DEVELOPING A HINDI VERSION OF THE CONSENSUS AUDITORY-PERCEPTUAL EVALUATION OF VOICE (CAPE-V)

A Thesis

Presented to

The Faculty of the Department

of Communication Sciences and Disorders

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

By

Isha Baheti

May, 2019

DEVELOPING A HINDI VERSION OF THE CONSENSUS AUDITORY-PERCEPTUAL EVALUATION OF VOICE (CAPE-V)

_____

An abstract of a Thesis

Presented to

The Faculty of the Department

of Communication Sciences and Disorders

University of Houston

_____

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

_____

By

Isha Baheti

May, 2019

ABSTRACT

Objective: The purpose of the current study was to develop a Hindi version of the Consensus of Auditory Perceptual Evaluation of Voice (CAPE-V), a tool for speech language pathologists to assess voice quality and degree of hoarseness. The Hindi version of the CAPE-V will be vital in voice assessments conducted in India and across the world for Hindi speaking individuals.

Method: The Hindi stimuli matched the original CAPE-V in terms of sentence types and sounds tested. Data were collected in two groups- the first in English and Hindi in bilingual speakers (16 males and 15 females) with normal voice quality. The CAPE-V ratings in both languages were compared to assess the validity of the Hindi version. The second group consisted of 13 Hindi speakers (10 males, 3 females) with disordered voice quality. The ratings in the Hindi CAPE-V were also compared to the GRBAS scale (Grade, Roughness, Breathiness, Asthenia, and Strain) as it is the current standard of care in India.

Results: A strong correlation was found between the Hindi CAPE-V scores and the English CAPE-V scores in overall severity, roughness, breathiness, and pitch ($r > .5$, $p < .01$) in normal participants. A weak correlation was found between the English and Hindi versions for the variable of strain ($r < .3$, $p = .439$) in the normative group. A strong correlation ($r > .5$, $p < .01$) was found between the overall severity/grade, roughness, and breathiness scores in GRBAS scale and the CAPE-V scale in normal and disordered voice samples. Significant interrater reliability ($r > .75$) in overall severity and breathiness was noted.

Conclusion: The Hindi CAPE-V can be administered reliably to evaluate features of voice quality in Hindi.

Acknowledgement

Foremost, I would like to express my sincere gratitude to my advisor Dr. Ashwini Joshi for the continuous support of my thesis and research, for her patience, motivation, and immense knowledge. Her guidance helped navigate me through the finest details of my research.

I would like to thank Dr. Lynn M. Maher for enlightening me with the first glance of research and for constantly encouraging me. My sincere thanks to Dr. Bhavya Tiwari for her insightful comments.

I am extremely grateful to Ms. Kamalika Chowdhury and Mr. Nirnay Kumar Kesree for helping me in the collecting the data from India, and not only encouraging me, but constantly pushing me on this path.

Nobody has been more important to me in the pursuit of this work than my family. I am forever indebted to my mother, Kavita Baheti, and my brother, Palash Baheti, whose love and guidance are with me in whatever I pursue. I am also profoundly grateful to Goutam Venkatesh, who was always encouraging and enthusiastic towards my work. Last, but not the least, I am grateful for my lovely friends in Houston who understood me and stood with me in everything that I did.

Thanks for all your encouragement!

TABLE OF CONTENTS

This work is dedicated to

*my dad, **Prafull Baheti**, who believed in me and made me who I am today, and*

*my grandfather, **R.L. Maheshwari**, who has been a pillar of strength to my entire family.*

INTRODUCTION

Voice Quality

Voice is significant for communication of ideas, messages, emotions, and thoughts, and we can identify individuals based on the distinctive features of their voice quality. When we hear somebody, we primarily perceive two things; the phonological information of what is being said, and characteristics of the voice when the phonetic information is removed, i.e., voice quality. Voice quality is considered to be a multidimensional perception of the human voice including a variety of components like pitch, loudness, resonance, breathiness, strain, roughness, harshness, hoarseness, sharpness etc. (Barsties & Bodt, 2015; Kreiman, Vanlancker-Sidtis & Gerratt, 2004). There are physiological definitions of voice quality that consider the perceived results of coordinated action from the subsystems like respiratory system, phonatory and resonance systems. (Kreiman et al., 2004). Similar to these definitions, Trask (1996) defined voice quality as the auditory-perceptual representation of laryngeal and supralaryngeal activity in an individual. Broadly, the voice is evaluated in one of two approaches, subjective (perceptual) or objective (quantifying the aspects of vocal production and physiological contributors) (Barsties & Bodt, 2015). The next section discusses the assessment of voice as a multidimensional phenomenon and various measures adapted for its evaluation.

Assessment of Voice

Voice is a complex and multidimensional phenomenon and a small change in the thickness, length or elasticity of the vocal folds could result in a change in multiple dimensions of voice quality (Ziwei, Zheng, & Pin, 2014). According to the American Speech-Language-Hearing Association (ASHA), altered voice quality can

be characterized by auditory-perceptual symptoms like breathiness (audible air escape); roughness (perception of abnormal vocal fold vibration); strain (perception of excessive tension in vocal folds); asthenia (weak voice or reduced power in voice); abnormal resonance (hyponasal, hypernasal, cul-de-sac resonance); hoarseness (audible aperiodic sound); pulsed voice (fry register, audible creaks or pulses in sound); shrill (high, piercing sound); tremulous voice (shaky voice); rhythmic pitch and loudness undulations; abnormal pitch (too high, too low, limited pitch range); and abnormal loudness (too loud, too soft, limited range).

An ideal voice assessment should evaluate multiple parameters of the voice including pitch, loudness, and quality. This type of assessment would help in determining the severity of a disorder, results of a treatment (Carding, Wilson, MacKenzie, & Deary, 2009; Ziwei et al., 2014), prognosis, comparison of results from different treatment approaches, follow up observations (Ziwei et al., 2014), and course of treatment.

Typically, a voice assessment battery includes five domains- visual imaging, perceptual analysis, aerodynamic assessment, acoustic analysis, and patient self-evaluation (Barsties & Bodt, 2015). Visual/physiological assessment, aerodynamic assessment and acoustic analysis are grouped as instrumental analysis of voice. (Özcebe, Aydinli, Tiğrak, İncebay, & Yilmaz, 2017). Visual analysis refers to the inspection of the vocal folds and other structures in the vocal tract, in order to observe any anatomical or vibratory abnormality. Aerodynamic assessment documents the change in airflow in the vocal tract. Any change in airflow, air pressure, and air volume can directly affect laryngeal functioning, vocal fold closure, valving activity, and nasal airflow, thus affecting the voice quality of the individual. Acoustic analysis targets evaluation of voice to obtain objective data of different parameters of voice

quality, like fundamental frequency, intensity, jitter, shimmer, harmonics to noise ratio, cepstral peak prominence etc. Acoustic analysis is performed on sustained vowels and continuous speech whereby the voice signals are evaluated in time, amplitude and frequency. Auditory-perceptual analysis of the voice involves subjective judgment of an individual's voice by a trained listener. Finally, self-evaluation of the voice documents the impact of the voice disorder on a patient's quality of life. For example, the Voice Handicap Index assesses the effect of the voice disorder on functional, physiological, and emotional domains. Other self-evaluation tools include Voice-Related Quality of Life, Vocal Performance Questionnaire, Voice Symptom Scale (VoiSS), and Vocal Tract Discomfort Scale (Behlau, Zambon, Moreti, Oliveira, & Couto Jr, 2016). These five domains of assessment are independent of each other (Barsties & Bodt, 2015), but allow for an integrated approach targeting both subjective and objective assessments to effectively and systematically evaluate voice quality (Ziwei et al., 2014). Amongst these domains of voice evaluation, the auditory-perceptual evaluation is considered important for a number of reasons. The next section focuses on auditory-perceptual evaluation, its importance and commonly used tools to carry out the evaluation.

Auditory- Perceptual Assessment

Voice quality is essentially a perceptual phenomenon that is provided as a response to acoustic stimuli (Oates, 2009). A listener recognizes any deviation in voice quality as unpleasant or inadequate when compared to normal voice perception (Zraick et al., 2011). Since voice is fundamentally perceptual in nature, it is reasonable to evaluate and share the information regarding features of voice quality in terms of their perception. For example, it will be more comprehendible if we know

that someone's voice is breathy instead of knowing harmonics- to- noise ratio of the same person's voice (Oates, 2009). Moreover, a patient can also directly relate to his voice using the descriptions of perceptual assessment. This implies that auditory-perceptual characteristics are intuitive and are interpretable by healthcare professionals, patients as well as their caregivers.

Auditory-perceptual voice assessment is hence a significant part of voice assessment. In addition to providing a reliable source to differentiate normal and disordered voices and determining prognosis, auditory-perceptual assessment facilitates the correlation of underlying pathophysiology and objective measures (Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2009). Moreover, it is efficient and easy to administer these tests (Barsites & Bodt, 2014; Oates, 2009; Özcebe et al., 2017).

Some of the auditory-perceptual assessment measures that are available include the Grade, Roughness, Breathiness, Asthenia, Strain (GRBAS) scale; the Consensus of Auditory Perceptual Evaluation (CAPE-V), The Stockholm Voice Evaluation Consensus Model, Vocal Profile, Analysis Scheme, Buffalo Voice Profile, and Hammarberg scheme (Barsties & Bodt, 2005). Severity is usually rated on a four-point Likert scale from 0 (normal) to 3 (severe). The CAPE-V, developed by ASHA, assesses voice using a visual analog scale. The examiner uses a 100 mm long horizontal line and makes a mark on it to indicate the perceived severity in different domains of voice.

There are different ways to measure and rate the severity of abnormalities of the different features of voice quality. The equal appearing interval scale involves assigning a number to a feature to indicate its severity. For example, the GRBAS, proposed by the Japan Society of Logopedics and Phoniatrics, is used to assess an

individual's voice in five domains- grade (G), roughness (R), breathiness (B), asthenia or power (A), and strain (S). When using the GRBAS scale, the clinician assigns a number between 0-3 to each of the features to indicate different levels of severity. When using direct magnitude estimation as a method, the listener assigns a number to indicate the extent to which a particular quality is present. It can be anchored (with a reference) or unanchored. A paired comparison is done to compare the severity of two different stimuli. Finally, a visual analog scale involves marking the perceived severity of a voice feature on a straight line that extends from 0 (normal) to usually 100 (very severe). For example, the CAPE-V scale uses a 100 mm long line where the listener marks the perceived severity of voice quality features (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993). Clinically, out of all these scales, the more commonly used scales for auditory-perceptual evaluation are GRBAS and CAPE-V (Nemr et al., 2012; Kempster et al., 2916; Wuyts et al., 1999; Zewei et al., 2014). The next section will provide further details on the CAPE-V.

Consensus Auditory-Perceptual Evaluation (CAPE-V)

Development of CAPE-V

The CAPE-V was developed after the Consensus Conference on Auditory-Perceptual Evaluation of Voice (June 2002), sponsored by Special Interest Division (now referred to as Special Interest Group) 3, Voice and Voice Disorders of ASHA at the University of Pittsburgh. This assessment protocol caters to the need for a consistent clinical tool to perceptually judge voice quality and determine the severity of the disorder. Additionally, it contributes to the information about the anatomical and physiological bases of the voice disorder, and thus indicates a need for an additional testing (American Speech-Language-Hearing Association, 2002). In the consensus meeting, the speech-language pathologists implemented knowledge of psychophysical and psychoacoustic measurement in the clinic to assess auditory-perceptual features of voice quality. Psychoacoustic measures were presented by describing influences of the outer and middle ear in voice perception, details of cochlear biomechanics, and integration of spectral information that takes place at higher cortical levels (Kempster et al., 2009). The researchers presented psychophysical measurements by describing differential limens, additional measures for scaling, partition, ratio and multidimensional scaling. The psychophysical discussion also determined the characteristics of stimuli that contribute to perception, the correlation of these perceptual characteristics, and the processes (cognitive, sensory etc.) that contribute to perception (Kempster et al., 2009). After considering different perspectives from the scientific data to clinical perspectives, the authors finally incorporated all of these perspectives to develop a protocol to evaluate and document auditory-perceptual features of voice. This led to the development of the CAPE-V.

Administration of the CAPE-V

To administer the CAPE-V, the individual to be assessed is asked to sustain two vowels (/a/, /i/) for 3-4 seconds, read six sentences in English, and answer a standard question for a continuous speech sample. Vowel prolongations make it possible to evaluate voice without any articulatory influence. The sentences in English are constructed in a way that each sentence has a particular phonetic characteristic. The first sentence (The blue spot is on the key again) includes all the vowel sounds in English. The second sentence (How hard did he hit him?) provides a context of glottal sounds for easy onset. The third sentence (We were away a year ago) consists of all voiced sounds and judges an individual's ability to maintain voicing across words. The fourth sentence (We eat eggs every Easter) gives an opportunity to elicit hard glottal attacks. The fifth sentence (My mama makes lemon muffins) has many nasal consonants thus providing an opportunity to check for hyponasality. The last sentence (Peter will keep at the peak) provides an opportunity to check for intraoral pressure since the sentence is loaded with voiceless plosives. The conversational task provides a natural sample for analyses.

The six features of vocal quality assessed are overall severity (overall impression of voice), roughness (irregularity in voice), breathiness (air escape that is audible), strain (perceived hyperfunction, excessive tension), pitch (the perceptual correlate of frequency), loudness (the perceptual correlate of intensity). These features are assessed with a visual analog scale that includes a line that is 100 mm in length. The extreme left represents normal and the extreme right represents very severe dysphonia. The clinician marks on the line according to the perceived severity for each of the six features which makes it easier to rate the client's voice. Additionally, the clinician marks 'I' for inconsistent presence of a feature or 'C' for consistent

presence of a feature of voice quality. Finally, an additional space at the bottom of the form allows the clinician to note any other significant feature in the voice.

Utility of this scale is often compared with the GRBAS. The GRBAS is an ordinal scale with a limited choice for rating; therefore, it is limited in its ability to identify small variations in voice quality (Nemr et al., 2012; Wuyts et al., 1998). This may frustrate the raters because of less degrees of freedom in choosing the severity level, and this calls for a need to introduce more subdivisions in the scale to indicate the severity of the disorder more accurately and on a continuum. The visual analog scale in the CAPE-V provides an opportunity to mark and identify small changes in voice quality. The CAPE-V is found to be more sensitive in determining dysphonia because of this continuous nature than the GRBAS, an ordinal scale with only four choices (normal, mild, moderate, severe), which limits the opportunities to apply statistical operations on collected data (Kempster et al., 2009). Other advantages of using a visual analog scale in CAPE-V include the same predetermined tasks to serve as samples for judgement across participants, a detailed and standard administration protocol, subjective evaluation of pitch and loudness in addition to other qualitative measures (Nemr et al., 2012), and an assessment of the voice in three different contexts. It has been reported that visual scales measure multidimensional features better than ordinal scales. (Chan & Yiu, 2002). Thus, CAPE-V can be considered as an effective clinical tool to evaluate auditory-perceptual features of voice. A number of studies are present where the researchers assessed the validity and reliability of this test and transadapted it in different languages.

Psychometric analyses and transadaption into other languages

In a study done by Zraick et al. (2011) that aimed to determine the reliability of the CAPE-V, it was found that the reliability was the same as or exceeded the reliability of the GRBAS scale. Intrarater reliability coefficients for the CAPE-V ranged from r =.35 for strain to r = .82 for breathiness. More than half of the raters achieved a reliability rating of r =.70 or greater in at least three measures (breathiness, roughness, and pitch). The other three measures (loudness, overall severity, and strain) had lower *r*-values and were reported to be difficult for reliable use by the raters. Intrarater reliability for GRBAS ranged from r = .53 (strain) to r = .69 (asthenia), and less than half of the raters achieved a score of r = .70 or greater in any of the six measures. In the same study, the interrater reliability coefficient of the CAPE-V ranged from .28 (pitch) to .76 (overall severity). Overall, the CAPE-V had improved interrater and intrarater reliability than the GRBAS, and thus it was reported as a potential tool to be used for clinical perceptual voice assessment.

The CAPE-V has been used and adapted into different languages including Turkish (Özcebe et al., 2017), Spanish (Núñez-Batalla, Morato-Galán, García-López, & Ávila-Menéndez, 2014), Portuguese (de Almeida, Mendes, & Kempster, 2018), and Mandarin (Chen, Fang, Zhang, Ge, Zhuang, Chou, & Jiang, 2018). In the study that adapted CAPE-V into Turkish, the intra-rater and inter-rater reliability coefficients were greater than 0.82 for all six components, for all four raters, indicating a strong correlation. In addition to this, classification results showed that the sensitivity was 83%, the specificity was 89%, the positive predictive value was 93%, and the negative predictive value was 75%. Finally, concurrent validity of the test was obtained by comparing results of the CAPE-V and the GRBAS on the same participants. The correlation coefficient was in the range of moderate (0.62) to high

(0.80) correlation (Özcebe et al., 2017). Núñez-Batalla et al. (2014) developed the Spanish version of the CAPE-V and found that the intra-observer reliability was very high for all the parameters (Intraclass Correlation Coefficient, ICC > .85). Inter-observer reliability coefficient ranged from 0.54-0.93, but most of the components (12 out of 18) had ICC > 0.75. Validity was assessed by comparing the results of the CAPE-V and GRBAS, and all of the parameters except breathiness had strong correlations. The authors concluded that this version could be used as an individual tool to assess voice by auditory-perceptual judgment. A comparison was done between the CAPE-V and the GRBAS scores for the voice of the subjects. The CAPE-V was found to be more sensitive to change in all the parameters of vocal quality. de Almeida et al. (2018) performed a similar study by adapting the CAPE- V in European Portuguese. Inter-rater reliability was found to be high (ICC > .84) for all the six parameters, intra rater reliability was moderate (r > .61) to high (r > .87) for all the parameters. Additionally, concurrent validity analysis revealed a high correlation between subscales of GRBAS and CAPE- V results (r > .89) except strain (r = 0.47), suggesting that CAPE- V can be used for perceptually evaluating voice. In the study done by Chen et al. (2018), intrarater reliability was found to be excellent for the overall correlation between the six parameters of Mandarin CAPE-V. Inter-rater reliability was found to be excellent for overall severity, roughness, and breathiness; good for loudness (ICC= 0.60- 0.74); and fair for strain and pitch (ICC= 0.40-0.59). For the common parameters in CAPE-V and GRBAS, interrater reliability was found to be more in CAPE-V than in GRBAS. All the above studies show that these adaptations of CAPE-V are valid and reliable to use in clinics for the auditory perceptual evaluation of voice of people speaking these languages. In a similar

manner, this current work aims at adapting the test to Hindi since there is no equivalent Hindi adaptation of the CAPE-V.

Perceptual assessment in India/ Hindi

According to the 2001 Census, 53.6% of the Indian population spoke Hindi as their first or second language and 41% people were native speakers of Hindi. Since Hindi is one of the most widely spoken languages in India (Census 2001), an assessment tool that evaluates the voice quality of Hindi speakers will serve a large population. A review of the literature revealed that perceptual voice assessment tools in Hindi are scarce. Most of the clinical studies with auditory-perceptual assessments have been done using the GRBAS (Balasubramanium, Karuppali, Bajaj, Shastry, & Bhat, 2018; Boominathan, Anitha, Shenbagavalli, & Dinesh, 2010; Fernandes, Balasubramanium, Pitchaimuthu, & Bhat, 2014; Jaykumar & Savitri, 2012; Munjal, Alam, & Panda, 2018). There is no language specific test available for clinicians to evaluate the voice of Hindi speakers perceptually. In India, the CAPE-V can be used only for the 10-12% of the general population that speaks English; primarily in the urban areas (Census 2001). The GRBAS rating scale is used for the remainder of the population. Since the CAPE-V is believed to be an important tool in the voice assessment battery, a Hindi adaptation of the CAPE-V will make this tool accessible to a larger clinical population.

Purpose of the study

The purpose of the current study was to develop the Hindi version of the CAPE-V adapting it to cultural and linguistic demands. The newly created Hindi stimuli were compared to the existing English stimuli for measures of reliability and validity. The same stimuli were used to rate the voice quality using GRBAS and the CAPE-V.

METHODOLOGY

Participants

Participants fulfilling the following requirements were included in the study: Adult males and females, over 18 years of age, with a normal voice quality, fluent in English and Hindi, non-smokers, with no history of hearing, speech and language disorders. Participants were included in the study based on their self-report and perceptual assessment of voice quality by the study personnel. Individuals who did not fulfill the inclusion criteria and spoke only one of the two languages were excluded. Thirty-one participants (16 males, 15 females) with normal voice quality and an age range of 21 years-33 years were recruited for this study. For the disordered voice sample, thirteen participants (10 males, 3 females) were recruited from Sri Aurobindo Institute of Medical Sciences, India. Table 1 shows the details of 13 voice disorder patients with a voice disorder in the age range of 19-78 years. Patients were diagnosed by an otolaryngologist as having dysphonia (n=4), vocal fold polyp (n=2), presbyphonia (n=2), vocal fold injury (n=1), vocal fold nodule (n=1), glottal stenosis (n=1), papilloma (n=1), and leukoplakia (n=1).

Previous research studies using the CAPE-V, including the original study, have had a similar sample size (20-40 participants) to have adequate power. Potential participants were recruited via word of mouth and e-mail. These participants included students, staff, and members of the Indian Students Association, as well as others in the community. E-mails were sent out via list-servs and through the contact person at the department/association.

Table 1: Demographic details of subjects with disordered voice

Participant	Sex	Age	Diagnosis
1	M	47	Vocal fold polyp
2	M	26	Papilloma
3	M	27	Dysphonia
4	M	19	Dysphonia
5	M	27	Dysphonia
6	M	45	Vocal fold polyp
7	F	59	Vocal fold injury
8	M	22	Dysphonia
9	M	70	Presbyphonia
10	F	19	Glottal stenosis
11	F	52	Vocal fold nodule
12	M	63	Leukoplakia
13	M	78	Presbyphonia

Development of the Hindi CAPE-V sentences

The sentence types in the Hindi CAPE-V were in keeping with the original CAPE-V and include sounds loaded with vowels, easy vocal fold onsets, voiced consonants, hard glottal attacks, nasal sounds and voiceless plosives. The developers of the English CAPE-V did not specify criteria when designing the stimuli, besides targeting certain sounds. In the absence of these guidelines, the percent of target sounds was calculated for each of the English sentences, and this number was matched in the Hindi sentences. The following are the original English sentences and their counterpart Hindi sentences:

1. Vowel sentence

English: The blue spot is on the key again.

Hindi: आज कई दिनों के बाद सौ रुपए मिले। (/ɑdʒ kaɾ̪ ðɾ̪ɲõ ke bað sɔ rupɛ mɪle/)

2. Easy Onset

English: How hard did he hit him?

Hindi: हमारे हाथों में हीरा नहीं है। (/ɦemɑɾe haθõ mẽ hiɾɑ ɲəhĩ hæ̃/)

3. Voiced Sentence

English: We were away a year ago.

Hindi: यहाँ बादल गरज रहे है। (/jəhɑ baðəl gəɾədʒ ɾəhẽ hæ̃/)

4. Hard glottal attack

English: We eat eggs every Easter.

Hindi: इतनी बड़ी एक ईमारत इधर खड़ी है। (/ɪt̪ɲi bəɖi imɑɾət̪ iðʰər kʰəɖi hæ/)

5. Nasal sentence

English: My mama makes lemon muffins.

Hindi: मेरे मामा ने मुझे मिठाई दी। (/mere mɑmɑ ɲe mʊʊdʒʰe mɪt̪ʰaɾ̪ ðĩ/)

6. Voiceless Plosives

English: Peter will keep at the peak.

Hindi: प्रीती के पास पता है। (/pɾiʈi ke pɑs pəʈɑ hæ/)

7. Spontaneous speech

English: Tell me about your family.

Hindi: अपने दिन के बारे में बताइये। (/ʌpne ðɪn ke bare mẽ bʌtaɪe/)

PROCEDURE

This study was approved by the Institutional Review Board at the University of Houston. This was a prospective cohort design and participants were involved in a one-time assessment of their voice quality. The following tasks were conducted and recorded (using CSL 4500, MDVP for normative sample and Merantz 6000 for disordered voice sample) by study personnel:

Sustained vowel: The participants were required to sustain the vowels /a/ and /i/ for 3-5 seconds.

Sentences: The participants with normal voice completed the sentences and spontaneous speech tasks for the original CAPE-V in English (Appendix 1). Participants from both the groups (normal and disordered voice) read six Hindi sentences.

Spontaneous Speech: Participants with normal voice were required to produce a minimum of two spontaneous sentences in Hindi and English, while participants with a disordered voice were required to produce two spontaneous sentences in Hindi.

The Hindi stimuli were rated on a separate scoring sheet but with the original CAPE-V blank scoring form/visual analog scale. The English and Hindi recorded samples were coded separately to keep each sample independent of the other. The study personnel rated each of these samples using the original CAPE-V rating scale for both English and Hindi. The order of tasks was counterbalanced to prevent an order effect.

To assess the concurrent validity of the test, GRBAS was administered on all the voice samples. Data collected include participant sex, age, and recording of the

above stimuli. All data were analyzed in the voice lab in the Dept. of Communication Sciences and Disorders, 119 Clinical Research Services at the University of Houston.

DATA ANALYSES

A graduate student clinician and an experienced speech language pathologist rated the recorded voice samples. English and Hindi samples of the same participant were not linked and were randomly assigned to the raters to minimize bias. All data were entered in an Excel spreadsheet and SPSS 24.0 (IBM Corp. Released 2016) was used to perform statistical analysis. Pearson's Product Moment correlation was performed to assess the strength of the relationship between the English and Hindi CAPE-V scores, and between the common parameters of the Hindi CAPE-V and GRBAS. Inter-rater reliability was also measured using Pearson's correlation.

RESULTS:

Normal voice quality

Mean scores of all the parameters of the English and Hindi versions of the CAPE-V are shown in Table 2. The mean overall severity was 3.87 (SD = 5.638) in Hindi version and 4.65 (SD = 5.930) in English version of CAPE-V.

Table 2: CAPE-V scores for normal subjects in Hindi and English

	Hindi		English	
	Mean	SD	Mean	SD
Overall Severity	3.87	5.63	4.65	5.93
Roughness	3.77	4.74	5.52	4.93
Breathiness	1.16	4.37	0.84	3.72
Strain	0.39	1.14	0.61	1.54
Pitch	0.29	1.13	0.16	0.90
Loudness	0.00[a]	0.00	0.00[a]	0.00

a. The correlation and t cannot be computed because the standard error of the difference is zero.

Correlation

Table 3 shows the correlational analysis of parameters of the perceptual voice analysis using English and Hindi Versions of CAPE-V. Overall severity, roughness, breathiness, and pitch showed a significantly strong correlation($r > 0.7$, $p < .01$) between both the versions. Strain showed a weak correlation between the two versions ($r < .3$, $p > .01$). Loudness ratings for all individuals for both Hindi and English stimuli were normal, leading to a score of 0 for this variable. Hence correlational analyses could not be performed for the loudness variable and is not listed in the table below.

Table 3: Correlation between English and Hindi CAPE-V scores in normal subjects

Parameter	Pearson Correlation	Significance level
Overall Severity	0.819**	P<.01
Roughness	0.836**	p <.01
Breathiness	0.883**	p <.01
Strain	0.144	p=.439
Pitch	0.773**	p <.01

**. Correlation is significant at the 0.01 level (2-tailed)

Hindi CAPE-V and GRBAS: To determine the concurrent validity, scoring for the perceptual parameters from Hindi version were compared to the scores of equivalent parameters in GRBAS. Hindi version of CAPE-V and GRBAS were administered on the same voice sample, but at different points of time. As seen in Table 4, overall severity, roughness, and breathiness in Hindi version of CAPE-V strongly correlate with grade, roughness, and breathiness in GRBAS ($r > .6$ $p < .01$).

Table 4: Correlation of Hindi CAPE-V and GRBAS scoring in normal subjects

CAPE-V	GRBAS	Pearson Correlation	Significance level
Overall severity	Grade	.751**	<.01
Roughness	Roughness	.758**	<.01
Breathiness	Breathiness	.632**	<.01

**. Correlation is significant at the 0.01 level (2-tailed).

Interrater reliability

Interrater reliability was established by comparing the CAPE-V score by a graduate student clinician and a CCC-SLP for 12 randomly selected samples in both the languages. The correlation scores for overall severity, roughness, and breathiness can be seen in table 5.

Table 5: Interrater reliability for Consensus of Auditory Perceptual Evaluation of Voice (CAPE-V)

Parameter	Pearson's coefficient
Overall severity	.829**
Roughness	-.369
Breathiness	.921**

**. Correlation is significant at the .01 level (2-tailed).

Disordered Voice quality

The following table shows mean scores for overall severity/grade, roughness, breathiness, and strain when the Hindi version of the CAPE-V and GRBAS were administered on the disordered voice samples from 13 individuals.

Table 6: GRBAS and Hindi CAPE-V scores in disordered voice samples

Parameter	CAPE-V		GRBAS	
	Mean	SD	Mean	SD
Overall severity/Grade	36.615	32.737	1.385	1.261
Roughness	14.231	24.830	.692	.947
Breathiness	25.769	30.698	.923	1.038
Strain	22.308	26.631	.923	.954

Correlation

Correlational analyses of the common parameters of the Hindi CAPE-V and GRBAS were performed using Pearson's correlation. There was a strong, significant correlation for all parameters i.e., overall severity, roughness, breathiness, and strain.

Table 7: Correlation between Hindi CAPE-V and GRBAS scores in disordered voice sample

CAPE-V	GRBAS	Pearson Correlation	Significance level
Overall severity	Grade	.955**	<.01
Roughness	Roughness	.843**	<.01
Breathiness	Breathiness	.865**	<.01
Strain	Strain	.778**	<.01

**. Correlation is significant at the 0.01 level (2-tailed).

DISCUSSION

In addition to the description of quantitative variables of voice, the assessment of quality of a person's voice is equally important (Zraick et al., 2011). In the multidimensional assessment of voice, auditory-perceptual evaluation plays a significant part in describing a patient's voice quality because of the easy and efficient procedure (Barsites & Bodt, 2014). Assessment of voice quality also helps in determining the direction of voice therapy and setting goals for treatment. The CAPE-V is one clinical tool used for auditory-perceptual evaluation of voice. A number of studies have been performed to determine the validity and reliability of the test. The CAPE-V was found to be equally reliable or more reliable than the GRBAS in rating the severity of voice disorders in a study done by Zraick et al. (2011). The CAPE-V has been adapted in a Turkish, Portuguese, Spanish, and Mandarin with significant reliability and validity scores (de Almeida et al., 2018; Núñez-Batalla et al., 2014 Özcebe et al., 2017).

The aim of this study was to develop a reliable Hindi version of the CAPE-V as a clinical tool for voice assessment of Hindi speakers. The Hindi version was developed by taking the English version of the CAPE-V as the reference. It required more than a direct translation of the English version because of the linguistic differences between the languages. The stimuli in the Hindi version were constructed to account for the characteristics of the Hindi language. The six sentences in the original CAPE-V were each loaded with a particular group of sounds- vowels, glottal sounds, plosive sounds, all voiced sounds, easy onsets and nasals. All the sentences in Task 2 of the Hindi version were designed with the same phonetic focus but with semantic and linguistic content grammatically and culturally appropriate for the Hindi speaking population. For example, the sentence "My mama makes lemon muffins" is

not appropriate for Hindi speakers because the word “muffin” is not a commonly used word in Hindi. The sentence “We eat eggs every Easter” is culturally biased, and not all Hindi speakers would relate to “Easter”. Thus, all the sentences were culturally and linguistically adapted while serving the purpose of each sentence in assessing different functions of the vocal tract and maintaining the phonetic content at the same time. While developing the sentences, care was taken to maintain the proportion of the target sounds in each sentence between the English and Hindi version to keep the Hindi version as authentic to the original CAPE-V as possible. The third task required a continuous speech sample and the original prompt in CAPE-V was “Describe your voice problem”. Since the present study included participants with a normal voice quality, the prompt question was changed to “Tell me about your family” in English and “Tell me about your day” in Hindi.

Sixteen males and 15 females with normal voice quality, and 10 males and 3 females with disordered voice quality participated in the present study. Participants with a normal voice completed the English and Hindi versions of the CAPE-V. The results compared the ratings on these versions and there was a strong correlation between the scores in overall severity, roughness, breathiness, and pitch. A significant difference was seen in the mean scores for strain in the English and the Hindi CAPE-V. GRBAS was also administered on the Hindi speech samples and there was a strong correlation found between the Hindi GRBAS scores and the Hindi CAPE-V scores in overall severity, roughness, and breathiness. The range of strain scores was 0-6 in the English version, and 0-7 in the Hindi version of CAPE-V. Since these subjects had normal voice quality, the strain scores were very low (within normal limits) in both the versions of CAPE-V. In the disordered voice group, Hindi CAPE-V and GRBAS

were administered by an experienced speech-language pathologist. A strong correlation was observed in overall severity, breathiness, roughness, and strain scores.

Reliability is the measure of consistency and accuracy of scores obtained after administering a test (Paul & Norbury, 2012). Interrater reliability measures the consistency of scores given by two or more different examiners administering the same test on the same subject/sample. According to the guidelines by Cicchetti and Sparrow (1981), it can be stated that the scores indicate significant excellent interrater reliability ($r > .75$) in overall severity and breathiness while non-significant ($p > .01$) reliability in roughness scores. In the current study, two different examiners used the Hindi version of CAPE-V for auditory perceptual assessment of normal voice samples. A strong correlation was observed in overall severity and breathiness, while a negative weak correlation was found between roughness scores by the examiners. This difference can be attributed to differences in listeners' expertise in perceptual analysis of voice.

Validity refers to the extent to which a test measures the behavior/quality that it claims to measure. Content validity refers to the extent to which the items on a test are fairly representative of the entire domain the test seeks to measure. Content validity of the CAPE-V Hindi was maintained by ensuring that the stimuli fulfil the purposes and maintain phonetic content. A Hindi linguistic expert was involved in ensuring the content validity of the test. Concurrent validity refers to the extent to which the scores determined by a test match the scores by an established test. For concurrent validity, a strong correlation was found between Hindi CAPE-V scores and English CAPE-V scores in overall severity, roughness, breathiness, and pitch ($r > 0.5$, $p < .01$) in normal subjects. A weak correlation was found between the English and Hindi versions for the variable of strain ($r < .3$, $p = .439$) in the normative group.

The GRBAS and CAPE-V were administered on the samples but at different points of time. Both the tests were administered at least a week apart, to prevent cross over effect. Strong correlation ($r > .5$, $p < .01$) was found between the overall severity/grade, roughness, and breathiness scores in GRBAS scale and CAPE-V scale in normal and disordered voice samples. This is consistent with the findings in other studies that involved adaptation of CAPE-V in other languages (Chen et al., 2018; de Almeida et al, 2018; Özcebe et al., 2017) and comparison between GRBAS scores and CAPE-V scores (Zraick et al., 2011).

Limitations and future implications

The disordered group was limited by the small sample size (n= 13). A larger sample size across different disorders would be able to better demonstrate the validity of the stimuli. There are multiple known dialects of Hindi and the multilingual nature of the Indian subcontinent leads to various Hindi accents. The effects of the various accents on the CAPE-V scores should be studied to determine the consistency of the stimuli across accents. The current version of the Hindi CAPE-V includes instructions and scoring in English with only the stimuli in Hindi. This was done keeping in mind that all SLPs in India complete their education in speech language pathology in English and have adequate English reading skills. However, to make the Hindi version more accessible to this population the entire form including the instructions and scoring will be translated into Hindi.

CONCLUSION

This is the first known study and Hindi version of the CAPE-V. It provides culturally and linguistically appropriate stimuli to assess the voice quality of the Hindi speaking population. These Hindi stimuli were found to be reliable and valid and, will prove to be a valuable tool for the auditory-perceptual evaluation of voice quality of Hindi speakers. The degree of interrater reliability, concurrent validity, and construct validity have been shown to be in agreement with other studies in the literature and comparable to GRBAS, the measure currently used as standard of practice. Future studies will aim refine the protocol to adjust for any effects of accent and dialect.

REFERENCES

de Almeida, S. C., Mendes, A. P., & Kempster, G. B. (2018). The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) Psychometric Characteristics: II European Portuguese Version (II EP CAPE-V). *Journal of Voice*.

American Speech-Language-Hearing Association, & American Speech-Language-Hearing Association. (2002). Consensus auditory-perceptual evaluation of voice (CAPE-V). *Rockville: ASHA Special Interest Division, 3*.

Balasubramanium, R. K., Karuppali, S., Bajaj, G., Shastry, A., & Bhat, J. (2018). Acoustic-Perceptual Correlates of Voice in Indian Hindu Purohits. *Journal of Voice.* https://doi.org/10.1016/j.jvoice.2018.03.006.

Barsties, Ben, and Marc De Bodt (2015). "Assessment of Voice Quality: Current State-of-the-Art." *Auris, Nasus, Larynx* 42 (3): 183–88.

Behlau, M., Zambon F., Moreti F., Oliveira G., & Couto Jr., E.B.2017. "Voice Self-Assessment Protocols: Different Trends Among Organic and Behavioral Dysphonias." *Journal of Voice: Official Journal of the Voice Foundation* 31 (1): 112.e13–112.e27.

Boominathan, P., Anitha, R., Shenbagavalli, M., & Dinesh, G. (2010). Voice characteristics and recovery patterns in indian adult males after vocal loading. *Journal of the All India Institute of Speech & Hearing*, *29*(2).

Carding, P., Wilson, J., MacKenzie, K., & Deary, I. (2009). Measuring voice outcomes: State of the science review. *The Journal of Laryngology & Otology, 123*(8), 823-829. doi:10.1017/S0022215109005398

Chen, Z., Fang, R., Zhang, Y., Ge, P., Zhuang, P., Chou, A., & Jiang, J. (2018). The Mandarin Version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) and Its Reliability. *Journal of Speech, Language, and Hearing Research, 61*(10), 2451-2457.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.

Fernandes, J. B., Balasubramanium, R. K., Pitchaimuthu, A. N., & Bhat, J. S. (2014). Nonlinear dynamic analysis of voice&58; A normative study in the Indian population. *International journal of medical research and health sciences*, *3*(1), 128-132.

IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

Jayakumar, T., & S. R. Savithri. 2012. “Effect of Geographical and Ethnic Variation on Dysphonia Severity Index: A Study of Indian Population.” *Journal of Voice: Official Journal of the Voice Foundation* 26 (1): e11–16.

Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, *18*(2), 124-132.

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.

Munjal, S., Alam, M. N., & Panda, N. K. (2018). Subjective Evaluation of Voice Characteristics of School Aged Children in a Basket Ball Team. *Indian Journal of Otolaryngology and Head & Neck Surgery*, 1-4.

Nemr, K., Simoes-Zenari, M., Cordeiro, G. F., Tsuji, D., Ogawa, A. I., Ubrig, M. T., & Menezes, M. H. M. (2012). GRBAS and Cape-V scales: high reliability and consensus when applied at different times. *Journal of Voice*, *26*(6), 812-e17.

Núñez-Batalla, F., Morato-Galán, M., García-López, I., & Ávila-Menéndez, A. (2015). Validation of the Spanish adaptation of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Acta Otorrinolaringologica (English Edition)*, *66*(5), 249-257.

Oates, J. 2009. “Auditory-Perceptual Evaluation of Disordered Voice Quality.” *Folia Phoniatrica et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics* 61 (1): 49–56.

Özcebe, E., Aydinli, F.E., Tiğrak, T.K., İncebay, Ö., & Yilmaz, T. 2018. “Reliability and Validity of the Turkish Version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V).” *Journal of Voice: Official Journal of the Voice Foundation*, January. https://doi.org/10.1016/j.jvoice.2017.11.013.

Paul R. and Norbury C. (2014). *Language Disorders from Infancy through Adolescence: Listening, Speaking, Reading, Writing, and Communicating.* Missouri: Elsevier.

Rogerson-Revell, P. (2011). *English phonology and pronunciation teaching*. New York, NY:Bloomsbury Publishing

Tiwari, M., & Tiwari, M. 2012. “Voice - How Humans Communicate?” *Journal of Natural Science, Biology, and Medicine* 3 (1): 3–11.

Wuyts, F. L., De Bodt, M. S., & Van de Heyning, P. H. (1999). Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *Journal of voice, 13*(4), 508-517.

Ziwei, Y., Zheng, P., & Pin, D. 2014. “Multiparameter Voice Assessment for Voice Disorder Patients: A Correlation Analysis between Objective and Subjective Parameters.” *Journal of Voice: Official Journal of the Voice Foundation* 28 (6): 770–74.

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., ... & Glaze, L. E. (2011). Establishing validity of the consensus auditory-perceptual evaluation of voice (CAPE-V). *American Journal of Speech-Language Pathology, 20*(1), 14-22.

Appendix 1

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

Name:_____ **Date:**_____

The following parameters of voice quality will be rated upon completion of the following tasks:
1. Sustained vowels, /a/ and /i/ for 3-5 seconds duration each.
2. Sentence production:
   a. The blue spot is on the key again.
   b. How hard did he hit him?
   c. We were away a year ago.
   d. We eat eggs every Easter.
   e. My mama makes lemon muffins.
   f. Peter will keep at the peak.
3. Spontaneous speech in response to: "Tell me about your voice problem." or "Tell me how your voice is functioning."

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

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

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

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

Clinician:_____