AN OBJECTIVE METHOD OF QUANTIFYING VOCAL BREATHINESS IN NORMAL AND SIMULATED CONDITIONS – A PRELIMINARY STUDY

Mr. Mohamed Abuthalha

Register No: 19SLP022

A Dissertation Submitted in Part Fulfillment of Degree of Master of Science

(Speech-Language Pathology)

University of Mysore

Mysuru

ALL INDIA INSTITUTE OF SPEECH AND HEARING

MANASAGANGOTHRI, MYSURU-570 006

September 2021

CERTIFICATE

This is to certify that this dissertation entitled "**An Objective Method Of Quantifying Vocal Breathiness In Normal And Simulated Conditions – A Preliminary Study"** is a Bonafide work submitted in part fulfillment for degree of Master of Science (Speech-Language Pathology) of the student Registration number 19SLP022. 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 award of any other Diploma or Degree.

Mysuru
September 2021

Dr. M. Pushpavathi
Director
All India Institute of Speech and Hearing
Manasagangothri
Mysuru- 570006

CERTIFICATE

This is to certify that this dissertation entitled "**An Objective Method Of Quantifying Vocal Breathiness In Normal And Simulated Conditions – A Preliminary Study"** is a Bonafide work submitted in part fulfillment for degree of Master of Science (Speech-Language Pathology) of the student Registration number 19SLP022. This has been carried out under my supervision and guidance. It is also certified that this dissertation has not been submitted earlier to any other University for award of any other Diploma or Degree.

Mysuru
September 2021

Guide
Dr. K. Yeshoda
Associate Professor in Speech Sciences
Department of Speech-Language Sciences
All India Institute of Speech and Hearing,
Manasagangothri, Mysuru- 570006

DECLARATION

This is to certify that this dissertation entitled "**An Objective Method Of Quantifying Vocal Breathiness In Normal And Simulated Conditions – A Preliminary Study"** is the result of my own study under the guidance of Dr. K. Yeshoda, Associate Professor in Speech Sciences, Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysuru, and has not been submitted earlier to any other University for award of any other Diploma or Degree

Mysuru ***Registration number: 19SLP022***
September 2021

TABLE OF CONTENTS

Chapter	Content	Page number
I	Introduction	1
II	Review of Literature	5
III	Method	16
IV	Results	18
V	Discussion	20
VI	Summary and Conclusion	21
	References	23

List of tables

Table No.	Title	Page number
1	Summary of studies reporting correlation between APQ and breathiness	6
2	Summary of studies reporting correlation between PPQ and breathiness	12
3	Median rating of breathiness and VFSD	18

List of Figures

Table No.	Title	Page number
1	Plot of individual VFSD values in each severity group	19

CHAPTER-1

INTRODUCTION

The pursuit of finding measures that accurately correlate with perceptual parameters of voice have been of interest for many researchers. The main reason for this preference towards objective measures is the issues with inter-rater and intra-rater reliability of subjective perceptual assessments (Kreiman et al., 1993; Latoszek, et al., 2018). In a clinical assessment, the voice can be described in terms of both subjective and objective measurements. Subjective measures comprise of perceptual judgment of the parameters and quantify them according to the descriptions by the listener. Objective measures such as acoustic analysis are based on voice recordings, which under controlled condition will provide consistent quantification of the parameter, hence it avoids the differences in how a voice is being perceived and interpreted among different examiners (Little et al., 2011).

1.1 Normal Voice Production

When voice is produced the subglottic pressure increases as air is trapped by the vocal folds coming together (adduction), this increases the pressure beneath them and pushes the vocal folds to open apart (abduction). Once the vocal folds are abducted the pressure drops and due to elastic recoil forces, the vocal folds adduct again and the process continually repeats at a constant rate which is termed as fundamental frequency (F_o). These alternating compression and rarefaction of air stream by the vocal folds is perceived by the ears as the voice having a specific pitch corresponding to the F_o. Any disruption in the cycle of vocal fold vibration such as closure, periodicity, instability, etc will lead to abnormal perception of voice known as “dysphonia”.

1.2 Breathiness

Breathiness, roughness and strain are the three primary vocal qualities that are used to describe an abnormal voice. These qualities have been routinely used in subjective auditory perceptual scales such as GRBAS (Hirano, 1981) and CAPE-V (ASHA, 2002) during clinical assessment of voice disorders. Breathy voice quality is a result of incomplete glottal closure due to any etiology, where the vocal folds do not approximate completely, and leaving turbulent airflow. Breathiness is defined in CAPE-V as, “audible air escape in the voice”. Breathiness is an acoustic-perceptual entity that is observed in cases of vocal nodules or polyps, vocal fold paralysis, etc. (Boone et al,, 2014). Breathiness is characterized by audible air leakage, turbulent, high frequency noise during auditory sensation along with whispery or airy voice in severe cases (Rontal & Rolnick, 1975).

1.3 Acoustic Analysis

Acoustic analysis of voice is a common practice in the clinical settings because of its objective nature and its ability to detect subtleties in the voice signal. It is a non-invasive method of assessing the nature of voice signal which is equivalent to the sound wave that reaches the ear. In recent years, computers programmed to calculate the acoustic parameters from digitized sound input devices using mathematical algorithms are incorporated for acoustic analysis of voice (Baken & Orlikof, 2000). A few commonly used systems are: Praat (Boersma, 2001), Kay Electremetrics' computerized speech lab (CSL), Voice and Speech Systems- Vaghmi, Tiger Electronics' Doctor Speech (DRS).

The vocal characteristics are assessed in terms of frequency, intensity, perturbation, noise related parameters. Sustained vowels are used commonly for acoustic analysis, however sentences are also used in order to obtain an actual representation of phonation that takes place in real-life condition (Latoszek et al., 2018). A number of parameters in acoustic measurements have been correlated with perceptual judgments like breathiness, roughness and hoarseness etc. E.g. frequency perturbation (jitter), amplitude perturbation (shimmer), and the harmonics-to-noise (or noise-to-harmonics) ratio, Cepstral peak prominence, and spectral tilt etc. (Wolfe & Martin, 1997; Latoszek et al, 2018).

1.4 Spectrography

Spectrography has been incorporated in voice analysis measurements to display the frequency and harmonic components of speech over time. A spectrograph is a three-dimensional representation of a sound wave with frequency along the Y-axis, time along the X-axis and energy represented as density of the tracing (Koenig et al., 1946). The spectrogram is typically obtained with several filters with either narrow bandwidth or wide bandwidth. Generally, a narrow-band filtering will enable to visualize individual harmonics while inspecting the vocal acoustic signal. The spectrography measures has been found to correlate with various voice qualities in dysphonic voices (Rontal & Rolnick, 1975). The formant structure of a speech or speaker can be determined from a spectrogram.

1.5 Acoustics of Vowel Production

A speech signal will contain two components: (i) the glottal source, and (ii) resonance from the vocal tract. The glottal source is the energy emitted by the air passing through the larynx that is modulated by the vocal fold vibrations. The vocal tract consists of the supraglottis, pharynx, nasal cavity and the oral cavity. The vibrations resulting from the resonances of the structures in the vocal tract will depend on their size, shape and configuration (Raphael et al., 2007). These vocal tract resonances are called “formants”.

In a phonation sample of vowel with normal voice quality, there are regular periodic vibration representing the vocal fold vibration at F_0 and clear formants present in the spectrogram (Kent, 1993). When the voice quality is breathy, the excess air flow through the vocal tract, will exhibit a noise component in the spectrogram i.e. spread of energy across frequency spectrum (Wolfe & Martin., 1997). It is accompanied by attenuation of formants by the noise and diminished intensity particularly at the region of fundamental frequency because of insufficient glottal closure amplitude. In cases where the breathiness is very severe, the F_0 is virtually absent, the spectrum will contain only noise-like energy embedded within formants produced because of the resonance of vocal tract (Södersten et al., 1991). The spectrogram of the voice which is at the maximum extent of breathiness would tend to be similar to the spectrogram of consonant /h/.

1.6 Relationship between the spectrum of /h/ and breathiness

The phoneme /h/ is an unvoiced glottal fricative produced by airflow through the larynx without voicing, which causes a turbulence with marked peaks mimicking the formants of a vowel, because during the production of /h/ the vocal tract assumes the configuration of the neighboring vowel due to coarticulatory effect (Strevens, 1960). It can be seen that as the breathiness in a vowel increases its frequency pattern shifts toward the spectrogram of the consonant /h/ (spectrogram of consonant /h/ and breathy vowel tend to be similar).

Hence it can be assumed that the difference between spectra of a vowel in relation with the /h/ produced in same phonetic context could provide us the with the information about extent of breathiness involved during the phonation of the vowel i.e. smaller the difference between the vowel and the fricative, may indicate that breathiness is present in the vowel. The consonant /h/ can be produced with no interference of the vocal pathology owing to its voiceless nature.

1.7 Need for the Study

Despite profuse acoustic measures that are found to be correlated with auditory –perceptual measures, no single measure gives a true estimation of the degree of pathology that is present in the phonatory system. Latoszek et al. (2018) had concluded in their meta-analysis of various acoustic measures that there is a need to develop more measures that are conceptually promising in this area.

The subjective nature of auditory-perceptual ratings may induce inter-rater and intra-rater variability due to several reasons attributed to the signal, listener or the scale that is being used (Kreiman et al., 1993). An objective and non-invasive measure such as acoustic analysis, done along with perceptual ratings will help overcome its pitfalls and improve reliability of the clinical voice assessment for various purposes like screening, diagnosing and documenting, etc. Hence the present study was planned to explore the possibility of an objective method based on spectral relationship between vowel and fricative, to predict the severity of breathiness.

CHAPTER -2

REVIEW OF LITERATURE

There are a number of researches documented in the literature which has shown several acoustic parameters that correlates with perceived breathy voice quality. The following section summarizes some of those studies for each acoustic parameter.

2.1 Amplitude Perturbation Quotient (APQ)

The amplitude Perturbation Quotient, or APQ, is the ratio of a moving average of amplitude perturbations in fundamental period to the mean amplitude. It represents the degree of deviation from the mean amplitude. The moving average can be performed to any number of sample periods, denoted as APQn, where n is the number of pitch periods analysed. APQ is frequently measured with a length five (APQ5) or eleven (APQ11) pitch periods are found in literature.

Several studies have reported that APQ correlated with perceived severity of breathiness in the voice of dysphonic patients. The summary of the studies reporting relationship between APQ and breathiness are given in table 1.

Smoothed APQ is a variation of APQ is one of the measures reported by Wolfe et al (1997) to correlate with perceptual judgment of breathiness. They studied the voice sample of 20 men and 31 women with dysphonia. The samples were rated on a 7-point rating scale for deviation in voice quality. It was found that smoothed APQ correlated with breathiness rating with a coefficient of r = .65.

Table 1: Summary of studies reporting correlation between APQ and breathiness

Authors (year)	Number and nature of subjects.	Voice sample used	Rating type	Correlation coefficient
Prosek et al. (1987)	19 men, 35 women[d] (mean age 40.1 yrs); 8 men, 8 women[n] (mean age 36.2)	Vowel /i:/	7 point- rating scale	r = .59
Hirano et al. (1988)	30 males[d] (30 to 62 years)	Phonation	4-point rating scale	r = .70
Wolfe et al. (1997)	20 men, 31 women[d] (15 to 79 years)	Vowels /a:/ and /i:/	7 point- rating scale	r = .65
McAllister et al. (1998)	50, ten year old children[n]	Vowel /a:/	10 cm VAS	r = .40
Heman-Ackah et al. (2002)	10 females and 9 males[d] (25 to 87 years)	Vowel /a:/ and connected speech.	12 cm VAS	r = .52
Moers et al. (2012)	29 men 49 women[d] (19 to 85 years)	Vowel /e:/	4-point rating scale	ρ = .46

Note: Moers et al. and McAllister et al. reported for APQ11; others reported for APQ5

[n] Normal voice samples; [d] Dysphonic voice samples

2.2 Breathiness Index (BRI)

Breathiness is described as the ratio of second derivative to the non derived energy of a signal. It was first described by Fukazawa et al. (1988) who reported BRI values for 31 patients and 24 normal voice samples. They reported that BRI obtained correlation coefficient, ρ of .73 with the perceptual rating of breathiness.

Hillenbrand and Houde (1996) studied the voice samples of 20 patients and 5 normal adults who produced sustained phonation of /a/ and reading "The rainbow passage". These samples were rated for severity of breathiness on direct magnitude estimation method. The ratings were then scaled within 10 points. They found that BRI correlated well with breathiness rating and obtained correlation coefficients of .62 and .83 for vowel and reading samples respectively.

2.3 Cepstral Peak Prominence (CPP) Measures

The Cepstral Peak Prominence (CPP) is the difference in amplitude between the cepstral peak and the corresponding value of the predicted cepstral magnitude for the quefrency at the cepstral peak (regression line). Smoothing of the cepstrum across time and across quefrency yield a smoothed cepstral peak prominence known as CPP-S. Cepstral measures have been shown to be promising and relatively robust acoustic measures in several studies for the evaluation of dysphonia (Fraile & Godino-Llorente, 2014).

Hillenbrand and Houde (1996) found that CPP and CPP-S correlated well with breathiness rating and obtained correlation coefficients of -.89 and -.88 for vowel and reading samples respectively for CPP; and -.96 and -.92 respectively for CPP-S.

Heman-Ackah et al. (2002) reported that CPP-S correlated well with breathiness rating and obtained correlation coefficients r of -.70 and -.71 for vowel and reading respectively.

Shrivastav and Sapienza (2003) analysed 27 vowel (/a:/) samples of dysphonia patients. The judgements of perceived severity of breathiness was rated on a 5-point scale. They reported that CPP correlated (r = -.87) with the severity of breathiness.

Moers et al. (2012) reported that CPP and CPP-S correlated well with breathiness rating and obtained correlation coefficients ρ of -.54 and -.66 for vowel and reading samples respectively for CPP; and -.46 and -.64 respectively for CPP-S.

2.4 Coefficient of Excess (EX)

The coefficient of excess (EX) is a kurtosis measure of amplitude in a residue signal of inverse filtering. It increases with addition of noise in the speech signal. This parameter has been reported to correlate with breathiness and hoarseness with coefficients *r* of -.31 and -.19 respectively (Prosek et al. 1987).

2.5 Glottal-to-noise excitation ratio (GNE)

The glottal-to-noise excitation ratio (GNE) is designed to detect noise component (i.e.) breathiness in voice based on correlation among Hilbert envelopes of different frequency bands each with a bandwidth (eg. 3000 Hz). The vibration of the vocal folds lead to a synchronous oscillation of different frequency bands since they all originate from the same glottal source. The synchronization is expressed in terms of correlations between envelopes of different frequency bands. Turbulent noise are asynchronous. (Michaelis et al., 1997).

Lopes et al. (2012) explored the correlation between various acoustic parameters and perceptual evaluation from 71 children (3-9 years) on phonating vowel /ɛ/ and used 100mm visual analog scale to rate the degree of perceptual voice qualities. They found that GNE correlated with breathiness with ρ of -.62 and it also correlated with grade of roughness with coefficient ρ of -.45.

Latoszek et al. (2018) reported a weighted average correlation coefficient r_w of .73 based on meta-analysis of studies.

2.6 Harmonic-to-noise ratio (HNR)

Moers et al. (2012) reported that HNR correlated negatively with the degree of breathiness in patients with dysphonia resulting a correlation coefficient(ρ) of -.51. It also correlated with ρ of -.47 for roughness and -.61 hoarseness.

2.7 H1-H2

Klatt & Klatt (1990) found that the differences between the amplitude of the first harmonic and second harmonic in the long-term average spectrum H_1-H_2 is an acoustic measure correlated with a coefficient of .83 on the sustained vowels for the estimation of breathiness in voice samples from 16 participants. The difference between the amplitude of the first and second harmonics (H_1-

H_2) measures the relative length of the open phase of the glottal oscillation. In breathy voices, the first harmonic amplitude is relatively high compared to following relatively weaker harmonics.

Shrivastav & Sapienza (2003) also reported the H_1-H_2 to correlate with perceived breathiness rating with coefficient of .55.

2.8 Jitter

Wolfe et al. (1997) reported that jitter percent correlated positively with the degree of breathiness in patients with dysphonia resulting a coefficient(r) of .61. It also correlated with r of .57 for both roughness and hoarseness.

Wolfe and Martin (1997) reported that breathy voice severity correlated with the standard deviation of jitter with r of .59.

In the study by Moers et al. (2012), jitter (percent) and absolute jitter (ms) yielded correlation coefficients ρ of .58 and .53 respectively with breathiness. Roughness also correlated in this study with ρ of .42 and .49 for jitter percent and absolute jitter respectively. In addition, both the acoustic parameters correlated with coefficient of .60 with hoarseness measures.

2.9 Log of period standard deviation (LNPSD)

LNSPD is the natural logarithm of period standard deviation (SD of mean period length). Wolfe and Steinfatt (1987) obtained samples of 102 vowel sounds (/a/ and /i/) from 51 subjects having dysphonia secondary to various laryngeal pathologies. The severity of deviation of voice quality was rated on a 7-point scale. The LNPSD correlated very strongly with breathiness and obtained a correlation coefficient $r = .95$ and stated it as “the best single predictor of severity of breathy voices”.

Kreiman et al. (1990) obtained phonation samples of vowel /a/ from 18 males with voice problems and 18 normal voices. The pathological voices were rated on a seven-inch visual analog scale for each characteristics. There was a correlation $r = -.63$, between LNPSD and breathiness judgement.

2.10 Pearson *r* at Auto-correlation Peak (RPK)

Hillenbrand & Houde (1996) reported that RPK correlated well with breathiness rating and obtained correlation coefficients of -.84 and -.85 for vowel and reading samples respectively.

Wolfe et al. (2000) reported from analysis of sustained phonation of /a/ from 182 samples that RPK correlated with *r* of -.58.

2.11 Noise-to-Harmonic ratio (NHR)

Dejonckere & Lebacq (1996) studied the voice samples of 34 men and 53 women (mean age 38.7 years) with voice problems. They obtained samples of sustained phonation /a:/ and they were rated perceptually for severity of voice quality and found that NHR correlated with breathiness severity with correlation coefficient of .70.

Wolfe et al. (1997) reported that NHR correlated positively with the degree of breathiness in patients with dysphonia resulting a coefficient(*r*) of .58. It also correlated with *r* of .54 for roughness and .63 hoarseness.

Heman-Ackah et al. (2002) reported that NHR correlated with breathiness rating and obtained correlation coefficient *r* of and .54.

Moers et al. (2012) found that the NHR yielded correlation coefficient ρ of .52, .50 and .63 respectively with breathiness, roughness, hoarseness.

2.12 Normalized Noise Energy (NNE)

The normalized noise energy (NNE) is a spectral measure which reflects the relative degree of turbulent noise with respect to the total energy. The NNE is based on adaptive comb filtering of specific particular frequency range (e.g. 0–5 kHz, 1–5 kHz, etc.) It gives an estimates of the average noise level appearing between the harmonics.

Hirano et al. (1988) reported that NNE at 1-5kHz correlated postively with breathiness with a coefficient of .73.

Feijoo & Hernandez (1990) obtained samples from 121 subjects (64 normals and 57 dysphonics) for sustained vowel /a:/ . The severity of deviation of voice quality was rated on a 4-point scale. The NNE at 1-4kHz correlated very strongly with breathiness and obtained a correlation coefficient $r = .84$.

McAllister et al.(1998) studied the voices of 50, ten year old children to identify the prevalence of various voice quality in these children and their relationship with four acoustic parameters. The have found that NNE at 0-5kHz and NNE at1-5 kHz correlated positively with coefficients(r) of .73 and .44 respectively.

2.13 Partial Loudness Measures

Shrivastav and Sapienza (2003) analysed their voice stimuli based on a nonlinear auditory model of sound perception. The partial loudness the aspiration noise which was measured in ‘sones’ was found to correlate well (r= .84) with the perceptual ratings of breathiness. This approach represents the psychoacoustic percepts of the auditory system. The spectrum of the signal is represented as loudness as a function of frequency after several stages of filtering and nonlinear transformation based on data from psychoacoustic data experiments, mimicing transformations as perceived in the auditory system.

2.14 Pitch Amplitude (PA)

Prosek et al.(1987) used sixty vowel samples of normal and dysphonic patients and rated the severity of breathiness on a seven point equal appearing interval scale. They assessed six parameters’ usefulness in predicting the severity vocal quality differences. They reported that the amplitude of the peak in the auto-correlation wave of residue signal (i.e. waveform resulted from inverse filtering), known as pitch amplitude, did correlate with breathiness measures with coefficient r of -.74. It can be seen that the amplitude of pitch is expected to decrease with increasing breathiness. Also, this parameter correlated with hoarseness and harshness with r of -.71 and -.68 respectively.

2.15 Pitch Perturbation Quotient (PPQ)

Pitch Perturbation Quotient, or PPQ, is the ratio of a moving average of perturbations in fundamental period to the mean fundamental period. It represents the degree of deviation from the mean period. The moving average can be performed to any number of sample periods, denoted as PPQn, where n is the number of pitch periods analysed. PPQ is frequently measured with a length five pitch periods (PPQ5) in literature.

Several studies has reported that PPQ5 to be correlated with perceived severity of breathiness in the voice of dysphonic patients. The summary of the studies reporting relationship between PPQ and breathiness are given in table 2.

Table 2: Summary of studies reporting correlation between PPQ and breathiness

Authors (year)	Number and nature of subjects	Voice sample used	Rating type	Correlation coefficient
Prosek et al. (1987)	19 men, 35 women[d] (mean age 40.1 yrs); 8 men and 8 women[n] (mean age 36.2)	Vowel /i:/	7 point-rating scale	$r = 0.38$
Hirano et al. (1988)	30 males[d] (30 to 62 years)	Phonation	4-point rating scale	$r = 0.66$
Wolfe et al. (1997)	20 men, 31 women[d] (15 to 79 years)	Vowels /a:/ and /i:/	7 point- rating scale	$r = 0.62$
Mcallister (1998)	50, ten year old children[n]	Vowel /a:/	10 mm VAS	$r = 0.61$
Moers et al. (2012)	29 men 49 women[d] (19 to 85 years)	Vowel /e/	4-point rating scale	$\rho = 0.55$

[n] Normal voice samples; [d] Dysphonic voice samples

2.16 Relative Amplitude Perturbation (RAP)

Wolfe et al. (1997) reported that RAP correlated positively with the degree of breathiness in patients with dysphonia resulting a coefficient(*r*) of .61. It also correlated with *r* of .56 for roughness and for .55 hoarseness.

Heman-Ackah et al. (2002) reported that RAP correlated well with breathiness rating and obtained correlation coefficient *r* of .54.

In Moers et al. (2012) the RAP yielded correlation coefficient ρ of .54, .36 and .55 respectively with breathiness, roughness, hoarseness.

2.17 Relative Spectral Energy

Klich (1982) found correlation for relative energy in the 100-500 Hz and 3500-4500Hz with the rating of breathiness of vowels in ten young adult females. The relative energy of 100-500 Hz showed negative correlation (r = -.67). This region encompasses the F_o which is found to be diminishing with increasing severity of breatiness. The relative energy of 3500-4500 Hz is positively correlated (*r = .58)* which consists of the concentration of high frequency in the acoustic signal, which is also a characteristic of the consonant /h/ (Strevens, 1960).

2.18 Relative Amplitude Perturbation (RAP)

Wolfe et al. (1997) reported that RAP correlated positively with the degree of breathiness in patients with dysphonia resulting a coefficient(*r*) of .61. It also correlated with *r* of .56 for roughness and for .55 hoarseness.

Heman-Ackah et al. (2002) reported that RAP correlated well with breathiness rating and obtained correlation coefficient *r* of .54.

In Moers et al. (2012) the RAP yielded correlation coefficient ρ of .54, .36 and .55 respectively with breathiness, roughness, hoarseness.

2.19 Relative Spectral Energy

Klich (1982) found correlation for relative energy in the 100-500 Hz and 3500-4500Hz with the rating of breathiness of vowels in ten young adult females. The relative energy of 100-500 Hz showed negative correlation (r = -.67). This region encompasses the F_o which is found to be diminishing with increasing severity of breatiness. The relative energy of 3500-4500 Hz is positively correlated (*r = .58)* which consists of the concentration of high frequency in the acoustic signal, which is also a characteristic of the consonant /h/ (Strevens, 1960).

2.20 Shimmer

Wolfe et al. (1997) reported that shimmer correlated positively with the degree of breathiness in patients with dysphonia resulting in a coefficient(r) of .67. It also correlated with r of .46 for roughness and .59 for hoarseness. Shimmer percent also was also reported in the study with the same coefficients for the three perceptual parameters.

Wolfe and Martin (1997) reported that breathy voice severity correlated with the standard deviation of shimmer with r of .55.

In Moers et al. (2012) the shimmer (dB) and shimmer percent yielded correlation coefficients ρ of .49 and .48 respectively with breathiness. Roughness also correlated in this study with ρ of 0.40 and 0.39 for shimmer (dB) and shimmer percent respectively. In addition, both the acoustic parameters correlated with coefficient of .57 and .56 respectively with hoarseness.

2.21 Spectral Flatness of Residue Signal (SFR)

Inverse filtering methods are useful to identify the excitatory pattern of the glottis. The residue signal is the output consisting of harmonics of fundamental frequncy, that is obtained as a result of filtering the waveform of a vowel with linear prediction coeffecients. This method yields a residual signal which mimics the vibrations of vocal folds and from this method, the abnormal vibratory aspects of can be viewed in an affected pathological voice.

Prosek et al.(1987) reported that the spectral flatness of residue signal (SFR) was one of the six variables studied correlated with severity of breathy voice quality (r = .59) and hoarseness r = .60, respectively.

2.22 Spectral tilt measures

Hillenbrand and Houde (1996) calculated spectral tilt on average spectral energy above 4000 Hz and below 4000 Hz (i.e. high frequency/ low frequency) from 25 samples of vowel phonation and reading passage each. They found that spectral tilt measure correlated well with perceptual judgement of breathiness with correlation coefficients of .64 and .84 for phonation and reading respectively.

To summarize the above studies, several parameters were found to correlate with the perceptual degree of breathiness however no such measure has very strong correlation in predicting vocal breathiness. In addition, these are some measures which were found to correlate with the vocal quality of roughness along with breathiness. This lack of independence between two vocal qualities would make it difficult to interpret the values based solely on such measure.
In this regard, an attempt is made to derive an objective measure to quantify vocal breathiness.

2.23 Aim of the Study

To validate an objective measure of breathiness of voice based on the spectral relationship with glottal fricative.

2.24 Objectives of the study:

a) To measure and document the difference of spectra between /h/ and /a/ in normal and breathy voices.
b) To determine the correlation between spectral difference measure and perceptual rating of breathiness scale in CAPE-V.

CHAPTER- 3

METHOD

3.1 Participants

In accordance with guidelines for conduct of dissertation work for the academic year 2020-2021, issued by AIISH in the light of Covid-19 pandemic, Speech-Language Pathologists were recruited as subjects with simulated breathiness as samples for the study. Twenty postgraduate SLP students (10 males and 10 females) with experience in evaluation and management of clients with voice disorders were included in the study. The participants' ages ranged between 21 to 26 years (Mean = 24.05 years; SD = 0.94). Participants with a history of upper respiratory tract infection or dysphonia were excluded.

3.2 Instruments

Computerised Speech Lab (CSL) Model 4500 (KAY Pentax, New Jersey, USA) was used to to record audio samples from all participants. Praat version 6.1.38 and RStudio version 1.4.1106 were used for analysis of the recorded audio data.

3.3 Procedure

The samples were recorded using in a quiet, acoustically treated room. The participants were instructed about the procedure. They were instructed to say the sentence "I told a joke and said hahaha" in normal voice quality and three increasing levels of severity of breathiness (i.e. mild, moderate and severe, breathiness). The severity targets were modeled before recording and the participants were given sufficient time to practice. The target syllable was /ha:/ was obtained in a sentence context in order to practice and carry over the simulated breathiness over the target. Five iterations of the sentence were obtained in each severity level.

3.4 Analysis

The samples were perceptually analyzed for the severity of breathiness and rated using CAPE-V's VAS by three SLP post-graduate students. The inter-rater reliability for the three raters was calculated using Chronbach's alpha. The mean rating of the three judges were taken for the analysis. The samples were grouped as normal, mild, moderate severe based on the severity of the breathiness according to CAPE-V rating.

Praat software was used to view, play and cut the audio samples for analysis. Out of five iterations of the target sentence, the middle portion was selected. In case the middle part does not contain stable portion of spectrogram, the next portion which is stable was selected. The segments of /h/ and /a:/ were selected with equal length in duration. The spectral comparison of the vowel-fricative was calculated in statistical computing, programming language R version 4.0.4 using packages tuneR (Ligges et al. 2018) and seewave (Seur, Aubin &Simonis, 2008). The selected segments of the voice sample were loaded in to the environment using readWave() function. The mean spectrum of segments were computed using mspec() function. The Vowel-fricative spectral distance (VFSD) difference between the mean spectra of /h/ and /a/ was computed with Itakura-Saito distance formula using itakura.dist() function. (Seur, 2018). VFSD is the Itakuro-Saito distance of mean spectra of /a:/ with respect to mean spectra of /h/. VFSD measurement is the highlight of this study.

$$VFSD = \sum_{i=1}^{f} \frac{h_i}{a_i} - \log\left(\frac{h_i}{a_i}\right) - 1$$

3.5 Statistical analysis

Statistical analysis was done in IBM SPSS 26. Descriptive statistics was used to report the mean, median of breathiness ratings and VFSD. Friedman's test was done to identify the difference in VFSD among the severity groups and Spearman's rank correlation was done to measure the relationship between the two variables.

CHAPTER - 4

RESULTS

The median of breathiness severity rating and the VFSD values are shown in Table 3. Inter-judge reliability for the rating which was calculated using Chronbach's alpha was 0.6 and above for the severity groups of breathiness.

Table 3: Median rating of breathiness severity groups and VFSD

Groups	Breathiness rating	VFSD
Normal	1.67	2.30
Mild	21.67	0.72
Moderate	51.50	0.69
Severe	84.67	0.19

To detect the difference between the each severity group for VFSD , Friedman's test (N=20) was done which revealed statistical significance ($\chi^2 = 43.980$; degree of freedom = 3; $p < 0.01$). Pairwise comparison with bonferroni correction were significant for the following pairs: Normal-mild, normal- moderate, normal-severe, mild- severe.

The relationship between severity rating and VFSD using Spearman's rank correlation (ρ) revealed negative correlation with coefficient (ρ) of -0.771 significant at 0.01 level. However on comparison of the two variables in each severity groups, significant correlation was found only in moderate and severe groups with coefficients of -.457 and -.575 respectively.

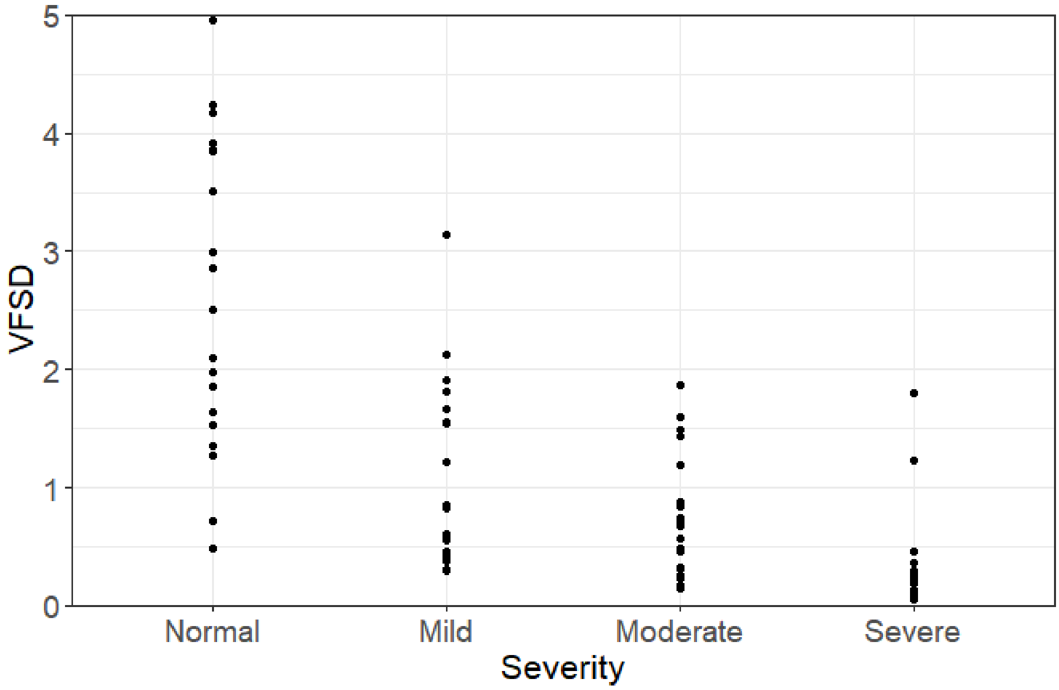

Figure 1: Plot of individual VFSD values in each breathiness severity group

Figure 1 shows the individual VFSD values plotted for the severity groups of breathiness. It can be seen that the VFSD decreases with increasing severity of breathiness, which is also indicated by negative correlation coefficient Spearman's rho (ρ). It can also be observed that the range of values of VFSD for each groups appear to be distinct from other groups, particularly in those pairs which obtained significance in Friedman's test.

Highlight of the results

- The inter-rater reliability of perceptual judgment of severity of breathiness was at acceptable level.
- Statistically significant difference was obtained for the values of VFSD among the four severity groups.
- Strong correlation was obtained between VFSD and severity ratings.

CHAPTER - 5

DISCUSSION

To summarize the results, the VFSD obtained strong correlation on Spearman's rank correlation with coefficient ρ of -0.771 with the perceptual ratings of severity of breathiness. Significance was reported by Friedman's test for the values of VFSD among the severity groups and post-hoc test revealed that the difference was significant for normal-mild, normal-moderate, normal-severe, mild-severe pairs.

The VFSD was able to correlate strongly with the perceptual judgement of severity of simulated breathiness rating scale. A similar degree of correlation can be found for various acoustic measures discussed in the review of literature. Negative correlation was expected as the value of VFSD decreases when there is less difference between the two spectra as discussed previously, with breathiness, the high frequency region of the vowel becomes similar to the glottal consonant /h/.

In addition, the VFSD was able to differentiate the normals from dysphonic groups, however it has failed to seperate the moderate from severe and mild groups in post- hoc comparison. The reason could likely be due to distribution cut-off values defined in VAS of CAPE V for description of severities that may have induced a variability in the differentiating these severity levels (Wewers & Lowe, 1990). The cut-off values which have better accuracy for discriminating deviations of voice were suggested based on ROC in other studies (Simberg et al., 2000; Martins et al. 2015; Yamasaki et al., 2017).

As seen from the results, VFSD mat be useful in quantifying the severity of breathiness and in differentiating normal voice quality from breathy voice quality.

CHAPTER - 6

SUMMARY

The need for developing new methods to accurately predict the severity of dysphonic voice qualities has been put forward in the literature. The spectrum of a vowel with breathy vocal quality has similarity with the spectrum of phoneme /h/. Thus it was hypothesized that the difference between the spectrum of a vowel and phoneme /h/ can be a predictor of the degree of breathiness present in the vowel. The present study aimed to explore the relationship of a parameter VFSD in with the perceptual rating of breathiness severity rating. Twenty particpants between the ages 21 to 26 years produced a sentence with normal voice quality, mild, moderate and severe breathy voice in simulated condition. The samples were recorded and analyzed for breathiness severity. VFSD was calculated for each sample from difference between the specrum of /a:/ and /h/. Results revealed that VFSD correlated strongly with the perceptual rating of breathiness. The VFSD was able to differentiate normals from dysphonic severity groups.

6.1 Conclusion

The VFSD described in this study is based in Itakuro-Saito distance formula. Future studies can also explore the possibilities of using other possible spectral dissimilarity measures such as those described in Suer (2018) can be carried out to find out better possible alternatives.

Further, investigating the application of VFSD with the use of filtering to determine the relationship of VFSD in low frequencies and VFSD in high frequencies as there are differences found among various frequency bands of voice with breathiness (Klich, 1982; Hilelndrand & Houde, 1996) and pre-processing techniques to refine the spectrum which would tune the audio signal to better assess the state of dysphonia as heard during real time listening as opposed to digitized recorded voice could be done. Such pre-processing may include methods that apply auditory excitation patterns from auditory models to predict the loudness of the sound (Moore et al., 1997; Chen et al, 2011). The partial loudness noise which is based on this auditory excitation patterns have been found useful in predicting the degree of breathiness. (Shrivastav & Sapienza, 2003).

6.2 Limitations and Future direction

Caution should be taken while interpreting based on the values of acoustic parameters since it can be seen in the review of literature that some parameters that correlated with breathiness also correlated with the degree of roughness, hoarseness and strain. However it can be noted that breathiness is predominated with noise component which is reflected in spectral measures, roughness is characterized by cycle-to-cycle variation, which is reflected mainy in time domain rather than frequency domain (Kojima et al., 1980)

Since the voice samples were obtained through simulating breathiness, the aspects of larynx involved in natural condition may not be mimicked (Hillenbrand et al 1994). Hence, validating the VFSD on patients with voice disorders has to be done along with studying the involvement of roughness on this parameter.

6.3 Clinical Implication

The clinical application of this VFSD is that it could be used to predict breathiness level of voice in evaluation of voice patients. Although any acoustic parameter is not able to predict the perceptual voice quality with absolute accuracy, this can be utilized along with subjective rating to confirm and support the rating (Hirano et al., 1988).

REFERENCES

American Speech-Language-Hearing Association. (2002). Consensus Auditory–Perceptual Evaluation of Voice (CAPE-V): ASHA Special Interest Division 3, Voice and Voice Disorders. *Rockville, MD: Author. Retrieved from asha.org/siteassets/uploadedfiles/asha/sig/03/cape-v-procedures-and-form.pdf.*

Baken, R., & Orlikoff, R. (2000). Clinical Measurement of Speech and Voice Singular Thomson Learning. San Diego, CA, 79.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9), 341-345.

Boone, D. R., McFarlane, S. C., Von Berg, S. L., & Zraick, R. I. (2014). The Voice and Voice Therapy. Brazil:Pearson.

Chen, Z., Hu, G., Glasberg, B. R., & Moore, B. C. (2011). A new method of calculating auditory excitation patterns and loudness for steady sounds. *Hearing research*, 282(1-2), 204-215. https://doi.org/10.1016/j.heares.2011.08.001

Feijoo, S., & Hernández, C. (1990). Short-term stability measures for the evaluation of vocal quality. *Journal of Speech, Language, and Hearing Research*, *33*(2), 324-334. https://doi.org/10.1044/jshr.3302.324

Fraile, R., &Godino-Llorente, J. I. (2014). Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control*, 14, 42–54. https://doi.org/10.1016/j.bspc.2014.07.001

Fukazawa, T., El-Assuooty, A., & Honjo, I. (1988). A new index for evaluation of the turbulent noise in pathological voice. *The Journal of the Acoustical Society of America*, 83(3), 1189-1193. https://doi.org/10.1121/1.396012

Godino-Llorente, J. I., Osma-Ruiz, V., Sáenz-Lechón, N., Gómez-Vilda, P., Blanco-Velasco, M., & Cruz-Roldán, F. (2010). The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders. *Journal of Voice*, 24(1), 47–56. https://doi.org/10.1016/j.jvoice.2008.04.006

Heman-Ackah, Y. D., Michael, D. D., & Goding Jr, G. S. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, 16(1), 20-27. https://doi.org/10.1016/S0892-1997(02)00067-X

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37(4), 769–778. https://doi.org/10.1044/jshr.3704.769

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. https://doi.org/10.1044/jshr.3902.311

Hirano M. Clinical examination of voice. Wien: Springer, 1981

Hirano, M., Hibi, S., Yoshida, T., Hirade, Y., Kasuya, H., & Kikuchi, Y. (1988). Acoustic analysis of pathological voice: Some results of clinical application. *ActaOto-Laryngologica*,105(5-6), 432–438. https://doi.org/10.3109/00016488809119497

Kasuya, H., Ogawa, S., Mashima, K., & Ebihara, S. (1986). Normalized noise energy as an acoustic measure to evaluate pathologic voice. *The Journal of the Acoustical Society of America*, 80(5), 1329-1334. https://doi.org/10.1121/1.394384

Kojima, H., Gould, W. J., And, A. L., & Isshiki, N. (1980). Computer analysis of hoarseness. *Acta Oto-Laryngologica*, 89(3-6), 547-554. https://doi.org/10.3109/00016488009127173

Kent, R. D. (1993). Vocal tract acoustics. *Journal of Voice*, *7*(2), 97-117. https://doi.org/10.1016/S0892-1997(05)80339-X

Kent, R. D., & Read, C. (2002). Acoustic analysis of speech. 2nd. Albany, NY: Singular/Thomson Learning.

Klich, R. J. (1982). Relationships of vowel characteristics to listener ratings of breathiness. *Journal* *of Speech, Language, and Hearing Research*, *25*(4), 574-580.

Koenig, W., Dunn, H. K., & Lacy, L. Y. (1946). The Sound Spectrograph. *The Journal of the Acoustical Society of America*, 18(1), 19–49. https://doi.org/10.1121/1.1916342

Kreiman, J., Gerratt, B. R., & Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech, Language, and Hearing Research*, 33(1), 103-115. https://doi.org/10.1044/jshr.3301.103

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech, Language, and Hearing Research*, 36(1), 21-40. https://doi.org/10.1044/jshr.3601.21

Latoszek, B. B., Maryn, Y., Gerrits, E., & De Bodt, M. (2018). A meta-analysis: Acoustic measurement of roughness and breathiness. *Journal of Speech, Language, and Hearing Research*, 61(2), 298-323. https://doi.org/10.1044/2017_JSLHR-S-16-0188

Little, M. A., Costello, D. A. E., & Harries, M. L. (2011). Objective dysphonia quantification in vocal fold paralysis: Comparing nonlinear with classical measures. *Journal of Voice*, *25*(1), 21–31. https://doi.org/10.1016/j.jvoice.2009.04.004

Ligges U, Krey S, Mersmann O, Schnackenberg S (2014). tuneR: Analysis of music. R package

Lopes, L. W., Lima, I. L. B., Almeida, L. N. A., Cavalcante, D. P., & de Almeida, A. A. F. (2012). Severity of voice disorders in children: correlations between perceptual and acoustic data. *Journal of Voice*, 26(6), 819-e7. https://doi.org/10.1016/j.jvoice.2012.05.008

Martins, P. C., Couto, T. E., & Gama, A. C. C. (2015). Auditory-perceptual evaluation of the degree of vocal deviation: correlation between the Visual Analogue Scale and Numerical Scale. In *Codas(*27), 279-284. Sociedade Brasileira de Fonoaudiologia. https://doi.org/10.1590/2317-1782/20152014167

Mcallister, A., Sundberg, J., & Hibi, S. R. (1998). Acoustic measurements and perceptual evaluation of hoarseness in children's voices. *Logopedics Phoniatrics Vocology*, 23(1), 27-38. https://doi.org/10.1080/140154398434310-1

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. https://doi.org/10.1016/j.jvoice.2011.05.001

Moon, K. R., Chung, S. M., Park, H. S., & Kim, H. S. (2012). Materials of acoustic analysis: sustained vowel versus sentence. *Journal of Voice*, *26*(5), 563-565. https://doi.org/10.1016/j.jvoice.2011.09.007

Michaelis, D., Fröhlich, M., & Strube, H. W. (1998). Selection and combination of acoustic features for the description of pathologic voices. *The Journal of the Acoustical Society of America*, 103, 1628–1639. https://doi.org/10.1121/1.421305

Prosek, R. A., Montgomery, A. A., Walden, B. E., & Hawkins, D. B. (1987). An evaluation of residue features as correlates of voice disorders. *Journal of communication disorders*, *20*(2), 105-117. https://doi.org/10.1016/0021-9924(87)90002-5

Raphael, L. J., Borden, G. J., & Harris, K. S. (2007). Speech science primer: Physiology, acoustics, and perception of speech. Lippincott Williams & Wilkins.

Rontal, E., Rontal, M., & Rolnick, M. I. (1975). Objective evaluation of vocal pathology using voice spectrography. *Annals of Otology, Rhinology & Laryngology*, 84(5), 662-671. https://doi.org/10.1177%2F000348947508400518

Shrivastav, R., & Sapienza, C. M. (2003). Objective measures of breathy voice quality obtained using an auditory model. *The Journal of the Acoustical Society of America*, *114*(4), 2217-2224. https://doi.org/10.1121/1.1605414

Simberg, S., Laine, A., Sala, E., & Rönnemaa, A. M. (2000). Prevalence of voice disorders among future teachers. *Journal of voice,* 14(2), 231-235. https://doi.org/10.1016/S0892-1997(00)80030-2

Södersten, M., Lindestad, P. A., &Hammarberg, B. (1991). Vocalfold closure, perceived breathiness, and acoustic characteristics innormal adult speakers. In J. Gauffin& B. Hammarberg (Eds.),Vocal fold physiology: Acoustic, perceptual, and physiologicalaspects of voice mechanisms (pp. 217–224). London, England:Whurr Publishers.

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and speech*, *3*(1), 32-49.

Sueur, J. (2018). Sound analysis and synthesis with R. Cham: Springer.

Sueur J, Aubin T, Simonis C (2008) seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics* 18(2), 213–226. https://doi.org/10.1080/09524622.2008.9753600

Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in nursing & health*, 13(4), 227-236. https://doi.org/10.1002/nur.4770130405

Wolfe, V., Fitch, J., & Martin, D. (1997). Acoustic measures of dysphonic severity across and within voice types. *Folia Phoniatrica et Logopaedica*, 49(6), 292–299. https://doi.org/10.1159/000266469

Wolfe, V., & Martin, D. (1997). Acoustic correlates of dysphonia:Type and severity. *Journal of Communication Disorders*, 30(5),403–415. https://doi.org/10.1016/S0021-9924(96)00112-8

Wolfe, V. I., Martin, D. P., & Palmer, C. I. (2000). Perception of dysphonic voice quality by naive listeners. *Journal of Speech, Language, and Hearing Research*, 43(3), 697-705. https://doi.org/10.1044/jslhr.4303.697

Wolfe, V. I., & Steinfatt, T. M. (1987). Prediction of vocal severity within and across voice types. *Journal of Speech and Hearing Research*, 30(2), 230–240. https://doi.org/10.1044/jshr.3002.230

Yamasaki, R., Madazio, G., Leão, S. H., Padovani, M., Azevedo, R., & Behlau, M. (2017). Auditory-perceptual evaluation of normal and dysphonic voices using the voice deviation scale. *Journal of Voice,* 31(1), 67-71. https://doi.org/10.1016/j.jvoice.2016.01.004