Use of Cepstral Analyses for Differentiating Normal From Dysphonic Voices: A Comparative Study of Connected Speech Versus Sustained Vowel in European Portuguese Female Speakers

*Lilia F. Brinca, *Ana Paula F. Batista, *Ana Inês Tavares, †Ilídio C. Gonçalves, and †Maria L. Moreno, *Faro and †Portimão, Portugal

Summary: Objective. The aim of this study was to investigate the use of cepstral peak prominence (CPP) and CPP-smoothed (CPPs) to differentiate dysphonic from nondysphonic voices, using two speech tasks: sustained vowel /a/ and connected speech.

Study Design. A retrospective study was based on data selected from an archival database of recorded voices.

Methods. Sixty age- and occupation-matched individuals (30 participants with dysphonia and 30 controls) were recorded producing the sustained vowel /a/ and reading the European Portuguese version of "The Story of Arthur the Rat." Recorded voices were analyzed acoustically by measuring CPP and CPPs and auditory-perceptual ratings were related to the acoustic measurements.

Results. For the sustained vowel, both CPP and CPPs measures were significantly different between dysphonic and control groups. For connected speech, only CPP values revealed significant differences between the two groups, both in direct and narrative speech. Acoustic measurements correlated with the auditory-perceptual classifications in both sustained vowel and connected speech, although the strongest correlation ($0.6 < r < 0.7$) was obtained between CPP and the perception of breathiness.

Conclusions. The results of this study suggest that analysis of CPP and CPPs is a promising tool in clinical practice with European Portuguese speakers.

Key Words: Dysphonia–Cepstral analysis–Auditory-perceptual rating–Female–Sustained vowel–Connected speech.

INTRODUCTION

Clinical assessment of dysphonia often relies on a combination of auditory-perceptual and acoustic measurement techniques. Acoustic measures have been developed to quantify voice characteristics and, whenever possible, to use them as predictors of dysphonia. A normal voice signal is periodic and should have very little cycle-to-cycle variability in frequency or amplitude. Pathologic lesions of the vocal fold often include breathiness and roughness. Breathiness is often a result of incomplete glottis closure which results in a turbulent airflow. This turbulent flow of air produces a nonperiodic acoustic signal, with an unpredictable cycle-to-cycle variation in intensity. Roughness is derived from the lack of periodicity in vocal fold vibration that results from differences in muscle tension. This aperiodicity produces an acoustic signal with a high variability in both amplitude and frequency.

The degree of signal periodicity can be measured using *time-based* analysis[1–4] or *frequency*-based methods.[5–10] Among the second type of methods, cepstral peak prominence (CPP) correlates most robustly to breathiness in sustained vowels as well as in connected speech.[11] Moreover, because CPP does not depend on the accuracy of fundamental frequency extraction, which is difficult to establish for severely disordered voices, it tends to be a more reliable indicator than other approaches.[5,7] According to Hillenbrand and Houde[12] "…a cepstrum is a log power spectrum of a log power spectrum. For periodic signals, the first power spectrum will show energy at harmonically related frequencies and the second spectrum will show a strong component corresponding to the regularity of the harmonic peaks." The rationale behind the measurement of CPP is that periodic voice signals display well-defined harmonic configuration in the spectrum and thus a more prominent cepstral peak.[5] Therefore, a decrease in overall CPP is indicative of voice abnormality.[5,12,13] Evidence suggests that measurements of CPP derived from the acoustic spectrum correlate best with auditory-perceptual classifications of dysphonia.[5,6,11,12]

Because CPP is a measure of periodicity, it should also predict voice roughness, although some authors have failed to identify a strong correlation between these factors.[14] In some cases,[15] the measurement of CPP significantly predicted all auditory voice quality judgments except the roughness in vowels. A simple modification to the CPP algorithm, the CPP-smoothed (CPPs), produced a noticeable improvement in prediction accuracy of this approach for assessment of voice quality. An additional processing step is used to generate CPPs and involves smoothing the individual cepstra before extracting the cepstral peak and calculating the peak prominence.[12]

Most auditory-perceptual studies of voice have focused more on sustained vowels than on connected speech, mainly because vowels are easily elicited and less affected by articulation and dialectal influences. However, connected speech is more

From the *Laboratory of Audiology and Speech Therapy, Centre for Research and Development in Health, School of Health, University of Algarve, Faro, Portugal; and the †Service of Otolaryngology, Algarve Central Hospital, Portimão, Portugal.

Address correspondence and reprint requests to Lilia Figueiredo Brinca, School of Health, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal. E-mail: lbrinca@ualg.pt

Journal of Voice, Vol. 28, No. 3, pp. 282-286
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2013.10.001

representative of a person's daily voice, and therefore, the inclusion of both stimulus, sustained vowel and connected speech, is important. Moreover, some vocal inconstancies are more clearly observed in connected speech rather than in sustained vowels, which is very important for an auditory-perceptual voice evaluation.[16,17] The aim of this study was to investigate the importance of CPP and CPPs to differentiate dysphonic from nondysphonic female voices, using two different speech tasks: sustained vowel /a/ and connected speech.

METHOD

Participants

The present study was based on the data selected from an archival database of voices recorded during a voice survey conducted in our laboratory. Subjects were divided into two groups. The clinical group consisted of 30 females with vocal dysfunction, ranging in age from 19 to 66 years (Table 1), with a mean age of 43.9 years. Diagnosis of vocal pathology was based on the laryngoscopic examination conducted by an experienced otorhinolaryngologist, through an endoscopy procedure. The control group comprised 30 individuals for whom laryngologic examination revealed normal structure and function of the vocal folds. Informed consent was obtained from all the subjects who participated in the study.

TABLE 1.
Summary of Subject Information for Participants in the Dysphonic Group

Subject No.	Age	Diagnosis
1	29	LG
2	63	VFN
3	55	LG*
4	43	LG
5	43	LG*
6	41	LG
7	29	VFE*
8	46	LG
9	41	LG
10	48	LG
11	61	LG
12	52	LG
13	34	LG
14	51	LG
15	43	VFN
16	58	LG
17	42	LG
18	54	VFN
19	38	LG
20	66	VFN
21	35	UVFP
22	58	LG
23	48	VFN
24	36	LG
25	36	VFN
26	19	VFN
27	27	VFE
28	29	VFC
29	72	LG
30	21	VFN

Abbreviations: LG, longitudinal gap; VFC, vocal fold cyst; VFE, vocal folds edema; VFN, vocal folds nodules; UVFP, unilateral vocal fold paralysis.
* Mild edema (gastroesophageal reflux).

Procedures

Signal processing. The subjects were seated on a comfortable chair. Voices were recorded in a soundproof booth using an omnidirectional microphone (DPA 4006-TL, P48) (Harman International Industries Ldt., Hertfordshire, UK), pointed to the speaker's mouth at a distance of approximately 30 cm.[18] Outside the soundproof booth, the microphone was connected to a computer through a mixer (Soundcraft COMPACT 4) and a USB audio interface (Edirol UA-1EX) (DPA Microphones A/S, Denmark). All vocal recordings were digitized at a sampling rate of 44.1 kHz and a resolution of 16 bits and saved in WAV-format, using *Audacity* 2.0.2, a freeware program (GNU General Public License/GPL).

The speech materials used for the auditory-perceptual/acoustic measurements were (1) the sustained vowel [a] and (2) two sentences extracted from the oral reading of the European Portuguese version of "The Story of Arthur the Rat."[19] Because the vowel /a/ is produced with a low height of the tongue and with an open mouth, its utterance provides a relatively stable condition of the phonatory system and, therefore, gives a reliable assessment of larynx stability.[18] The rationale for using oral reading was to obtain uniform samples (with the same linguistic content) of connected speech.[19] Subjects were asked to sustain the vowel and read aloud at a comfortable pitch, loudness, and speaking rate.

Acoustic measures. Values of CPP and CPPs were computed for the sustained vowel /a/ and the connected speech, using the algorithms described in the literature.[12]

The mean CPP and CPPs values of the sustained vowel /a/ were extracted from the central portion of the recordings of the second utterance. The length of the mid-vowel segment chosen for analysis was 1 second and choice was based on the reported use and appropriateness of this time in a meta-analysis[11] of previous studies. Connected speech analysis was focused on two sentences extracted from the oral reading of the European Portuguese version of "The Story of Arthur the Rat"[19]—"…não podemos esperar todo o dia até que te decidas. Toca a andar! Marchar!…" and "…nessa noite houve uma grande tempestade que fez tremer a terra …." The first sentence is an example of direct speech (that sounds more like natural conversation) and the second was chosen from a narrative part of the text. Using both natural and narrative types of text gives a better idea of the speaker's voice characteristics. The silent intervals in the recordings were not removed.

Auditory-perceptual rating. The evaluation of voice quality was made using the GRBAS scale, which comprises five parameters. Each parameter represents a dimension of phonation: G represents the degree of overall voice abnormality, R represents roughness, B represents breathiness, A represents asthenia (weakness), and S represents strain. The GRBAS scale uses a 4-point Likert scale of 0 (normal) to 3 (extreme) for all five parameters.

Three speech and language therapists, with more than 3 years of clinical experience in judging dysphonia severity, performed an auditory-perceptual evaluation of voice quality by listening to the taped sustained vowel /a/and oral readings from the European Portuguese version of "The Story of Arthur the Rat."[19] Voice samples for evaluation were presented in a random order. The auditory-perceptual evaluation took place in an ordinary room, and all recordings were presented from a computer using good quality headphones.

The percentage agreement between raters was calculated according to the number of diagnoses where there was total agreement between the three therapists. An overall percentage rate of agreement of 70% was obtained between the three speech and language therapists. The overall coherence of listeners was measured using the intraclass correlation coefficient.

Analyses

All samples were labeled manually in the digitized sound wave using the computer program *Praat* (GNU General Public License/GPL).[20] CPP and CPPs values were obtained using Hillenbrand software, available from http://homepages.wmich.edu/~hillenbr/. Comparisons of the mean values of CPP and CPPs from dysphonic and control groups were made applying a Student t test for independent samples. Significance was set at $P < 0.05$. The linear correlation between acoustic measures and auditory-perceptual ratings was measured using the Spearman product-moment correlation coefficient. All statistical analyses were completed using *SPSS* for Windows version 17.0 (SPSS, Inc, Chicago, IL).

RESULTS

Sustained vowel /a/

A Student t test for independent samples was used to compare the CPP and CPPs means from clinical and control groups (Figures 1 and 2). The results revealed that a significantly lower value of CPP ($P = 0.031$) and CPPs ($P = 0.055$) were obtained in dysphonic individuals.

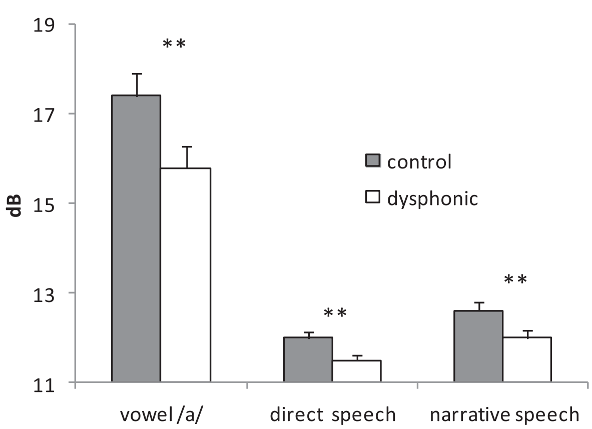

FIGURE 1. Mean ± standard error of the CPP values from dysphonic and control groups; **significant difference at $P \leq 0.05$.

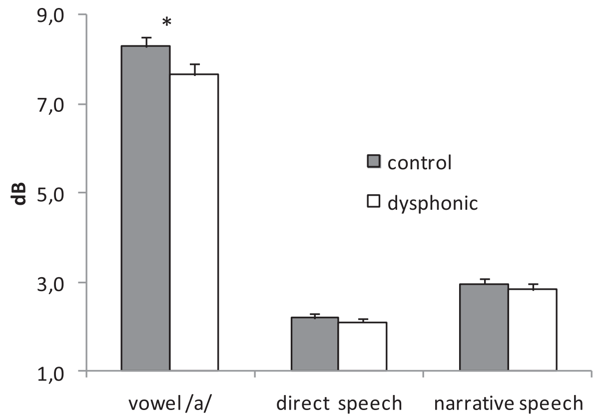

FIGURE 2. Mean ± standard error of the CPPs values from dysphonic and control groups; *significant difference at $P \leq 0.10$.

Connected speech

Figures 1 and 2 show the results obtained for CPP evaluation in clinical and control groups for both direct and narrative speech types. The CPP values were significantly different between the two groups in both direct ($P = 0.029$) and narrative speech ($P = 0.032$). No significant differences were obtained for CPPs values in the dysphonic groups compared with the control for both direct and narrative speech.

Auditory-perceptual measures

A Spearman r correlation was computed to determine the strength of the relationship between acoustic and auditory-perceptual measures using the sustained vowel and connected speech samples. The results indicated that values obtained for CPP and CPPs correlated with the auditory-perceptual classifications (Table 2). The strongest correlation was obtained between values of CPP and the perceived degree of breathiness for sustained vowels, direct speech, and narrative speech recordings.

Measures of interrater reliability for breathiness, the voice characteristic more closely related to the acoustic measures, are included in Table 3. The results indicated that the listeners could differentiate between the different levels of breathiness.

DISCUSSION

As a whole, the results of the present study are in agreement with the conclusions from several other studies.[5,7,9,10,12,14,15,21–24] Our results confirmed that cepstral measures are a promising acoustic measure of dysphonia severity, although Metha and Hillman[25] drew attention to the fact that "…the interpretation of cepstral measures relative to the underlying physiology of vocal fold vibration are not as intuitive as more traditional measurements. Clearly more studies are required to better delineate the relationship between cepstral measures and vocal fold function".

TABLE 2.
Spearman Correlation (r) Between Auditory-Perceptual and Acoustic Measures

Speech Materials	Acoustic Measure	G	R	B	A	S
Sustained vowel /a/	CPP	−0.608**	−0.231	−**0.665****	−0.509**	−0.080
	CPPs	−0.588**	−0.404	−0.461*	−0.435*	−0.023
Oral reading (direct speech)	CPP	−0.231	−0.116	−**0.565****	−0.362	0.217
	CPPs	−0.131	−0.025	−0.080	−0.088	0.362
Oral reading (narrative speech)	CPP	−0.491**	−0.215	−**0.583****	−0.383	−0.261
	CPPs	−0.450*	−0.267	−0.431*	−0.090	−0.311

Notes: Values in bold indicate the best correlations between acoustic and auditory-perceptual measures.
*Correlation significant at the 0.05 level.
**Correlation significant at the 0.01 level.

In the present study, the values obtained for CPP consistently emerged as a significant factor in predicting dysphonia in both sustained vowel and connected speech. The best correlation coefficients ($0.6 < r < 0.7$) were obtained between CPP and grade B of the GRBAS scale. Breathiness is often a result of incomplete glottis closure, and approximately 80% of the sample in the present study was constituted of speakers with vocal nodules or a longitudinal gap between the vocal folds.

There are several limitations inherent to the methodology used in the present study that need to be addressed in future work. One is related to the fact that the study only assessed female voices. Female voices were specifically selected because the majority of our patients seeking help for voice difficulties are women. It will be important in the future to verify if the cepstral values measured in this study are the same when the study is repeated with male voices. Future studies comparing females versus males may provide further information about the importance of these acoustic measures for studies of dysphonic voices in European Portuguese speakers.

Another limitation of the study is related to the lack of homogeneity of voice disorders in the clinical group. The most frequent laryngeal pathology was a functional disorder (longitudinal gap) followed by discrete mass lesions of the vocal folds (mainly nodules). But participants with other laryngeal pathologies were also included in the sample. In future studies, the dysphonic group should be as homogenous as possible to minimize interspeaker variability. The validity of comparing different pathologies in the same group with a normal group should be emphasized in future work. Moreover, most of the sample comprised speakers with slight-to-moderate dysphonia, which could explain the poorer differentiation between dysphonic and normal voices.

The difference found in this study between the dysphonic and control groups concerning connected speech could also be contaminated by other factors. According to Baken and Orlikoff[18] "…measures of connected speech are dependent on the proportion of the sample occupied by silent or quasi-silent intervals." It was noted during the study that the participants read the text with different levels of fluency, and this difference was related to the heterogeneity in academic and social backgrounds. A region-dependent accent was also detected in some of the study participants. Considering the following statement made by Ladefoged and Disner[26] "…even without considering differences of accent, the range of human voices is enormous…," with this in mind, it will be important in the future to use a more homogenous sample of speakers that share similar linguistic and academic backgrounds.

CONCLUSIONS

In relation to the sustained vowel /a/, the results of the present study revealed that CPP and CPPs had significantly lower values in the dysphonic group. For the connected speech, significantly lower values of CPP were obtained in the dysphonic group compared with the control group, but no differences were obtained with CPPs. With both speech tasks, sustained vowel and connected speech, the values obtained for CPP correlated strongly with breathy voices. The results of the study suggest that CPP and CPPs are promising tools in clinical practice with the European Portuguese speakers. However, as suggested by some authors,[22] it will be important in future studies to incorporate different time-based and spectral-based acoustic measures to make a comprehensive analysis of the voice signal in the Portuguese population.

TABLE 3.
Measures of Interrater Reliability of a Single Auditory-Perceptual Quality—Breathiness

Intraclass Correlation Coefficient	Sustained Vowel /a/	Oral Reading (Direct Speech)	Oral Reading (Narrative Speech)
Reliability of mean rating	0.676	0.832	0.876
95% Confindence interval	Lower = 0.400	Lower = 0.689	Lower = 0.771
	Upper = 0.837	Upper = 0.916	Upper = 0.938

Acknowledgment

The authors thank the speech and language therapist Patrícia Nogueira Pinto for her contribution to the perceptual evaluation of voice quality.

REFERENCES

1. García MJV, Cobeta I, Martín G, Alonso-Navarro H, Jimenez-Jimenez FJ. Acoustic analysis of voice in Huntington's disease patients. *J Voice*. 2011;25:208–217.
2. Ma EP, Yiu EM. Multiparametric evaluation of dysphonic severity. *J Voice*. 2006;20:380–390.
3. Niebudek–Bogusz E, Fiszer M, Kotylo P, Sliwinska-Kowalska M. Diagnostic value of voice acoustic analysis in assessment of occupational voice pathologies in teachers. *Logoped Phoniatr Vocol*. 2006;31:100–106.
4. Olszevvski AE, Shen L, Jiang JJ. Objective methods of sample selection in acoustic analysis of voice. *Ann Otol Rhinol Laryngol*. 2011;120:155–161.
5. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Hear Res*. 1994;37:769–778.
6. Awan SN, Roy N. Acoustic prediction of voice type in women with functional dysphonia. *J Voice*. 2005;19:268–282.
7. Kumar BR, Bhat JS, Prasad N. Cepstral analysis of voice in persons with vocal nodules. *J Voice*. 2010;24:651–653.
8. Maryn Y, De Bodt M, Roy N. The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord*. 2010;43:161–174.
9. Alpan A, Schoentgen J, Maryn Y, Grenez F, Murphy P. Assessment of disordered voice via the first rahmonic. *Speech Commun*. 2012;54:655–663.
10. 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;54:1525–1537.
11. Maryn Y, Roy N, De Bodt M, Paul MD, Van Cauwenberge P, Corthals P. Acoustic measurement of overall voice quality: a meta-analysis. *J Acoust Soc Am*. 2009;126:2619–2634.
12. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Hear Res*. 1996;39:311–321.
13. Blankenship B. The timing of nonmodal phonation in vowels. *J Phon*. 2002;30:163–191.
14. Heman-Ackah YD, Michael DD, Goding GS. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice*. 2002;16:20–27.
15. Eadie TL, Baylor CR. The effect of auditory-perceptual training on inexperienced listeners' judgments of dysphonic voice. *J Voice*. 2006;20:527–544.
16. Yiu E, Worrall L, Longland J, Mitchell C. Analysing vocal quality of connected speech using Kay's computerized speech lab: a preliminary finding. *Clin Linguist Phon*. 2000;14:295–305.
17. Zraick RI, Wendel K, Smith-Olinde L. The effect of speaking task on auditory-perceptual judgment of the severity of dysphonic voice. *J Voice*. 2005;19:574–581.
18. Baken RJ, Orlikoff RF. *Voice Clinical Measurement*. San Diego, CA: Singular Publishing Group; 2000.
19. Guimarães I, Abberton E. Fundamental frequency in speakers of Portuguese for different voice samples. *J Voice*. 2005;19:592–606.
20. Boersma P, Weenink D. Praat: Doing Phonetics by Computer [v. 5.3.16]. 2005. Available at: www.fon.hum.uva.nl/praat.
21. Heman-Ackah YD, Heuer RJ, Michael DD, et al. Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol*. 2003;112:324–333.
22. Awan SN, Roy N. Outcomes measurement in voice disorders: application of an acoustic index of dysphonia severity. *J Speech Lang Hear Res*. 2009;52:482–499.
23. Lowell SY, Colton RH, Kelley RT, Hahn YC. Spectral- and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker. *J Voice*. 2011;25:223–232.
24. Moers C, Bernd M, Rosanowski F, Elmar N, Ulrich E, Tino H. Vowel- and text-based cepstral analysis of chronic hoarseness. *J Voice*. 2012;26:416–424.
25. Mehta DD, Hillman RE. Voice assessment: updates on auditory-perceptual, acoustic, aerodynamic, and endoscopic imaging methods. *Curr Opin Otolaryngol Head Neck Surg*. 2008;16:211–215.
26. Ladefoged P, Disner SF. *Vowels and Consonants*. 3rd ed. Chichester, UK: Wiley-Blackwell; 2012.