CEPSTRAL PEAK PROMINENCE IN NORMAL AND DYSPHONIC VOICE

Kavya S

Student Register Number
13SLP012

This Dissertation is submitted as partfullfillment
for the Degree of Master of Science in Speech-Language Pathology
University of Mysore, Mysore

May, 2015

All India Institute of Speech and Hearing
Manasagangothri, Mysore -570006

CERTIFICATE

This is to certify that this dissertation entitled "**Cepstral Peak Prominence in Normal and Dysphonic Voice**" is a bonafide work in part fulfillment for the degree of Master of Science (Speech-Language Pathology) of the student (Registration No. 13SLP012). This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Mysore

May, 2015

Prof . S. R. Savithri
Director
All India Institute of Speech and Hearing
Manasagangothri
Mysore- 570 006

CERTIFICATE

This is to certify that this dissertation entitled "**Cepstral Peak Prominence in Normal and Dysphonic Voice** " is a bonafide work in part fulfillment for the degree of Master of Science (Speech-Language Pathology) of the student (Registration No. 13SLP012). 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.

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 dissertation entitled "**Cepstral Peak Prominence in Normal and Dysphonic Voice**" is the result of my own study under the guidance of Mr. Gopikishore 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 any other University for the award of any Diploma or Degree.

Mysore

May, 2015 **Register No.** ***13SLP012***

*****

Dedicated to

My Daddy, Mom and my grandma

*****

Acknowledgement

First and foremost, praise and thanks to the God. You have given me the power to believe in myself and pursue my dreams. I could never have done this without the faith I have in you, the Almighty.

I would like to express my deepest gratitude to my wonderful guide Mr. Gopikishore. Sir, thank you so much for your excellent guidance, encouragement, patience and providing me with an excellent opportunity to learn. Your creative thinking and hard-working nature has been a great inspiration to me. It was a great learning curve to work under you sir. THANK YOU.

I would also like to Thank Navya ma'am for her constant support and timely help throughout my research work.

My heartfelt thanks to our beloved Director, Prof. Savithri. S. R. for giving me an opportunity to carry out this research.

I thank the participants of the study, who were kind enough to be a part of this study.

My sincere thanks to our statisticians Dr. Vasanthalakshmi and Mr. Santhosh.

I would like to thank all teachers at AIISH and Special thanks to Dr. Pusphavathi, Dr.Yeshoda, Dr. Prakash, Dr. Sundaraju, Mr. Rajasudhakar and Mr. Gopishakar.

Thank you to our Library staffs and AIISH.

Mr. M.K. thank you for your love, wishes, encouragement, trust, friendship, support and motivation. These words feel short to express the feelings…….but they carry the sound from the strings of my heart!!!!!!!!Thank YOU SO MUCH!!!!!

I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating me for my future. Also, I express my thanks to My brothers Ravi and Raghu. "Kulla" though you are younger to me, U r always der fr cheering me up and stood by me through the good times and bad. luv u loads n wish u very best fr ur future.

Am really lucky to have seniors like u Madhu sir, Beena akka and Chaithanya sir thanks a ton for helping me with analysis.

I Would like to thank all my juniors who helped me with my data collection, special thanks to nikil, sumanth, rakshu, aishu, rashmi, pooja, kavita, harisha, nivanka, deepu and akshaya.

Special thanks to d one who stood with me in tough times, Dear friends thank you so much….. Darshu, Jaan, nimmi, shri, kappu, usha, guna, suji, sneha, Sandra, himansh, PG, supreeth, avengers group …. *Special thanks to* " harika, manjula, navya, suman and Padma"…..

I Would like to thank all my classmates who directly and indirectly helped me to complete my work.

Each one we meet in our life definitely has a role in our life. Some hurts and some loves you. Some stays back and some leaves for ever. Thanks to all the wonderful people in my life. U made my life blissful and stronger… and **special thanks for all those difficult people too.. for they made me bolder**. Life gets harder as we progress … and sometimes the most beautiful relation does not need a daily conversation… a glimpse makes it all… and I believe at the end we sit back in our easy chair seeing the beautiful portrait and then close ur eyes and believe strongly **"if it means to be, it will be"**!!!!!!!!

Table of Contents

List of Tables

List of Figures

CHAPTER I

Introduction

Voice is a powerful tool to express both emotionality and spoken language. It serves the melody of our speech and provides expression, feeling and mood to our articulated thoughts. Voice disorders occur as a result of faulty structure or function of the larynx which results in the change in quality, pitch, or loudness of the voice that is different from what is expected from someone of the same age or sex (Smith, Verdolini, & Gray et al 1996). Boone, Mcfarlane, Von Berg andZraick (2010) classified the voice disorders under three kinds based on the etiological factors. The first kind is the organic voice disorders which includes any laryngeal structural deviations that affect vocal fold vibration. The second kind of voice disorder is neurogenic voice disorder, related to neurological condition that causes faulty vocal fold closure from either paralysis (or weakness) or from neurological disease. The third kind of voice disorder is functional voice disorders: psychogenic voice disorders which are caused by psychosocial factors, and muscle tension voice disorders (Muscle Tension Dysphonia), which can develop from excessive muscle usage.

Clinical voice evaluation includes both auditory perceptual evaluation and instrumental evaluation which further includes various voice measurement techniques such as acoustic analysis, aerodynamic measurement and laryngeal imaging. Among the three domains of instrumental analysis, acoustic analysis of voice is more commonly used in clinical practice due to its noninvasive procedure, increased objectivity, automaticity and the use of speech signal for acoustic analysis. Another reason for the frequent clinical use of acoustic analysis, is the availability of various freely downloadable softwares in market.

Acoustic analysis of voice includes both time based measures and frequency based measures which are used to quantify voice characteristics, predict dysphonia and to track the intervention program. Historically, time-based measures used perturbation parameters, like jitter and shimmer, and other measures like Noise to Harmonic ratio (NHR) for acoustic analysis of voice. However, these traditional time based measures have a substantial disadvantage. These measures require exact identification of the cycle-to-cycle boundaries in the acoustic waveform. Although this boundary identification can be reliable and accurate for voices that are reasonably periodic, whereas reliability breaks down for a reduced amount of periodic voices that characterize dysphonic individuals as it is difficult to identify the cycle-to-cycle boundaries because they lack periodicity (Lowell, Colton, Kelly, & Hahn, 2011). As these measures rely 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, these measures are useful in analysis of steady state phonation, however, they have limitations when applied to connected speech (Awan, Roy, Jette, Meltzner, & Hillman, 2010). Many studies in the past have shown that these traditional time based measures are unreliable predictors of dysphonia (Heman-Ackah, Ostrowski, Horman, Baroody, & Hillenbrand et al., 2003). Because of these validity and reliability issues in analyzing voice in sustained vowels and connected speech across a continuum of dysphonia severity, robust acoustic measurements which do not depend on time-based analyses are necessary(Hillenbrand, Ronald, Cleveland and Erickson, 1994; Herman-Ackah et al., 2003; Maryn, 2009; Awan, Roy, Jette, Meltzner, & Hillman, 2010). Therefore, the current practice is moving towards spectral based measures, which overcome the drawback of time based measures. As they do not require cycle boundary detection, they can be used to analyze severe dysphonic voice.

Spectral based measures such as Cepstral measures have been used as an alternative to the traditional measures for voice analysis as these measures estimate aperiodicity or additive noise without locating the cycle boundaries (Awan, Giovinco, & Owens, 2012). “Cepstrum is described as a discrete Fourier transform of the logarithm power spectrum; i.e. it is a log power of a log power spectrum” (Hillenbrand et al., 1994; Hillenbrand & Houde, 1996). To produce a cepstrum, first, an acoustic signal is Fourier transformed to create a spectrum, and the voice signal is converted from the time domain to the frequency domain. Thus, the intensity of each frequency within the signal is represented in the spectrum. Performing a Fourier transformation of the spectrum then produces the cepstrum. In doing so, the signal is transformed from the frequency domain to the quefrency (which equals 1/frequency) domain and a better visual picture of the degree of harmonic organization is produced. A linear regression line is fitted relating quefrency to cepstral magnitude. One among the cepstral measures is Cepstral Peak Prominence (CPP). It is the difference in amplitude between the cepstral peak and the corresponding value on the regression line that is directly below the peak. CPP is, thus, a measure of the degree of harmonic organization, which tells how far the cepstral peak emanates from the cepstral “background noise” (Hillenbrand et al., 1994). Another measure under cepstral analysis is the smoothened Cepstral Peak Prominence (sCPP) in which the individualized cepstra of voice signal are averaged over a given number of frames before extracting the cepstral peak and calculating the peak prominence (Hillenbrand et al., 1996).

The Cepstrum graphically shows the extent to which the dominant rahmonic is individualized. A voice signal that shows disturbed periodicity or increased spectral noise as seen in dysphonic voice is associated with a decrease in amplitude of the cepstral peak i.e., lower harmonic energy. Cepstral-based measures profit from the fact that they are computed via frames of signal data rather than cycle boundary identification. Periodic voice signals display well defined harmonic configuration in the spectrum and thus, a more prominent cepstral peak is obtained (Hillenbrand et al., 1994). Decrease in overall CPP shows voice abnormality (Hillenbrand et al., 1994; Hillenbrand et al., 1996; Blankenship, 2002).

The CPP can be measured in different ways including the use of CSL model or Speech tool program (HIillenbrand, et al., 1994; Hillenbrand et al., 1996). Speech tool program is advantageous over the other methods (Heman-Ackah, Michael, &Goding, 2002) as it uses particular algorithm which is developed by Hillenbrand et al., 1994 and 1996. Speech tool program automatically calculates the CPP and sCPP, whereas in CSL it is a manual basis where the clinician has to perform different operations to obtain CPP and sCPP. Speech tool is a more reliable program to measure CPP and sCPP (Heman-Ackah et al., 2002; Heman-Ackah et al., 2003).

Figure 1 represents the voice cepstrum of a normal voice and figures 2 and 3 represents a cepstrum of dysphonic voice. When the voice becomes dysphonic the CPP tends to be reduced. The better the voice signal better will be the CPP. If the voice lacks periodicity, then cepstrum becomes relatively flat and cepstral peak reduces. Signals which lack a well-defined harmonic structure have reduced CPP (Hillenbrandet al., 1996).

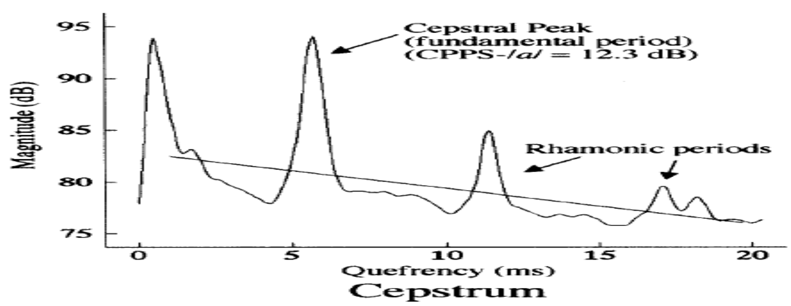

*Figure 1.*Cepstrum representing a normal voice signal.
(Source: Heman-Ackah, 2003)

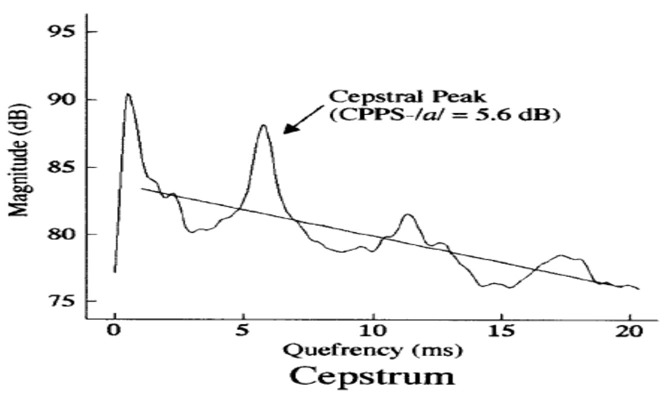

*Figure 2.*Cepstrum representing moderately dysphonic voice signal.
(Source: Heman-Ackah, 2003)

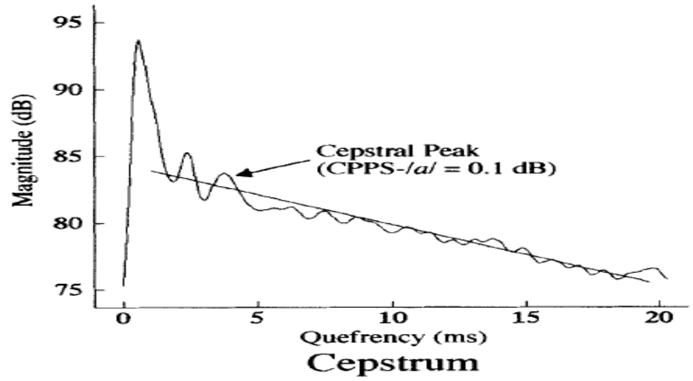

Figure 3.Cepstrum representing a severely dysphonic voice signal.
(Source: Heman-Ackah 2003)

Studies has shown that CPP and the CPPS correlate with perception of breathiness, with sCPP being the better predictor (Hillenbrand, et al., 1994; Hillenbrand, Cleveland, & Erickson, 1994; Olson, Goding, & Michael,1998).Unlike perturbation measures and Noise to Harmonic Ratio (NHR), CPP and sCPP measures do not rely on the accurate identification of the fundamental frequency; they are based on a peak-to-average calculation. For this reason, measures of CPP and CPPS tend to be more consistent than other measures of periodicity (Hillenbrand et al., 1996).

Studies have also shown that CPP is more reliable indicator of dysphonia than any other approaches (Hillenbrand et al., 1994; Hillenbrand et al., 1996; Olson, Goding, and Michael, 1998; Kumar, Bhat, & Prasad, 2010; Kumar, Bhat, Fahimand, &Raju, 2011) because CPP doesn’t depend on the accuracy of fundamental frequency (fo) extraction which is difficult to establish in severely disordered voices.It is more reliable measure to analyze both phonation and connected speech (Heman-Ackah et al., 2003). Studies have shown that measurement of CPP derived from the acoustic spectrum correlates best with auditory perceptual classification of dysphonia (Hillenbrand et al., 1994; Hillenbrand et al., 1996; Awan, 2005; Maryn, 2009; Moers, 2012).

Need for the study

Literature indicates potential clinical applications of Cepstral measures in voice evaluation. Most of these studies reported that Cepstral measures have good correlation with the perceptual evaluation of voice and aids in discriminating normal from dysphonic voices. However, the studies (Kumar et al., 2010, Balasubramanium et al., 2011 & Brincaet al., 2013) have not investigated the discrimination ability of cepstral measures in categorizing the individuals with dysphonia based on the perceived severity. Another limitation is most of the researchers evaluated and compared the cepstral measures in normals and individuals with dysphonia due to vocal nodules (Kumar et al., 2010) and vocal paralysis (Balasubramanium et al., 2011). These pathologies often lead to glottic chink resulting in breathy voice and hence reducing the amplitude of vocal harmonics. This in turn can reduce the Cepstral measures in individuals with vocal pathologies (vocal nodules and vocal paralysis) compared to normals. With this limited research evidence, it is difficult to comment upon the effect of dysphonia on measures of CPP as the studies have considered only specific voice disorders. Therefore, it is essential to evaluate and compare cepstral measures in individuals with dysphonia (irrespective of voice disorder) with normals. Hence the present study is aimed to investigate the relation between the CPP and perceived severity in individuals with dysphonics.

Aim of the study: To investigate Cepstral measures across different levels of perceptually based dysphonia severity**.**

Objectives of the study: The objectives of the study are 1. To document Cepstral Peak Prominence(CPP) and Smoothened Cepstral Peak Prominence(sCPP) in individuals with normal voice quality, mild, moderate dysphonia and severe dysphonia. 2. To verify whether Cepstralmeasures can differentiate levels of perceptually based dysphonia severity.

CHAPTER II

Literature Review

The primary objectives of the diagnostic voice evaluations are to discover the etiologic factors associated with the voice disorder, to describe the deviant vocal symptoms and to develop an understanding of how the disorder is affecting the sub-systems of voice production which are respiration, phonation and resonance (Stemple, 2010). In the area of measuring voice, there are many methods or approaches in the literature, these can be classified grossly as subjective and objective methods of voice evaluation. Objective methods of evaluation include the usage of various instruments in order to assess the voice, whereas subjective evaluation depends on the perceptual assessment of voice.

Perceptual assessment is the core foundation for voice evaluation and for treatment outcomes in both surgical and behavioral intervention of voice disorders**.** It involves describing the voice solely through listening by an experienced by clinician. It can be performed in either formal or informal way. Informal assessment takes place through the conversation between the clinician and client. Formal evaluation involves the use of standardized protocols and is performed systematically using standardized procedures. Some of the standard protocols used for perceptual analysis of voice include GRBAS (Hirano, 1981), and Consensus Auditory perceptual evaluation of voice (CAPE-V). GRBAS is a four point rating scale whereas CAPE-V is a visual analog scale. The perceptual features of voice quality are also likely to have greater shared reality among a wide range of listeners including clinicians, clients, employers and other associates of those clients. Therefore, perceptual evaluation is often considered as a gold standard in voice assessment. However, perceptual analyses are criticized for their reliably**.** It is because they are influenced by several factors such as experience of the clinician and stimulus used for elicitation. Some of these challenges can be overcome by substantiating perceptual analysis of voice by objective measurements of voice.

Objective evaluation of voice is one of the best assessment methods for clinical voice evaluation. It includes both invasive and non-invasive methods for assessment. Invasive methods include vocal imaging techniques such as stroboscopy, videolaryngoscopy, high- speed digital videoendoscopy, videolaryngostroboscopy, or videokymography and they are used for 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.

Non-invasive methods of voice analysis include acoustic analysis of voice. These are widely used to assist perceptual analysis. The acoustic analysis is usually presumed easy to administer. These are non invasive procedures and less time consuming when compared to laryngeal imaging procedures. These methods help in clinical diagnosis of various voice disorders and also in monitoring and documenting treatment outcomes. Acoustic analysis of voice provides information regarding the stability or variability of the vocal fold movement through perturbation measures of amplitude and frequency. It also yields information about harmonics and noise components of the voice, thus, helps in understanding the turbulence at the level of vocal folds. There are many acoustic measures available for measurement of voice as mentioned in the literature. These can be broadly classified further into time based measures and frequency based measures.

The traditionally used voice measures for voice analysis are the time based measures along with the noise measures. These include fundamental frequency and its variability, as well as perturbations measures that is jitter (frequency perturbation), shimmer (amplitude perturbation), and noise-to-harmonic ratio (NHR). These measures have the ability to accurately identify and track changes in fundamental frequency. These measures are effective in identifying and classifying dysphonia in mild dysphonic. However, these traditional measures have certain limitations. These measures have not met success in their abilities to consistently and reliably quantify the voice.

"Validity of these traditional measures in analyzing voices of individuals with moderately to severely disordered voices, has been recently called into question because cycle boundary identification can be exceptionally difficult" (Awan & Roy, 2005). The reliability and clinical practicality of particular perturbation measures are still unsure for moderate or severely disordered voices (Maryn, Roy, De Bolt, Van Cauwenberge, &Corthals, 2009). The reliability and validity issues are because these perturbation measurements are influenced by accurate identification of cycle boundaries (i.e. where the cycle of vibration begins and ends in the signal) in the voice signal. So, it is increasingly noticeable that the existence of significant noise in the signal makes it less periodic and difficult to accurately track these cycle boundaries. Thus, these measures are possible only in mildly dysphonic voice signal which is reasonably periodic. The other limitation is that the time-based measures are suitable for analysis of sustained vowels only. When it comes to connected speech, these measures are shown to be less accurate (Maryn, Roy, De Bolt, Van Cauwenberge, &Corthals, 2009). It is because 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 (Maryn et al., 2009).

Awan, Roy, Jette, 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. And also, traditional measures are influenced by few extraneous variables such as distance between mouth to microphone, the loudness of the voice signal, the type of 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. Thus, in order to reliably analyze voice in connected speech, sustained vowels, and across a continuum of dysphonia severity, acoustic measurements other than those that are time-based are necessary.

Other set of acoustic measures are frequency based measures /spectral measures. These measures overcome the limitation of traditional measures. In literature, several investigators have reported that measures derived from spectral analysis strongly predict presence of additive noise in the signal, perceived severity of dysphonia and type of voice disorder (Hillenbrand et al., 1994; Hillenbrand et al., 1996; Dejonckere&Wieneke 1996; Callan, Kent, & Roy,1999; Wolfe, Martin, & Palmer, 2000). Some of the spectral measures include long term average spectrum, low to high spectral ratio, measures of spectral tilt, amplitude of the first spectral harmonic, and reductions in spectral harmonic-to-noise ratios. These measures have been reported as effective indices of dysphonic type and severity. In addition to the measures of the spectrum, derivation of the Cepstrum has also been investigated as a useful method for describing the dysphonic voice. Cepstrum was originally described by Noll (1964). It was derived via a Fourier transform of the power spectrum of the voice signal, and it graphically displays the extent to which the spectral harmonics and, in particular, the vocal fundamental frequency, are individualized and emerge out of the background noise level. It is the degree to which the cepstral peak relates to extraneous vocal frequencies that theoretically provides an effective method of quantification for the disordered voice (Hillenbrand et al., 1994).

One among the cepstral measures is Cepstral Peak Prominence (CPP). It is the difference in amplitude between the cepstral peak and the corresponding value on the regression line that isdirectly below the peak. CPP is, thus, a measure of the degree of harmonic organization, which tells how far the cepstral peak emanates from the cepstral“background noise”(Hillenbrand et al., 1994). Another measure under cepstral analysis is the smoothened Cepstral Peak Prominence (sCPP) in which the individualized cepstra of voice signal are averaged over a given number of frames before extracting the cepstralpeakand calculating the peak prominence (Hillenbrand et al., 1996). The rationale behind these measures is, periodic voice signals displays well-defined harmonic configuration in the spectrum and thus obtains a prominent cepstral peak in the selected signal. Voices with less aperiodicity will have robust CPP and the voices with severe dysphonia/ severely distorted voice will have flat cepsral peak representation. Advantages of cepstral measures over traditional measures are: cepstral measures quantify the voice signal without relying on frequency or its intensity or on any other variables that can affect the accuracy of the measurement, these measures are reliable and reproducible, these measures are based on peak to peak calculation and not on accurate calculation of fundamental frequency. CPP and sCPP can be used for analyzing voices with severe dysphonia. 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 (Halberstam, 2004; Heman-Ackah, 2004; Awan, Roy, Jetté, Meltzner, & Hillman, 2010). These two measures can be incorporated to evaluate the dysphonia severity in continuous speech and sustained vowel voice signal, thus measures overall severity (Maryn, Roy, De Bodt, Van Cauwenberge, & Corthals, 2009).

Maryn et al. (2009) performed meta-analysis on acoustic measurement of voice to identify measures of acoustic analysis that have good correlation with perceptual rating

(Heman-Achah, 2004). The meta-analysis reviewed a total of 25 studies; 21 studies examined sustained vowels using 69 acoustic markers and seven studies examined connected speech using 26 acoustic markers. The meta-analysis identified six acoustic parameters that were determined to correlate reasonably well with listener ratings: (1) Pearson *r* at autocorrelation peak, (2) spectral flatness of residue signal, (3) pitch amplitude, (4) cepstral peak prominence (CPP), (5) smoothed cepstral peak prominence, and (6) signal-to-noise ratio. Most of these measures are not time-based; thus they do not require cycle boundary identification to determine fundamental frequency and estimate aperiodicity. Among the six measures, CPP and smoothened CPP were found to be the best predictors of dysphonia severity as compared to listener ratings. These cepstral-based measures can also be used to evaluate the severity of voice for both continuous speech and sustained vowels, thus, they appear to be ideal for evaluating overall severity. In the present study these two measures were used.

Instrumentation for Cepstral analysis

Voice analysis using cepstral measures can be carried out using 1. SpeechTool program (Hillenbrand, Western Michigan University, Kalamazoo, MI) 2.CSL *Multi-Dimensional Voice Program* [(*MDVP*), Kay Elemetrics]. The difference is CSL program does not separate voiced portion and unvoiced portion of the voice signal thus, consists of both computations and for analysis the selections are completely manual based. Also, CSL method of calculating CPP does not correlate with perceptions of dysphonia (Heman- Achak, 2004). In CSL method of calculating CPP, the ability of clinician to separate the average energy within the cepstrum visually from the graphic representation is not as accurate as mathematically calculating the linear regression line of the average energy, as it is performed in Speech Tool program. It is possible to mathematically calculate the linear regression line by hand from the CSL and then determine the magnitude of the cepstral peak prominence. This process is observed to be cumbersome and time consuming. In comparing this, speech tool is reported to be more advantageous (Heman-Achah, 2004) because CPP and sCPP can be measured automatically and it directly provides the mean CPP and mean sCPPvalues, making the task easy and less time consuming. Further, Speech tool is freely downloadable program. Hence, speech tool program is used in the present study. In recent literature Awan, Solomon, Helou, and Stojadinovic (2013) have used automated voice detection algorithm (Analysis of Dysphonia in Speech and Voice (ADSV); KayPENTAX, Montvale, NJ), for measuring CPP. The ADSV model was established for the Cepstral and Spectral index of Dysphonia Severity (CSID). It uses both Cepstral and spectral measures for voice analysis and also can be used for both connected and spontaneous speech sample for analysis. ADSV program is a purchasable program. Hence in this study, Speech tool program is used for voice analysis.

Studies related to the application of Cepstral analysis of voice.

Considering the robustness and varied clinical applications of cepstral analysis, there are several studies which have compared normal and dysphonic voices (Kumar et al., 2010; Kumar et al., 2011; &Brincaet al., 2013). Studies have compared correlation of cepstral measures with perceptual evaluation (Heman-Ackah, 2004; & Moers et al., 2012) and few studies have used cepstral measures to track therapy/ surgical intervention (Heman-Ackah et al., 2002; & Gillespie et al., 2014). Some of them have assessed the reliability and validity in assessing voices with high level of aperiodicity (Heman–Ackah, 2014). Some of the studies have predicted dysphonia severity (Wolfe & Martin, 1974; &Awan et al., 2010) The following section describes some of the above mentioned studies in detail.

Kumar, Bhat and Prasad (2010) analyzed CPP in 50 subjects (25 males and 25 females) with vocal nodules and 50 age and gender matched controls. All the participants phonated samples of vowel /a/ at their habitual pitch and loudness. These samples were directly recorded into CSL 4150 hardware using a dynamic microphone, which was maintained at a constant distance of 10 cm from the participant's mouth. The results of the study revealed significantly lowered mean CPP values in dysphonic group. Authors attributed lower values of CPP in the vocal nodule group to the presence of flat harmonic structure which is due to the presence of glottic chink secondary to vocal nodule in clinical group. The authors opined that the air escape through the glottic chink during voicing would have contributed to increased level of noise in acoustic signal which lessened the CPP value. The results of the study emphasize on the significance of cepstral analysis as a diagnostic measure for voice evaluation.

Balasubramanium, Bhat, Fahimand, and Raju (2011) measured CPP in 60 participants (30 males and 30 females) with unilateral adductor vocal fold palsy and compared it with age and gender matched controls using Computerized Speech Lab for phonation of vowel /a/. The results of study revealed significantly lower mean CPP values in the clinical group. The authors attributed lower values of CPP in the clinical group to the presence of a flat harmonic structure which was evidenced because of the presence of phonatory gap due to vocal cord palsy. Results also revealed significant differences between males and females in both the groups of participants, indicating lower values of cepstral peak in females. It is attributed to the fact that, for about 80% of females have a posterior glottic chink. The above studies have used only phonation task for cepstral analysis. They had not included connected speech tasks.

Brinca, Batista, Tavares, Gonc¸ Alves, and Moreno (2014**)** studied the importance of CPP and sCPP to differentiate dysphonic from nondysphonic female voices, using two different speech tasks: sustained vowel /a/ and connected speech. The study included 30 females in the age range of 19-66 years with vocal dysfunction due to various vocal pathologies like edema, nodule, laryngeal gap and UVFP and 30 individuals as the control group. Participants phonated vowel and read aloud at a comfortable pitch, loudness, and speaking rate. The Acoustic analysis was carried out using Hillenbrand speech tool program and GRBAS was used for perceptual evaluation of voice quality.

Results of the study revealed that CPP and sCPP had significantly lower values in the dysphonic group. For the connected speech, significantly lower values of CPP were obtained in the dysphonic group compared with the control group and no differences were obtained with sCPP. With both speech tasks the values obtained for CPP correlated strongly with breathy voices. The results of the study thus suggest that CPP and CPPs are promising tools in clinical practice. However, a limitation of the study was that it assessed only female voice and a second limitation was the lack of homogeneity of voice disorders in the clinical group. Therefore, this study recommends for future research on cepstral analysis in both male and female voice groups and also suggests to carry out research in homogeneous clinical population to minimize the inter speaker variability. This study also emphasizes to check the validity of comparing different pathologies in the same group with a normal group.

Studies correlating perceptual analysis with Cepstral analysis.

Heman-Ackah, (2004) studied the perceptual correlation of dysphonia with CPP, which is obtained through two different methods. It is measured using CSL program and Speech tool program separately. Voice samples of running speech and sustained vowel phonation was obtained from 150 participants. The samples were subjected to perceptual and acoustic analysis. Perceptual analysis was carried out by ten speech-language pathologists using GRBAS –like scale and quantified the severity in each category by a mark on a 100-mm line from most normal (0) to most abnormal (100). CPP was determined separately using two different procedures i.e., manual method to obtain CPP in CSL program and automatic calculation of CPP in Speech tool program. Pearson correlation was used to correlate CPP and perceptual dysphonia. Results revealed that cepstral peak prominence was shown to be a reliable predictor of dysphonia. The cepstral peaks derived from CSL MDVP software program do not show correlation with perception of dysphonia, whereas CPP derived using Speech tool correlated with dysphonia.

Moers, Mobius, Rosanowski, Noth, Eysholdt, and Haderlein (2012) conducted an ex post facto research where they analyzed voice of 73 German participants with chronic hoarseness (24 Men and 49 Women). They analyzed phonation sample of vowel /i/ and reading sample of German text. The study had included both perceptual and acoustic evaluation in which perceptual evaluation on the reading sample was performed by five speech therapists and physicians using the German Roughness-Breathiness-Hoarseness (RBH) scale. In acoustic evaluation, noise based measures were analyzed using Praat software and cepstral analysis was carried out using Speech tool software. Results revealed that cepstral measures correlated with perceptual analysis of a representative group of chronically hoarse subjects. However, following the exclusion of participants with unreliable perturbations, the correlation was moderate. The cepstral measures CPP and sCPP outperformed all introduced perturbation measures.This is the first study to compare text based cepstral measures with perceptual analysis and to analyse the impact of "unreliable" measures on the voice evaluation. Hence, this study recommends to use text based recording and cepstral measures along with other acoustic measures and perceptual analysis.

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 participants as control group, between 25–32 years of age and 12 male and female participants between 21–78 years of age with varying dysphonia severity (mild, moderate, and severe). The sustained vowel and CAPE-V sentences are recorded using

KayPentax CSL model 4400. Results indicated that there is 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 and to the sustained vowel /ɑ/.

Alpan, Schoentgen, Maryn, Grenez, and Murphy (2012) carried out a cepstral analysis in 28 normophonic and 223 dysphonic speakers using 2 tasks, concatenation of two Dutch sentences with sustained vowel /a/. Dysphonic group comprised of various diagnosed pathologies. The obtained samples were further given for perceptual analysis by five judges using GRABS scale where they rated parameter "grade",(G) using 4 point rating scale where 0 (normal) to 3 (severe). The five perceptual scores per stimulus of the each samples were averaged. Acoustic analysis was carried out using speech tool to obtain CPP. The results showed significant correlation between CPP and perceptual ratings.

Wolfe and Martin (1997) did a study to explore acoustic categorization of commonly occurring dysphonic voice qualities and also to predict severity based on acoustic analysis. Study included 51 participants in dysphonic group with 20 males and 31 females in the age range of 15 to 79 years. Voice samples included phonation of vowel /a/ and /i/ for 2 or many trials for obtaining consistent phonatory sample. Totally 102 samples were recorded. These were given to 5 experienced clinicians to classify voice. The four classifications of voice were breathy, hoarse, strained and normal, and these achieved maximum agreement. Hence, these four were included for perceptual analysis. The samples which were classified were given to 11 undergraduate SLP students for further perceptual analysis. The raters had 2 short training sessions prior to rating of experimental sample. After training period, 11 SLP's rated 101 vowel samples using 7 point equal appearing scale where 1-*5* normal and *5-7* severely abnormal and classified the sample as primarily normal," "primarily breathy," "primarily hoarse," or "primarily strained".

For acoustic analysis two measures of perturbation and two measures of spectral noise as well as fundamental frequency were included. Two spectral measures included were signal-to-noise ratio (SNR) and Cepstarl peak prominence. CPP was measured using speech tool program. Results showed that there was no significant difference found between hoarse and breathy voice types on the basis of spectral noise measures, CPP or SNR. However, strained voice differed from the other two voice types in having lower levels of CPP and SNR. Strained voice type also differed from breathy in having lower values of jitter. Both breathy and hoarse type voices had lower values of CPP than the strained voice type. Furthermore, there was no significant difference between hoarse and breathy voice types on CPP. The authors concluded that cepstral measures couldn't differentiate voice quality types.

Lowell, Colton, Kelley, and Hahn (2011) did a study to determine whether cepstral and LTAS-derived acoustic measures could differentiate a group of 27 mixed, laryngeal-based dysphonic participants in the age range of 19–86 years from a group of 27 controls in the age range of 26–55 years using continuous speech task and also to see correlation between acoustic measures and auditory-perceptual voice quality ratings.

Dysphonic group included various laryngeal lesions such as mass lesions of the vocal folds, paresis/paralysis, keratosis/ leukoplakia, vocal fold edema, presbyphonia and laryngeal web. Acoustic analysis was carried out using Kay PENTAX CSL 4500 system for analyzing LTAS measures (spectral mean, spectral SD, kurtosis, and skewness) and Speech tool for Cepstral measures. Perceptual analysis was carried out by three judges with extensive experience in voice disorders. They rated the dysphonic samples using a 100-mm VAS with definition of overall severity as provided under the CAPE-V for three voice qualities (roughness, breathiness, and strain).

Results indicated that LTAS and cepstral-based measures strongly differentiated the dysphonic voice samples from normal speaker group. For the dysphonic group, spectral mean, spectral SD, CPP, and sCPP were lower. Both CPP and sCPP were lower for the dysphonic speakers relative to the normal speakers. Cepstral measures (CPP and sCPP) showed moderate to strong relationships with overall voice severity.There was a negative correlation of overall voice severity and CPP or sCPP. As the severity of voice disorder increased, there was a decrease in CPP and sCPP. This study recommends future studies checking reliability of cepstral-and LTAS-based measures when intervention or other functional change has not occurred.

Heman-Ackah, Michael, and Goding (2002) studied thirty-eight samples of connected speech and sustained vowel phonation preoperatively and postoperatively from 19 participants in the age range of 25 to 87 years with unilateral recurrent laryngeal nerve paralysis. They had undergone surgical intervention (seven patients had a type I thyroplasty and 12 patients underwent reinnervation) to improve their voice quality. Pre and postoperative voice samples were subjected to three separate analyses, in which first analysis was perceptual analysis by two Speech-language pathologists who rated the grade (overall dysphonia), roughness, breathiness, asthenia, and strain. Severity in each category was quantified by a mark on a 120-mm line from least abnormal (0) to most abnormal (120). In the second analysis, the sustained vowel samples were analyzed using the Multi-Dimensional Voice Profile (MDVP) (CSL model 4305, Kay Elemetrics, Lincoln Park, NJ). The relative average perturbation (RAP), smoothed pitch perturbation quotient (sPPQ), amplitude perturbation quotient (APQ), and NHR were measured from each of the samples. In the third analysis, connected speech (25-kHz sampling rate) and sustained vowel samples were analyzed by measuring sCPP-s and sCPP-/i/, respectively, using Speech tool software.

Results reveal that CPPS for both connected speech (CPPS-s) and sustained vowel phonation (CPPS-/i/) correlated inversely and strongly with perceptions of overall dysphonia (grade) and breathiness, with the greatest correlation seen with overall dysphonia. This study concludes that measurement of the cepstral peak prominence from samples of running speech provide the most reliable predictor and objective correlate of dysphonia of the available methods of objective voice measurements.

Heman-Ackah, Ostrowski, Horman, Margaret, Baroody, Hillenbrand, and Sataloff (2003) did a study with the following aim; to determine the ability of the sCPP to predict severity of dysphonia reliably, to determine the reliability of the sCPP in predicting dysphonia relative to other acoustic measures, and to determine the range of normal and abnormal values of sCPP. Study included voice samples from 281 participants with voice problem, who ranged in age from 7 to 80 years with the mean age of 43 years. The samples included sustained phonation of /a/ and running speech task included reading of "Marvin Williams" passage. Ten individual speech-language pathologists and voice specialists who had minimum of 3 years of professional experience rated the running speech sample using GRBAS-like scale in which Severity in each category was quantified by a mark on a 100-mm line from most normal (0) to most abnormal (100). In which perception of grade, which was defined as overall dysphonia, was used as the standard against which the acoustic measures were compared. Acoustic analysis was carried out using MDVP(model 4300B, Kay Elemetrics) and Speech tool program. The conventional measures included jitter, shimmer, and NHR for sustained vowel /a/ and the sCPP using Speech tool program for connected speech and the sustained vowel of /a/.

Results revealed that the range of values for the sCPP-/a/ was 0 to 16.99 dB. The criterion for positivity for the sCPP-/a/ was 10 dB or lower, with values above 10 dB falling within the normal range. This criterion resulted in a sensitivity for the sCPP-/ a/ of 89% and a specificity of 77%. The positive predictive value of the CPPS -/a/ was 69%, and the negative predictive value was 80%. For the sCPP-s, the criterion for positivity was 5.0 dB or lower. The range of CPP-s values was 0.76 to 8.13 dB. All values above 5.0 dB were deemed to be within the normal range. By these criteria, the sensitivity of the CPP-s in detecting overall dysphonia was 87%, and the specificity was 90%. The corresponding positive predictive value was 81%, and the negative predictive value was 77%. Overall, the sensitivity of the sCPP-s and sCPP-/a/ were similar to the sensitivity of the conventional measures. Whereas when compared with predictive values of CPP and conventional measures, the CPP-s and CPP-/a/ were better measures. Finally, when compared CPP of connected speech with CPP of /a/ in comparison with conventional measures, the CPP-s has better sensitivity, specificity, and positive and negative predictive values. The above study has been carried out in a systematic way with including large samples of dysphonics to validate the cepstral measures in predicting the dysphonia and the study has included perceptual analysis by 10 number of judges which is a very good method to validate cepstral analysis. However, the study has certain limitations, the above study has given criteria for CPP without considering the normal population. If it had included normal participants as one more group for sensitivity and specificity measures, it would have more reliability. This study concludes that measurement of the cepstral peak prominence from samples of running speech provide the most reliable predictor and objective correlate of dysphonia of the available methods of objective voice measurements. (Heman-Ackah, 2002 & 2003).

Watts and Awan (2011) performed cepstral measurements on both continuous speech and phonation of vowel. Sixteen hypofunctional participants (mean age of 52 years; 11 females, 5 males) and 16 controls (mean age of 53 years; 11 females, 5 males) were asked to sustain /a/ and read the Rainbow Passage. Two speech-language pathology graduate students served as perceptual judges. The students identified the speakers' voice quality type 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 1-s 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 (CPPsd). Spectral measures included were L/H spectral ratio, and L/H spectral ratio standard deviation (L/H spectral ratio sd). Results showed that among the measures used in this study, CPP and L/H spectral ratio showed significant differences between groups in both speaking conditions. By demonstrating CPP and L/H spectral ratio as effective discriminatory measures of normal versus abnormal voice qualities, this study provides further evidence of the clinical value of cepstral/spectral-based measures.

Heman-Ackah, Sataloff, Laureyns, Lurie, Michael, and Heuer et al., (2014) did a study aiming at identifying cut- off values for normal voice that can accurately diagnose a dysphonic voice using Speech tool program for CPP and using perceptual measurements of dysphonia as the gold standard for comparison, and to define the sensitivity and specificity of sCPP in screening for dysphonia.

Voice samples from 835 patients with voice problem and 50 participants with normal voice were collected using analog tape recorder. All voice recordings were performed with the microphone positioned 6 inches from the mouth. The task used is reading “ Marvin Williams’’ passage. The samples were edited to consist only first sentence of passage for acoustic analysis.

For all the original recordings, perceptual analysis was carried out by eight laryngologists and four speech-language pathologists in a blinded fashion. Each rater was asked to rate each individual voice sample on the degree of dysphonia/normality using an analog scale in which one end of the scale represented ‘‘normal’’ and the other end represented ‘‘profound dysphonia’’. The analog scale was a single line that measured

100 mm, onto which the raters were instructed to place a tic mark at the point at which he/she felt the voice sample fell in the continuum between normal and abnormal. Objective analysis for all the samples were carried out using the smoothing algorithm of the CPP for running speech designed by Hillenbrand. The mean perceptual rating was then used as the gold standard against which values of CPP were compared. In this study, a voice was considered normal if the mean perceptual rating fell below the 10th percentile of the distribution of mean perceptual rating scores. Dysphonia was defined as the perceptual rating above the 90th percentile of the distribution of mean perceptual ratings. An ROC analysis was performed to determine the cut-off value for CPPS for positivity that has the highest sensitivity and specificity for discriminating between normal and dysphonic voice. The results of this study establish 4.0 as the cut-off value for normal for Hillenbrand CPPS smoothing algorithm for running speech, with values below 4.0 suggesting the presence of dysphonia. Using 4.0 as the cut-off value for normal, this version of CPP has a sensitivity of 92.4%, a specificity of 79%, a predictive value positive of 82.5% and a predictive value negative of 90.8%.

The advantage above study is that it has included both perceptual and acoustic analysis and has included large data with 12 members of raters to rate the samples. It has included both SLPs and laryngologist from different regions to rate the voce samples.

However, this has certain limitations. The study could have used phonation task along with connected speech. Rather using 100 mm scale to rate the sample,it could have included any standardized rating scales for perceptual analysis.

Studies using Spectral andCepstral measures

Awan and Dromey (2009) were able to identify spectral/cepstral measures that most effectively predicted dysphonia severity in pre- and post-treatment continuous voice recordings of female speakers. Pre- and post-treatment speech samples were selected from an archival database of patients with muscle tension dysphonia, with 104 female speakers chosen for analysis (mean of 46.4 years of age). Voice therapy for the patients consisted of a single extended session of manual laryngeal re posturing maneuvers and/or circum laryngeal massage, which stimulated an improved voice. The female speakers were asked to read the Rainbow Passage at a comfortable pitch and loudness. Afterwards the speech samples were edited to include only the 2nd and 3rd sentences.

All samples were analyzed for CPP, low/high (L/H) spectral ratio that the authors referred to as the DFT ratio (DFTR), and DFTR standard deviation (DFTR SD). Five master's degree students in communication disorders served as auditory-perceptual judges of the 104 speakers, with 208 total samples judged. Judges were asked to rate the continuous speech samples on a 100-point visual analogue scale. One end of the scale was labeled normal, and the opposite side was labeled profoundly abnormal, with higher numbers suggesting increased severity of dysphonia.

Step-wise linear regression analysis revealed a three-factor model consisting of CPP, DFTR SD, and DFTR, strongly correlating with perceived dysphonia severity (mean of R = .85; R2 = 73%). CPP was the strongest contributor to the three-factor predictive model, in addition to being the strongest individual correlate of listener perceived dysphonia severity (r = -.81; r2 = 66%). Paired t-tests were conducted to establish whether significant pre- versus post-treatment changes occurred in any of the three spectral/cepstral-based components (CPP, DFTR, and DFTR SD) of the predictive dysphonia severity model. Results indicated significant differences in all pre- versus post treatment comparisons, with significant increases in all variables following treatment.

An additional series of paired t-tests was performed to determine whether significant differences existed between pre- versus post-treatment mean perceived severity ratings and pre- versus post-treatment predicted severity ratings. In both instances, post-treatment mean perceived severity and post-treatment predicted values were significantly lower than pre-treatment observations. Last, treatment change scores were computed by subtracting post-treatment from pre-treatment ratings, showing a reduction in dysphonia severity. This study shows that strong predictions of listener perceived dysphonia severity can be made from L/H spectral ratio and cepstral measures.

The purpose of the study by Stranik, Cmejla, and Vokral (2014) was to objectively classify breathiness in continuous speech according to a subjective evaluation of voice based on the GRBAS scale. The database was recorded from the 1970s to the 1990s as a common element of the voice examination of healthy and pathologic voices. All records were performed using professional recording equipment and were recorded in a soundproof booth with a level of ambient noise lower than 18 dB SPL. The database was originally recorded on tapes and then digitalized with a sampling frequency 44.1 kHz and 16 bit resolution. A total of 593 records containing readings of a standard phonetically unbalanced text were selected from the database. The recordings were evaluated by means of the GRBAS scale by five experts from the Department of Phoniatrics. Each expert made two assessments of each recording with a delay of at least 2 weeks between assessments. During the subjective assessment, the records were identified by a random ID, and the IDs were different for the second assessment. The final grades according to the GRBAS scale were determined in two ways (a) using the modus from the final 10- element (five raters, two sessions) set of grades for discrete classification and (b) using the mean value from the final 10-element set of grades for continuous classification. Only the breathiness B from the whole GRBAS scale was taken into account for this experiment. The records were subsequently subjected to acoustic analysis using Cepstral Peak Prominence (CPP), Pearson r at autocorrelation peak (RPK), Breathiness Index (BRI), ratio of high- to mid/low-frequency energy (HLR), and the first harmonic amplitude. These parameters were subsequently analyzed and a total of 92 features were created for each record. After feature space reduction based on Correlation Feature Selection and Information Gain, the feature space was reduced to four parameters. These four parameters were used for classification of breathiness In the final set of four, the acoustic parameters have significantly different mean ranks in every grade of breathiness according to the GRBAS scale (Kruskal-Wallis test [$P < 0.001$]). The accuracy of classifier for objective evaluation of level of breathiness based on the discrete scale of breathiness reached 77%. Assuming continuous grades of breathiness, the classifier reached r ¼ 0.92 ($P < 0.001$). They concluded that the level of breathiness in continuous speech can be effectively described by automatic system–based analysis of acoustic measures. They also opined that the proposed automatic system was able to determine the level of breathiness in continuous speech with sufficient precision.

Gillespie, Dastolfo, Magid, and Gartner-Schmidt (2014) analyzed 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 patient records from January 2009 to July 2013 were included if records indicated the following inclusion criteria: age older than 18 years, primary diagnosis of benign mid membranous vocal fold lesion(s) , primary muscle tension dysphonia (MTD-1), vocal fold atrophy, or unilateral vocal fold paralysis (UVFP). Only patients with single-category diagnoses were included (ie, atrophy alone, not atrophy and UVFP). Diagnoses were determined via a team consisting of a fellowship-trained laryngologist and a voice-specialized SLP. Data were specifically chosen as pre and post intervention measures. Participants were asked to phonate vowel /a/ and read the sentence from the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) protocol at their most comfortable pitch and loudness and the recordings of these tasks were completed using the Analysis of Dysphonia in Speech and Voice (ADSV; KayPENTAX, Montvale, NJ) and Multi-Dimensional Voice Profile (MDVP) software from the Computerized Speech Lab (KayPENTAX, Montvale, NJ). ADSV was used to provide measures of the CPP in the sentence (CPP speech) and vowel (CPP vowel) and respective SDs, low-high spectral ratio in the sentence (L/H ratio speech) and vowel (L/H ratio vowel) and respective SDs. In addition, a multifactorial estimate of dysphonia severity, referred to as CSID was calculated for the all-voiced sentence. Auditory-perceptual evaluations were made during the clinical evaluation from the CAPE-V sentences using the zero to three Grade, Roughness, Breathiness, Asthenia and Strain scale.

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. These findings support the hypothesis that a one-size-fits-all approach to voice outcomes may not be appropriate. 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 were the most consistent indicators of change in response to treatment. The robust and significant changes in CSID provide an example whereby a multifactor formula appears to be more sensitive to change than individual measures. In addition, both measures are taken from connected speech, which indicates that ecologically valid measures—such as those taken during speech and not a single phoneme—may be 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. The improvement in L/H ratio in patients with UVFP indicates a more consistently stable voice postreatment than pretreatment and possible reduction in high-frequency spectral noise. Participants who improved in CSID also improved in VHI-10.

The authors concludes by saying that CPP- speech may be a worthwhile measure to regularly collect and analyze, especially on patients with pathologies more likely to be treated with surgery, such as large lesions not appropriate for voice therapy and UVFP and also recommends to analysis of CPP speech, CPP vowel, L/H ratio SD in speech, and CSID as a regular battery of analysis for patients with lesions. Assessment of both vowels and connected speech in frequency-based measuresand vowels in time-based measures may be appropriate for these patients.

Lowell, Colton, Kelley, & Mizia, (2013) usedcepstral and spectral measures to discriminate perceptual based voice quality in 28 dysphonic speakers and 14 normal speakers. Based on the perceptual analysis, further groups were devided into 14 breatiness 14 roughness and 14 normal speakers were included in this study. Cepstral and spectral analyses of the first and second sentences of the Rainbow passage were performed. Discriminant analysis determined the combination of variables that optimally differentiated the three voice quality type. Cepstral peak prominence (CPP) showed the greatest predictive contribution to dysphonia severity in the regression model. The discriminant analysis produced two discriminant functions that included both CPP and its standard deviation (CPP SD) as significant contributors($P < 0.001$), with an overall classification accuracy for the combined functions of 79%. Cepstral-based measures showed the highest capacity to discriminate voice quality types, with better classification accuracy for normal and dysphonic-breathy than for dysphonic-rough voices. The relative contributions ofmultiplecepstral/spectral acoustic measures in their ability todiscriminate voice qualityfunction.The classification results for the combined discriminant functions 1 and 2 were as follows: 100%, 78.6% and 57.1% respectively for normal, breathy and rough groups respectively and the overall classification rate when averagingall voice quality groups was 78.6%.The cepstral-based measures of CPP and CPP SD showed the greatest ability to differentiate normal, rough, and breathy voice qualities relative to the spectral-based measures.

On review of literature, it is observed that there are very few studies are carried out in comparing cepstral measures across dysphonia severity and there are no documents of CPP and sCPP in mild, moderate and severe group specifically. There are studies reporting correlation between perceptual evaluation and cepstral measures but there are no study to see whether cepstral parameters can classify perceptual based dysphonia severity. Hence, the present study was aimed to document CPP and sCPP across dysphonia severity and to verify whether these measures discriminate perceptually based dysphonia severity.

CHAPTER III

Method

Participants

The present study aimed at documenting the CPP and sCPP values in individuals with normal voice and dysphonia and to verify whether these parameters could differentiate the perceptual levels of dysphonia severity. For the purpose of the study a total of 88 participants in the age range of 20-40 years were recruited. The participants included 45 individuals with dysphonia (group I) and 43 age and gender matched controls (group II). Table1 indicates the details of average age and distribution of participants across the groups.

Table 1.

Details of average age and distribution of participants across groups

Group	Average age	Number of participants
Group I	31.5	45
Group II	27	43

Following are the details of the participants under two groups and the criterion followed for their inclusion.

Group I: group I consisted of 4 participants who were diagnosed with voice disorder by a team consisting of Speech language pathologist and an Otorhinolaryngologist experienced in dealing with clients with voice disorders. The battery of assessments used to arrive at diagnosis included perceptual evaluation, electroglottography and videostroboscopy. The final diagnosis following all the domains of assessment was based on the consensus across the team members. The vocal pathology of participants are described in table 2. Further, all the participants were native Kannada speakers in the age range of 20-40 years and their mean age was 31.5 years.

Table 2.

Distribution of vocal pathology in dysphonic group

Vocal pathology	**Number of participants**
Glottic chink with MTD	12
Dysphonia plicaventricularis	2
Polyp	7
Nodule	8
Puberphonia	2
Vocal cord paralysis	8
Functional dysphonia	3
Granulomatosis	1
Laryngeal papilloma	1
Spasmodic dysphonia	2

*Inclusion criteria for dysphonic group:*The participants who visited the voice clinic at the institute were selected for the study based on the following inclusion criteria. (1) Individuals with complaint of voice problems, (2) individuals who were diagnosed to have dysphonia (Hoarse/Harsh/Breathy) by team consisting of three Speech language pathologists and an Otorhinolaryngologist through clinical examination and vocal imaging. (3) Individuals without any associated speech problems such as resonant disorders due to velopharyngeal dysfunction or motor speech disorders. (4) Individuals with no endocrinal disturbances or with hearing loss.

Group II: This group consisted of 43 individuals with no complaints of voice problems and with Kannada as their native language. All the participants were in the age range of 20-40 years with an average age of 27 years. All the participants were randomly selected from in and around of Mysore city.

Inclusion criterion followed for selection of controls: Individuals with perceptually normal voice as judged by a Speech language pathologist were included for the study. It was ensured that the participants were free from sensory problems such as hearing loss, motor speech disorders such as dysarthria or apraxia of speech. 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. Participants without any history of chronic smoking and alcohol consumption,intubation, neurological disorders, systemic illness, and surgery /accident / trauma related to head and neck were included for the study. Further, to be included in the study the participants had to be free from upper respiratory tract infections or allergies on the day of testing.

Ethical issues: The participants were informed regarding the purpose of the study and procedures involved. An informed consent was obtained from all the participants before the initiation of recording.

Procedure

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 were instructed to phonate vowels /a/, /i/ and /u/ for five seconds each. Following this the participants were instructed to read Kannada voiced passage and spontaneous speech samples at their habitual pitch and loudness. For illiterate speakers under dysphonic group repetition of first three sentences of voiced passage was used instead of reading task. At end of the each task participants gave opinion on their voice as having soft, loud or habitual loudness and pitch. The task was repeated whenever the participant revealed their voice as either too loud or too soft or different from their habitual voice. All voice samples were directly recorded to Speech tool program using head mounted microphone (Logitech H110 Headphone Microphone) with 44 kHz sampling rate and with a constant mouth to microphone distance of five cm. Thus recorded samples were saved in the hard disk using save option in speech tool program for further perceptual analysis.

Data Analysis

Auditory-perceptual analysis

Tokens: The dysphonic voice samples were checked for duration and clarity. The tokens for perceptual analysis sample consisted of phonation samples of /a/, /i/ and /u/, reading of first three lines from Bengaluru passage and spontaneous speech. A total of 45 tokens were coded separately in order to blind fold the judges from participant's complaint, diagnosis and other details. This list of tokens were copied to three pendrives and were given to three judges to rate the samples independently.

*Perceptual tool:*The GRBAS scale (Japan Society and Logopedics and Phoniatrics, Hirano, 1981) was used in the study for rating the perceptual severity of the voice samples. GRBAS is a four point rating scale consists of voice based parameters; overall grade of severity, roughness, breathiness, astheny and strain. Each parameter is rated on four point rating scale ranging from 0 to 3 corresponding to normal, mild, moderate, extreme respectively.

*Perceptual analysis:*Three qualified Speech-Language Pathologists (SLPs) who had minimum of two years clinical experience in diagnosing voice quality disorders formed the judges. The samples were coded and given to the judges and they had freedom to listen as many times as required and were asked to rate both the phonation and connected speech sample of each participant individually and to give overall rating for the voice sample using GRBAS. Out of all the parameters rated using GRBAS, "Grade" was considered for the further analysis. Score sheets with printed GRBAS rating scale was prepared and handed over to the judges for documenting their ratings of each participant.

Acoustic analysis

Tokens: Phonation of vowel /a/ and reading or repetition of second sentence of the voiced Bengaluru passage were used for the acoustic analysis. The samples were edited to retain middle and stable portion of vowel /a/ for a duration of 5 sec and the second sentence (out of the three recorded sentences) of the voiced passage for acoustic analysis.

Instrument/software used for acoustic analysis: Speech tool program (Hillenbrand, Cleveland and Erickson, 1994) version 1.65 was used to analyze the Cepstralparameters that isCepstral Peak Prominence (CPP) and smoothened Cepstral Peak Prominence (sCPP) for both phonation and reading tasks. This program is a freely downloadable software which is available from the site http;//homepages.wmich.edu/_hillenbr/. The trimmed samples of phonation of /a/ for 5 sec and second line of voice passage were retrieved from the PRAAT software (Version 5.2.36) and were opened using Speech tool program and the CPP and sCPP were calculated for each sample. By clicking analysis icon, the Speech tool program automatically calculates CPP, sCPP and meanfo values.

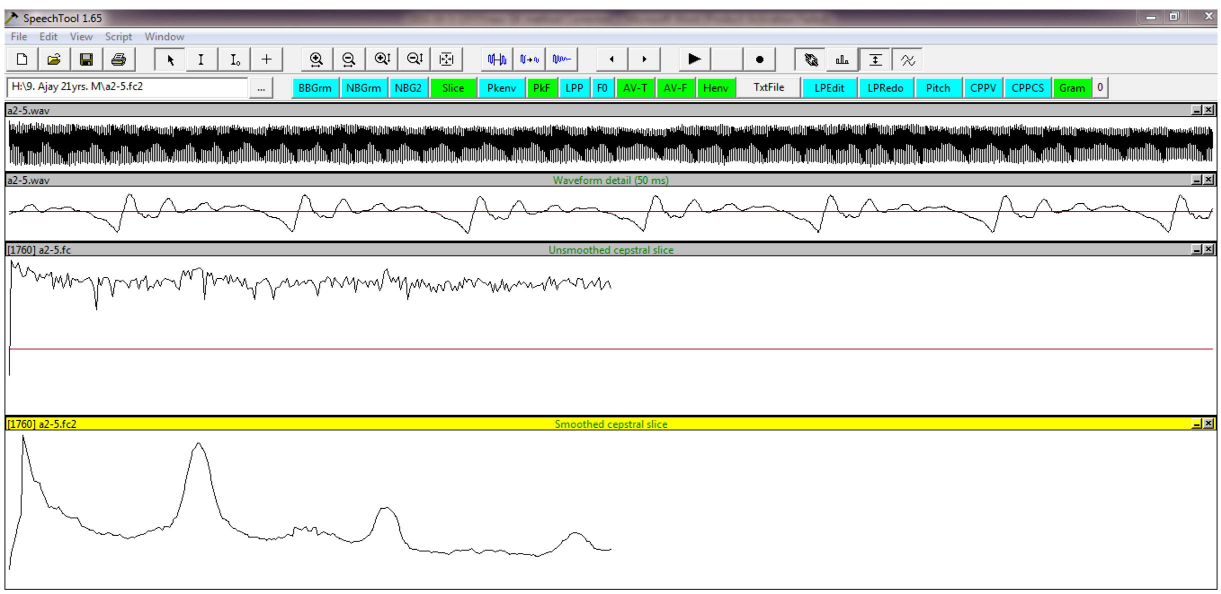

*Figure 4.*Illustration of unsmoothed and smoothened Cepstrum for the phonation of vowel /a/ in speech tool software

.

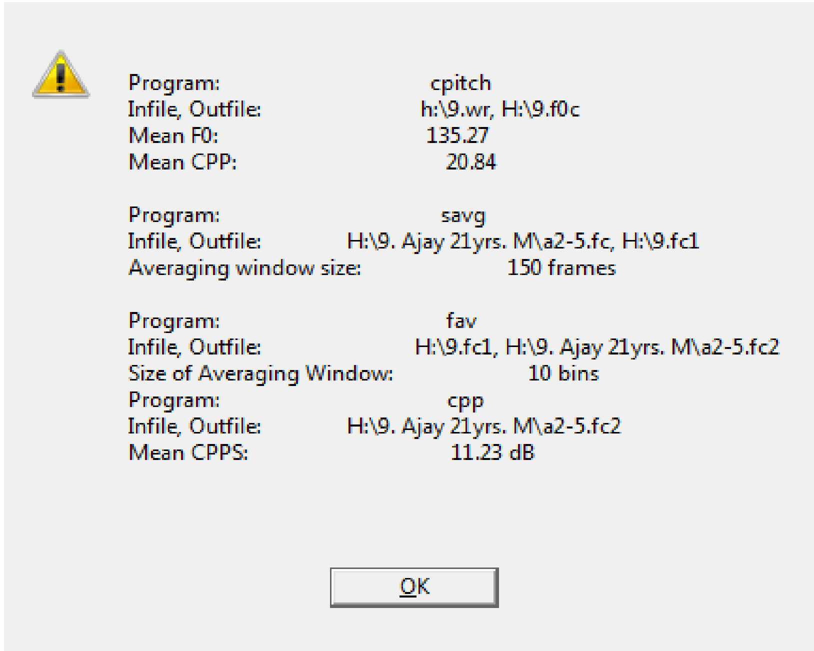

*Figure5***.**Illustration of obtained CPP and sCPP values for the phonation of vowel /a/ in Speech tool software.

Statistical analysis

The following statistical measures were performed using Statistical Package for Social Sciences (SPSS) software 20.0 version on the data to meet objectives. Kappa test was performed to measure the agreement across the judges and reliability of perceptual ratings. Descriptive and dispersion statistic measures were performed to obtain the mean, standard deviation, median in normal, mild , moderate and severe groups for CPP-P, CPP-R, sCPP-P and sCPP-R . One way MANOVA, Kruskal Wallis and Mann Whitney-u tests were performed to compare across severity and Discriminant function analysis was carried out to see whether Cepstral measures discriminated perceptual based dysphonia severity.

CHAPTER IV

Results

The present study was conducted with the aim of documenting Cepstral Peak Prominence across normal and different levels of perceptually based dysphonia severity and to see whether cepstral measures can differentiate perceptually based severity.To achieve this, speech data was obtained from 88 participants, in which 45 were dysphonic and 43 were controls in the age range of 20-40 years. The voice samples were collected individually for both phonation and connected speech task.

The results of the study will be presented and discussed under the following headings.

I. Perceptual evaluation of voice
II. CPP and sCPP in Normal and dysphonic group.
III. Range of CPP and sCPP across the groups.
IV. Comparison of CPP and sCPP across Normal and Dysphonic groups.
V. Discriminant function analysis to see whether CPP can discriminate perceptual based dysphonia severity.

Perceptual evaluation of voice

The voice samples were rated by the judges based on four point rating scales. These ratings were subjected to reliability measures to evaluate the agreement between the judges using Kappa coefficient. The measures of agreement between judges were carried out in pairs Judge1-Judge2, Judge2-Judge3 and Judge3-Judge1. Kappa coefficient was calculated for all the pairs. The results showed that Kappa coefficient is closer to 0.5 for all the three pairs and $p<0.05$showing significant agreement between them. The group membership of the participants was assigned based on the consensus obtained in the judgment by two or more judges. Finally all the participants with dysphonia were divided in to three groups based on severity. Among 45 dysphonics 13 were rated as mild dysphonia, 22 were rated as moderate dysphonia and 10 were rated as severe dysphonia.

CPP and sCPP in Normal and Dysphonic group

Table 1 and 2 depicts mean, SD, and median values of CPP and sCPP across the groups derived using descriptive statistics. There were differences in mean and SD values of all parameters across the groups. The mean values of CPP for phonation and reading task indicated higher values in normals than dysphonics. There was a reduction in CPP with the increase in severity of dysphonia. Table 2 depicts the sCPP measures across the normal and dysphonic groups. The measure of sCPP also indicated increased values in normal group followed by dsyphonics. Similar to the measures of CPP, with the increase in severity of dysphonia sCPP values reduced. Both sCPP and CPP measures for phonation were higher than reading across all the groups except in the severe dysphonic group. Figure 1 and 2 depicts graphical representation of CPP and sCPP across normal, mild, moderate and severe groups for phonation of vowel and reading task.

Table 3.

Mean, standard deviation (SD) and Median of CPP for phonation and reading across the four groups

Groups	Subjects	CPP-Phonation			CPP-Reading		
	N	Mean	SD	Median	Mean	SD	Median
Normal	43	18.21	2.52	18.21	14.59	2.12	14.82
Mild	13	16.29	2.47	17.04	13.98	1.50	13.99
Moderate	22	13.60	2.34	12.86	12.98	1.91	13.11
Severe	10	10.03	0.84	9.88	10.63	1.1	10.30

Note. SD – Standard Deviation; N – Number of subjects; CPP – Cepstral peak prominence.

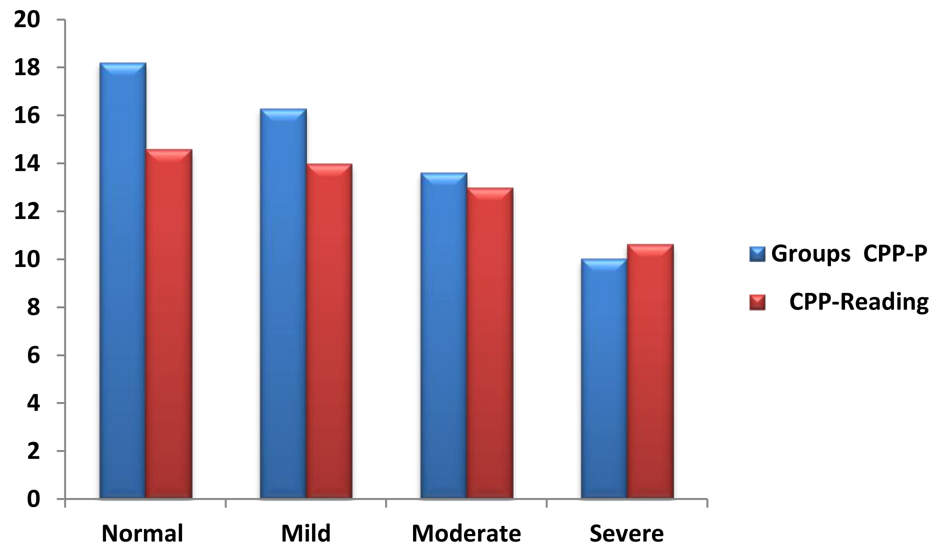

*Figure 5.*CPP for phonation and reading across groups

.

Table 4.

Mean, standard deviation (SD) and Median of sCPP for phonation and reading across the four groups

Groups	Subjects	sCPP-Phonation			sCPP-Reading		
	N	Mean	SD	Median	Mean	SD	Median
Normal	43	8.72	1.51	8.43	6.08	1.11	6.23
Mild	13	7.95	1.54	8.07	5.55	0.86	5.55
Moderate	22	5.83	2.01	6.08	4.96	1.46	4.99
Severe	10	1.17	1.50	0.73	2.16	0.54	2.01

Note. SD – Standard Deviation; N – Number of subjects; sCPP – Smoothened Cepstral peak prominence.

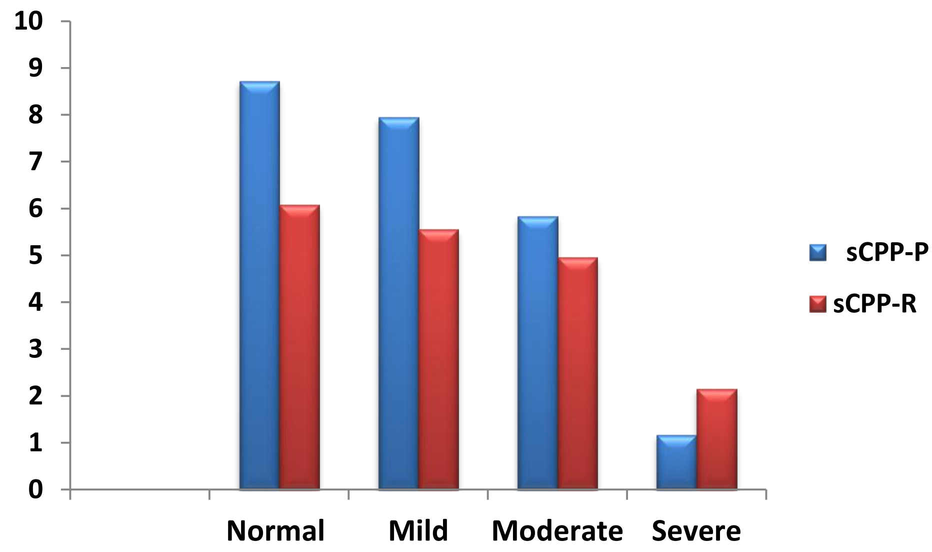

Figure 6: sCPP for phonation and reading across four groups

Range of CPP and sCPP across the groups

Using descriptive statistics at 95% confidence interval of mean, lower limit and upper limit of CPP and sCPP measures were estimated across the groups. Table 5 indicates that Control group had higher upper and lower limits for CPP and sCPP across the contexts than dysphonics. Whereas, both the groups exhibited narrow range (upper limit–lower limit) of CPP and sCPP across the contexts. Even though, both the groups have a limited range, the participants in dysphonic group exhibited relatively slightly wider range than controls across the variables except CPP-R. .

Table5.

Range of normal and abnormal values of CPP and sCPP

	Normative range		Dysphonic range	
	Lower limit (dB)	Upper limit (dB)	Lower limit (dB)	Upper limit (dB)
CPP-P	17.43	18.97	12.66	14.51
CPP-R	13.94	15.24	12.14	13.36
sCPP-P	8.26	9.18	4.50	6.32
sCPP-R	5.73	6.42	3.99	5.02

Note. CPP-P: Cepstral peak prominence for phonation, CPP-R: Cepstral peak prominence for reading, sCPP-P: Smoothened cepstral peak prominence for phonation, sCPP-R: Smoothened cepstral peak prominence for reading.

Comparison of CPP and sCPP across the groups.

The CPP and sCPP were compared across the groups using non parametric (Kruskal Wallis) test. The non parametric statistics were used as the sample size between groups are not similar and SD is more in sCPP-P than mean, even though all the parameters followed normal distribution. Results indicated significant differences in CPP and sCPP measures across the groups as shown in Table 6.

Table 6.

Kruskal Wallis test results for CPP and sCPP measures across the contexts and groups

	χ^2	df	Sig
CPP-P	50.283*	3	.000
CPP-R	29.525*	3	.000
sCPP-P	47.900*	3	.000
sCPP-R	34.612*	3	.000

Note. *Significant difference at $p<0.05$, χ^2 – Chi square, **df**– Degrees of freedom, CPP-P: Cepstral peak prominence for phonation, CPP-R: Cepstral peak prominence for reading, sCPP-P: Smoothened cepstral peak prominence for phonation, sCPP-R: Smoothened cepstral peak prominence for reading.

To see which all two groups significantly different from other for all four variables, Mann Whitney–U test was performed. Results revealed that there was significant difference ($p>0.05$) across all groups for all variables except in normal versus mild group for CPP-P [/Z/= 2.183; $p>0.05$], CPP-R [/Z/= 1.194; $p>0.05$], sCPP-P

[/Z/= 1.572; *p*>0.05], sCPP-R [/Z/= 1.912; *p*>0.05], and Mild versus moderate group for CPP-R [Z= 1.912; *p*>0.05], sCPP-R [/Z/= 1.280; *p*>0.05]. The results obtained from Mann Whitney-U test for all four variables CPP-P, CPP-R, sCPP-P and sCPP-R comparing all the groups are depicted in table 7.

The Results obtained from non-parametric were cross checked for three variables CPP-P, CPP-R and for sCPP-R across four groups using one way MANOVA. It revealed similar results as obtained that of Non parametric tests.

Table 7.

Results of Mann Whitney test for all four variables

	CPP-P	CPP-R	sCPP-P	sCPP-R
Normal Vs Mild	2.18	1.19	1.57	1.91
Normal Vs Moderate	5.38*	3.19*	5.14*	3.24*
Normal Vs Severe	4.89*	4.47*	4.89*	4.80*
Mild Vs Moderate	2.94*	1.91	3.11*	1.28
Mild Vs Severe	4.03*	3.85*	4.03*	4.03*
Moderate Vs Severe	4.07*	3.33*	4.15*	4.23*

Note. * Significant difference across the groups$p < 0.05$, CPP-P: Cepstral peak prominence for phonation, CPP-R: Cepstral peak prominence for reading, sCPP-P: Smoothened cepstral peak prominence for phonation, sCPP-R: Smoothened cepstral peak prominence for reading.

Discriminant function analysis

In the present study discriminant analysis was used to determine whether CPP and sCPP measures during phonation and reading can predict the groups with varying levels of perceptual based dysphonia severity. The data under all the groups was subjected to discriminant function analysis. As there were four groups, discriminant analysis provided three discriminant functions.

Wilk's Lambda was used to test the significant differences between the groups on individual predictors and their functions. For discriminant function one (DF1) Wilk'sLambaλ was highly significant at 0.199, $\chi^2(12, N=88) = 134.027$, $p<0.001$. Discriminant function two (DF2) Wilk'sLambaλ is significant at 0.761, $\chi^2(6, N=88) = 22.624$, $p<0.001$. However, discriminant function three (DF3) Wilk'sLambaλ was not found to be statistically significant 0.969, χ^2 (2, N=88)= 2.649, $p > 0.001$ (Table 9). As both DF1 and DF2 are significant, hence these can be chosen to discriminate the groups. The percentage of variance in DF1 and DF2 are accounted for 90.3% and 8.7% respectively (Table 8).

Table 8.

Eigenvalues for all four groups

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.827[a]	90.3	90.3	.859
2	.272[a]	8.7	99.0	.462
3	.032[a]	1.0	100.0	.177

Note. [a] First 3 canonical discriminant function were used in the analysis.

Table 9.

Wilks Lambda values

Test of Function(s)	Wilks' Lambda	χ^2	Df	Sig.
1 through 3	.199	134.027	12	.000
2 through 3	.761	22.624	6	.001
3	.969	2.649	2	.266

Note. * Significant difference at $P < 0.05$

Table 10.

Standardized Canonical discriminant function coefficient

	Function		
	1	2	3
CPP-P	-.018	1.578	-.966
CPP-R	-.549	.188	1.811
sCPP-P	.736	-.892	1.336
sCPP-R	.899	-.594	-1.912

Table 11.

Structure Matrix of variables obtained using step wise discriminant analysis

	Function		
	1	2	3
CPP-P	.884*	.212	.375
CPP-R	.647*	-.067	-.293
sCPP-P	.401*	.151	.020
sCPP-R	.695	.710*	.099

On standardized discriminant function coefficient in groups DF1 was heavily weighted on CPP-P, CPP-R, sCPP-P and DF2 was found to be heavily weighted on sCPP-R (table 11). In order to interpret the first discriminant function DF1, standardized discriminant function coefficients were considered. The table 12 depicts the group centroids for normal, mild, moderate and severe dysphonic groups as 1.1, .53, -.57 and -4.17 on DF1 respectively. The centroids based on DF2 are .35, -.31, -.77 and .58 for control, mild, moderate and severe groups respectively.

Table 12.

Functions at Group Centroids

	Function		
Groups	1	2	3
Normal	1.103	.352	-.061
Mild	.534	-.312	.405
Moderate	-.574	-.766	-.138
Severe	-4.174	.580	.040

Unstandardized canonical discriminant functions evaluated at group means

Table 13.

Discriminant functions predicting group membership

			Classification Results				
			Predicted Group Membership				
		Groups	Normal	Mild	Moderate	Severe	Total
Original	Count	Normal	28	12	3	0	43
		Mild	5	5	3	0	13
		Moderate	3	2	16	1	22
		Severe	0	0	1	9	10
	%	Normal	65.1	27.9	7.0	.0	100.0
		Mild	38.5	38.5	23.1	.0	100.0
		Moderate	13.6	9.1	72.7	4.5	100.0
		Severe	.0	.0	10.0	90.0	100.0

Note. 65.9% of original grouped cases correctly classified

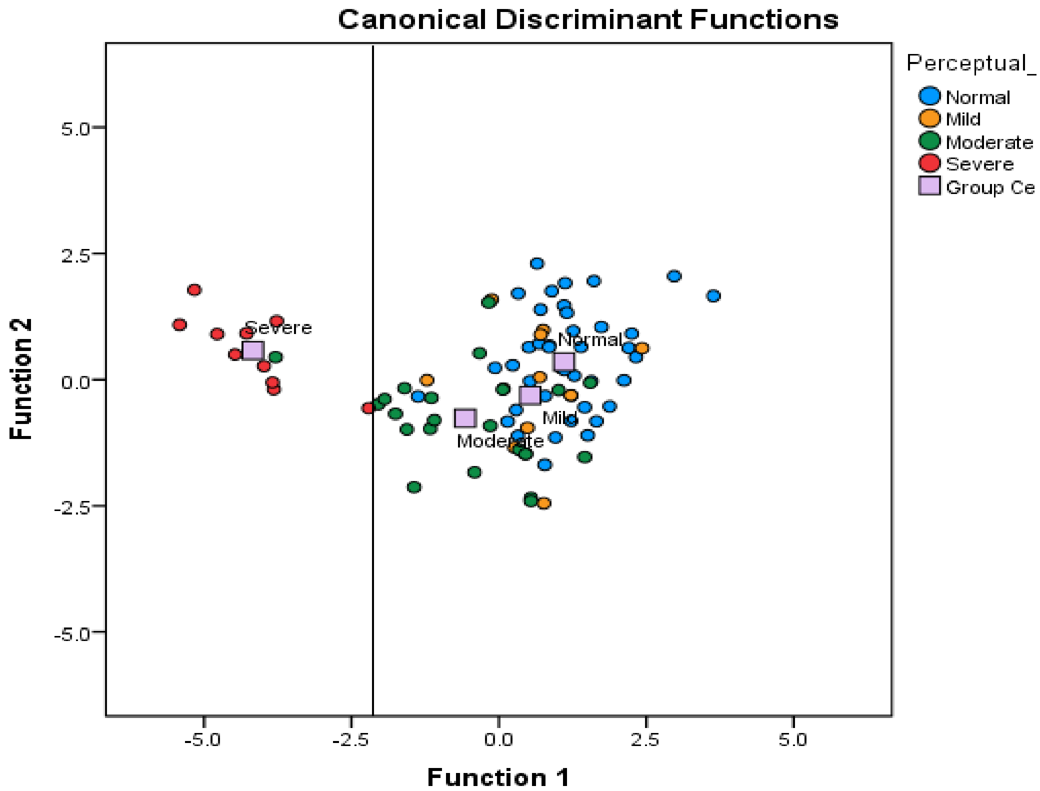

Figure 8: Combined group plot for canonical discriminant functions in groups

The cutoff score is the one that results in the fewest classification errors. For the unequal group sizes the cutoff score is calculated from the weighted means of the centroids (Meyers et al. 2006). The cutoff score of DF1 to differentiate severe group and other group is -2.37.This indicates the index score below -2.37 indicates severe dysphonic group and above indicates the other three group. Under functions at group centroids, group means on each of the discriminant functions are presented. Based on DF1 severe group was discriminated significantly from other three groups, whereas there is some overlap of data points as shown in the Figure 8 for normal, mild and moderate groups.

Classification results based on discriminant functions revealed that overall 65.9% of the participation were correctly classified (Table 13). Specific to individual groups, 65.1% of participants in normal group and 38.1% of participants in mild group and 72.5 % of participants in moderate group and 90% of participants in severe group were correctly classified and predicted group membership based on the Cepstral measures.

CHAPTER V

Discussion

The present study was conducted with the aim of documenting Cepstral Peak Prominence and smoothened Cepstral Peak Prominence across normal group, mild, moderate and severe dysphonic group and to verify whether cepstral measures could differentiate these levels of perceptually based dysphonia severity. To achieve this voice samples were obtained from a total of 88 participants including 45 individuals with dysphonia and 43 phononormic individuals in the age range of 20-40 years. The voice samples was collected individually for both phonation and connected speech task. Appropriate statistical analysis was carried out. The results obtained are discussed under the following headings.

Perceptual evaluation of voice

Perceptual analysis of dysphonic voice was evaluated by three judges using GRBAS rating scale separately. Inter-rater reliability between three judges was carried out using the Kappa measures of agreement. Kappa coefficient was found to be 0.5, which showed moderate reliability.

There are many scales for perceptual evaluation of voice mentioned in the literature. Some of the standard protocol used for perceptual analysis of voice include (Wilson, 1987), The vocal profile analysis protocol (Laver, 1980), Buffalo II voice profile (Wilson, 1987), CAPE-V (ASHA, 2002) and GRBAS (Hirano, 1981). Among all the other perceptual scales, the GRBAS scale was reported to be widely used for judging disordered voice quality (Carding, Wilson, Mackenzie & Deary, 2009). It gives more objectivity regardless of the type of speech sample (Nemr, Zenari, Codeiro, Tsuji, Ogawa, Ubrig & Menezes, 2012) and it was reported to be reliable, easy, valid and offers no discomfort or inconvenience to the judge. It also has strong correlation with acoustic measurements for dysphonic population (Dejonckere, Remacle, Elbaz, Woisard, Buchman & Millet, 1996). So in the current study GRBAS rating scale has been used.

There are few studies in literature that have showed good inter judge relaibilty. One among such study is the study carried out by Dejonckere, obbens, de Moor & Wieneke (1993) reported correlation between judges as (0.7) for 10 judges. Nemr et al (2012) has reported strong correlation between the intra-judge consensus analysis for three judges.

The inter-judge reliability is moderate in the study. It can be attributed to various factors such as the judges in the current study had minimum of 2 years of clinical experience. If the study had included more experienced listeners, the ratings could have been consistent and the results could have been correlated with each other. And the other factor being the samples included in the study were phonation of /a/, /i/ and /u /along with connected speech sample, this included spontaneous speech and reading. In a study carried out by Bele, (2004). The reliability of connected speech was higher compared to phonation task. So in the current study we can expect variations with respect to overall perceptual rating by judges. Few authors might have only considered phonation task as the main to consider for perceptual rating and few might have considered connected speech. So the study recommends considering separate ratings for phonation and connected speech task.

CPP and sCPP in Normal and Dysphonic group

The results indicated that mean values of CPP and sCPP in normal voice are higher during both phonation and reading compared to dysphonic group. Among the participants in dysphonic group, there was reduction in Cepstral peak prominence with the increase in severity of dysphonia. The reduced CPP and sCPP with the increase in severity of dysphonia can be attributed to vocal pathologies exhibited by the participants in the group of dysphonics. The voice quality gets affected due to vocal pathologies. When there is dysphonia, the pattern of vocal fold movements and the vibratory nature of cords get altered. This can lead to aperiodicity in vocal fold vibration resulting in increased noise component at the level of glottis. The presence of noise in the signal can reduce the harmonic organization in acoustic signal. As, CPP is the measure of the degree of harmonic organization, so the presence of noise in the voice signal can reduce harmonic organization in acoustic signal and due to this there will be reduction of CPP values. The more the dysphonia severity, lesser/ flat will be the CPP. Similarly, the voice which has good quality will have higher CPP values. These results are in consonance with study done by Balaubamanya et al., 2009; Balasubarmanya et al., 2010; Lowell et al., 2011; Brinca et al., 2014.

Balasubramanya et al (2010) investigated CPP values in vocal nodule group and control group in the age range of 20-40 yrs. The results revealed that mean of CPP for males and females are 1.72 and 1.68 respectively in control group. The CPP values of males and females in dysphonic groups are 1.38 and 1.34 respectively. The study concluded that reduced CPP measures in dysphonics can be due to the presence of vocal fold pathologies. As dysphonic group exhibited glottic chink secondary to vocal nodules, this can result in huge volumes of air escape during voicing lead to increased noise levels and reduced CPP. In a similar study Balasubramanya et al. (2011) also reported mean lower CPP values in participants with vocal cord paralysis than compared to control group. The CPP of control group was 1.77 for males and 1.70 for females which was higher than the dysphonics exhibiting VC palsy (1.45 for males and 1.33 for females). The mean values of CPP in the current study were higher than these two studies . These differences could be due to the methodological variations as the study by Balasubramanya et al (2010) and Balasubramanya et al (2011) measured CPP by using Computerized Speech Lab (CSL) and the current study used speech tool software. The CPP measures in CSL are based on manual analysis whereas speech stool software consists of logarithmic operations to calculate the CPP. Therefore the dissimilarities in signal processing between the two measures could have lead to differences in the obtained Cepstral measures.. Despite these methodological variations, similar trend of lower CPP in dysphonic compared to normal voice was observed in the current study. Similar results were reported by Brinca et al. (2014) indicating lower CPP-P and sCPP-P in dysphonic group compared to control group.

Lowell et al., (2011) compared CPP and sCPP among 27 dysphonics and 27 controls. The results revealed that cepstral measures strongly differentiated dysphonic from normal speakers. CPP and sCPP were lower for dysphonic group compared to normal group and showed moderate to strong relationships with overall voice severity. There was negative correlation of overall voice severity and CPP or sCPP, i.e. decrease in CPP and sCPP amplitude was observed with severity of dysphonia. . The similar finding has been observed in the current study with lowest cepstral values in severe dysphonia, a highest value for normal with rising trend in between. similarlyHeman-achah (2002) analyzed pre and post operativeCepstral measures from 19 individuals with Recurrent Laryngeal Nerve Palsy paralysis. The investigators reported that sCPP- R and CPP-P for /i/ correlated inversely and strongly with perceptions of overall dysphonia.

The current study also considered measurement of CPP in two contexts those are phonation and reading. The results indicated higher CPP and sCPP in phonation than reading context. This could be attributed to increased vocal stability during vowel phonation compared to reading task. The inflections in vocal tone during reading task would have lead to the inconsistency in vocal fold vibration, in turn to phonatory instability. This phonatory instability while reading could affect the periodicity, harmonic organization and could have leaded to reduced CPP during reading task compared to phonation. Sustaining a sound without instability is difficult for individuals with dysphonia than in controls and this is more challenging during continuous phonation than reading. Reading requires wide variation in the phonatory mechanism, and hence even the individuals with dysphonia can produce voiceless sounds at least some point of time. These variations and presence of variety of phonemic contexts would in turn might make the reading task less challenging even for severe dysphonics in terms of maintaining vocal stability, thus leading to better CPP values with reading than in phonation task.

Range of CPP and sCPP across the groups

The results of the current study indicated increased upper and lower limits of CPP and sCPP for control group than dysphonics. The abnormal range of sCPP for phonation and reading were 3.99- 5.02 and 4.50- 6.32 respectively. These results are consistent with the studies done by Heman-Ackah et al. (2003), the abnormal range of sCPP were within 0.76 - 8.13 and above 5.0 were indicated for normal group.

In another study by Heman-Ackah et al., (2014) reported that sCPP for connected speech in normal range starts at 4.0 and above. If the sCPP value is less than 4 then the voice is considered as dysphonic or abnormal. There are wide variations in the range of CPP and sCPP values mentioned across the studies. These variations can be attributed to differences in methodology used in studies with respect to participants characteristics in terms of the vocal pathologirs considered, the sample size, instrument used to measure CPP, stimuli used analysis.

Comparison of CPP across Severity of dysphonia

In the present study CPP-P, CPP-R, sCPP-P, sCPP-R were significantly different across normal, mild, moderate and severe groups. The similar finding is reported by Brinca et al., 2014, where they reported significant difference between normal and dysphonic groups with CPP-P, sCPP-P CPP-R measures.. However, the the authors reported contradicting finding to the current studyas no significant difference between the groups with parameter sCPP-R.

The results obtained from Mann Whitney - U test revealed that all four groups were significantly different across four variables, except in Normal vs mild for all four parameters and mild vs moderate for sCPP-R and sCPP-P. This can be explained with respect to the overlapping of values across mild and normal groups. As mentioned in literature (Heman-ackha et al., 2014), subtle changes in voice quality are difficult to identify, this may have lead to inclusion false negative in mild group or false positive in normal group which has caused overlap of values.

Further, the results indicated effective differentiation between normal vs severe, normal vs moderate and moderate vs severe with cepstral measures in . It is because the severity of dysphonia leads to lesser values of CPP. and hence more significant difference across the groups. This finding is in coherence with the study by Watts and Awan (2011) that revealed significant difference between normal and hypofunctional dysphonic group for CPP with both phonation and reading tasks..

Discriminant function analysis

The discriminant analysis tried to predict group membership and classified correctly 65.9% of participants of the study. Specific to individual groups, 65.1%, 38.1%, 72.5 % and 90% correctly classified in control, mild, moderate and severe dysphonic groups respectively. The severe group was differentiated from other groups. 90% of the participants were classified as severe dysphonic similar to perceptual analysis. The other groups are not differentiated based on the cepstral parameters. There is a need to include other acoustic variables such as spectral measures to differentiate the groups. The results are supporting that Cepstral measures are more useful for differentiating severe dysphonia from other groups (Hillenbrand, 1994).

There are few studies in literature supporting, that the CPP can classify vocal quality types along with other acoustic measures (Wolfe & Martin, 1997, Awan et al., 2005; Lowell et al., 2013). However, the current study is first of its sort in attempting to differentiate across the levels of perceptual dysphonia severity using Cepstral measures. Wolfe and Martin (1997) explored the ability of various acoustic measures to classify dysphonic patients. Using discriminant function analysis, 45 dysphonic participants were classified with 92% accuracy into breathy, hoarse, and strained voice types using a four-parameter model consisting of jitter standard deviation, fundamental frequency, SNR standard deviation, and cepstral peak prominence (CPP). Under similar lines Awan and Roy (2005) had explored the ability of Cepstral/ spectral measures and traditional analysis measures to classify 140 individuals including both normal, rough, hoarseness and breathiness voice quality types using discriminant function analysis. The five variable model classified voice type with 79.9% accuracy. Among all five variables, Cepstral based measures consistently emerged as a significant factor in predicting voice type.

Lowell et al., (2013) used Cepstral and spectral measures to discriminate perceptual based voice quality in 28 dysphonic speakers and 14 normal speakers. Based on the perceptual analysis, the participants were further divided into breathiness (14), roughness (14) and normal speakers (14). On discriminant analysis, predicted membership of groups revealed that100%, 78.6% and 57.1% for normal, breathy and rough dysphonic were classified respectively. The overall classification rate when averaging all voice quality groups was 78.6%. The cepstral-based measures of CPP and standard deviation CPP showed the greatest ability to differentiate normal, rough, and breathy voice qualities relative to the spectral-based measures.

To summarize, the results of the present study indicates Cepstral measures CPP and sCPP as appropriate for analysis of voice and even the voices with severe dysphonia. These measures were also effective in differentiating across the perceptual levels of dysphonia severity, however with varying degree of efficiency. The consistent trend of decrease in CPP amplitude with increase in dysphonia severity was observed. However, the differences between severe versus normal, mild and moderate groups were more significant compared to normal versus mild, normal versus moderate and mild versus moderate groups. This may be due to the uniparametric analysis with only using Cepstral measures which are consistenly reported to be sensitive for breathy voices. Therefore, future studies may focus on verifying the efficacy of multiparametric assessments in correlating with perceptual dysphonia and differentiating across the levels of perceptual dysphonia severity.

CHAPTER VI

Summary and Conclusion

Cepstrum is a measure of the degree of harmonic organization, which indicates how far the cepstral peak emanates from the cepstral "background noise". Cepstral based measures are based on peak to peak calculation and not on accurate calculation of fundamental frequency and are independent of cycle boundary identification, hence suitable for analysis of severely dysphonic and aperiodic voices. CPP and the smoothed CPP are two cepstral measures which are found to be the best predictors of dysphonia severity, have good correlation with perceptual analysis of voice. These parameters were also reported to be valid outcome measures to track therapeutic/surgical effects. Considering the robustness, varied clinical applications, good correlation with perceptual analysis of voice, it was hypothesized in the present study that the cepstral measures CPP and sCPP could differentiate across the levels of perceptual dysphonia severity.

To accomplish the objectives of the study 45 individuals diagnosed with dysphonia by a team of speech language pathologist and Otorhinolaryngologist and 43 age and gender matched controls were recruited for the study. Voice samples of phonation of vowel /a/ and reading first three lines from "Bengaluru" passage and spontaneous speech were recorded at their habitual pitch and loudness. For illiterate speakers under dysphonic group repetition of first three sentences of voiced passage was used instead of reading task. The samples from dysphonic group were given to three trained SLP's to rate the sample using GRBAS scale and for further analysis only

“Grade” was considered. The obtained rating was subjected to statistical measure of agreement; results indicated that there was moderate agreement among judges. Based, on this the samples from dysphonic group were further grouped under mild, moderate and severe category. The steady state of phonated samples of /a/ for five seconds and reading samples of second line from all the groups including control group were subjected to acoustic analysis using *Speech Tool program* (Hillenbrand, 1994). The acoustic parameters considered in the study were CPP-P, CPP-R, sCPP-P and sCPP-R.

The obtained acoustic data was subjected to statistical analysis. Cepstral parameters for reading and phonation were documented across the groups using descriptive statistical measures. There was overall decrease in Cepstral measures with increase in dysphonia severity. The decrease in CPP for dysphonic group can be attributed to following reasons majorly, the varied vocal pathologies in dysphonic group, which causes irregular vocal fold vibration and there will be aperiodicity and there will be increase in noise component more than signal, which leads to decrease in overall CPP. Similar findings were reported in the earlier studies and attributed this finding to presence of vocal fold pathologies in dysphonic group such as vocal nodule and paralysis. Due to this dysphonic group exhibited glottic chink can result in huge volumes of air escape during voicing leading to increased noise levels and reduced CPP(Radish Kumar et al., 2010 & Balasubramanium et al., 2011). Wolfe and Martin (1974) attributed lower values of CPP in dysphonic group is to breathy, hoarse and strained quality of voice which dysphonic grop compared to normal group.

To compare severity across four cepstral parameters, Kruskall-Wallis test was performed. There was overall significance difference across groups for all four parameters with $p<0.05$. To find out which all groups were significantly different, Mann Whitney–u pairwise test was performed. The results revealed significant difference across normal vs moderate, normal vs severe, mild vs severe and severe vs moderate with CPP-P, CPP-R, sCPP-P, sCPP-R and for mild vs moderate groups significant difference was obtained for CPP-P and sCPP-P with $p< 0.05$. Normal vs mild with all the parameters and mild vs moderate for CPP- R and sCPP- R did not reveal significant difference.

Further, to verify whether cepstral measures could differentiate perceptually based dysphonia severity step wise canonical discriminant function analysis was carried out. Results showed that overall 69.9% of the data was correctly classified based on four Cepstral measures. The severe dysphonic group was 90% correctly classified as severe dysphonia with centroid placing at the far end of other groups. Centroids of normal, mild and moderate groups were' not correctly classified. Thus indicating that the Cepstra lmeasure are effective in discriminating severe groups from normal, mild, moderate groups.

Overall findings from the present study indicate that organic manifestation of voice disorder can cause abnormally low values for cepstral parameters compared to normal. An inversely proportional relation was observed between severity of dysphonia and amplitude of CPP. Also results revealed that cepstral measures could differentiate groups based on the severity and overall discrimination was better for severe dysphonic group compared to other groups. Based on this, study recommends using Cepstral measures as a routine acoustic measure for evaluation of voices with severe dysphonia.

Implications of the study: The study documented CPP values of normal voice, mild, moderate and severe dysphonic voices of individuals in the age range of 20-40 years. Results of the present study indicated that the cepstral measures CPP and sCPP are effective in discriminating normal from dysphonic voice. This is especially true with severely dysphonic voice which is often impossible to assess using time domain measures such as perturbations.

Limitations and future directions of the study

The present study considered limited number of participants with varying number under each subgroup of perceptual dysphonia severity (a total of 45 participants under clinical group with 13, 22, and 10 under mild, moderate, and severe categories respectively). This limits the generalization of the results and limits the applicability. Therefore further studies considering large number of participants under each degree of perceptual dysphonia may be conducted. Further, future studies considering larger number of participants under specific voice pathologies may be conducted in order to verify whether the accuracy of these measures is maintained across the conditions as each pathological condition characterizes peculiar features. The present study investigated the effectiveness of cepstral measures in isolation, however, the multivariate measures including spectral, time domain and cepstral measures might reveal better discriminating ability across the groups of perceptual dysphonia severity and needs further investigations.

References

Alpan, a., Schoentgen, J., Maryn, Y., Grenez, F., & Murphy, P. (2012). Assessment of disordered voice via the first rahmonic. *Speech Communication*, *54*(5), 655–663. http://doi.org/10.1016/j.specom.2011.04.001

Awan, S. N., & Roy, N. (2005). Acoustic prediction of voice type in women with functional dysphonia. *Journal of Voice*, *19*(2), 268–282. http://doi.org/10.1016/j.jvoice.2004.03.005

Awan, S. N., Roy, N., Jetté, 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. http://doi.org/10.3109/02699206.2010.492446

Awan, S. N., Solomon, N. P., Helou, L. B., & Stojadinovic, A. (2013). Spectral-cepstral estimation of dysphonia severity: external validation. *The Annals of Otology, Rhinology, and Laryngology*, *122*(1), 40–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23472315

Awan, S. N., Roy, N., Jetté, 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.

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.e15–670.e20. http://doi.org/10.1016/j.jvoice.2011.12.001

Awan, N. S., Roy,N., &Christopher Dromey, C.(2009).Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. Clinical Linguistics & Phonetics, *23(11)*, 825-841.

Baken, R. J., &Orlikoff, R.F.(2007).*Clinical measurement of speech and voice*. Second edition.

Balasubramanium, R. K., Bhat, J. S., Fahim, S., & Raju, R. (2011). Cepstral analysis of voice in unilateral adductor vocal fold palsy. *Journal of Voice*, *25*(3), 326–329. http://doi.org/10.1016/j.jvoice.2009.12.010

Brinca, L. F., Batista, A. P. F., Tavares, A. I., Gonçalves, I. C., & Moreno, M. L. (2014). 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. http://doi.org/10.1016/j.jvoice.2013.10.001

Bele,I. V. (2004).Reliability in perceptual analysis of voice quality. *journal of voice*.19(4). 555-573

Dejonckere, P.H., Remacle, M., Elbaz, E., Woisard, V., Buchman, L. & Millet, B. (1996). Differentiated perceptual evaluation of pathological voice quality:reliability and corelations with acoustic measurements, *Revue de LaryngolieOtologie Rhinology*, 117, 219-224.

Dejonckere, P. H., obbens, C., de Moor, G. M. &Wieneke, G.H. (1993).Perceptual evaluation of dysphonia:Reliability and relevance. *Folia Phoniatrica*,*45*, 76-83.

Carding, N. P., J A Wilson, A. J., MacKenzie, K, &Deary, I.J (2009).Measuring voice outcomes: state of the science review.*The Journal of Laryngology & Otology.123(8).* 823- 829.

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 : Official Journal of the Voice Foundation*, 14–16. http://doi.org/10.1016/j.jvoice.2014.02.002

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

Heman-Ackah, Y. D., Sataloff, R. T., Laureyns, G., Lurie, D., Michael, D. D., Heuer, R., Hillenbrand, J. (2014). Quantifying the Cepstral Peak Prominence, a Measure of Dysphonia. *Journal of Voice*, *28*(6), 783–788. http://doi.org/10.1016/j.jvoice.2014.05.005

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., … Sataloff, R. T. (2003). Cepstral Peak Prominence: A More Reliable Measure of Dysphonia. *Annals of Otology, Rhinology & Laryngology*, *112*(4), 324–333. http://doi.org/10.1177/000348940311200406

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.

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

Lowell, S. Y., Colton, R. H., Kelley, R. T., & Hahn, Y. C. (2011). Spectral- and cepstral-based measures during continuous speech: Capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice*, *25*(5), e223–e232. http://doi.org/10.1016/j.jvoice.2010.06.007

Lowell, S. Y., Colton, R. H., Kelley, R. T., & Mizia, S. A. (2013). Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech. *Journal of Voice*, *27*(4), 393–400. http://doi.org/10.1016/j.jvoice.2013.02.005

Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: a meta-analysis. *The Journal of the Acoustical Society of America*, *126*(5), 2619–2634. http://doi.org/10.1121/1.3224706

Moers, C., Möbius, B., Rosanowski, F., Nöth, E., Eysholdt, U., & Haderlein, T. (2012). Vowel- and text-based cepstral analysis of chronic hoarseness. *Journal of Voice*, *26*(4), 416–424. http://doi.org/10.1016/j.jvoice.2011.05.001

Nemr,K., Zenari,M. S., Codeiro, G.F., Tsuji, D., Ogawa,A. I., Ubrig, M. T., &Menezes, M.,(2012). GRBAS and Cape-V Scales: High Reliability and Consensus When Applied at Different Times.Journal of voice. 26(6) 812.e17-812.e22.

Olson, E. L. D., Goding, S. G. & Deirdre D. Michael, D. D. (1998).Acoustic and perceptual evaluation of laryngeal reinnervation by ansacervicalis transfer.*The Laryngoscope, 108(12),* 1767-1772.

Radish Kumar, B., Bhat, J. S., & Prasad, N. (2010). Cepstral analysis of voice in persons with vocal nodules. *Journal of Voice*, *24*(6), 651–653. http://doi.org/10.1016/j.jvoice.2009.07.008

Stráník, A., Cmejla, R., & Vokřál, J. (2014). Acoustic Parameters for Classification of Breathiness in Continuous Speech According to the GRBAS Scale. *Journal of Voice : Official Journal of the Voice Foundation*, 1–9. http://doi.org/10.1016/j.jvoice.2013.07.016

Sapienza, C. &Hoffman-Ruddy, B. (2009). *Voice disoders*. Plural Publishing. San diego oxford brisbane

Stemple, J.C. & Fry, L. T. (2010).*Voice therapy-Clinical case studies*.third edition. Plural Publishing. San diego oxford brisbane

Boone, D. R. (1977).*The voice and voice therapy*.Eaglewood Cliffs, NJ;Prentice-Hall

Watts, R. C., & Awan, N. S.(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, 201*

Wolfe, V., & Martin, D. (1997). Acoustic correlates of dysphonia: Type and severity. *Journal of Communication Disorders*, *30*(5), 403–416. http://doi.org/10.1016/S0021-9924(96)00112-8