Acoustic and Perceptual Classification of Within-sample Normal, Intermittently Dysphonic, and Consistently Dysphonic Voice Types

*,[1]Christopher S. Gaskill, †Jordan A. Awan, ‡Christopher R. Watts, and §,[1]Shaheen N. Awan, *Montevallo, Alabama, †Waltham, Massachusetts, ‡Fort Worth, Texas, and §Bloomsburg, Pennsylvania

Summary: Objectives. Intermittent dysphonia within an utterance is common, but presents difficulty for both perceptual and objective voice evaluation. This study examined the ability of measures from the within-sample cepstral peak prominence (CPP) distribution to differentiate normal voices from intermittently and consistently dysphonic voices.

Study Design. Exploratory design.

Methods. Sixty samples of the sentence "We were away a year ago" were classified as normal, intermittently dysphonic, or consistently dysphonic by four judges. Measures of CPP within each sample were obtained, and further analysis with examined CPP distribution variability and patterns of CPP outliers.

Results. Whereas the mean CPP was the strongest single discriminator among the three voice types, the normal and intermittent dysphonia groups were not significantly different on CPP distribution skewness and measures of CPP distribution outliers. Both the normal and intermittently dysphonic voices differed significantly from the consistently dysphonic samples on these variables. A combination of measures of the CPP distribution was effective for a linear prediction of percent dysphonia duration for the speech samples (r = 0.825; rho = 0.81). The CPP standard deviation significantly improved the use of the mean CPP in discriminant function analyses and also the classification of the intermittently dysphonic voices.

Conclusions. Auditory-perceptual judgment of dysphonic segments and the typically robust acoustic measurement of mean CPP are both ineffective for classifying intermittently dysphonic voices. However, dysphonia duration may be effectively predicted via measures of the CPP distribution, and acoustic classification of dysphonic voice types via cepstral methods may be improved with an analysis of the CPP distribution across an utterance.

Key Words: cepstrum–cepstral analysis–cepstral peak prominence–Statistical Process Control–intermittent dysphonia.

INTRODUCTION

Commonly used measures of dysphonia, such as jitter and shimmer, provide indices of deviations from periodicity (ie, perturbation) by focusing on cycle-to-cycle variations in period (jitter) or amplitude (shimmer). Unfortunately, these measures are hampered by the facts that they become invalid with more than mild cases of vocal quality-based dysphonia in which cycles of vibration become indeterminate. Titze[1] has indicated that perturbation measures, such as jitter and shimmer, may only be useful in the description of type I signals (those that were nearly periodic in nature), and urged that types II and III category signals (those that contained modulations and/or subharmonics or approached random, aperiodic vibration) be analyzed and described via methods other than time-based acoustic measures, such as spectrographic analysis. Fortunately, cepstral measures of the voice have emerged in recent years as valuable measures of dysphonia that provide benefits beyond those of traditional perturbation measures, such as jitter and shimmer. The cepstrum (produced via a spectral analysis of an initial spectrum) tends to show a prominent peak corresponding to the dominant rahmonic (an anagram of *harmonic*) in the voiced sound wave. This prominent cepstral peak is generally associated with the fundamental frequency of the voice, with the relative amplitude of this peak used as a measure of periodicity (typical, relatively periodic voices tend to have high cepstral peaks, whereas voices characterized by quality disturbances [eg, breathiness] tend to have reduced or absent cepstral peaks). Because cepstral measures are obtained from frames of data and not cycles of vibration, they may be obtained for even moderate-to-severely dysphonic voices. In addition, cepstral measures have been frequently reported to have validity as measures of dysphonia severity in both continuous speech and sustained vowel contexts. The relative amplitude of the cepstral peak has been referred to by Hillenbrand and colleagues as the *cepstral peak prominence (CPP)*,[2,3] and has been used in numerous studies to characterize various aspects of both normal and disordered voice function.[4–14]

The majority of studies that have reported on cepstral measures of voice have focused on the *mean CPP* computed across numerous, often smoothed cepstra computed from multiple frames of data (eg, 1024 pt. frames). Among various acoustic measures of the voice, the mean CPP has been reported to be possibly the strongest and most robust measure of dysphonia severity currently available.[15] Although the mean is one of the most commonly used measures of central tendency, measures of the variability

Accepted for publication April 28, 2016.

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

From the *Department of Communication Science and Disorders, University of Montevallo, Montevallo, Alabama; †Brandeis University, Waltham, Massachusetts; ‡Texas Christian University, Fort Worth, Texas; and the §Bloomsburg University of Pennsylvania, Bloomsburg, Pennsylvania.

[1]These authors contributed equally to this manuscript.

Address correspondence and reprint requests to Christopher S. Gaskill, Department of Communication Science and Disorders, University of Montevallo, Station 6720, Montevallo, AL 35115. E-mail: cgaskill@montevallo.edu

Journal of Voice, Vol. 31, No. 2, pp. 218–228
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.04.016

within the CPP distribution have also been reported to have value in describing dysphonic voice. A number of studies have reported that the standard deviation of the CPP distribution over time can also be a valuable measure of dysphonia whether in combination with the mean CPP or as an individual measure.[7,16–18] In addition, studies that have examined variability of spectral characteristics of voice (eg, via long-term average spectra) have also reported that distribution measures such as skewness and kurtosis may also show promise in the description of dysphonia.[19–22]

The focus of this study was to further investigate the ability of the mean CPP and additional measures of the within-sample CPP distribution (eg, measures of CPP variability) to discriminate between typical and disordered voices. In particular, this study focused on the ability to differentiate within-sample typical, non-dysphonic voices from intermittently and consistently dysphonic voices. Intermittently dysphonic voice, in which voice change occurs within an utterance, has been reported for a diversity of disorders, including vocal fold irritation secondary to laryngopharyngeal reflux,[23] mutational falsetto,[24] adductor and abductor spasmodic dysphonia,[25–28] and vocal cord dysfunction.[29] Any condition in which characteristics such as the stiffness or tension of the vocal fold(s) are mildly or temporarily disrupted may result in difficulty in sustaining periodic vibration[30] and result in intermittent, within-sample vocal quality change. Although within-sample intermittently dysphonic voices are frequently observed, they may present difficulty for both perceptual and objective methods of voice evaluation. In particular, voices in which transient periods of dysphonia occur may be perceptually salient to some listeners but not to others will likely result in poor interjudge reliability for perceptual ratings. In addition, these intermittent dysphonia cases may result in a lack of reliability between perceptual and acoustic measures of voice. As an example, perceptually salient dysphonic transients may have a relatively lesser effect on the overall mean CPP because the mean will be weighted toward the greater number of analysis frames that are representative of typical, non-dysphonic voice production. It may be in cases such as these that other measures from the CPP distribution that reflect variability may provide value. With this in mind, the purpose of the present study was to investigate the acoustic characteristics of perceptually categorized normal (N), intermittently dysphonic (I), and consistently dysphonic (C) voices by analyzing measures of the CPP distribution (including measures of both CPP central tendency and variability). Additionally, the ability to discriminate normal versus intermittently dysphonic versus consistently dysphonic voice types, and the strength of correlation between measures of the CPP distribution and within-sample dysphonia duration, were also investigated. It was hypothesized that measures of CPP variability would provide benefit beyond measures of the mean CPP in typical versus dysphonic group description and classification.

METHODOLOGY

Participants

Sixty recorded samples of "We were away a year ago" from the Consensus Auditory Perceptual Evaluation of Voice (CAPE-V)[31] sentence list were used as the stimuli for both perceptual and acoustic analysis. This particular utterance from the CAPE-V was chosen for analysis as it has been shown to be both particularly sensitive to the physiologic and acoustic changes associated with dysphonia, as well as having the most robust CPP of all the CAPE-V sentences for normal voices.[7,32,33] A 60-sample corpus was drawn from the personal collections of the authors of this study to include reasonably equal contributions of within-sample normal, intermittently dysphonic, and consistently dysphonic voice samples. The corpus included samples from 20 speakers with no history of or current dysphonia (6 male, 14 female) recruited for participation by the first author, and 40 speakers (12 male, 28 female) with varying degrees of dysphonia who had previously sought treatment at two different voice clinics (evaluated by the third and fourth authors).

All voice recordings were conducted using a high-quality microphone held at a consistent 5-cm distance mouth-to-microphone distance. All recordings were made in a room with background noise below 45 dB SPL (sound pressure level), with the microphone signal input directly into the *Computerized Speech Lab* hardware module (CSL; PENTAX Medical, Montvale, NJ). The CSL input gain was held constant, and all recordings were sampled at a rate of 44.1 kHz for analysis.

Acoustic analysis of CPP variation

The CSL software module *Analysis of Dysphonia in Speech and Voice (ADSV)* was used to provide an initial analysis of the recorded samples. A complete description of the analysis algorithms used in the *ADSV* program is available in a number of previously published works by Awan and colleagues.[32,34,35] Instead of obtaining average values of the cepstral/spectral data as typically provided by the *ADSV* analysis protocol,[32] for each recorded sample the smoothed CPP raw data for each set of analysis frames across the utterance were saved and exported to a second independent custom software program. The custom software (written in Java by author J.A.A.) allowed for detailed analysis of the CPP distribution characteristics beyond that available in the current *ADSV* program.

Using the custom software, the CPP raw data were displayed graphically as a time series, and an *average moving range* (XmR) was computed by calculating the mean of the absolute CPP to CPP differences. Next, the upper and lower control limits (UCL and LCL, respectively) around the mean CPP were computed as UCL = average + 2.66 × average moving range, and LCL = average − 2.66 × average moving range.[36] This method of data analysis was originally developed for quality control in manufacturing applications and is referred to as Statistical Process Control (SPC).[37,38] This data analysis method allows for the detection of individual or groups of data points that represent a statistically significant deviation from the mean of the data,[36,37] and has shown promise as an analysis tool for behavioral research, particularly for displaying time series data in single-subject designs.[37] Because the control limits are calculated from an average moving range instead of a global standard deviation, the method is considered to be statistically robust for detecting meaningful changes across data points in a time series.[38] This method was chosen for the present study to provide a means