Relationship of Cepstral Peak Prominence-Smoothed and Long-Term Average Spectrum with Auditory–Perceptual Analysis

Angélica Emygdio da Silva Antonetti * , **Larissa Thais Donalonso Siqueira** ,
Maria Paula de Almeida Gobbo, Alcione Ghedini Brasolotto and Kelly Cristina Alves Silverio

Speech-Language Pathology and Audiology Department, Faculdade de Odontologia de Bauru,
Universidade de São Paulo, Bauru, SP 17012-901, Brazil; larisiqueira@usp.br (L.T.D.S.);
maria.gobbo@usp.br (M.P.d.A.G.); alcione@usp.br (A.G.B.); kellysilverio@usp.br (K.C.A.S.)
* Correspondence: angelica.antonetti@usp.br

check for updates

Abstract: Cepstral peak prominence-smoothed (CPPs) and long-term average spectrum (LTAS) are robust measures that represent the glottal source and source-filter interactions, respectively. Until now, little has been known about how physiological events impact auditory–perceptual characteristics in the objective measures of CPPs and LTAS (alpha ratio; L1–L0). Thus, this paper aims to analyze the relationship between such acoustic measures and auditory–perceptual analysis and then determine which acoustic measure best represents voice quality. We analyzed 53 voice samples of vocally healthy participants (vocally healthy group-VHG) and 49 voice samples of participants with behavioral dysphonia (dysphonic group-DG). Each voice sample was composed of sustained vowel /a/ and connected speech. CPPs seem to be the best predictor of voice deviation in both studied populations because there was moderate to strong negative correlations with general degree, breathiness, roughness, and strain (auditory–perceptual parameters). Regarding L1–L0, this measure is related to breathiness (moderate negative correlations). Hence, L1–L0 provides information about air leak through closed glottis, assisting the phonatory efficiency analysis.

Keywords: acoustic signal processing; speech signal processing; voice quality; communication; dysphonia; auditory–perceptual analysis

1. Introduction

According to Zhang [1], the voice is a suprasegmental aspect of speech that communicates meaning, ideas, opinions, and intentions by its different modulations. Thus, a balanced phonation system must convert aerodynamic energy into acoustic energy. Hence, the vocal folds (source) modulate the air through the glottis and produce acoustic energy, which will propagate and be modulated by the vocal tract (filter), thus amplifying or attenuating frequencies/harmonics. The harmonic energy modulation determines the speech sounds and their meanings, so the harmonic energy must be above the noise level for effective communication and its meanings to happen [1].

Voice disorders (i.e., dysphonia) include pitch, loudness, and vocal quality disturbances, varying from mild to complete vocal loss. Voice disorders decrease speech intelligibility and communication [2] as well as impact verbal and emotional messages and social interactions [3]. Thus, speech-language pathologists must assess the vocal quality so they can understand vocal disorders and then choose the best treatment (i.e., suppress the communicative handicap).

Acoustic analysis of vocal quality, along with auditory–perceptual analysis, is one of the components of the multidimensional voice assessment. It adds information on the voice signal

that, in nearly normal voices, has low perturbations of the acoustic signal [4]. The source signal periodicity or aperiodicity can be analyzed by using time-based measures [5,6], i.e., jitter and shimmer, that measure the cycle-to-cycle boundary deviation of frequency and intensity, respectively [7–9]. It can also be analyzed by using frequency-based measures, i.e., cepstral peak prominence (CPP) [10,11].

Since time-based measures require nearly periodic signals (type I) and exact measurement of the fundamental frequency, their reliability is questionable in more than mild disturbed vocal signals (types II and III). However, most vocal disorders have a greater aperiodicity level, making the use of these measures questionable [12,13].

Among frequency-based measures, according to Hillenbrand and Houde [14], CPP is calculated by using an inverse Fast Fourier Transform (FFT) of the log power spectrum of a voice signal. That means it is determined by measuring the amplitude (dB) difference from the most prominent cepstral peak to the regression line drawn direct below this peak. In other words, the cepstrum generated through spectral analysis of a spectrum produces a peak that corresponds to the dominant rahmonic (i.e., an anagram for harmonic) associated with the fundamental frequency (F_0), the first harmonic, of the voice [9]. Thus, voices with nearly periodic signals have a better harmonic definition, which means a higher CPP [9,11,15]. Furthermore, since CPP is extracted from frames instead of cycles, it is more reliable for analyzing voices with moderate to intense deviations [10], i.e., type II and III signals. A CPP variation is called cepstral peak prominence-smoothed (CPPs) and has different calculation algorithms, such as smoothing the cepstrum before extracting the peak and using 1024 frames every 2 ms instead of 10 ms, which makes CPPs more robust and with no influence of artifacts [15]. On the other hand, one advantage of CPP and CPPs regarding time-based measures is the possibility of applying them in continuous speech [4,15–17].

Brinca et al. [4] analyzed CPP and CPPs differences between vocally healthy and dysphonic women. They found that dysphonic women had lower CPP and CPPs values. As a complementary analysis, they discovered that the acoustic analysis had a moderate negative correlation with general degree and breathiness parameters of vocal quality. At last, they suggest including men in these analyses to better understand these measures. Heman-Ackah, Michael, and Goding [17] analyzed the correlations between CPPs and auditory–perceptual parameters in participants with vocal fold unilateral paralysis. They found that CPPs had a strong negative correlation with general degree and breathiness parameters of auditory–perceptual analysis. At last, Lopes et al. [18] analyzed the correlations between CPPs and auditory–perceptual parameters in a dysphonic group from vowel emission. They found that CPPs had a moderate negative correlation with general degree, roughness, and breathiness parameters of vocal quality, and weak negative correlation with strain parameter. Furthermore, they highlight how important it is to incorporate continuous speech to the analysis.

Alternatively, a voice signal analysis can be performed by using long-term average spectrum (LTAS) measures. LTAS is a fast Fourier transform (FFT) analysis generated from an acoustic signal frequency spectrum. Equally to CPP and CPPs, LTAS does not require a nearly periodic signal to be reliable [19]. Also, LTAS can be applied to sustained vowel and continuous speech by analyzing the source and filter signals and their interaction in voice production [20,21], and these measures are very accurate to detect vocal disorders [22]. Among LTAS measures, we highlight the alpha ratio and L1–L0, which have been increasingly used in studies [23–26].

The alpha ratio represents the spectral slope, therefore, the steeper spectral slope, the lower intensity of high harmonics concerning the low harmonics [27]. Previous studies have found that the inefficient or too slow glottal closure produces a turbulent and irregular airflow, reducing the energy of the high harmonics [24,28]. Thus, when there is increased glottal closure, which quickly interrupts the airflow, there will be decreased spectral slope and increased vocal projection [29,30]. However, the less steep spectral slope is also present in hyperfunctional voices due to the increased closed phase of the glottal cycle. Thus, it is important to add another LTAS measure, such as the L1–L0.

L1–L0 is the difference between the first formant (F1) energy level and the fundamental frequency (F_0) energy level, which represents the phonation mode, and may be associated with

vocal quality. Hence, when L0 (F_0 energy level) is higher regarding L1 (F1 energy level), there is a glottal hypoadduction, whereas when L0 is lower regarding L1, there is a glottal hyperadduction [22,31]. Furthermore, some studies observed that close values of L1 and L0 are associated with economic voice production [32,33].

Guzmán et al. [23] found that strained or projected emotional patterns, such as anger and joy, have higher (less negative) alpha ratio and L1–L0 values. Emotions with a higher level of breathiness, such as erotism, have lower values.

We observed that previous studies only analyzed CPPs in vocally healthy and dysphonic (behavioral dysphonia) women [4], participants with unilateral vocal fold paralysis (organic dysphonia) [17], and participants with behavioral dysphonia by only using sustained vowel [18]. At last, one study analyzed how emotional patterns affect LTAS measures. Hence, until now little has been known about how the physiological events that create the auditory–perceptual characteristics (general degree, roughness, breathiness, and strain) impact the objective measures of CPPs, alpha ratio, and L1–L0.

By analyzing the same variables in both groups (vocally healthy and dysphonic) and both genders, and with the sustained vowel and connected speech, this study aims to contribute to scientific progress by implementing objective voice assessment measures. The auditory–perceptual analysis has low reproducibility and depends on individual characteristics and speech pathologists' previous experiences (i.e., the examiner). Therefore, using robust algorithms with higher reproducibility and reliability to measure the voice signal such as CPPs and LTAS could be complementary to auditory–perceptual analysis, thus improving the voice assessment. Hence, it is key that we understand the correlations that lie between these acoustic measures and the auditory–perceptual parameters.

Thus, this study mainly aims to analyze the correlations between acoustic measures (CPPs, alpha ratio, and L1–L0) and auditory–perceptual analysis (general degree, roughness, breathiness, and strain). Moreover, its second aim is to determine which acoustic measures best represent the vocal quality of vocally healthy participants and of those who have behavioral dysphonia.

2. Materials and Methods

2.1. Study Design

This was an observational, transversal, and retrospective study.

2.2. Ethical Aspects

This research followed the rules of the National Health Council/National Research Ethics Committee. It was further approved by the Committee for Ethics in Research on Human Beings (no. 3.284.883).

2.3. Research Team

The research team consisted of four speech pathologists who were blinded to procedures:

- Researcher 1: responsible for measuring the acoustical parameters.
- Researchers 2, 3, and 4: responsible for auditory–perceptual analysis.

2.4. Sample

All participants registered in the database completed an initial assessment protocol form (i.e., vocal complaints, self-referred health issues) and have been diagnosed with normal voice and larynx or behavioral dysphonia [34] (i.e., vocal complaints, benign mass lesion, vocal folds edema, glottic gap, and minor structural alterations).

We selected voices samples from the database of participants aged 18–50 years who had no vocal complaints and were vocally healthy to create the vocally healthy group (VHG), and voices samples of participants with behavioral dysphonia to create the dysphonic group (DG). Participants who reported

endocrine, vascular, or pulmonary alterations or who were smokers were excluded. Thus, we selected a total of 102 voice samples, which were distributed as follows:

(1)   VHG—25 females, 28 males; 53 total participants (mean age = 31.0 years);
(2)   DG—29 females, 20 males; 49 total participants (mean age = 26.4 years).

2.5. Procedures

2.5.1. Voice Recording

Participants had their voices recorded in an acoustically treated room at the Speech-Language Pathology and Audiology Clinic. We used an AKG microphone (model C 44 PP; AKG Acoustics GmbH, Viena, Austria) positioned at a 45° angle from the participants' mouth and 2 cm away from their labial commissure. The microphone was connected to a computer with a Creative Sound Blaster sound card (model Audigy II; Creative Technology Ltd., Singapore). We used the Sound Forge program (version 10; MAGIX, Munich, Germany) with a sample rate of 44,100 Hz. A 16-bit.wav extension was used to record the voices samples. Participants were asked to sustain for approximately 5 s, at a comfortable pitch and loudness, the vowel /a/ and then count from 1 to 10 (i.e., assessing continuous speech).

2.5.2. Acoustic Measures

CPPs, alpha ratio, and L1–L0 were measured by using the Praat software (version 6.0.43; https://www.fon.hum.uva.nl/praat) from the central 3 s of vowel /a/ emission and the entire continuous speech.

CPPs were measured according to previous studies [35,36] as follows:

1.   Open Praat and the voice sample and then select "Analyze Periodicity";
2.   Click "To PowerCepstrogram";
3.   A new window will open. Keep the standard values of the software: Pitch floor (Hz) = 60, Timestep (s) = 0.002, Maximum frequency (Hz) = 5000, and Pre-emphasis (Hz) = 50;
4.   Click on the new generated file, select "Query", and then click "Get CPPS";
5.   On the new window, deselect the "Subtract tilt before smoothing" box. Then adjust Time-averaging window (s) = 0.01, Quefrency-averaging window (s) = 0.001, Peak search pitch range (Hz) = 60–330, Tolerance (0–1) = 0.05, Interpolation = Prabolic, Tilt line quefrency range (s) = 0.001–0.0 (=end), Line type = straight, and Fit method = robust;
6.   A new window will open with the CPPs value.

To measure LTAS, the researchers used a spectral analysis with pitch correction in the standard values of the software. In the alpha ratio measure (spectral slope), they performed a difference in the harmonic distribution between the ranges of 50 to 1000 Hz and 1000 to 5000 Hz. In the L1–L0 measure (phonation mode), that difference was from 50 to 300 Hz and 300 to 800 Hz [23].

2.5.3. Auditory–Perceptual Analysis

Three experienced speech pathologists, who were blinded to the groups, analyzed the vowel /a/ emission and continuous speech with the following parameters: general degree, roughness, breathiness, and strain. They were instructed to complete a visual analogical scale (VAS) with 100 mm length at any point, indicating deviation intensity. Left indicated no deviation and right the highest deviation.

To obtain the deviation intensity, we calculated the mean values of each speech pathologist. Furthermore, they reanalyzed 20% of the samples to verify the intra-examiner reliability.

2.6. Data Analysis

The sample homogeneity assessment showed that all variables had a normal distribution, except for the auditory–perceptual parameter strain. Thus, the Pearson correlation coefficient was used to analyze the correlations, and the Spearman correlation coefficient was used to analyze those involving the strain parameter. Both tests had their significance level fixed at 5% ($p < 0.05$).

The systematic and random error calculation was used to determine auditory–perceptual analysis intra-examiner and inter-examiner reliability.

3. Results

According to Fleiss, Levin, and Paik [37] classification, intra-examiner reliability for vowel /a/ was excellent for all examiners (0.85, 0.88, and 0.88). For continuous speech, the reliability for two examiners was excellent (0.75 and 0.79), and the third was satisfactory (0.73). Regarding inter-examiner reliability, it was satisfactory among all examiners in vowel /a/ (0.51, 0.59, and 0.68) and continuous speech (0.64, 0.70, and 0.72).

Tables 1 and 2 show a complimentary analysis aiming to compare the groups regarding both acoustic and auditory–perceptual analyses. We used a *t*-test with a significance level fixed at 5% ($p < 0.05$).

Table 1. Analysis of acoustic measures (dB) from vowel /a/ and connected speech (number counting 1–10) in both groups: vocally healthy and dysphonic.

| Acoustic Measures	Vowel			Connected Speech		
	VHG Mean ± SD	DG Mean ± SD	*p* Value	VHG Mean ± SD	DG Mean ± SD	*p* Value
CPPs	16.447 ± 2.92	14.991 ± 2.65	0.001 *	7.769 ± 1.68	7.448 ± 1.38	0.297
Alpha ratio	−18.225 ± 5.26	−18.181 ± 4.64	0.965	−23.687 ± 4.10	−24.136 ± 3.74	0.565
L1–L0	−6.691 ± 4.50	−5.245 ± 4.36	0.103	−6.149 ± 3.31	−5.612 ± 3.19	0.407

t-Test ($p < 0.05$). Abbreviations: * = statistically significant result; VHG = vocally healthy group; DG = dysphonic group; CPPs = cepstral peak prominence-smoothed; SD = standard deviation.

Table 2. Analysis of the intensity of the auditory–perceptual parameter (mm) from vowel /a/ and connected speech (counting 1–10) in both groups: vocally healthy and dysphonic.

Auditory–Perceptual Parameters	Vowel			Connected Speech		
	VHG Mean ± SD	DG Mean ± SD	*p* Value	VHG Mean ± SD	DG Mean ± SD	*p* Value
General degree	32.881 ± 8.03	50.510 ± 10.54	<0.001 *	22.925 ± 6.08	50.510 ± 10.54	<0.001 *
Roughness	21.308 ± 10.44	42.939 ± 14.23	<0.001 *	15.969 ± 8.83	42.939 ± 14.23	<0.001 *
Breathiness	23.648 ± 11.92	37.803 ± 12.19	<0.001 *	12.994 ± 7.67	37.803 ± 12.19	<0.001 *

t-Test ($p < 0.05$). Abbreviations: * = statistically significant result; VHG = vocally healthy group; DG = dysphonic group; SD = standard deviation.

As mentioned before, the strain parameter of auditory–perceptual analysis does not have a normal distribution, so we used the Mann–Whitney test ($p < 0.05$) to complete the complementary analysis (Figure 1).
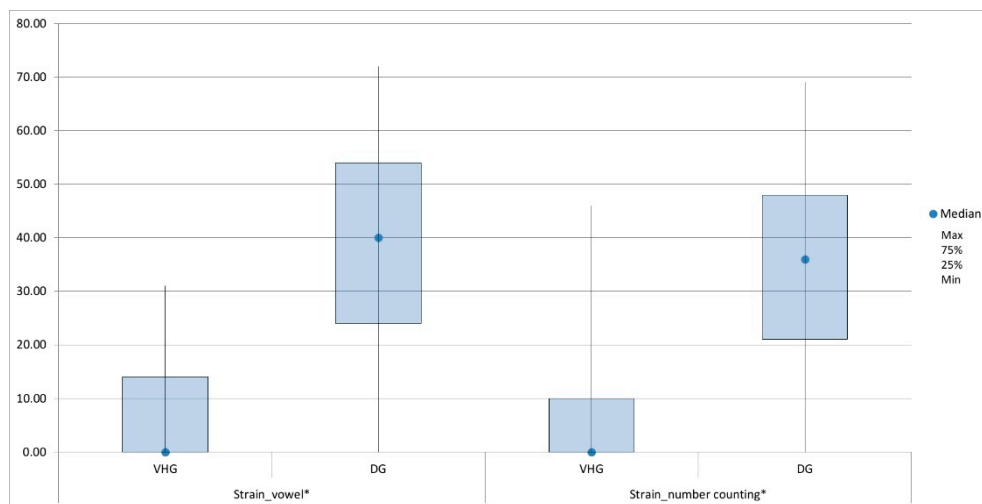
Figure 1. Analysis of perceptual-auditory parameter strain intensity (mm) measured from sustained vowel and connected speech (number counting) in both groups: vocally healthy and dysphonic. * = statistically significant result ($p < 0.05$)—Mann–Whitney test.

Tables 3 and 4 show all the correlations results. The statistically significant results were highlighted with an *.

Table 3. Correlations analysis (r-value) between the intensity of auditory–perceptual parameters (mm) and acoustic measures (dB) for vowel /a/ emission in both groups: vocally healthy and dysphonic.

Auditory–Perceptual Parameters	VHG			DG		
	CPPs r-Value	Alpha Ratio r-Value	L1–L0 r-Value	CPPs r-Value	Alpha Ratio r-Value	L1–L0 r-Value
General degree	−0.664 *	−0.100	−0.385 *	−0.570 *	−0.088	−0.218
Roughness	−0.065	−0.010	−0.019	−0.231	−0.113	−0.110
Breathiness	−0.789 *	−0.230	−0.512 *	−0.846 *	−0.107	−0.554 *
Strain **	−0.149	0.322 *	−0.163	−0.416 *	0.134	−0.120

Pearson Correlation Coefficient. Abbreviations: * = statistically significant correlation ($p \leq 0.05$); ** = Spearman Correlation Coefficient; VHG = vocally healthy group; DG = dysphonic group; CPPs = cepstral peak prominence-smoothed.

Table 4. Correlations analysis (r-value) between the intensity of auditory–perceptual parameters (mm) and acoustic measures (dB) for connected speech (counting 1–10) emission in both groups: vocally healthy and dysphonic.

Auditory–Perceptual Parameters	VHG			DG		
	CPPs r-Value	Alpha Ratio r-Value	L1–L0 r-Value	CPPs r-Value	Alpha Ratio r-Value	L1–L0 r-Value
General degree	−0.366 *	−0.059	0.109	−0.108	−0.242	−0.296 *
Roughness	−0.444 *	−0.242	0.206	−0.134	−0.352 *	−0.283 *
Breathiness	−0.161	−0.271 *	−0.241	0.053	−0.079	−0.143
Strain **	−0.107	0.220	−0.121	−0.001	0.118	−0.129

Pearson Correlation Coefficient. Abbreviations: * = statistically significant correlation ($p \leq 0.05$); ** = Spearman Correlation Coefficient; VHG = vocally healthy group; DG = dysphonic group; CPPs = cepstral peak prominence-smoothed.

Between the acoustic variables and auditory–perceptual analysis, and for vowel /a/ emission (Table 3), VHG showed the following statistically significant correlations according to the Hinkle et al. [38] classification: moderate negative between CPPs and general degree ($p < 0.001$); strong negative between CPPs and breathiness ($p < 0.001$); weak positive between the alpha ratio and strain ($p = 0.019$); weak negative between L1–L0 and general degree ($p = 0.004$); and moderate negative between L1–L0 and breathiness ($p < 0.001$). DG had the following statistically significant correlations: moderate negative

between CPPs and general degree ($p < 0.001$); strong negative between CPPs and breathiness ($p < 0.001$); moderate negative between CPPs and strain ($p = 0.003$); and moderate negative between L1–L0 and breathiness ($p < 0.001$).

Regarding continuous speech (Table 4), VHG had the following statistically significant correlations according to Hinkle, Wierma, and Jurs [38]: weak negative between CPPs and general degree ($p = 0.007$); moderate negative between CPPs and roughness ($p = 0.001$); and weak negative between the alpha ratio and breathiness. DG had the following statistically significant correlations: weak negative between the alpha ratio and roughness ($p = 0.013$); weak negative between L1–L0 and general degree ($p = 0.039$); and weak negative between L1–L0 and roughness ($p = 0.049$).

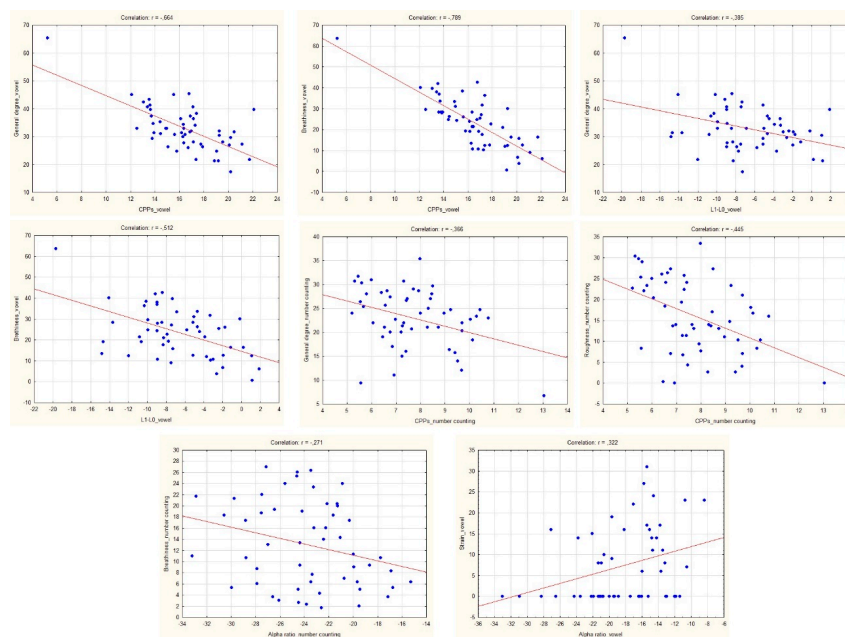Figures 2 and 3 show the scatter plots for VHG and DG statistically significant correlations, respectively.

Figure 2. Scatter plot with regression line regarding the statistically significant correlations in VHG.
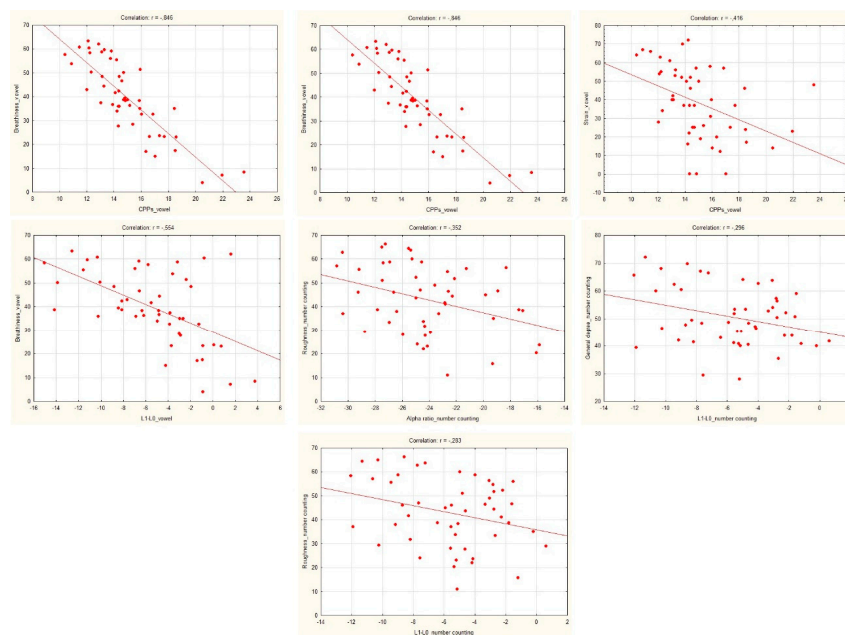

Figure 3. Scatter plot with regression line regarding the statistically significant correlations in DG.

4. Discussion

The human voice is a product of the aerodynamic energy (i.e., subglottic pressure provided by the lungs) conversion into acoustic energy [39]. The transglottal airflow pulse is determined by how long vocal folds remain closed, which is associated with the intensity and definition of harmonics produced in the glottal source [40]. Hence, incomplete glottal closure provides turbulent airflow, which reduces the acoustic signal harmonic structure and increases the amplitude variability [4,41], which results in the audible vocal quality disturbances classified as breathiness [42]. Furthermore, differences in muscle tone and mass in vocal folds cause the variability of frequency and amplitude in the acoustic signal and result in an audible vocal quality disturbance classified as roughness [4]. Roughness is described as an unpleasant low-pitched noise [42]. At last, strain and general degree indicate the auditory perception of vocal effort and global characteristics of the voice, respectively [43].

Regarding auditory–perceptual analysis, since the VAS scores were below 35.5 points, the VHG scores indicate a normal variation of voice quality [44]. DG had a mild to moderate vocal quality deviation (35.6 to 50.5 points). So, the complementary analysis evidenced that VHG had statistically inferior values than DG.

Since auditory–perceptual analysis may be inconsistent and biased due to its subjectivity [19], acoustic measures (i.e., robust and objective method for voice analysis) must be associated with the auditory–perceptual analysis to assist the understanding of phonation system mechanisms. Besides, CPPs and LTAS measures use robust algorithms, are more accurate, allow a complete analysis, are not invasive, are low cost as well as more used in studies to assess the vocal quality [16,23–26,42,45–47]. Thus, this study contributes to evidence by associating robust measures with auditory–perceptual analysis.

CPPs amplitude provides information on the dominant rahmonic periodicity, which corresponds to F_0 [9]. The harmonics are multiples of F_0 and are produced at glottal source in a one-second period (i.e., the occurrence moment of cepstral peak), called quefrency, which determines the period of a glottal cycle [15] and can be expressed by:

$$T = \frac{1}{F_0}$$

where T is the period. In this case, it corresponds to quefrency (s), and F_0 is the fundamental frequency value (Hz). We highlight that rahmonic and quefrency are anagrams to harmonic and frequency, respectively.

When CPPs have increased amplitude, there is a better harmonic definition, which is a characteristic of vocal fold cycle regularity [11]. We could observe that in our study as, for vowel /a/ emission, there were negative correlations with breathiness and general degree in both groups and with strain in DG. Regarding continuous speech, there were negative correlations with general degree and roughness in VHG. In other words, we believe that voices with an increased deviation of general degree, roughness, and breathiness have a less regular acoustic signal produced in the glottal source, thus decreased CPPs. Furthermore, excessive contraction of intrinsic and extrinsic laryngeal muscles that result in the auditory–perceptual parameter of strain [42], present with a higher deviation in dysphonics, can also reduce the acoustic signal periodicity.

The alpha ratio provides information on the spectral slope [27], and there was a negative correlation with breathiness for continuous speech in VHG. We observed that when there are increased alpha ratio values (less negative values), the deviation degree of breathiness decreases. A possible explanation is when the vocal tract (filter) is inertive and the flow declination rate is faster than the glottis area declination rate, the voice is brighter and more projected, leading to a less steep spectral slope (i.e., more energy in harmonics above 1 kHz) [33,48,49]. It is important to highlight that the glottis must be normotensive and close properly.

On the other hand, when there is a loudness or phonatory system rigidity increase, the subglottic pressure also increases to start phonation, making the closed phase of the glottal cycle longer [23] and increasing vocal strain. This situation also leads to a less steep spectral slope [23,28,42]. Our finds corroborate this information as we observed positive correlation (weak) of the alpha ratio and strain in

VHG, indicating that the increase of vocal strain (even with a normal variation of voice quality) also leads to a less steep spectral slope.

Even though participants in DG have behavioral dysphonia, characterized by increased loudness and closed phase of the glottal cycle, there were no significant correlations of the alpha ratio with breathiness and strain. However, for continuous speech, there was a negative correlation between the alpha ratio and roughness. We believe that the vocal quality aperiodicity of frequency and amplitude reduced the intensity of the harmonics above 1 kHz, thus decreasing alpha ratio values.

As previously described [23,28,42], the alpha ratio can indicate a less steep spectral slope in two situations: higher vocal projection and increased vocal strain. This fact was found in the complementary analysis of the current study. There was no significant difference between both groups, VHG and DG, for alpha ratio. We believe that people with behavioral dysphonia have vocal strain to compensate for other deviations of vocal quality and the inefficiency of the phonatory system. The absence of a significant difference between groups was also observed for L1–L0.

L1–L0 measure provides information on the phonation mode and is the difference between the F1 intensity level (L1) and the F_0 intensity level (L0) [31]. According to the non-linear source-filter coupling theory [50], the various filter configurations influence in vocal fold vibrations and emphasize different harmonics. Thus, the formants represent the mode that supraglottal airflow is molded by the filer [51]. Hence, the F1 intensity is determined by the mandibular lowering [51] and by the emission intensity [25]. At last, previous studies found that, when L0 is near or just below L1, the source has an efficient vibratory pattern due to the reactive vocal tract, and there is an improved vocal economy [32,52,53].

The significant correlations with L1–L0 were negative, which indicates that when that measure has lower values, there is a high auditory–perceptual parameter deviation. This could be seen in the current study regarding general degree and breathiness in both groups and roughness for DG. Breathy vocal patterns reduce L1–L0 [23], which was expected and corroborated our findings. Gauffin and Sundberg [28] suggest that the spectrum of voices with high glottal adduction L0 has lower energy than L1, however, the spectrum of voices with insufficient glottal adduction L0 has higher energy than L1. It is important to highlight that normal voices can also be breathy as a normal variation of voice quality.

The negative correlation of L1–L0 with roughness in DG was a surprising finding. We believe that roughness, a perceptual-auditory parameter generated from differences in muscle tone and mass in vocal folds (benign mass lesion; condition of DG), may have affected the phonation mode, making it less efficient with lower values of L1–L0.

Although continuous speech better represents the daily vocal use [54,55], this voice sample may have higher variability on its auditory–perceptual analysis due to the influences of each individual's articulation, prosody, and accent [4,34]. However, a sustained vowel, which is not biased in that way, has its auditory–perceptual parameter easily analyzed and may provide more information about the source. Thus, we think that was the reason for the weakest correlation degree with acoustic measures and auditory–perceptual analysis in continuous speech.

We encourage future studies to analyze the correlations of voice intensity with CPPs and LTAS measures. Thus, a complete and objective analysis will be possible. Also, analyzing these measures in different vocal conditions may provide a better understanding of phonatory adjustments among voice pathologies and source-filter interactions.

At last, the current study corroborates previous studies that analyzed other populations, such as vocally healthy and dysphonic women [4], participants with unilateral vocal fold paralysis [17], and analysis of sustained vowel in only dysphonic participants (women and men) [18]. Thus, the current study contributes to scientific evidence regarding the relationship and complementarity of the analysis of robust acoustic measures (CPPs, alpha ratio, and L1–L0). Hence, it provides a complete analysis of the vocal sample (sustained vowel and continuous speech) in vocally healthy participants and ones with behavioral dysphonia.

Since CPPs and LTAS measurements are more accurate, enable a complete analysis, are not invasive, and are low cost, they may be used in the clinical practice to assess clients and as research outcomes.

5. Conclusions

CPPs appear to be the best predictor of voice deviation in both studied populations because there was moderate to strong negative correlations with general degree, breathiness, roughness, and strain. Regarding L1–L0, because of moderate negative correlations, this measure is related to breathiness. Hence, L1–L0 provides information on air leakage through the closed glottis, assisting the phonatory efficiency analysis.

At last, the results which permeate alpha ratio analysis do not objectively reflect the physiological events of vocal production. Therefore, due to most of the weak correlations and their ambiguity, we suggest that this measure cannot stand alone. Thus, there needs to be other measures, such as CPPs and L1–L0, to complete it and provide a better understanding.

References

1. Zhang, Z. Mechanics of human voice production and control. *J. Acoust. Soc. Am.* **2016**, *140*, 2614–2635. [CrossRef] [PubMed]
2. Ramig, L.O.; Verdolini, K. Treatment efficacy. *J. Speech Lang. Hear. Res.* **1998**, *41*, S101–S116. [CrossRef] [PubMed]
3. Sataloff, R.T.; Abaza, M.M. Impairment, disability, and other medical-legal aspects of dysphonia. *Otolaryngol. Clin. N. Am.* **2000**, *33*, 1143–1152. [CrossRef]
4. Brinca, L.; Batista, A.P.F.; Tavares, A.I.; Gonçalves, I.C.; Moreno, M.L. Use of cepstral analyses for differentiating normal from dysphonic voices: A comparative study of connected speech versus sustained vowel in European Portuguese female speakers. *J. Voice* **2014**, *28*, 282–286. [CrossRef] [PubMed]
5. García, M.J.V.; Cobeta, I.; Martin, G.; Alonso-Navarro, H.; Jiménez-Jiménez, F.J. Acoustic analysis of voice in huntington's disease patients. *J. Voice* **2011**, *25*, 208–217. [CrossRef]
6. Olszewski, A.E.; Shen, L.; Jiang, J. Objective methods of sample selection in acoustic analysis of voice. *Ann. Otol. Rhinol. Laryngol.* **2011**, *120*, 155–161. [CrossRef]
7. Karnell, M.P.; Scherer, R.S.; Fischer, L.B. Comparison of acoustic voice perturbation measures among three independent voice laboratories. *J. Speech Lang. Hear. Res.* **1991**, *34*, 781–790. [CrossRef]
8. Bielamowicz, S.; Kreiman, J.; Gerratt, B.R.; Dauer, M.S.; Berke, G.S. Comparison of voice analysis systems for perturbation measurement. *J. Speech Lang. Hear. Res.* **1996**, *39*, 126–134. [CrossRef]
9. Gaskill, C.S.; Awan, J.A.; Watts, C.R.; Awan, S.N. Acoustic and perceptual classification of within-sample normal, intermittently dysphonic, and consistently dysphonic voice types. *J. Voice* **2017**, *31*, 218–228. [CrossRef]
10. Maryn, Y.; Roy, N.; De Bodt, M.; Van Cauwenberge, P.; Corthals, P. Acoustic measurement of overall voice quality: A meta-analysis. *J. Acoust. Soc. Am.* **2009**, *126*, 2619–2634. [CrossRef]
11. Hillenbrand, J.; Cleveland, R.A.; Erickson, R.L. Acoustic correlates of breathy vocal quality. *J. Speech Lang. Hear. Res.* **1994**, *37*, 769–778. [CrossRef] [PubMed]
12. Kumar, B.R.; Bhat, J.S.; Prasad, N. Cepstral analysis of voice in persons with vocal nodules. *J. Voice* **2010**, *24*, 651–653. [CrossRef] [PubMed]

13. Hunter, E.J.; Titze, I.R. The voice use profile; illustrating actual voice use from long term monitoring from the National Center for Voice and Speech voice dosimeter. *J. Acoust. Soc. Am.* **2007**, *121*, 3201. [CrossRef]

14. Hillenbrand, J.M.; Houde, R.A. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J. Speech Lang. Hear. Res.* **1996**, *39*, 311–321. [CrossRef] [PubMed]

15. Sujitha, S.P.; Pebbili, G.K. Cepstral analysis of voice in young adults. *J. Voice* **2020**. [CrossRef]

16. Hasanvand, A.; Salehi, A.; Ebrahimipour, M. A cepstral analysis of normal and pathologic voice qualities in Iranian adults: A comparative study. *J. Voice* **2017**, *31*, 508.e17–508.e23. [CrossRef]

17. Heman-Ackah, Y.D.; Michael, D.D.; Goding, G.S. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J. Voice* **2002**, *16*, 20–27. [CrossRef]

18. Lopes, L.W.; Sousa, E.S.D.S.; Da Silva, A.C.F.; Da Silva, I.M.; De Paiva, M.A.A.; Vieira, V.J.D.; De Almeida, A.A. Medidas cepstrais na avaliação da intensidade do desvio vocal. *CoDAS* **2019**, *31*, e20180175. [CrossRef]

19. Tanner, K.; Roy, N.; Ash, A.; Buder, E.H. Spectral moments of the long-term average spectrum: Sensitive indices of voice change after therapy? *J. Voice* **2005**, *19*, 211–222. [CrossRef]

20. Master, S.; De Biase, N.; Pedrosa, V.; Chiari, B.M. O espectro médio de longo termo na pesquisa e na clínica fonoaudiológica. *Pró-Fono Revista Atualização Científica* **2006**, *18*, 111–120. [CrossRef]

21. Linville, S.E.; Rens, J. Vocal tract resonance analysis of aging voice using long-term average spectra. *J. Voice* **2001**, *15*, 323–330. [CrossRef]

22. Kitzing, P. LTAS criteria pertinent to the measurement of voice quality. *J. Phon.* **1986**, *14*, 477–482. [CrossRef]

23. Guzmán, M.; Correa, S.; Muñoz, D.; Mayerhoff, R. Influence on spectral energy distribution of emotional expression. *J. Voice* **2013**, *27*, 129.e1–129.e10. [CrossRef]

24. Guzmán, M.; Higueras, D.; Fincheira, C.; Muñoz, D.; Aud, C.G.; Dowdall, J. Immediate acoustic effects of straw phonation exercises in subjects with dysphonic voices. *Logop. Phoniatr. Vocology* **2013**, *38*, 35–45. [CrossRef]

25. Da Silva, P.T.; Master, S.; Andreoni, S.; Pontes, P.A.D.L.; Ramos, L.R. Acoustic and long-term average spectrum measures to detect vocal aging in women. *J. Voice* **2011**, *25*, 411–419. [CrossRef] [PubMed]

26. Master, S.; De Biase, N.; Chiari, B.M.; Laukkanen, A.-M. Acoustic and perceptual analyses of brazilian male actors' and nonactors' voices: Long-term average spectrum and the "Actor's Formant". *J. Voice* **2008**, *22*, 146–154. [CrossRef]

27. Frokjaer-Jensen, B.; Prytz, S. Registration of voice quality. *Bruel Kjaer Tech. Rev.* **1976**, *3*, 3–17.

28. Gauffin, J.; Sundberg, J. Spectral correlates of glottal voice source waveform characteristics. *J. Speech Lang. Hear. Res.* **1989**, *32*, 556–565. [CrossRef]

29. Laukkanen, A.-M.; Kankare, E. Vocal loading-related changes in male teachers' voices investigated before and after a working day. *Folia Phoniatr. Logop.* **2006**, *58*, 229–239. [CrossRef]

30. Nordenberg, M.; Sundberg, J. Effect on LTAS of vocal loudness variation. *Logop. Phoniatr. Vocology* **2004**, *29*, 183–191. [CrossRef]

31. Löfqvist, A. The long-time-average spectrum as a tool in voice research. *J. Phon.* **1986**, *14*, 471–475. [CrossRef]

32. Andrade, P.A.; Wood, G.; Ratcliffe, P.; Epstein, R.; Pijper, A.; Švec, J.G. Electroglottographic study of seven semi-occluded exercises: Laxvox, straw, lip-trill, tongue-trill, humming, hand-over-mouth, and tongue-trill combined with hand-over-mouth. *J. Voice* **2014**, *28*, 589–595. [CrossRef] [PubMed]

33. Titze, I.R.; Laukkanen, A.-M. Can vocal economy in phonation be increased with an artificially lengthened vocal tract? A computer modeling study. *Logop. Phoniatr. Vocology* **2007**, *32*, 147–156. [CrossRef]

34. Behlau, M.; Zambon, F.; Moreti, F.; Oliveira, G.; Couto, E.D.B. Voice self-assessment protocols: Different trends among organic and behavioral dysphonias. *J. Voice* **2017**, *31*, 112.e13–112.e27. [CrossRef] [PubMed]

35. Watts, C.R.; Awan, S.N.; Maryn, Y. A comparison of cepstral peak prominence measures from two acoustic analysis programs. *J. Voice* **2017**, *31*, 387.e1–387.e10. [CrossRef]

36. Phadke, K.V.; Laukkanen, A.-M.; Ilomäki, I.; Kankare, E.; Geneid, A.; Švec, J.G. Cepstral and perceptual investigations in female teachers with functionally healthy voice. *J. Voice* **2020**, *34*, 485.e33–485.e43. [CrossRef]

37. Fleiss, J.L.; Levin, B.; Paike, M.C. *Statistical Methods for Rates and Proportionsi*, 3rd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2003; pp. 598–626.

38. Witz, K.; Hinkle, D.E.; Wiersma, W.; Jurs, S.G. Applied statistics for the behavioral sciences. *J. Educ. Stat.* **1990**, *15*, 84. [CrossRef]

39. Titze, I.R. Acoustic interpretation of resonant voice. *J. Voice* **2001**, *15*, 519–528. [CrossRef]

40. Doval, B.; D'Alessandro, C.; Henrich, N. The spectrum of glottal flow models. *Acta Acust. United Acust.* **2006**, *92*, 1026–1046.

41. Awan, S.N.; Krauss, A.R.; Herbst, C.T. An examination of the relationship between electroglottographic contact quotient, electroglottographic decontacting phase profile, and acoustical spectral moments. *J. Voice* **2015**, *29*, 519–529. [CrossRef]

42. Watts, C.R.; Awan, S.N. 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*, 1525–1537. [CrossRef]

43. American Speech-Language-Hearing Association. Available online: https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589942600&section=Assessment (accessed on 15 March 2020).

44. Yamasaki, R.; Madazio, G.; Leão, S.H.; Padovani, M.; Azevedo, R.; Behlau, M. Auditory-perceptual evaluation of normal and dysphonic voices using the voice deviation scale. *J. Voice* **2017**, *31*, 67–71. [CrossRef] [PubMed]

45. Leino, T. Long-term average spectrum in screening of voice quality in speech: Untrained male university students. *J. Voice* **2009**, *23*, 671–676. [CrossRef] [PubMed]

46. Burk, B.R.; Watts, C.R. The effect of parkinson disease tremor phenotype on cepstral peak prominence and transglottal airflow in vowels and speech. *J. Voice* **2019**, *33*, 580.e11–580.e19. [CrossRef] [PubMed]

47. Antonetti, A.E.D.S.; Ribeiro, V.V.; Brasolotto, A.G.; Silvério, K.C.A. Effects of performance time of the voiced high-frequency oscillation and lax vox technique in vocally healthy subjects. *J. Voice* **2020**. [CrossRef] [PubMed]

48. Titze, I.R. Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation. *J. Speech Lang. Hear. Res.* **2006**, *49*, 439–447. [CrossRef]

49. Sundberg, J.; Salomão, G.L.; Scherer, K.R. Analyzing emotion expression in singing via flow glottograms, long-term-average spectra, and expert listener evaluation. *J. Voice* **2019**. [CrossRef]

50. Titze, I.R. Nonlinear source–filter coupling in phonation: Theory. *J. Acoust. Soc. Am.* **2008**, *123*, 2733–2749. [CrossRef]

51. Sundberg, J. *Ciência da Voz: Fatos Sobre a Voz na Fala e no Canto*, 1st ed.; EdUSP: São Paulo, Brazil, 2018; pp. 41–47.

52. Story, B.H.; Laukkanen, A.-M.; Titze, I.R. Acoustic impedance of an artificially lengthened and constricted vocal tract. *J. Voice* **2000**, *14*, 455–469. [CrossRef]

53. Titze, I.R. The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.* **1988**, *83*, 1536–1552. [CrossRef]

54. Maryn, Y.; Weenink, D. Objective dysphonia measures in the program praat: Smoothed cepstral peak prominence and acoustic voice quality index. *J. Voice* **2015**, *29*, 35–43. [CrossRef] [PubMed]

55. Maryn, Y.; Corthals, P.; Van Cauwenberge, P.; Roy, N.; De Bodt, M. Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *J. Voice* **2010**, *24*, 540–555. [CrossRef] [PubMed]