An Examination of Pre- and Posttreatment Acoustic Versus Auditory Perceptual Analyses of Voice Across Four Common Voice Disorders

*,[1]Amanda I. Gillespie, *Jackie Gartner-Schmidt, *Ali Lewandowski, and †,[1]Shaheen N. Awan, *Pittsburgh and †Bloomsburg, Pennsylvania

Summary: Objective. The objective, instrumental acoustic measure of Cepstral Spectral Index of Dysphonia (CSID) correlates with audio-perceptual measures, is sensitive in detecting voice abnormalities, and tracks change following treatment. The goals of the current study were to (1) test the agreement between CSID versus auditory perceptual measures of pre- versus posttreatment voice change, and (2) investigate whether change in scores is based on voice disorder or phonemic structure of sentence stimuli.

Methods. Forty patients with benign voice disorders produced sentences and a sustained /a/ vowel from the Consensus Auditory Perceptual Evaluation of Voice protocol before and after treatment. CSID was calculated, and overall audio-perceptual voice severity was judged by 7 blinded, trained raters using a 100-mm visual analog scale. Differences between CSID and audio-perceptual measures of voice change across voice disorder and stimuli, and correlation between change in CSID and perceptual rating scores were assessed pre- and postintervention.

Results. Across all subjects, there were significant correlations between CSID and perceptual ratings change scores ($P < 0.001$), and no significant differences in pre- and posttreatment change. Disorder-specific analyses indicated that all tested sentence/vowel contexts are effective measures for pre- versus posttreatment change in atrophy and paralysis cases. Acoustic versus perceptual measures of voice change were significantly correlated in lesion cases for the sentence "How hard did he hit him" and with all sentences combined. There were no significant findings observed for muscle tension dysphonia.

Conclusion. CSID provides an accurate objective correlate to auditory-perceptual posttreatment change in overall voice severity ratings. Implications for outcomes testing and disorder-specific findings are discussed.

Key Words: Voice–Acoustics–Cepstral–Auditory perceptual–Analysis.

INTRODUCTION

Voice clinicians and researchers have at their disposal a plethora of voice assessment tools. Acoustic, aerodynamic, auditory-perceptual, and patient-perceptual are the most common outcome measures employed in the clinic and voice research laboratory.[1] However, many of these outcome measures have demonstrated poor ability to do more than differentiate disordered voices from normal ones.[2–4] In addition, these different outcome measures have shown an inability to discriminate degrees of severity of a disorder, to discriminate one disorder from another, or to similarly demonstrate reliable change following treatment.[5,6]

To date, the gold standard voice outcome has been considered to be auditory-perceptual analysis of voice.[1,7–9] This method involves training listeners to rate the overall severity of a voice, as well as individual component parts. The Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) was developed to standardize auditory-perceptual voice evaluations.[10,11] The protocol consists of six sentences, sustained vowels, and an extemporaneous speech sample. The sentences were designed specifically to elicit certain vocal behaviors. The sentences are, "How hard did he hit him" (voice to voiceless transitions, soft glottal attacks), "We were away a year ago" (all voiced sentence, elicits the presence of spasms or voice stops or the inability to maintain voicing), "We eat eggs every Easter" (hard glottal attacks), and "Peter will keep at the peak" (transition between voiceless stops and voiced vowels). Individual voice disorders are characterized by different perceptual characteristics; therefore, theoretically, these sentence stimuli should elicit these disordered productions and assist in diagnosis and outcomes evaluations.[11] However, the clinical use of this type of assessment may be adversely influenced by a number of limitations associated with the perceptual assessment of voice, including the potential influence of clinician inexperience or bias on the reliability and validity of perceptual ratings; the potential for shifting definitions of dysphonia severity; the need for consistent training to achieve focused and reliable perceptual ratings; and (particularly pertinent to research contexts) the need for multiple trained judges from which mean or median ratings may be acquired.[7,12]

Because of the aforementioned limitations with auditory-perceptual judgments, a clinical need exists to have an easily obtained objective measure that may be able to provide an automatic categorization of the normal versus disordered state as well as dysphonia severity.[7] Because they are readily available at relatively low cost compared with other methods of voice analysis, applicable to treatment as well as diagnosis, and are

Accepted for publication April 19, 2017.

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 & Voice (ADSV)* program from Dr. Awan.

[1]These authors contributed equally to this work.

From the *University of Pittsburgh Voice Center, Department of Otolaryngology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; and the †Bloomsburg University of Pennsylvania, Bloomsburg, Pennsylvania.

Address correspondence and reprint requests to Amanda I. Gillespie, UPMC Mercy Hospital, 1400 Locust St. Building B, Suite 11500H, Pittsburgh, PA 15219. E-mail: gillespieai@upmc.edu

Journal of Voice, Vol. ■■, No. ■■, pp. ■■-■■
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2017.04.018

supported by a substantial body of literature,[13] acoustic analysis methods provide a viable option for dysphonia categorization. Over the last decade, frequency-based acoustic measures, specifically cepstral analyses, have moved to prominence in the acoustic evaluation of voice.[7,11,14] Cepstral analyses overcome many of the challenges found with traditional time-based acoustic measures (such as frequency and amplitude perturbation). Cepstral measures can be collected in connected speech, whereas traditional perturbation measures are most applicable to isolated vowels produced with steady pitch and loudness.[11] Further, cepstral analyses do not require definitive cyclic behavior in the voice signal and therefore can be used to evaluate any voice type regardless of severity.[5,11] Cepstral measures have shown good discrimination between normal and disordered voices and have been demonstrated to be effective correlates of dysphonia severity.[11,15–18]

The Cepstral Spectral Index of Dysphonia (CSID) is a multiparameter algorithm for measuring dysphonia severity.[7] The algorithm mathematically combines measures of cepstral peak prominence (CPP) and its standard deviation, and low/high spectral ratio (4 kHz cutoff) and its standard deviation. The CSID has been shown to discriminate normal from disordered voices[7] and correlate with perceptual ratings of overall voice severity on a visual analog scale as per the CAPE-V across voice severity ratings.[11,19] The CSID has been reported to be sensitive to change following treatment in patients with adductor spasmodic dysphonia, unilateral vocal fold paralysis (UVFP), bilateral vocal fold lesions (lesions), and vocal fold atrophy (atrophy),[19] and unlike traditional acoustic measures, has demonstrated low to moderate correlation with a patient's perception of voice handicap as measured by the Voice Handicap Index and Voice Handicap Index-10.[14,20,21]

Despite the availability of acoustic assessment options with improved sensitivity to measure dysphonia and track treatment change, the battery of subjective and objective tests used to measure voice outcomes remain the same across patient groups. In other words, voice clinical care and research often employ a "kitchen-sink" approach to outcomes testing. Regardless of the voice disorder being studied, the same acoustic, aerodynamic, and auditory-perceptual voice rating analyses are used. Although a hypothesis-driven outcomes testing is preferred in research, little data exist on what measures are most appropriate to track change following treatment in patients with specific voice disorders.

Recent work has attempted to create disorder-specific protocols for acoustic and aerodynamic voice outcomes testing.[5,22] Findings demonstrated improvements in CSID measured from the all-voiced sentence "We were away a year ago" following treatment in patients with muscle tension dysphonia (MTD) and UVFP. It is unknown if these improvements correlated with improved auditory-perceptual ratings of voice severity in these patients. Despite the design of the CAPE-V sentences to elicit certain vocal behaviors, it is unclear if the specific sentences are more sensitive to change following treatment for different voice disorders, and if those sentence- and disorder-specific changes correspond in both agreement and relationship with listener-perceived ratings of change in voice severity. The purpose of the current study was to determine if correlation and agreement between CSID and auditory-perceptual ratings of posttreatment change in dysphonia severity exist, if the degree of correlation is stimulus-specific, and finally, if findings differ depending on voice disorder.

METHODS

Participants

The participants with voice disorders in this study were the same as those in a companion paper.[5] Data from 40 patient participants, diagnosed with vocal fold atrophy (n = 10), vocal fold lesions (n = 10), MTD (n = 10), and UVFP (n = 10) were included. All diagnoses were made by a team consisting of a laryngologist and voice-specialized speech language pathologist). Patient records were assessed retrospectively from a clinical research database. Patients included were over 18 years old with only one diagnosis (ie, atrophy alone, not occurring with another laryngeal disorder or diagnosis) and had complete records, including voice laboratory voice recordings before and after a specific intervention at a preset time point. The follow-up time points for patients with atrophy, MTD, and UVFP were baseline and 6 months after treatment, whereas for lesions it was baseline and 12 months after treatment. Intervention was injection augmentation with calcium hydroxyapatite or lipoinjection for patients with atrophy, phonomicrosurgery for patients with lesions, voice therapy for patients with MTD, and thyroplasty medialization with Gore-Tex (Gore, Newark, DE) for patients with UVFP.

Procedures

Recording procedures and acoustic analysis

All participants provided pre- and posttreatment readings of the sentences from the CAPE-V and a sustained vowel /a/ production at his/her most comfortable pitch and loudness level. Recordings were captured with a Shure Beta-54 WBH54 head-mounted microphone (SHURE, Chicago, IL) and analyzed using the *Analysis of Dysphonia in Speech and Voice* (ADSV) program (ADSV, KayPENTAX, Montvale, NJ). Each recording was then saved in a designated folder on the computer within the voice research laboratory. Only four of the six sentences were analyzed ("How hard did he hit him"; "We were away a year ago"; "We eat eggs every Easter"; "Peter will keep at the peak") because the initial CSID validation was completed using these sentence stimuli.[11] Each individual sentence was then isolated from the whole sample and saved as a separate file. This was carried out by placing the cursor before and after each sentence, extracting the recording within the selected range, and saving under a new name. Each sample was precisely edited so that additional cursor placement was not required for acoustic analysis of each stimulus. For each sample, a multifactorial estimate of dysphonia severity known as the CSID was automatically calculated in the *ADSV* program. The CSID is a multiparametric acoustic measure that incorporates different cepstral and spectral measures into a single acoustic estimate of dysphonia severity that approximates perceptual ratings reported via the 0–100 mm CAPE-V scale. For CAPE-V sentences, the CSID is calculated