for dysphonic speakers. Watts et al (2011) reported CPP measure in an average of 11.08 dB for the vowel /a/ in control group (including both males and females) of their study. The only other available form of normative data for these measurements was by Garrett (2013), were the CPP for females and males averaged across age as 10.929 dB and 12.544 dB respectively. Also, considering the sCPP measures Awan et al (2012) reported CPP values in vowels as /a/ 7.56, /i/ 6.53, /u/ 6.78. All these reference values are lower compared to the current study (as given in table 2). These variations in results can be due to physiological and topographical variations in the participants in earlier and current study and moreover as reported by Awan et al (2012), CPP has a significant effect on loudness and this loudness factor was not controlled instrumentally in the current study.

Table 2

*Mean and Standard deviation of CPP and sCPP across age and gender*

| Cepstral Parameters | 20-30 years<br>Male (N = 23)<br>Mean (±SD) | <br>Female (N = 23)<br>Mean (±SD) | 30-40 years<br>Male (N = 23)<br>Mean (±SD) | <br>Female (N = 25)<br>Mean (±SD) |
|---|---|---|---|---|
| CPPa | 18.49 (3.05) | 18.35 (2.18) | 18.27 (1.96) | 18.37 (2.32) |
| CPPi | 17.18 (3.04) | 17.62 (2.07) | 17.42 (2.42) | 17.78 (1.96) |
| CPPu | 17.06 (2.85) | 14.65 (1.49) | 15.74 (2.09) | 15.70 (1.81) |
| CPPs1 | 15.25 (1.03) | 15.24 (0.96) | 14.32 (0.99) | 15.76 (1.15) |
| CPPs2 | 14.73 (0.99) | 14.79 (0.93) | 13.68 (0.62) | 14.79 (1.04) |
| sCPPa | 8.64 (1.05) | 7.37 (0.93) | 7.73 (1.09) | 7.63 (1.27) |
| sCPPi | 6.24 (1.66) | 5.64 (0.84) | 6.28 (1.36) | 5.55 (0.80) |
| sCPPu | 6.95 (1.97) | 5.49 (1.29) | 6.02 (1.55) | 6.35 (1.57) |
| sCPPs1 | 5.65 (0.79) | 5.70 (0.85) | 4.93 (0.73) | 6.07 (1.12) |
| sCPPs2 | 5.22 (0.60) | 5.25 (0.61) | 4.49 (0.53) | 5.37 (0.98) |

With respect to the vowel, it is evident from the table 2 that CPP for /a/ has the highest value (/a/ > /i/ > /u/) with when compared to /i/ and /u/. The similar trend was observed in sCPP, where vowel /a/ is higher than /i/ and /u/. Similar findings were reported by (Awan et al., 2012) who reported higher CPP for vowel /a/. The authors attributed it to the presence of open oral cavity in low vowel production resulted in increase of overall intensity along with a low frequency emphasis which may contribute to the increase in CPP observed for low vowels. However, there is a different trend observed in younger male and older female group where /u/ > /i/.

The overall CPP and sCPP values of sentences are on a lower side compared to the vowels. Interestingly this trend is seen more prominent when the overall CPP and sCPP scores are considered across the age groups as depicted in figure 6. Females obtained consistently higher values in sentences for CPP and sCPP measures, whereas this trend is absent in case of vowels. The robust value of CPP for vowels compared to sentences was also reported in the earlier studies (Watts & Awan, 2011; Brinca et al, 2013; Reddy, 2014). The relatively higher CPP values among vowels can be due to the stability in phonation and relatively lesser variations compared to the sentence reading which has variations similar as in continuous speech. The authors (Watts & Awan, 2011; Brinca et al, 2013; Reddy, 2014) attributed the variations in sentences to the factors such as transition from vowel to consonant or consonant to vowel, changes in vowel spectrum and due to the intonation patterns leading to lower cepstral measures.

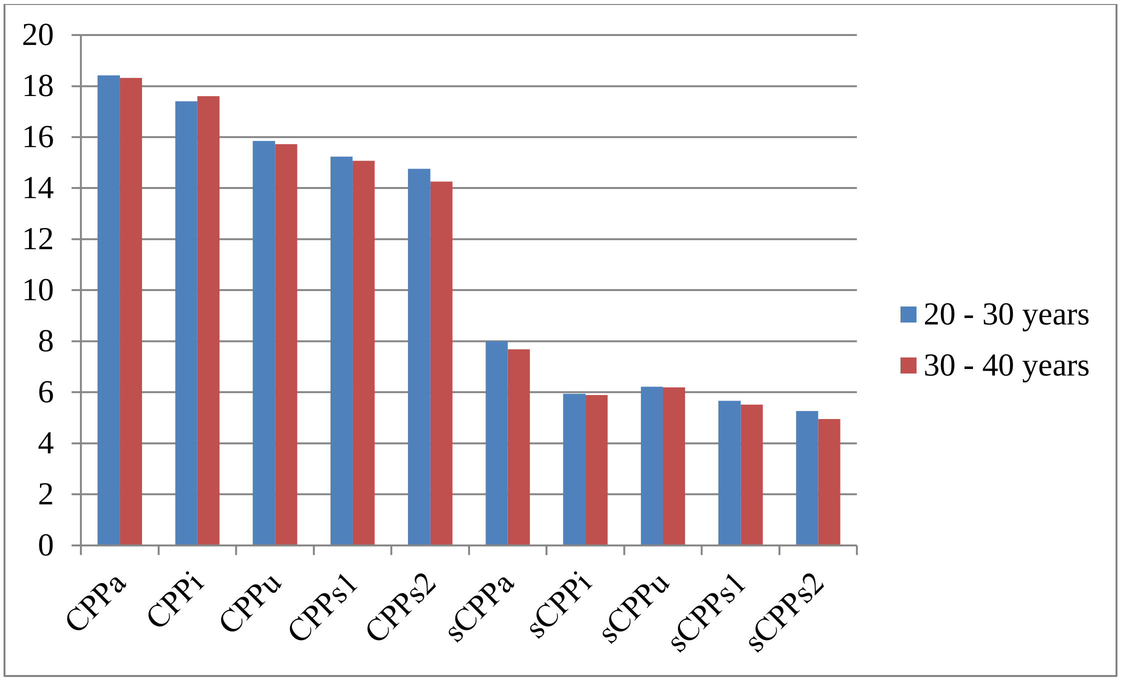


*Figure 6*: Overall CPP and sCPP measures in 20 to 30 years and 30 to 40 year participants.

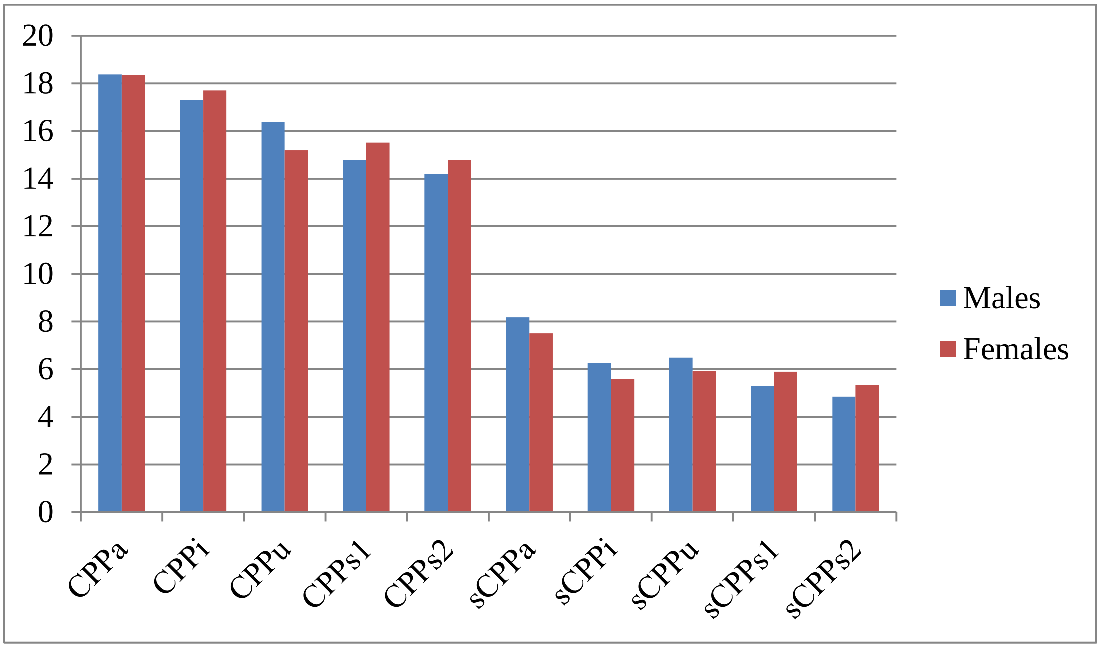


*Figure 7*: Overall CPP and sCPP measures in males and females.

Sentences from two different passages were considered so as to investigate the effect of stimulus, specifically the presence of complete voiced phonemes in stimulus as in Bengaluru passage (represented as S1) and with the first sentence taken from ‘300 word Kannada reading passage’ which does not specifically consist of voiced phoneme (represented as S2). Results indicated higher CPP and sCPP values for the Bengaluru passage with more voiced phonemes. The consistently lower CPP values with S2 in both smoothened and unsmoothed analysis can be due by the fact that this sentence consists of a significantly lower number of unvoiced phonemes compared to the voiced Bengaluru sentence, indicating that the speech tool software is more sensitive for voiced components. However, this difference is irrational as the speech tool is assumed to be considering only voiced phonemes for analysis while denying the unvoiced phonemes in the given stimulus as reported by Lowell et al (2011). Thus, this aspect regarding the precision of z-tool for selective analysis of voiced phonemes requires further investigations. It was noted during the study that the participants read the text with different levels of fluency, and this difference was related to the heterogeneity in academic and social backgrounds. A region-dependent accent was also detected in some of the study participants. Considering the following statement made by Ladefoged and Disner (2012) ‘‘even without considering differences of accent, the range of human voices is enormous,’’ with this in mind, it will be important in the future to use a more homogenous sample of speakers that share similar linguistic and academic backgrounds.

### Effect of gender on CPP and sCPP

To study the effect of gender on CPPa, CPPi and CPPu two way MANOVA was carried out. The results revealed a significant main effect of main effect of gender on CPPu [F(1, 90) = 7.803, $p<0.05$] and not in CPPa and CPPi. The CPPu scores were higher in male participants compared to the females. To evaluate the interaction effect of age on CPPu one way MANOVA was performed. The results revealed significantly higher CPPu [F(1, 44) = 12.839, $p<.05$] for males in the age range of 20 to 30 years, whereas there was no significant difference between males and females in the age range of 30 to 40 years.

Two way MANOVA was performed to check the gender effect on sCPPa, sCPPi, and sCPPs2. There was a significant difference between males and females in sCPPa [F(1, 90) = 8.997, $p<0.05$], sCPPi [F(1, 90) = 7.002, $p<0.05$], and sCPPs2 [F(1, 90) = 10.588, $p<0.05$]. The sCPPa and sCPPi has significantly higher scores in males and sCPPs2 was significantly higher in females. Further, to evaluate interaction effect of age on sCPPa, sCPPi and sCPPs2, one way MANOVA was performed. The results indicates higher sCPPa [F(1, 44) = 18.758, $p<0.01$] scores for males in the age range of 20 to 30 years and higher sCPPi (F(1, 46) = 5.229, $p<0.05$) scores for males in the age range of 30 to 40 years. Also, females in the age range of 30 to 40 years indicated significantly higher scores for sCPPs2 [F(1, 46) = 14.536, $p<0.01$] than males.

The effect of gender in both sentences on unsmoothed cepstral measures was checked using mixed ANOVA. This statistical analysis was performed to study the within subject (s1 & s2), between subjects effects and interaction (age & gender) effects.

Results of within subject effect in mixed ANOVA reveals significant differences [F(1, 90) = 50.911, $p<0.05$] in the measures of CPPs1s2 for both the sentences. No significant differences in the measures of CPPs1s2 across males, females (CPPs1s2*gender) and age (CPPs1s2*gender*age). The between subject effect in mixed ANOVA reveals significant differences in gender [F(1, 90) = 12.857, $p<0.05$], and there was a significant [F(1, 90) = 11.789, $p<0.05$] interaction between gender*age in the measures of CPPs1s2. Further one way MANOVA was done to check the interaction effect. Results indicated no significant effect of gender in CPPs1 and CPPs2 in the 20 to 30 years age group, whereas there is significant effect of gender in CPPs1 [F(1, 46) = 21.197, $p < .01$] and CPPs2 [F(1, 46) = 19.525, $p < .01$] in 30 to 40 years, where the females have higher vales than males.

The data based on cepstral measures sCPPu and sCPPs1 are not following to the normal distribution. Hence, Mann Whitney U test (non parametric test) was performed to evaluate the significant differences in sCPPu and sCPPs1 across the gender. The table 3 depicts significant differences across the gender for sCPPu in 20 to 30 years and sCPPs1 in 30 to 40 years age group respectively.

Table 3

*Results of Mann Whitney test for comparison of gender in sCPPu and sCPPs1*

| Cepstral Parameters | 20-30 years | | 30-40 years | |
|---|---|---|---|---|
| | Z | Asymp. Sig. (2-tailed) | Z | Asymp. Sig. (2-tailed) |
| sCPPu | -2.53 | 0.01 | -0.73 | 0.46 |
| sCPPs1 | -0.18 | 0.85 | -3.49 | 0.00 |

Overall, from the above results a major gender effect in males for vowels are realized and similarly a major gender effect in females for sentences. The decrement of measures in females can be attributed to the factor that females usually have softer habitual voice, hence reducing the CPP. Also it could be because of the posterior phonatory gap which increases the noise component in female voice compared to male voice. Balasubramanium et al (2011) supports this finding, they reported the lower CPP values in females and attributed it to the posterior phonatory gap which accounts for softer and less intense voice in females. Studies reported poor voice quality in young females related with habitual use of vocal fry phonation (Gottliebson, Lee, Weinrich, & Sanders, 2007; Wolk, Abdelli-Beruh, & Slavin, 2012). The significant effect in CPP measures at sentences level could be because the female connected speech has more suprasegmental variations than the male speech.

## Effect of age on CPP and sCPP

To study the effect of age on CPPa, CPPi and CPPu two way MANOVA was carried out. The results in the table 4 revealed no significant difference across the age groups in CPPa , CPPi, CPPu. Similarly, two way MANOVA was done again to check the age effect in sCPPa, sCPPi, and sCPPs2. There was a significant difference across age groups in the measures of sCPPs2 [F(1, 90) = 4.854, *p*<.05]. The sCPPs2 has significantly higher scores in the age range of 20 to 30 years than 30 to 40 years; whereas, there was no significant differences in sCPPa and sCPPi across the age group.

Table 4

*Results of Two may MANOVA age effect*

| Type of effect | Cepstral Parameters | *P* |
|---|---|---|
| Age | CPPa | .840 |
| | CPPi | .692 |
| | CPPu | .762 |

The effect of age on unsmoothed cepstral measures (CPPs1 and CPPs2) in both the sentences was evaluated using mixed ANOVA. The mixed ANOVA was performed to study the main effects and interaction of within subject (s1 & s2) and between subjects (age & gender). Results of within subject effect in mixed ANOVA revealed significant differences for measures of CPPs1s2 [F(1, 90) = 50.911, $p<0.05$] and there was no significant interaction effect of CPPs1s2*age and CPPs1s2*gender*age. However, the between subject effect of mixed ANOVA indicates significant main difference of age [F(1, 90) = 3.988]. However, there is no age interaction present.

The cepstral measures sCPPu and sCPPs1 are not following to the normal distribution. Hence, Mann Whitney U test (non parametric test) was done to evaluate the effect of age. The result indicates there is no age effect on sCPPu and sCPPs1.

Overall, the above results are indicating no age effect on the cepstral measures. Though, there is an effect of age in cepstral measures using sentences this can be attributed the factors such as participants in the study includes from various social life style which in turn vary how each individual project their voice and also the ease at which they read the sentences, the education level can also vary the results.. Though no supporting studies are available for this specific age groups (20 to 30 and 30 to 40 age groups), a study by Garret (2013) in 20 to 30 and 40 to 50 years age groups reported there is no significant age effect on the vowels /a/ and /i/ in CPP measures; however, for connected speech, age appeared to have a significant effect. Hence, a narrow age group in the current could be another reason for absence of age effect on cepstral measures.

### Effect of stimulus

In the present study two different sentences were considered to check the influence of the stimulus on speech tool. Sentences were taken from two different passages (Bengaluru passage and 300 word Kannada reading passage). S1 comprised the initial sentence of Bengaluru passage which was completely voiced phonemes and S2 comprised of the initial sentence of 300 word Kannada reading passage which does not specifically consist of voiced phoneme.

To find the effect of these two different sentences (S1 and S2) in smoothened CPP, Wilcoxon signed rank test was done in which pair wise comparison of two the sentences was done across age and gender. The results of the analysis indicated that the scores obtained in the age range of 20-30 years was significant for both the genders as well as the scores obtained in the age range of 30-40 years was also significant for both the genders.

Table 5

*Results of Wilcoxon Signed Ranks Test for pair wise comparison of sCPPs2 - sCPPs1*

| Cepstral Parameters | 20-30 years | | | | 30-40 years | | | |
|---|---|---|---|---|---|---|---|---|
| | Male | | Female | | Male | | Female | |
| | Z | Asym. Sig (2-tailed) | Z | Asymp. Sig (2-tailed) | Z | Asymp. Sig (2-tailed) | Z | Asymp .Sig (2-tailed) |
| sCPPs2 - sCPPs1 | 2.722 | .006 | 2.373 | .018 | 2.281 | .023 | 4.103 | .000 |

The results of the present study revealed that the speech tool software was able to significantly differentiate CPP and sCPP values obtained using S1 (sentence with more voiced phonemes) and S2 (sentence with less voiced phonemes) stimuli across gender in each age range. This could be attributed to the following factors. The significantly lower CPP values with S2 in both smoothened and unsmoothed analysis and across the age and gender can be due by the fact that this sentence consists of a significantly lower number of unvoiced phonemes compared to the voiced Bengaluru sentence. This finding is in consensus with the study conducted by Lowell et al (2011) who also found a significant effect of stimulus with voiced segments and unvoiced segments on Cepstral measures. These findings indicate that the speech tool software is more sensitive for voiced components. Nevertheless, the usage of the connected speech as a stimulus for analysis of quality of voice makes the analysis more naturalistic in nature thereby increasing the usage of this software tool for various clinical and research purposes.

# CHAPTER V

# SUMMARY AND CONCLUSIONS

Cepstral analysis is a measure of acoustic analysis to quantify the fundamental frequency and harmonic organization in voice. Cepstrum is defined as a discrete Fourier transform of the logarithm power spectrum (Hillenbrand et al 1994). The Cepstral Peak Prominence (CPP) and smoothened Cepstral peak prominence (sCPP) are the two commonly studied cepstral measures. The CPP is determined by measuring the amplitude difference from the highest peak of the cepstrum to the corresponding regression line which is drawn directly below to the cepstral peaks and when a smoothening factor is applied to that, sCPP is obtained. Cepstral analysis has been reported to be a reliable and valid measure for voice evaluation even for signals with high level of aperiodicity, which the time based measures fails to evaluate. Also, it permits use of speech as a stimulus, thus making the analysis more naturalistic which is not applicable to time base measures. Moreover, the in many studies NHR, jitter, and shimmer measures which are the traditional measures, are reported to be the unreliable predictors of dysphonia (Kumar et al 2010; Heman-Ackah et al 2003). Further, studies have reported the successful use of CPP as a parameter to differentiate normal and dysphonic voices and as a parameter to track therapy/surgical outcomes.

The vast applications and the ease involved in measuring makes CPP an essential part of the routine voice evaluation. However, as CPP is reported to be influenced by the factors such as age and gender of the participant, stimuli and algorithm used for analysis etc, it is essential to have age and gender specific norms to a particular geographic region to serve as reference. Therefore, considering the relevance of cepstral measures in evaluating quality of voice and the dearth of normative data for CPP and sCPP, the present study aims to investigate the CPP and sCPP in phononormic adults in the age range of 20-40 years. The objectives of the study are 1. To establish reference values for CPP and sCPP in phonation and connected speech in phononormic adults between 20 to 40 years of age. 2. To investigate the effect of age on CPP and sCPP. 3. To investigate the effect of gender on CPP and sCPP. 4. To investigate the effect of stimulus on CPP and sCPP.

The participants included in this study are hundred adults in the age range of 20-40 years. They are subdivided into two groups with age interval of ten years (20-30 years and 30-40 years with the upper limit excluded from the class interval selected). Each group will include 50 individuals with equal number of males and females. All participants fulfilling inclusion criteria are randomly selected. The participants for the current study were recruited through various methods. Recordings were performed in a quiet room in a single sitting for all the included participants. The participants were made to sit on a comfortable chair with their back straight and instructed to phonate vowels /a/, /i/ and /u/ for five seconds each and further to read '300 word Kannada reading passage' (Savithri, Jayaram 2007) and the Bengaluru passage at their habitual pitch and loudness. Sentences from two different passages were considered so as to investigate the variations in cepstral measures across stimulus. All the phonation and reading samples were recorded using a headphone to the speech tool software (version 1.65) by Hillenbrand et al (1994) and also by maintaining a constant mouth to microphone distance of 5 cm. The recordings were repeated if participant feels his/her loudness is altered compared to the daily conversation. To check the test retest reliability, recording were repeated for 10% of participants following a week of initial recording. The recorded samples were edited using Praat software (Version 5.3.53) given by Boersma et al (2012) to select the stable, middle three second segment of phonation of vowels and to select the target portion of both the sentences. Further, the CPP and sCPP measures were obtained using Speech tool for each vowel and sentence.

The collected data was subjected to descriptive statistics, parametric tests and non-parametric tests appropriately to analyze the data. The test retest reliability results indicated good reliability of CPP and sCPP measures. Based on the descriptive statistics of cepstral measures, the results indicated that the overall sCPP values are lower than CPP values. The lower sCPP values could be due to the effect smoothening, which averages and reduces the artifacts in cepstral peaks. Similar results were also reported by Brinca et al (2013). The normative values developed for CPP using z-tool software in the current study are higher such as CPPa is 18.49 for males and 18.32 in 20 to 30 years and 30 to 40 years respectively when compared to the previous studies 10.929 and 12.544 dB respectively (Garrett, 2013). This could be due to variations in the use of software for analysis and factors such as lack of control over the participants intensity during the recording of voice sample.

With respect to the cepstral measures of vowels, it is evident that CPP for /a/ was highest followed by /i/ and /u/. The same trend was observed in smoothened CPP. When compared to sentences, vowels had consistently higher cepstral peak measures across all the three vowels. Another possible explanation for relatively higher sCPP values among vowels can be due to the stability in phonation compared to the sentences which are similar to continuous speech with suprasegmental variations. When the cepstral measures from two sentences were compared, CPP and sCPP values are higher for s1 with more voiced phonemes than s2 sentence, indicating that the z-tool software is more sensitive for voiced components.

With respect to gender differences, the CPP values in vowel /u/ were higher in males than females in the age range of 20 to 30 years. The sCPP measures also indicated higher values for males across the age groups for vowel /a/ and /i/, whereas only sCPPs2 was higher in females in the age range of 30 to 40 years. Cepstral measure variations in terms of age groups, there was no significant differences across the parameters except that sCPPs2 is higher in 20-30 year age group. This indicates sCPP in sentences had significant variations with respect to gender across the age groups.

The present study provides the reference values for CPP and sCPP across various stimuli. The results of the study indicate cepstral measures are affected by factors such as stimulus used for the study in terms of vowels, sentences, completely voiced versus general sentences, age and gender of the participant.

**Limitations of the study and future directions:** The previous research results indicate that vocal loudness/ intensity and vowel type have a significant effect on measures of the CPP. Monitoring loudness instrumentally is not considered for the study. This can be a factor that could have resulted in higher cepstral measures and would have contributed to stimulus related variations in the current study. Therefore future studies could consider and control this factor. The stimulus used for sentence 2 is complex for individuals with low educational qualification which may affect their ease of reading which may inturn affect the results. Therefore future studies could consider repletion of sentences as one of the options. However, the influence of such presentation on Cepstral measures needs to be evaluated. Further, the two age groups considered in the present study were in a narrow range, thus revealing minimal effects on the cepstral measures. However, future studies are warranted with wider range of ages so as to verify the effect of the pediatric, adult and aging voice on cepstral measures and hence developing age specific reference data if necessitates.

**Implications:** The results of the present study would serve reference values for Cepstral peak prominence (CPP) and smoothened cepstral peak prominence to phononormic individuals in the age range of 20-40 years. These reference values can be utilized to verify whether post-therapeutic outcomes are approaching the normal voice characteristics. Further, the present study also addresses the effect of age, gender, and stimulus on the cepstral measures CPP and sCPP.

# References

Awan, S. N., Giovinco, A., & Owens, J. (2012). Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of Voice*, *26(5),* 670-e15.

Awan, S. N., Roy, N., Jette, M. E., Meltzner, G. S., & Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics, 24(9),* 742–758.

Baken, R. J. (1987). Clinical Measurement of Speech and Voice. Boston, MA: College Hill Press, Benjamins.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice,* 2nd edition, Taylor and Francis, NY.

Balasubramanium, R. K., Bhat, J. S., & Prasad, N. (2010). Cepstral Analysis of Voice in Persons with Vocal Nodules. *Journal of Voice, 24(6),* 651-653.

Balasubramanium, R. K., Bhat, J. S., Prasad, N., Fahim, S. & Raju, R. (2011). Cepstral Analysis of Voice in Unilateral Adductor Vocal Fold Palsy. *Journal of Voice, 25(3),* 326-329.

Behrman, A. (2005). Common practices of voice therapists in the evaluation of patients. *Journal of Voice, 19(3),* 454-469.

Benninger, M. S., Ahuja, A. S., Gardner, G., & Grywalski, C. (1998). Assessing outcomes for dysphonic patients. *Journal of Voice, 12(4),* 540-550.

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer.

Bogert, B. P., Healy, M. J., & Tukey, J. W. (1963, June). The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In *Proceedings of the symposium on time series analysis* (Vol. 15, pp. 209-243). chapter.

Brinca, L. F., Batista, A. P. F., Tavares, A. I., Goncalves, I. C., & Moreno, M. L. (2013). Use of Cepstral Analyses for Differentiating Normal from Dysphonic Voices: A Comparative Study of Connected Speech Versus Sustained Vowel in European Portuguese Female Speakers. *Journal of Voice, 28(3),* 282-286.

Casper, J. K. (2001). Treatment outcomes in occupational voice disorders; in Dejonckere, P. H (ed): *Occupational Voice: Care and Cure.* The Hague, Kugler, 187-199.

Colton, R. H., & Casper, J. (1996). Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment. 2nd ed. San Diego, CA: Singular Publishing Group.

De Bodt, M. S., Van de Heyning P. H., Wuyts, F., & Lambrechts, L. (1996). The perceptual evaluation of voice disorders. *Acta Otorhinolaryngol Belgica, 50,* 283-291.

Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., & Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Oto-rhino-laryngology, 258(2),* 77-82.

Deliyski, D. D., Shaw, H. S., & Evans, M. K. (2005). Influence of sampling rate on accuracy and reliability of acoustic voice analysis. *Logopedics Phoiatrica Vocology, 30(2),* 55-62.

Deliyski, D.D., Shaw, H.S., & Evans, M. K. (2006). Regression tree approach to studying factos influencing acoustic voice analysis. *Folia Phoniatrica Logopaedica, 58,* 274-288.

Eadie, T. L., & Baylor, C. R. (2006). The effect of perceptual training on inexperienced listeners judgments of dysphonic voice. *Journal of Voice, 20(4),* 527-544.

Eadie, T. L., Kapsner, M., Rosenzweig, J., Waugh, P., Hillel, A., & Merati, A. (2010). The role of experience on judgments of dysphonia. *Journal of Voice, 24(5),* 1-10.

Garrett, R. (2013). Cepstral- and Spectral-Based Acoustic Measures of Normal Voices. Thesis. University of Wisconsin Milwaukee. August 2013.

Gillespie, A. I., Dastolfo, C., Magid, N., & Gartner-Schmidt, J. (2014). Acoustic Analysis of Four Common Voice Diagnoses: Moving Toward Disorder-Specific Assessment. *Journal of Voice*, *28*(5), 582-588.

Glicklich, R. E., Glovsky, R. M., & Montgomery, W. W. (1999). Validation of a voice outcome survey for unilateral vocal cord paralysis. *Otolaryngology Head Neck Surgery, 120,* 153–158.

Gottliebson, R. O., Lee, L., Weinrich, B., & Sanders, J. (2007). Voice problems of future speech-language pathologists. *Journal of Voice, 21*(6), 699-704.

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

Heman-Ackah, Y. D. (2004). Reliability of calculating the cepstral peak without linear regression analysis. *Journal of voice*, *18*(2), 203-208.

Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S. (2002). The Relationship Between Cepstral Peak Prominence and Selected Parameters of Dysphonia. *Journal of Voice*, *16(1),* 20-27.

Heman-Ackah, Y. D., Michael, D. D., Baroody, M. M., Ostrowski, R., Hillenbrand. J., Heuer, R. J., Horman, M., & Sataloff, R. T. (2003). Cepstral Peak Prominence: A More Reliable Measure of Dysphonia. *Annals of Otology, Rhinology and Laryngology, 112,* 324-333.

Hillenbrand, J., & Houde, R. A. (1996). Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech. *Journal of Speech and Hearing Research, 39,* 311-321.

Hillenbrand, J., Ronald, A., Cleveland, R, A., & Erickson, R. L. (1994). Acoustic Correlates of Breathy Vocal Quality. *Journal of Speech and Hearing Research, 37,* 769-778.

Hogikyan, N. N., & Sethuraman, G. (1999) Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice, 13,* 557–569.

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The voice handicap index (VHI) development and validation. *American Journal of Speech-Language Pathology, 6*(3), 66-70.

Jiang, J. J., & Titze, I. R. (1993). A methodological study of hemilaryngeal phonation. *Laryngoscope, 103,* 872-882.

Jiang, J. J., Zhang, Y., & Ford, C. N. (2003). Nonlinear dynamics of phonations in excised larynx experiments. *The Journal of the Acoustical Society of America, 114(4),* 2198-2205.

Jiang, J. J., Zhang, Y., & McGilligan, C. (2006). Chaos in voice, from modeling to measurement. *Journal of Voice, 20(1)*, 2-17.

Johns, M. M., Garrett, C. G., Hwang, J., Ossoff, R. H., & Courey, M. S. (2004). Quality-of-life outcomes following laryngeal endoscopic surgery for non-neoplastic vocal fold lesions. *Annals of Otology, Rhinology and Laryngology, 113,* 597–601.

Kent, R. D. (1996). Hearing and believing: some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech Language Pathology, 5,* 7-23.

Kreiman, J., & Gerratt, B. R. (2000). Measuring vocal quality. In: Kent RD, Ball MJ, eds. *Voice Quality Measurement.* San Diego, CA: Singular Pub.

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. *Journal of Speech and Hearing Research, 36(1),* 21–40.

Ladefoged, P., & Disner, S. F. (2012). Vowels and Consonants. (3rd ed.). Chichester, UK: Wiley-Blackwell.

Leong, K., Hawkshaw, M. J., Dentchev, D., Gupta, R., Lurie, D & Sataloff, R. T. (2013). Reliability of Objective Voice Measures of Normal Speaking Voices. *Journal of Voice, 27 (2),* 170-176.

Llorente, G. J. I., Ruiz, O. V. N., Lechon, S. N., Marco, C. I., Herranz, G. R., & Calvo, R. C. (2008). Acoustic analysis of voice using WPCVox: a comparative study with Multi Dimensional Voice Program. *European Archives of Otorhinolaryngology, 265(4),* 465–476.

Lowell, S. Y., Colton, R. H., Kelley, R. T., & Hahn, Y. C. (2011). Spectral and Cepstral Based Measures During Continous Speech: Capacity to Distinguish Dysphonia and Consistency Within a Speaker. *Journal of Voice, 25(5),* 223-232.

Ludlow, C., Bassich, C. J., Connor, N. P., Coulter, D. C., & Lee, Y. J. (1987). The validity of using jitter and shimmer to detect laryngeal pathology. In: Baer T, Sasaki C, Harris K, eds. *Laryngeal Function in Phonation and Respiration: Vocal Fold Physiology Series.* (492-508). Boston, MA: College Hill,.

Ma, E. P. M., & Yiu, E. M. L. (2001). Voice activity and participation profile: Assessing the impact of voice disorders on daily activities. *Journal of Speech, Language and Hearing Research, 44(3),* 511-524.

Ma, E. P., & Yiu, E. M. (2005). Multiparametric evaluation of dysphonic severity. *Journal of Voice, 20(3),* 380–390.

Moers, C., Bernd, M., Rosanowski, F., Elmar, N., Ulrich, E., & Tino, H. (2012). Vowel- and text-based cepstral analysis of chronic hoarseness. *Journal of Voice. 26(4)*, 416–424.

Murry, T., & Rosen, C. A. (2000). Outcome measurements and quality of life in voice disorders. *Otolaryngologic Clinics of North America, 33(4),* 905-916.

Noll, A. M. (1967). Cepstrum Pitch Determination. *Journal of Acoustical Society of America, 41(2),* 293-309.

Orlikoff, R., Dejonckere, P., Dembowksi, J., Fitch, J., Gelfer, M. P., Gerratt, B. R., et al. (1999). Perceived role of voice perception in clinical practice. *Phonoscope, 2,* 89-108.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., & Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics, 28,* 233–272.

Reddy, M. S., Pebbili, G. K., Joanna, K. B., Jesnu, J. B. (2014). Smoothened Cepstral Peak Prominence in Untrained Choral Singers and Non Singers. Proceedings of the International Symposium Frontiers of Research on Speech and Music, 101/104.

Rosen, C. A., Lee, A. S., Osborne, J., Zullo, T., & Murphy, T. (2004). Development and validation of the Voice Handicap Index-10. *Laryngoscope, 114(9),* 1549-1556.

Sapienza, C. M. (1996). Glottal airflow: Instrumentation and interpretation. *The Florida Journal of Communication Disorders*, *16*, 3-7.

Savithri, S. R., & Jayaram, M. (2005). 300 word reading passages in Dravidian languages. ARF project undertaken at AIISH, Mysore.

Titze, I. R. (1995). Summary statement: Workshop on Acoustic Voice Analysis. Denver, CO: National Center for Voice and Speech.

Watts, C. R., & Awan, S. N. (2011). Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech, Language, and Hearing Research, 54*, 1525-1537.

Webb, A. L., Carding, P. N., Deary, I. J., MacKenzie, K., Steen, N., & Wilson, J. A. (2004). The reliability for three perceptual scales for dysphonia. *European Archives of Otorhinolaryngology, 261(8)*, 429-434

Wolk, L., Abdelli-Beruh, N. B., & Slavin, D. (2012). Habitual use of vocal fry in young adult female speakers. *Journal of Voice*, *26*(3), 111-116.

Yang, Z., Fan, J., Tian, J., Liu, L., Gan, C., Chen, W., & Yin, Z. (2014). Cepstral Analysis of Voice in Children With Velopharyngeal Insufficiency After Cleft Palate Surgery. *Journal of Voice*, *26 (6),* 789-792.

Zhang, Y., & Jiang, J. (2003). Nonlinear dynamic analysis in signal typing of pathological human voices. *Electronics Letters, 39,* 1021–1023.

Zhang, Y., & Jiang, J. J. (2005). Spatiotemporal chaos in excised larynx vibrations. *Physical Review E*, *72*(3), 035201.

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

Zyski, B., Bull, G. L., McDonald, W. E., & Johns, M. E. (1984). Perturbation analysis of normal and pathological larynges. *Folia Phoniatrica, 36,* 190-198.