A Comparison of Cepstral Peak Prominence Measures From Two Acoustic Analysis Programs

*Christopher R. Watts, †Shaheen N. Awan, and ‡Youri Maryn, *Fort Worth, Texas, †Bloomsburg, Pennsylvania, and ‡Antwerp, Belgium

Summary: Purpose. This study aimed to investigate the relationship and reliability of cepstral peak prominence (CPP) measures from two acoustic software applications, *Analysis of Dysphonia in Speech and Voice* (*ADSV*) and *Praat*.

Methodology. Flemish and English recordings of sustained vowels and connected speech samples were analyzed using *ADSV* and *Praat*. Correlational analyses and measures of the standard error of the estimate were applied to the vowel and connected speech data obtained from the two programs.

Results. Analyses revealed very strong relationships (eg, r > 0.88) between CPP measures derived from *ADSV* and those derived from *Praat*, regardless of context (vowel or connected speech) or language spoken. Average residual errors ranged from 0.55 to 1.1 dB for the prediction of *Praat* CPP data from actual observed *ADSV* CPP data, and average residual errors ranged from 0.57 to 1.58 dB for the prediction of *ADSV* CPP data from actual observed *Praat* CPP data.

Conclusions. Measurements of CPP derived from *ADSV* and *Praat* manifested strong parallel-forms reliability. Although CPP data values obtained *via* these programs will be different owing to algorithmic processing differences, this study found that estimated CPP values derived using regression equations could be transformed between programs with relatively small predictive error, regardless of language. The strong measurement relationships indicate that CPP values from either program have a high degree of shared variance and may be expected to differentiate across a wide range of voice signal periodicity in a relatively similar fashion. This finding supports the use of either program in clinical use and voice science research.

Key Words: Cepstrum–Cepstral peak prominence–*Analysis of Dysphonia in Speech and Voice*–*Praat*–Voice disorders.

INTRODUCTION

The cepstral peak prominence (CPP) is an acoustic measurement obtained from the cepstrum of a sound wave that has shown great promise as an acoustic marker of dysphonia. The cepstral peak is the dominant "rahmonic" (an anagram of "harmonic") within the cepstrum (an anagram of "spectrum"), which is computed *via* a Fourier transform of the logarithm power spectrum of a recorded sound wave. The relative amplitude of the cepstral peak in relation to the expected amplitude of the cepstral peak as estimated *via* linear regression has been referred to as the "cepstral peak prominence"[1,2] and represents the degree of periodicity in the voice signal. A periodic signal will show a prominent (ie, distinct, high amplitude) cepstral peak and high CPP values corresponding to a well-defined fundamental frequency and harmonic structure, whereas dysphonic voice signals with disturbed periodicity are associated with a decrease in amplitude of the cepstral peak (ie, lower harmonic energy) and lower CPP values.[3,4] The predictive accuracy of identifying the cepstral peak *via* automatic signal processing methods is improved by smoothing cepstra across cepstral frames and across quefrency, and most recent reports of computer algorithms used to calculate CPP for voice signals incorporate these smoothing procedures.[a]

Although measurements obtained from cepstral analysis have existed for many decades, only in recent years have they been applied to the acoustic profiling of dysphonia. Hillenbrand and colleagues demonstrated that measurements of CPP served as strong acoustic correlates for perceptions of breathiness in simulated and actual voice impairment.[1,2] Heman-Ackah and colleagues extended this work by reporting strong correlations of CPP with overall dysphonia severity compared with time-based measurements of jitter and shimmer,[5] and later demonstrated strong sensitivity and specificity for the CPP in differentiating normal from individuals with dysphonia.[6] Awan and colleagues have further established the robustness of CPP as an objective correlate and sensitive marker of dysphonic severity in both sustained vowel and continuous speech in various clinical populations, both as an individual measure and in combination with other spectral and cepstral acoustic parameters as part of the multidimensional estimate of dysphonia severity, referred to as the Cepstral Spectral Index of Dysphonia.[4,7–16] Maryn and colleagues have also similarly reported on the robust nature of the CPP as a measure of dysphonia and as the strongest of multiple acoustic predictors combined in the multidimensional Acoustic Voice Quality Index (AVQI).[14–16]

The CPP is emerging as a favored method for the acoustic profiling of disordered voice and as an objective supporting measure for the auditory-perceptual assessment of dysphonia severity. Reasons supporting the use of the CPP in the description of dysphonic voice quality include the following: (1) the

[a]Smoothing of cepstra before identifying the cepstral peak is available in both of the computer programs to be used in this study, either as a default setting or as an option. For consistency, the term "CPP" used throughout the remainder of this manuscript will assume that some degree of smoothing was incorporated in the calculation.

Accepted for publication September 13, 2016.

Disclosure: Dr. S.N. Awan is a consultant to Pentax Medical/KayPentax (Montvale, NJ) for the development of commercial computer software including cepstral analysis of continuous speech algorithms. KayPentax licenses the algorithms that form the basis of the *Analysis of Dysphonia in Speech and Voice* (*ADSV*) program from Dr. Awan.

From the *Texas Christian University, Fort Worth, Texas; †Bloomsburg University, Bloomsburg, Pennsylvania; and the ‡European Institute for ORL, Sint-Augustinus Hospital, Antwerp, Belgium.

Address correspondence and reprint requests to Christopher R. Watts, Davies School of Communication Sciences & Disorders, Texas Christian University, TCU Box 297450, Fort Worth, TX 76129. E-mail: c.watts@tcu.edu

Journal of Voice, Vol. 31, No. 3, pp. 387.e1–387.e10
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.09.012

availability of numerous studies demonstrating that measures of CPP correlate more strongly with perceptions of dysphonia severity than other univariate acoustic measurements; (2) measures of CPP are extracted from frames of voice signal data *vs.* the identification of cycle-to-cycle variations as measured using traditional perturbation measures such as jitter and shimmer and, therefore, can be applied to moderate and severe dysphonic voices in which definitive cycles of vibration are questionable or absent; and (3) in contrast to traditional perturbation measures that were designed for the measurement of sustained vowels produced with the intention of steady pitch and loudness, measures of CPP have been reported to provide valid measures of dysphonia in both sustained vowels and connected speech.[3,4,7,15,16] The CPP has also been reported to have strong sensitivity and specificity for the categorization of normal *vs.* dysphonic voice, and also to be a useful measure in documenting change in voice quality secondary to treatment.[11,17–19]

Acoustic analyses have been recommended as a standard component of voice assessment in clinical populations,[20–22] and the ability to compute CPP measurements has become readily available in recent years *via* software packages that allow for easily computed cepstral calculations from recorded voice signals. Two popular software packages for obtaining CPP measurements are *Analysis of Dysphonia in Speech and Voice* (*ADSV*, PENTAX Medical, Montvale, NJ) and *Praat* (Paul Boersma & David Weenink, Institute of Phonetic Sciences, University of Amsterdam, The Netherlands; www.praat.org). Both programs provide a smoothed version of the CPP (referred to as CPPS (Cepstral Peak Prominence Smoothed) in *Praat* and simply as the CPP in *ADSV*), and both programs have been included in a substantial body of clinical voice research.[9,12,13,15,16,23] However, there are a number of known differences between the algorithms used in these respective programs that may result in differences in reported CPP values:

1. The algorithms used in the *ADSV* incorporate a simple form of voicing activity detection (VAD) (similar to that used in the homomorphic pitch tracker) in which CPP values <0 dB (ie, dominant cepstral rahmonics that had an amplitude lower than the expected value as determined *via* subsequent linear regression analyses) are removed from further statistical analyses. This procedure was incorporated into *ADSV* algorithms as a means of reducing the potential effect of low-amplitude, highly aperiodic signals often associated with breath sounds and portions of unvoiced consonants.[2,3,24–26] In contrast, the CPPS algorithm in *Praat* does not incorporate any form of VAD.
2. Both programs compute the logarithm of the amplitude spectrum to construct a power cepstrum. The *ADSV* program then obtains the cepstrum *via* a forward Fourier transform of the log spectrum *vs.* taking the inverse Fourier transform of the log spectrum to obtain the cepstrum in *Praat*.
3. Simple least squares linear regression is used to compute a line of best fit for the cepstrum in *ADSV vs.* linear regression using the Theil robust fitting method (a nonparametric form of linear regression that is less affected by outlying data points).
4. In *ADSV*, the regression line is calculated using a lower quefrency value of .0001 s (equivalent to 10 kHz) *vs.* .001 s (equivalent to 1 kHz), tending to result in a greater negative slope to the regression line in *ADSV vs. Praat*.
5. Various minor differences including the choice of windowing function (ie, a Gaussian window in the *Praat* algorithm *vs.* a Hamming window in *ADSV*), use of parabolic interpolation to find the peak value in *Praat vs.* no interpolation in *ADSV*, and a sampling frequency independent pre-emphasis in *Praat*.[15]

Although both programs have been regularly used in clinical voice practice and voice science research, to date there has been no existing comparison of the measurement reliability between these two software applications regarding CPP calculations. The purpose of this study was to investigate the reliability of CPP measured from two software applications (*Praat* and *ADSV*) by examining the relationship between CPP values obtained *via* these programs in vowels and connected speech from speakers with a wide range of dysphonic severity (eg, nondysphonic through severely dysphonic) and of different languages. We conducted two experiments focused on the following research questions: (1) What is the strength of relationship between measures of CPP measured using *ADSV* and using *Praat*?; and (2) What is the magnitude of the average expected error in predicting the CPP values of one program (eg, *Praat*) from the CPP values obtained using an alternative program (eg, *ADSV*)? We studied these questions with two experiments, which differed based on the population sample. Experiment #1 included dysphonic and non-dysphonic speakers of Flemish, whereas Experiment #2 included dysphonic and non-dysphonic speakers of English. To answer the first research question, we applied correlational analyses to the measurements as a method of computing parallel-forms (ie, equivalent) reliability.[27–29] To answer the second research question, we applied regression analyses to the vowel CPP and connected speech CPP measurements obtained from *ADSV vs. Praat* to obtain predictive equations by which actual observed CPP values obtained with one program (eg, *ADSV*) could be transformed to predicted CPP values for an alternative program (eg, *Praat*) and *vice versa*. Measures of the standard error of the estimate (aka, standard error of the regression) were calculated to determine the average error that our predictive regression models would produce.

METHODS

Two experiments were conducted, both using CPP (in decibels) as the dependent variable. The first experiment used recordings of Flemish speakers obtained with *Praat* and subsequently analyzed with *Praat* and *ADSV*, whereas the second used recordings of English speakers obtained with *ADSV* and subsequently analyzed with *ADSV* and *Praat*. The analytical methodology was the same for both data sets, but owing to the fact that (1) stimulus sound structure influences acoustic measurements, (2) the *Praat* and *ADSV* recordings were obtained using different sampling rates, and (3) the recordings were obtained using equipment in separate clinics, correlations for the Flemish and English speakers were compared separately.