Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a Screening Tool for Voice Disorders: Development of Clinical Cutoff Scores

*Shaheen N. Awan, †Nelson Roy, ‡Dong Zhang, and §Seth M. Cohen, *‡*Bloomsburg, Pennsylvania,* †*Salt Lake City, Utah, and* §*Durham, North Carolina*

Summary: Objectives. The purposes of this study were to (1) evaluate the performance of the Cepstral Spectral Index of Dysphonia (CSID—a multivariate estimate of dysphonia severity) as a potential screening tool for voice disorder identification and (2) identify potential clinical cutoff scores to classify voice-disordered cases versus controls.

Methodology. Subjects were 332 men and women (116 men, 216 women) comprised of subjects who presented to a physician with a voice-related complaint and a group of non–voice-related control subjects. Voice-disordered cases versus controls were initially defined via three reference standards: (1) auditory-perceptual judgment (dysphonia +/−); (2) Voice Handicap Index (VHI) score (VHI +/−); and (3) laryngoscopic description (laryngoscopic +/−). Speech samples were analyzed using the Analysis of Dysphonia in Speech and Voice program. Cepstral and spectral measures were combined into a CSID multivariate formula which estimated dysphonia severity for Rainbow Passage samples (ie, the $CSID_R$). The ability of the $CSID_R$ to accurately classify cases versus controls in relation to each reference standard was evaluated via a combination of logistic regression and receiver operating characteristic (ROC) analyses.

Results. The ability of the $CSID_R$ to discriminate between cases and controls was represented by the "area under the ROC curve" (AUC). ROC classification of dysphonia-positive cases versus controls resulted in a strong AUC = 0.85. A $CSID_R$ cutoff of ≈24 achieved the best balance between sensitivity and specificity, whereas a more liberal cutoff score of ≈19 resulted in higher sensitivity while maintaining respectable specificity which may be preferred for screening purposes. Weaker but adequate AUCs = 0.75 and 0.73 were observed for the classification of VHI-positive and laryngoscopic-positive cases versus controls, respectively. Logistic regression analyses indicated that subject age may be a significant covariate in the discrimination of dysphonia-positive and VHI-positive cases versus controls.

Conclusions. The $CSID_R$ can provide a strong level of accuracy for the classification of voice-disordered cases versus controls, particularly when auditory-perceptual judgment is used as the reference standard. Although users often focus on a cutoff score that achieves a balance between sensitivity and specificity, more liberal cutoffs for screening purposes versus conservative cutoffs when cost or risk of further evaluation is deemed to be high should also be considered.

Key Words: Cepstrum–Cepstral analysis–Voice disorders–Sensitivity–Specificity–Voice screening.

INTRODUCTION

Accurate diagnosis of a medical condition is often the first step toward its treatment. The diagnostic process is fundamentally a process of categorization. As voice clinicians, we seek to classify individuals accurately with respect to (1) whether they should be evaluated further (ie, for the purpose of "screening"), (2) whether they have a voice disorder (for the purpose of "diagnosis"), or (3) whether they have a particular type of voice disorder (for the purpose of "differential diagnosis"). Screening for a voice disorder is a specific type of diagnostic test that aims to detect the presence of a possible voice disorder. What sets screening tests apart from other diagnostic tests is that (1) they are typically applied to "healthy or at risk" individuals on a large scale and therefore must be noninvasive and inexpensive and (2) a positive screening test is usually followed, not directly with treatment, but with additional focused diagnostic procedures that will help to confirm or refute the initial screening result and, in the case of confirmation, provide further detail regarding the presenting disorder.

Screening healthy or at risk populations for early signs of a voice abnormality is attractive because some forms of a voice disorder may be most successfully treated if detected early. For instance, teaching school is a high-risk profession for developing voice disorders. Given the intense voice demands of the profession, it is not surprising that voice disorders are a relatively common occupational hazard, with 11% of teachers reporting a current voice disorder and 58% reporting a history of a voice disorder during their lifetime. Vocal dysfunction interferes with job satisfaction, performance, and attendance, causing 18% of teachers to report missing work on a yearly basis. Teachers are more likely than nonteachers to have consulted a medical professional regarding a voice disorder and to consider future career change because of voice-related dysfunction.[1,2] Because of lost workdays and treatment expenses, the societal costs have been estimated at 2.5 billion dollars annually in the United States alone.[3] Thus, early detection and treatment may prevent the development of more intractable voice disorders which can threaten the career of a teacher.

Accepted for publication April 10, 2015.

From the *Department of Audiology and Speech Pathology, Bloomsburg University of Pennsylvania, Bloomsburg, Pennsylvania; †Department of Communication Sciences and Disorders, The University of Utah, Salt Lake City, Utah; ‡Department of Mathematics and Computer Science, Bloomsburg University of Pennsylvania, Bloomsburg, Pennsylvania; and the §Division of Otolaryngology–Head and Neck Surgery, Duke Voice Care Center, Duke University Medical Center, Durham, North Carolina.

Address correspondence and reprint requests to Shaheen N. Awan, Department of Audiology and Speech Pathology, Bloomsburg University of Pennsylvania, Centennial Hall, 400 East Second Street, Bloomsburg, PA 17815. E-mail: sawan@bloomu.edu

Journal of Voice, Vol. ■, No. ■, pp. 1-15

0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.04.009

The development and application of any voice disorder screening test must be initially based on a definition of how an individual will be categorized as "voice disordered" or "vocally normal." In medical circles, this is often referred to as distinguishing a "case" from a "noncase" or "control." For some medical diseases, a "gold standard" (definitive test) exists for diagnosing a case representative of the disease or disorder of interest. This "gold" or "reference" standard is then used to evaluate the diagnostic accuracy of any new medical test, with a reference standard defined typically as the "best available method for establishing presence or absence of the target condition (p. 8)."[4] Unfortunately, in the field of voice disorders, a perfect reference standard or definitive test seemingly does not exist and, as in many other fields of study, a variety of reference standards and potential ways to define a case are available.[5] For instance, the categorization of "normal" versus "voice disordered" may be based on a variety of reference standards (and sources of information) including but not limited to the following:

(1) The condition and characteristics of the underlying laryngeal structures that, in conjunction with respiratory flows and pressures, are responsible for phonation. This form of categorization is typically made via laryngeal imaging (eg, laryngoscopy, laryngeal videostroboscopy), with "cases" defined by visual evidence of abnormal laryngeal structure and/or function.[6] The American Academy of Otolaryngology–Head and Neck Surgery recommends that, at a minimum, when evaluating a patient, the basic protocol should include a rigorous clinical history, physical examination, and "visualization of the larynx via laryngoscopy."[7]

(2) An assessment of the psychosocial handicapping effects of the patient's voice condition on the basis of the patient's judgment about the relative impact of their voice condition on daily activities.[8] A number of patient-based instruments have been developed, including the Voice Activity and Participation Profile[9]; the Voice Handicap Index (VHI)[8]; the VHI-10[10]; the Voice-Related Quality of Life measure[11]; and the Voice Symptom Scale.[12] Of these instruments, the VHI[8] represents one of the most studied and popular. The VHI is a psychometrically validated tool developed for the measurement of the psychosocial handicapping effects of voice disorders. To define a "case" using such disablement measures would ultimately demand that some threshold level/score of voice-related handicap be exceeded, and multiple examples are available in the literature that confirm the use of these measures to define voice-disordered cases.[13–18]

(3) The auditory-perceptual characteristics of the voice are often regarded as the "gold standard" to define a voice disorder, although the actual definition of normal/typical voice can be elusive and variable. As examples, Eskenazi et al[19] defined normal voice "as a voice with no apparent pathology … and no unusual voice characteristics or habits (p. 33)." Fex[20] stated that normal voice quality is a concept based on subjective opinion that may vary with different cultures and represents a continuum in which a vast number of people may be judged as having normal but nevertheless individually differentiated voices. Awan[21] states that as long as a particular voice does not deviate substantially from the listener's expectations for age, gender, and body type in terms of parameters such as pitch, loudness, quality, and duration, it will be considered within the normal range. In contrast, when a voice is perceived as deviating from the normal range, it may be characterized as being dysphonic. Therefore, defining a "case" using auditory-perceptual criteria would require some minimum level of auditorily perceived dysphonia to be present and detected.

All the aforementioned methods are imperfect reference standards as they are primarily perception based, with an examiner or judge responsible for a perceptual rating or description of observed vocal characteristics (in the case of laryngeal imaging and auditory-perceptual descriptions) or a patient responsible for self-perception (in the case of patient-based handicap scales). Because of variability in any type of perceptual judgment due to factors such as experience, training, bias, shifting definitions, and so forth, it would be of clinical value to have an easily obtained objective correlate of these methods that may be able to provide an automatic categorization of the normal versus disordered state of the presenting patient. Acoustic analysis methods provide a viable option for this type of screening categorization as they are readily available at relatively low cost compared with other methods of voice analysis; applicable to treatment as well as diagnosis; and are supported by a substantial body of literature.[21] In addition, acoustic evaluation methods have the benefit that (1) they are noninvasive, (2) incorporate algorithms that will always analyze voice signals in a similar manner every time (ie, test results are reproducible), (3) they provide results in numerical format, thereby allowing for built-in scaling and ease of communication,[22] and (4) can be applied to a large number of voices in a relatively short period of time.

In recent years, spectral- and cepstral-based acoustic measures have been demonstrated to be strong predictors of dysphonia type and severity in both sustained vowel and continuous speech samples.[23–30] Key measures from spectral- and cepstral-based analyses have included estimates of the relative amplitude of the cepstral peak referred to as the cepstral peak prominence (CPP); ratios of low versus high-frequency spectral energy; and the respective standard deviations for these measures.[22,29] The CPP has been consistently reported as a particularly robust measure of both presence versus absence of dysphonia and dysphonia severity.[22,25,31] In addition, the CPP has been included in an automated multiple regression-based mathematical estimate of dysphonia severity referred to as the Cepstral Spectral Index of Dysphonia (CSID) which uses several of the aforementioned cepstral- and spectral-based measures described by Awan et al.[24,26] Because these spectral and cepstral acoustic measures may be obtained with relative efficiency and low cost, they may offer promise as potential dysphonia screening tools.