Leonardo Wanderley Lopes[1]
Estevão Silvestre da Silva Sousa[2]
Allan Carlos França da Silva[2]
Itacely Marinho da Silva[2]
Maxsuel Avelino Alves de Paiva[2]
Vinícius Jefferson Dias Vieira[3]
Anna Alice Almeida[1]

Cepstral measures in the assessment of severity of voice disorders

Medidas cepstrais na avaliação da intensidade do desvio vocal

ABSTRACT

Purpose: To analyze whether there is an association between the presence, intensity and type of voice disorder and the cepstral measures in samples of individuals with voice complaints. **Methods:** We used 376 vowel /ɛ/ samples from individuals of both genders that had voice complaints. An analogue-visual scale was used for the auditory-perceptual analysis of voices regarding the overall grade of dysphonia (G) and the grades of roughness (R), breathiness (B), and strain (S), including a determination of voice quality (rough, breathy or strained). Measures related to cepstral peak prominence smoothed (CPPS) and spectral decline of vocal samples were extracted. **Results:** There were differences in the CPPS values between the groups with or without voice disorders as well as between the different intensities and types of voice disorder. CPPS values were lower because of the presence and intensity of voice disorders. The CPPS values differentiated the following voices: rough x breathy, rough x strained, and breathy x strained. The spectral decline only differentiated breathy x strained voices. CPPS correlated positively and strongly with G and B; moderately and negatively with R, and negatively and weakly with S. The spectral decline had a moderate positive correlation with S and a weak negative correlation with B. **Conclusion:** There is association between voice disorder, G, predominant voice quality, and CPPS. In particular, G is strongly correlated with CPPS. Spectral decline is associated only with the parameters B and S.

RESUMO

Objetivo: analisar se existe associação entre a presença, a intensidade e o tipo de desvio vocal e as medidas cepstrais em amostras de indivíduos com queixa vocal. **Método:** Foram utilizadas 376 amostras da vogal /ɛ/ de indivíduos de ambos os sexos, com queixa vocal. Utilizou-se uma escala analógico-visual para análise perceptivo-auditiva das vozes quanto à intensidade do desvio vocal (GG), graus de rugosidade (GR), soprosidade (GS) e tensão (GT), incluindo-se a determinação da qualidade vocal predominante (rugosa, soprosa ou tensa). Foram extraídas as medidas relacionadas ao *Cepstral Peak Prominence-Smoothed* (*CPPS*) e o declínio espectral das amostras vocais. **Resultados:** Houve diferença dos valores do *CPPS* entre os grupos com e sem desvio vocal, assim como entre as diferentes intensidades e tipos de desvio vocal. Os valores do *CPPS* foram mais reduzidos em função da presença e intensidade do desvio vocal. Os valores do *CPPS* diferenciaram vozes rugosas x soprosas, rugosas x tensas e soprosas x tensas. O declínio espectral apenas diferenciou vozes soprosas x tensas. O *CPPS* se correlacionou de modo positivo e forte com os GG e GS, de modo negativo moderado com o GR, e de forma negativa fraca com o GT. O declínio espectral apresentou correlação positiva moderada com o GT e correlação negativa fraca com o GS. **Conclusão:** Existe associação entre a presença de desvio vocal, o GG, a qualidade vocal predominante e o *CPPS*. De modo especial, o GG é fortemente correlacionado ao *CPPS*. O declínio espectral está associado apenas aos parâmetros de soprosidade e tensão.

Correspondence address
Leonardo Wanderley Lopes
Departamento de Fonoaudiologia,
Centro de Ciências da Saúde
Cidade Universitária - Campus I,
Bairro Castelo Branco, João Pessoa
(PB), Brasil, CEP: 58051-900.
E-mail: lwlopes@hotmail.com

INTRODUCTION

Evaluation of voice disorders should take into account a multidisciplinary and multidimensional approach[1], including a detailed anamnesis to identify the risk factors and symptoms of the complaints; an auditory-perceptual analysis to identify the presence, type and intensity of voice disorder; a quantitative and qualitative acoustic evaluation of the voice signal; an aerodynamic assessment of the data on airflow control for phonation; a structural and functional visual examination of the larynx[1,2].

Auditory-perceptual analysis and visual laryngeal examination are the main methods used by speech-language pathologists and otorhinolaryngologists to evaluate voice disorders, respectively. Both methods have confounding factors associated with the subjectivity of the evaluator[3] who either makes an auditory judgment of voice quality (auditory-perceptual assessment) or a visual judgment based on the laryngeal examination.

Acoustic analysis is complementary to auditory-perceptual and laryngeal evaluation, providing quantitative and qualitative data on vocal function and presenting high reproducibility for patient monitoring[4,5]. One of its relevant aspects is the possibility of quantifying the disorder present in the signal and comparing it with normative data[6]. Validity of the acoustic measures depends on their capacity to represent the voice quality disorder that is aurally perceived and the physiological mechanisms underlying voice production. Thus, one of the challenges for clinicians and researchers is to understand to what extent each measure is associated with the auditory-perceptual assessment and visual laryngeal examination.

In general, acoustic analysis may involve extraction of measures that quantify a specific characteristic of the voice signals and a descriptive analysis of their visual patterns[7,8]. In extracting the classical measures of perturbation (jitter and shimmer) and noise (harmonics-to-noise ratio), it is necessary to estimate the fundamental frequency (F_0) values with clear determination of the glottic cycles, which is more likely be performed only on voice signals with mild disorders[1,9].

In dysphonic individuals, voice signals can range from almost periodic to completely aperiodic, so that the complexity of a signal with moderate and severe disorders may compromise the reliability of traditional measures based on linear models such as jitter and shimmer[1]. Thus, although these traditional measures show a moderate-to-strong correlation with the auditory perception of voice disorders[10], they may have a restricted application in the analysis of voices with more severe disorders.

In turn, cepstral analysis has proved to be an alternative for the evaluation of signals with greater deviation, because it is able to determine the F_0 and produce estimates of aperiodicity and/or additional noise without the identification of individual cycle thresholds, as recommended in the extraction of perturbation and noise measures[11]. In general, the cepstrum shows the extent to which the harmonics from F_0 are individualized and stand out in relation to the noise level present in the signal. Signals with greater regularity and less noise present greater definition and amplitude of the dominant cepstral peak[11]. Thus, cepstral measures are more reliable than the traditional perturbation and noise measures for the evaluation of voices with a wide range of disorders. Moreover, they have been shown to be strong predictors of the presence of voice disorder[11-13].

In this context, the objective of the present study was to analyze whether there is an association between the presence, intensity and type of voice disorder and the cepstral measures in samples of individuals with voice complaints.

METHODS

Study design

This descriptive, cross-sectional and observational study was approved by the Research Ethics Committee of the aforementioned Institution under opinion no. 52492/12.

Study sample

The study sample was composed of 376 patients with voice complaints, of both genders, who were assisted at the voice laboratory of the aforementioned Institution. All participants signed an Informed Consent Form (ICF) prior to study commencement.

For sample selection, the following eligibility criteria were considered: voice complaint verified by the positive response to the following question: "Do you consider that you have a voice problem currently?"; visual laryngeal examination for diagnostic confirmation of voice disorder within two weeks prior to or following the data collection session; no cognitive or neurological impairment that would prevent voice recording; absence of previous voice therapy or surgical treatment of the larynx.

A total of 376 individuals, 294 women and 82 men, with mean age of 41.20 ±14.04 years were selected. These patients presented the following medical diagnoses: 99 (26.30%) individuals without structural or functional changes in the larynx, 90 (23.90%) with vocal nodules, 42 (11.20%) with speech disorder secondary to laryngopharyngeal reflux, 38 (10.10%) with vocal cyst, 25 (6.66%) with mid-posterior triangular flap in cleft lip, 22 (5.85%) with unilateral vocal fold paralysis, 21 (5.60%) with voice disorder secondary to neuromuscular disease, 20 (5.30%) with vocal fold polyp, 11 (2.90%) with sulcus vocalis, and 8 (2.10%) with Reinke's edema.

All patients either sought assistance spontaneously or were referred by an otorhinolaryngologist; they were evaluated prior to voice therapy. Patients with voice disorder secondary to neuromuscular disease also presented a neurology medical report. Thus, all participants had voice complaints and received diagnostic confirmation of voice disorder by visual laryngeal examination. Considering the objective of this study, auditory-perceptual evaluation was chosen as reference standard for determining the outcome (presence/absence of voice disorder, intensity of disorder, and predominant voice quality), regardless of the outcome of visual laryngeal examination.

Data collection procedures

All data were collected in the voice laboratory of the aforementioned higher education Institution. Initially, patients filled in a form on demographic data and information on voice

complaints. Subsequently, they were submitted to recording of the sustained /ɛ/ vowel.

Voice collection was performed in laboratory using Fonoview, 4.5 (CTS, Informática) software, an all-in-one desktop computer (Dell, Inc.), and a unidirectional cardioid microphone (Senheiser, model E-835) on a stand and connected to a preamplifier (Behringer, U-Phoria UMC 204). The voices were collected in a recording booth with acoustic treatment, noise level <50 dB NPS, sampling rate of 44000 Hz, 16 bits per sample, at a distance of 10 cm between the microphone and the patient's mouth.

Voice collection occurred with patients standing in front of the stand at the recommended distance between mouth and microphone. The patients received the voice collection instructions, and the voice was recorded soon after. During recording, the patients were asked to emit the sustained /ɛ/ vowel at their normal frequency and intensity. The /ɛ/ vowel was selected for this study because it is an oral, open, unrounded vowel sound, and is considered to be the vowel with the most mid-position in Brazilian Portuguese, which allows a more neutral and intermediate position of the vocal tract. In addition, it is the vowel most commonly used for the evaluation of voice quality in Brazil.

Subsequently, the voices were edited using SoundForge 10.0 software; the two initial and final seconds of the sustained /ɛ/ vowel emission were deleted because of the greater irregularity existing in these sections, and a minimum time of three seconds for each emission was preserved.

Extraction of the acoustic measures was performed using the free-access Praat 5.3.84 (Paul Boersma and David Weenink, University of Amsterdam, The Netherlands) software,, and the cepstral peak prominence-smoothed (CPPS) and spectral decline of the vocal samples were obtained. CPPS is a modification of the cepstral peak prominence algorithm that enables a noticeable improvement in the accuracy of analysis of deviant voices. This modification involves smoothing the cepstrum before extracting the cepstral peak. Instead of calculating the cepstrum every 10 ms, in CPPS it is calculated every 2 ms, thus increasing the identification precision of the irregularities present in the signal[11].

The following commands and parameters were applied to generate CPPS in Praat:

1. Select "Analyze periodicity" and subsequently "To PowerCepstrogram";

2. In the "menu", select "Pitch floor (Hz) = 60", "Time step (s) = 0.002", "Maximum frequency (Hz) = 5000" and "Pre-emphasis from (Hz) = 50".

3. Select "Query", and "Get CPPS" in the "menu"; then proceed with "Subtract tilt before smoothing" and "Time averaging window (s) = 0.01", "Quefrency-averaging window (s) = 0.001", "Search peak in pitch range (Hz) = 60-330", "Tolerance (0-1) = 0.05", "Interpolation = Parabolic", "Tilt line quefrency range (s) = 0.001-0.0 (=end)", "Line type = Straight", and "Fit method = Robust".

4. The outcome of this procedure is the CPPS measurements as described in Maryn and Weenink[14].

The following commands and parameters were applied to obtain the spectral decline in Praat:

1. Select "Analyze spectrum" and choose "To Ltas";

2. Proceed to "bandWidth" with 100 Hz;

3. Select the "Ltas" signal and "Query";

4. Proceed with "Get slope". In "Low Band", change the values to 0 and 1250 Hz, and in "High Band", change the values to 1250 and 4000 Hz;

5. In "Query", obtain the values of Spectral Decline "Report spectral tilt".

6. The results of this procedure are the Spectral Decline measures, as described in Maryn and Weenink [14].

All CPPS and spectral decline values were manually assessed for outlier identification, and corresponded to spurious values derived from errors in the extraction of the analyzed measure. No outliers were identified in the data set for the evaluated signs.

For the analysis of auditory-perceptual measures, the voices were re-edited in SoundForge using the "normalize" control in peak level mode to obtain a standardization of the audio output from -6 to 6 dB for all signals so that the intensity of the audio signal did not influence the judgment of the evaluators regarding the intensity of voice disorder.

Auditory-perceptual evaluation was independently performed by three speech-language pathologists. Initially, the judges were trained with 16 anchor stimuli (sustained /ɛ/ vowel), containing four samples from individuals with normal voice quality variability (NVQV), four samples from individuals with mild-to-moderate voice disorder, four samples from individuals with moderate voice disorder, and four samples from individuals with severe voice disorder. The judges were instructed to listen to the anchor stimulus immediately before analysis of the individuals' voices. All selected samples for this training were previously analyzed by speech-language pathologists with experience in voice analysis, and were routinely used for auditory-perceptual training and as anchor stimuli in the laboratory where this research was conducted.

For the auditory-perceptual analysis, a visual analogue scale (VAS) from 0 to 100 mm[15] was used to evaluate the overall grade of dysphonia (G) and grades of roughness (R), breathiness (B) and strain (S) in the emission of the sustained /ɛ/ vowel. The judges were advised that the voices marked closest to 0 would represent more socially acceptable voices, which were produced more naturally, with less effort, noise, or unstable conditions[15]. In contrast, voices marked closer to 100 would represent those less socially accepted and with greater perception of effort, noise, or instability. They were also instructed that roughness would correspond to the presence of vibratory irregularity, breathiness would be related to audible impression of turbulent air leakage during voice emission, and strain would be associated with perception of vocal effort during voice emission.

The auditory-perceptual parameters of roughness, breathiness, and strain were chosen to characterize the signals in this study because they are universally used to describe voice quality disorders[16] and present known physiological and acoustic correlates.

For evaluation, each vocal emission of the sustained /ɛ/ vowel was presented three times through a loudspeaker at a comfortable intensity, self-reported by the evaluator. After each presentation, the judges evaluated the G, R, B and S, followed by the identification of voice quality (type of disorder) predominant in the deviated voices (rough, breathy, or strained).

At the end of the auditory-perceptual assessment session, 20% (76 signals) of the samples were randomly repeated to analyze judge reliability using the Cohen's kappa coefficient. The judge with the highest coefficient (0.80) was selected, which indicated good evaluator reliability[17].

VAS cut-off values[15] were used to classify the voices according to presence of voice disorder and overall grade of dysphonia. A total of 97 voice samples were classified as NVQV (G≤ 35.5 mm), and 279 voice samples were categorized as deviant (G>35.5 mm). All individuals with NVQV had no structural or functional laryngeal changes. Of the patients with deviant voices, only two showed absence of structural or functional changes in the larynx, whereas the remaining 277 presented the medical diagnoses previously mentioned. Further, G values in VAS were used to classify signals into four groups using the cut-off values described in the literature[15]: 97 voices showed NVQV (0-35.5 mm); 239 voices were grade 2 (35.6-50.5 mm), which corresponds to mild-to-moderate disorder; 165 voices were grade 3 (50.6-90.5 mm), moderate disorder; 27 voices were grade 4 (90.6-100 mm), severe disorder.

Notably, the reference study[15] used in the Brazilian context to determine the VAS cut-off values used only counting from 1 to 10 (connected speech) as speech task. Although this fact may constitute a limitation of the present study, a choice was made for the use of the cut-off values proposed by Yamasaki et al. (2017)[15], because they used only the four disorder levels that are internationally considered (healthy or NVQV, mild to moderate, moderate, and severe) and are the main references used in Brazil for the cut-off values in this classification.

DATA ANALYSIS

Descriptive statistical analysis was performed for all variables, including mean and standard deviation values. The nonparametric Mann-Whitney test was applied to compare the means of the cepstral measures between the groups with or without disorder. The Kruskal-Wallis test was used to compare the mean of the cepstral measures as a function of voice disorder intensity using the Nemenyi post-hoc test for paired comparison of the groups.

The Spearman's correlation test was applied to verify the correlation between voice disorder intensity and the cepstral measures. The correlation coefficients were used to evaluate and quantify the degree of linear relationship between the two variables, and it was observed whether the variables changed together and to what degree. The following values were considered for classification of the correlation coefficients in this study: 0.1 to 0.3 - weak correlation, 0.4 to 0.6 - moderate correlation, and >0.6 strong correlation between the variables[18].

All analyses were processed using the Statistical Package for Social Sciences (SPSS) 2.0. The level of significance was set at 5%.

RESULTS

The nonparametric Mann-Whitney test was initially used to compare the mean of the cepstral measures between the groups with and without voice disorder (Table 1). There was a difference in CPPS values between groups ($p<0.001$), and higher values were found for patients without voice disorder.

The Kruskal-Wallis test was applied to compare the mean of the cepstral measures as a function of voice disorder intensity (Table 2). Differences between the groups were observed for CPPS values ($p<0.001$). Subsequently, the Nemenyi post hoc test was used for paired comparisons of the groups. There was a difference between individuals in the NVQV group and the group with mild-to-moderate grade ($p=0.001$), with the NVQV group presenting higher values. Similarly, there was a difference between the mild-to-moderate and moderate grade groups ($p=0.001$), and the first presenting higher values. Difference was also observed between the moderate and severe grade groups ($p=0.001$), and the first showing higher values.

Table 1. Comparison of the means of cepstral measures between the groups with and without voice disorders

VARIABLES	NVQV		DEVIANT VOICES		p-value
	Mean	SD	Mean	SD	
CPPS	16.35	2.40	13.93	3.54	<0.001*
Spectral Decline	-13.63	6.25	-14.74	6.51	0.139

*Significant values (p<0.05) - Mann-Whitney test
Captions: NVQV = normal voice quality variability; SD = standard deviation; CPPS = cepstral peak prominence-smoothed

Table 2. Comparison of the means of cepstral measures as a function of intensity of the voice disorder

Variables	Intensity of Vocal Deviation								p-value
	NVQV		Mild to moderate		Moderate		Severe		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
CPPS	16.37	2.38	15.05	2.63	12.58	3.25	7.56	3.62	8.75
Spectral Decline	-13.48	6.67	-15.10	-15.10	-15.07	7.74	-12.78	<0.001*	0.479

*Significant values (p<0.05) - Kruskal-Wallis test
Captions: NVQV = normal voice quality variability; CPPS = cepstral peak prominence-smoothed; SD = standard deviation

Table 3. Comparison of cepstral measures as a function of predominant voice quality

| Variables | Predominant Voice Quality | | | | | | p-value |
| | Roughness | | Breathiness | | Strain | | |
	Mean	SD	Mean	SD	Mean	SD	
CPPS	13.75	3.62	12.5	3.15	15.5	3.11	<0.001*
Spectral Decline	-14.94	5.81	-16.74	7.24	-10.58	5.79	<0.001*

*Significant values (p<0.05) - Kruskal-Wallis test
Captions: SD = standard deviation; CPPS = cepstral peak prominence-smoothed

Table 4. Correlation between voice disorder intensity, grades of roughness, breathiness and strain, and cepstral measures

| Variables | G | | R | | B | | S | |
	Correlation	p-value	Correlation	p-value	Correlation	p-value	Correlation	p-value
CPPS	-0.629	<0.001*	-0.536	<0.001*	-0.618	<0.001*	-0.150	0.001*
Spectral Decline	0.050	0.272	-0.021	0.624	-0.145	0.001*	0.308	<0.001*

*Significant values (p<0.05) - Spearman's correlation test
Captions: CPPS = cepstral peak prominence-smoothed; G = general grade of dysphonia; R = grade of roughness; B = grade of breathiness; S = grade of strain

The nonparametric Kruskal-Wallis test was applied to compare the cepstral measures as a function of predominant voice quality. Differences in CPPS ($p<0.001$) and spectral decline ($p<0.001$) values were observed between the different types of voices (Table 3). In the post-hoc analysis, the CPPS values separated the rough voices from the breathy voices ($p=0.001$), and rough voices had higher CPPS mean values. There were differences in CPPS ($p=0.001$) and spectral decline ($p<0.001$) mean values between the rough and strained voices. Rough voices presented lower CPPS values and higher values of spectral decline compared with strained voices. CPPS ($p<0.001$) and spectral decline ($p<0.001$) values also differentiated the breathy voices from the strained voices. Strained voices presented higher CPPS values and lower values of spectral decline.

Finally, the Spearman's correlation test was used to compare the auditory-perceptual and cepstral measures (Table 4). CPPS showed strong negative correlation with G ($p<0.001$) and B ($p<0.001$), moderate negative correlation with R ($p<0.001$), and weak negative correlation with S ($p=0.001$). Regarding spectral decline, moderate positive correlation with S ($p<0.001$) and weak negative correlation with B ($p=0.001$) were observed.

DISCUSSION

In the context of voice assessment, clinicians and researchers have made an effort to identify measures that can reliably characterize and monitor voice quality disorders[19]. Thus, cepstral measures have shown potential to evaluate voices with a wide range of disorders, justifying the increase in studies using these measures, which aids in the understanding of their role in voice clinics[20].

In the present study, it was observed that CPPS is able to differentiate individuals with or without voice quality disorder, with the latter showing higher values. This difference can be explained by the fact that the signals of voice without disorder present greater periodicity, with well-defined harmonic configuration and, consequently, higher CPPS values. In contrast, the more deviated voices present a smaller proportion between harmonics energy and the components of noise and aperiodicity, with lower CPPS values[11].

In this study, most individuals with voice quality disorder presented structural and/or functional changes in the larynx. Physiologically, the presence of such changes may alter the vibratory patterns and glottic closure, resulting in aperiodicity and noise in the voice signal, respectively[21].

Some studies[20,22,23] have investigated the ability of cepstral measures to discriminate healthy voices from deviant voices. These studies found accuracy rates of 71-85% regarding classification into healthy and deviant signals. Those authors[20,23] used auditory-perceptual evaluation (accuracy=85%) as a reference standard, followed by visual laryngeal examination (accuracy=73%)[20,24], and voice self-assessment (accuracy=75%). The classification rates behaved differently according to the reference standard used, with better performance related to auditory-perceptual analysis compared with visual laryngeal examination and voice self-assessment. However, in all cases, the cepstral measures were able to differentiate between healthy and deviant voices.

When diagnostic confirmation is the objective of a test, interpretation of the classification rates should be based on the test objective, favoring sensitivity over specificity in the case of screening measures. A study conducted by Awan et al.[20] proposed the use of cepstral measures for the screening of voice disorders. Thus, the authors used lower cut-off values (19.09 dB, 19.01 dB, and 19.46 dB for auditory-perceptual assessment, laryngeal visual examination, and self-assessment, respectively) to classify the signals as healthy or deviant based on the recommended reference standards. Values below this cut-off point would indicate presence of change compared with the cited reference standards.

The CPPS values found in the present study for both groups (NVQV and deviant voice) are below the cut-off values recommended in the literature[20]. From this finding, one can discuss the methodological differences between the previous study[20] and the present research. There are three main differences, and these are associated with the auditory-perceptual judgment of the analyzed voices, the speech task for auditory-perceptual evaluation, and the allocation criteria of the individuals.

Awan et al.[20] used a binary/categorical evaluation in which the evaluators indicated only whether the voices were healthy

or deviant. In contrast, the present study used the cut-off point of VAS to categorize voices as healthy or deviant.

The previous study[20] used two speech tasks, including connected speech (reading of "The Rainbow Passage") and the sustained /a/ vowel, whereas the present survey used only the sustained /Ɛ/ vowel. According to the same author[4], there is still uncertainty about which speech tasks should be included in predictive models of the presence or absence of vocal disorders, especially when comparing sustained vowels and connected speech.

Connected speech is closer to everyday conversation; however, during voice quality classification, it seems to be more variable because there is perceptual focus on non-vocal phenomena, such as prosody, articulation of words, and the entire phonetic and phonological context. In turn, sustained vowels are less prone to this phonetic variability[25]. In addition, the use of vowels is one of the most cited procedures in clinical practice for voice quality evaluation. However, it is known that cultural differences, mainly with regards to language, can influence the results of voice quality assessments. There is still no cut-off point for cepstral measures in Brazilian Portuguese-speaking individuals.

Furthermore, in the previous study[20], none of the individuals without voice disorder presented voice complaints or underwent visual laryngeal examination. This fact may justify the presence of a higher cut-off point, because a more homogeneous group is created when a combined reference standard is used. The criteria used in that study[20] are justified by the fact that the authors sought to identify a cut-off point for voice disorder screening. Conversely, the present research seeks to investigate the relation between these measures regarding voice quality disorder using only auditory-perceptual evaluation as standard of reference.

Another study[3] used auditory-perceptual assessment as reference standard and the cut-off point for VAS for the allocation of individuals with and without voice quality disorder. The CPPS cut-off point to identify healthy individuals with voice disorders was 17.68 dB, which is closer to the results found in the present study.

Comparison of the acoustic measure means as a function of intensity of voice disorder showed difference between the groups with different grades of dysphonia, with lower values observed in the most deviant voices in each group (NVQV x mild to moderate, mild to moderate x moderate, moderate x severe). Thus, the higher the voice disorder intensity, the lower the acoustic energy of F_0 and its definition in relation to the total energy of the acoustic signal[26], which causes a decrease in the spectral peak as a function of voice disorder intensity[3,11,12].

Regarding the predominant voice quality, there was difference in CPPS values between the different types of voice disorder. Voices with predominance of strain presented higher CPPS values compared with those of predominantly rough and breathy voices. In turn, rough voices showed higher CPPS values than breathy voices. With respect to spectral decline, strained voices presented smaller values compared with those of rough and breathy voices.

Phonatory strain is commonly characterized by increased contraction of the intrinsic and extrinsic muscles of the larynx, which results in greater rigidity in the system and greater longitudinal pressure on the vocal folds, with increased subglottic pressure and increased time of the closed phase of the glottic cycle[27]. In general, such an adjustment produces signals with higher energy level and definition of F_0, which explains the higher CPPS and lower spectral decline values in the strained voices compared with those in the rough and breathy voices[27].

For differentiation between breathy voices and rough and strained voices, the physiological pattern typically associated with the first is characterized by greater separation between the vocal processes, lesser convexity of the free edge of the vocal folds, and shorter time of the closed phase of the glottic cycle. This physiological pattern leads to a decrease in energy below 2500 Hz and an increase in energy in the higher frequency bands, which explain the lower CPPS values in breathy voices, because the increase in noise at high frequencies is one of the factors that most influences the decrease in CPPS[27].

Rough voices have a higher noise component at low frequencies than at high frequencies, which may be related to higher CPPS values in rough voices compared with breathy voices. In previous studies, spectral decline[24,27,28] and CPP were the main parameters used to differentiate between breathy and healthy voices, although these studies did not differentiate between rough and breathy voices or did not select the main measure to differentiate between rough and healthy voices. In multivariate acoustic analysis, only the combination of shimmer and mean F_0 measures were able to differentiate rough voices from breathy voices[24,27,28].

strong negative correlation was observed between CPPS and G and B, where more deviant voices with greater B component showed greater decrease at the cepstral peak. Other studies[3,10,11] have demonstrated that there is strong correlation between voice disorder intensity and cepstral measures, as well as between perception of breathiness in vocal emission and these measures. In general, cepstral measures are more strongly correlated with voice disorder intensity compared with measures based on time domain (jitter and shimmer). In a previous study[8], moderate positive correlation was found between the jitter and shimmer measures and overall grade of dysphonia.

The spectral characteristics of the voice signal are closely related to changes in the duration of contact of the vocal folds[28]. There is a strong positive correlation between the opening quotient and degree of convexity of the vocal folds and the increase in energy in the region of 4 KHz. This explains the strong correlation found between B and CPPS in the present study, because the decrease in the closed-phase time of the glottic cycle is the main physiological correlate to the presence of breathiness in vocal emissions.

Regarding roughness, moderate negative correlation with CPPS was observed. Roughness corresponds to the vibrational irregularity of the vocal folds caused by changes in subglottic pressure or structural changes in the free edge of the vocal folds[24], producing an emission with presence of sub-harmonics, amplitude modulation, and increased signal perturbation. Roughness is characterized by the low frequency noise component[10], which is associated with decreased mean F_0 and an increase in its standard deviation[8].

Thus, CPPS, whose value is directly related to the difference between energy in the lower frequencies and presence of additional noise in the higher frequencies, seems to be less correlated to the R component compared with the B component[10]. Therefore, the presence of R is more adequately characterized by acoustic analysis methods that involve measures based on energy distribution and temporal aspects of the emission, such as the cepstral/spectral and jitter/shimmer measures, respectively[24]. Performance of the cepstral measures is lower than that of the time domain measures in the evaluation of the R parameter[8,21].

Weak negative correlation was observed between the S parameter and CPPS. Among the auditory-perceptual parameters, strain has been referred as the most controversial and difficult characteristic to be acoustically evaluated[8,29]. The presence of strain in vocal emission is physiologically associated with vocal fold longitudinal strain, increased subglottic pressure, greater contraction of the extrinsic and intrinsic muscles of the larynx, more verticalized position of the larynx, and increased time of the closed phase of the glottic cycle[29]. In acoustic terms, strained voices tend to present increased energy at high frequencies[26], which may also occur in vocally healthy individuals using a more projected voice.

A study[26] using multivariate acoustic analysis based on cepstral and spectral measures identified lower CPP values in dysphonic individuals with vocal strain compared with those in vocally healthy individuals, in addition to a shift from the dominant cepstral peak to higher frequencies. The authors also observed strong negative correlation between S and the cepstral measures in connected speech compared and weak negative correlation in the sustained vowel, corroborating the findings of the present study.

Results of the present study show that the cepstral acoustic measures are clear indicators of the presence and intensity of voice disorder, as well as B, in addition to contributing to the differentiation between different types of voice disorders. The findings regarding the evaluation of the R and S parameters reinforce the importance and current trend of using multivariate acoustic analysis, because no single measure is capable of providing a reliable analysis of signals with different components of concomitant irregularity, noise and strain. Overall, the data from this study demonstrate that the cepstral measures are a reliable tool for quantifying voice disorders and producing estimates of aperiodicity and/or additional noise without the need for individual identification of cycle thresholds[11].

In the present study, only the sustained /ɛ/ vowel was used for the evaluation of the relationship between the cepstral measures and the auditory-perceptual analysis. Thus, an evaluation using other tasks such as connected speech and the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) phrases is suggested, with identification of the best task for cepstral analysis in the context of Brazilian Portuguese. In addition, it is necessary to establish the cut-off point and discriminatory power of these measures for the different speech tasks in Brazilian Portuguese, as well as for different reference standards (laryngeal visual examination, auditory-perceptual evaluation, and vocal self-assessment).

One of the possible limitations of this study may also be associated with the reference values used to classify the voices at different grades of dysphonia, because the original validation study of the cut-off points[15] used connected speech, and the present survey used a sustained vowel. This reinforces the importance of further studies using CPPS with the same connected speech task used by Yamasaki et al.[15]. Thus, it would be possible to observe whether there is correspondence between the CPPS findings at different grades in the sustained vowel and connected speech, even if cut-off values not previously defined for the sustained vowel are used.

CONCLUSION

There is association between presence of voice disorder, G, predominant voice quality, and CPPS. Deviant voices have lower CPPS values compared with those of healthy voices. Voices with a predominance of strain present higher CPPS values compared with those of predominantly rough and breathy voices. Rough voices show higher CPPS values than breathy voices. Overall grade of dysphonia (G) and breathiness (B) show strong negative correlation with CPPS, whereas roughness (R) and strain (S) present moderate and strong negative correlations with CPPS, respectively. Spectral decline is associated only with B and S.

REFERENCES

1. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical treatments and evaluating new assessment techniques: Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). Eur Arch Otorhinolaryngol. 2001;258(2):77-82. http://dx.doi.org/10.1007/s004050000299. PMid:11307610.

2. Hunter EJ, Titze IR. Quantifying vocal fatigue recovery: dynamic vocal recovery trajectories after a vocal loading exercise. Ann Otol Rhinol Laryngol. 2009;118(6):449-60. http://dx.doi.org/10.1177/000348940911800608. PMid:19663377.

3. Awan SN, Roy N, Jetté ME, Meltzner GS, Hillman RE. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V. Clin Linguist Phon. 2010;24(9):742-58. http://dx.doi.org/10.3109/02699206.2010.492446. PMid:20687828.

4. Awan SN, Helou LB, Stojadinovic A, Solomon NP. Tracking voice change after thyroidectomy: application of spectral/cepstral analyses. Clin Linguist Phon. 2011;25(4):302-20. http://dx.doi.org/10.3109/02699206.2010.535646. PMid:21158501.

5. Uloza V, Verikas A, Bacauskiene M, Gelzinis A, Pribuisiene R, Kaseta M, et al. Categorizing normal and pathological voices: automated and perceptual categorization. J Voice. 2010;25(6):700-8. http://dx.doi.org/10.1016/j.jvoice.2010.04.009. PMid:20579842.

6. Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art. Auris Nasus Larynx. 2015;42(3):183-8. http://dx.doi.org/10.1016/j.anl.2014.11.001. PMid:25440411.

7. Lopes LW, Alves GAS, Melo LM. Evidência de conteúdo de um protocolo de análise espectrográfica. Rev CEFAC. 2017;19(4):510-28. http://dx.doi.org/10.1590/1982-021620171942917.

8. Lopes LW, Cavalcante DP, Costa PO. Intensidade do desvio vocal: integração de dados perceptivo-auditivos e acústicos em pacientes disfônicos. CoDAS. 2014;26(5):382-8. http://dx.doi.org/10.1590/2317-1782/20142013033. PMid:25388071.

9. Godino-Llorente JI, Osma-Ruiz V, Sáenz-Lechón N, Gómez-Vilda P, Blanco-Velasco M, Cruz-Roldán F. The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders. J Voice. 2010;24(1):47-56. http://dx.doi.org/10.1016/j.jvoice.2008.04.006. PMid:19135854.

10. Awan SN, Roy N. Outcomes measurement in voice disorders: application of an acoustic index of dysphonia severity. J Speech Lang Hear Res. 2009;52(2):482-99. http://dx.doi.org/10.1044/1092-4388(2008/08-0034). PMid:19339702.

11. Dejonckere PH, Wieneke GH. Cepstral of normal and pathological voices: correlation with acoustic, aerodynamic and perceptual data. In: Ball MJ & Duckworth M. editors. Advances in clinical phonetics. Amsterdam: John Benjamins; 1996. p. 217-226. http://dx.doi.org/10.1075/sspcl.6.13dej.

12. Awan SN, Roy N, Dromey C. Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model. Clin Linguist Phon. 2009;23(11):825-41. http://dx.doi.org/10.3109/02699200903242988. PMid:19891523.

13. Wolfe VI, Martin DP, Palmer CI. Perception of dysphonic voice quality by naive listeners. J Speech Hear Res. 2000;43(3):697-705. http://dx.doi.org/10.1044/jslhr.4303.697. PMid:10877439.

14. Maryn Y, Weenink D. Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index. J Voice. 2015;29(1):35-43. http://dx.doi.org/10.1016/j.jvoice.2014.06.015. PMid:25499526.

15. Yamasaki R, Madazio G, Leão SHS, Padovani M, Azevedo R, Behlau M. Auditory-perceptual evaluation of normal and dysphonic voices using the voice deviation scale. J Voice. 2017;31(1):67-71. http://dx.doi.org/10.1016/j.jvoice.2016.01.004. PMid:26873420.

16. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. Am J Speech Lang Pathol. 2009;18(2):124-32. http://dx.doi.org/10.1044/1058-0360(2008/08-0017). PMid:18930908.

17. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005;85(3):257-68. PMid:15733050.

18. Dancey C, Reidy J. Estatística sem matemática para psicologia: usando SPSS para Windows. Porto Alegre: Artmed; 2006.

19. Brockmann-Bauser M, Drinnan MJ. Routine acoustic voice analysis: time to think again? Curr Opin Otolaryngol Head Neck Surg. 2011;19(3):165-70. http://dx.doi.org/10.1097/MOO.0b013e32834575fe. PMid:21483265.

20. Awan SN, Roy N, Zhang D, Cohen SM. Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. J Voice. 2015;30(2):1-15. PMid:26361215.

21. McAllister A, Sederholm E, Ternström S, Sundberg J. Perturbation and hoarseness: a pliot study of six children's voices. J Voice. 1996;10(3):252-61. http://dx.doi.org/10.1016/S0892-1997(96)80006-3. PMid:8865096.

22. Watts CR, Awan SN. An examination of variations in the cepstral spectral index of dysphonia across a single breath group in connected speech. J Voice. 2015;29(1):26-34. http://dx.doi.org/10.1016/j.jvoice.2014.04.012. PMid:25108589.

23. Awan SN, Solomon NP, Helou LB, Stojadinovic A. Spectral-Cepstral estimation of dysphonia severity: external validation. Ann Otol Rhinol Laryngol. 2013;122(1):40-8. http://dx.doi.org/10.1177/000348941312200108. PMid:23472315.

24. Awan SN, Roy N. Acoustic prediction of voice type in women with functional dysphonia. J Voice. 2005;19(2):268-82. http://dx.doi.org/10.1016/j.jvoice.2004.03.005. PMid:15907441.

25. Barsties B, Maryn Y. External validation of the acoustic voice quality index version 03.01 with extended representativity. Ann Otol Rhinol Laryngol. 2016;125(7):571-83. http://dx.doi.org/10.1177/0003489416636131. PMid:26951063.

26. Lowell SY, Kelley RT, Awan SN, Colton RH, Chan NH. Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. Ann Otol Rhinol Laryngol. 2012;121(8):539-48. http://dx.doi.org/10.1177/000348941212100808. PMid:22953661.

27. Watts CR, Awan SN. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. J Speech Lang Hear Res. 2011;54(6):1525-37. http://dx.doi.org/10.1044/1092-4388(2011/10-0209). PMid:22180020.

28. Awan SN, Krauss AR, Herbst CT. An examination of the relationship between electroglottographic contact quotient, electroglottographic decontacting phase profile, and acoustical spectral moments. J Voice. 2014;29(5):519-29. http://dx.doi.org/10.1016/j.jvoice.2014.10.016. PMid:25795367.

29. Van Houtte E, Van Lierde K, Claeys S. Pathophysiology and treatment of muscle tension dysphonia: a review of the current knowledge. J Voice. 2011;25(2):202-7. http://dx.doi.org/10.1016/j.jvoice.2009.10.009. PMid:20400263.

AUTHOR CONTRIBUTIONS