Objective Dysphonia Measures in the Program Praat: Smoothed Cepstral Peak Prominence and Acoustic Voice Quality Index

*,†,‡Youri Maryn and §David Weenink, *Bruges, †Ghent, and ‡Antwerp, Belgium, and §Amsterdam, The Netherlands

Summary: Purpose. A version of the "smoothed cepstral peak prominence" (ie, CPPS) has recently been implemented in the program *Praat*. The present study therefore estimated the correspondence between the original CPPS from the program *SpeechTool* and *Praat*'s version of the CPPS. Because the CPPS is the main factor in the multivariate Acoustic Voice Quality Index (AVQI), this study also investigated the proportional relationship between the AVQI with the original and the second version of the CPPS.

Study Design. Comparative cohort study.

Methods. Clinical recordings of sustained vowel phonation and continuous speech from 289 subjects with various voice disorders were analyzed with the two versions of the CPPS and the AVQI. Pearson correlation coefficients and coefficients of determination were calculated between both CPPS-methods and between both AVQI-methods.

Results. Quasi-perfect correlations and coefficients of determination approaching hundred percent were found.

Conclusions. The findings of this study demonstrate that the outcomes of the two CPPS-methods and the two AVQI-methods are highly comparable, increasing the clinical feasibility of both methods as measures of dysphonia severity.

Key Words: Smoothed cepstral peak prominence–Acoustic voice quality index–SpeechTool–Praat–Feasibility–Accuracy.

INTRODUCTION

Acoustic methods have a long history in clinical voice assessment. Besides measures of fundamental frequency and sound intensity to objectify pitch and loudness, respectively, numerous acoustic markers have been proposed to objectify dysphonia type and severity.[1] Acoustic measurements are remarkably appealing because of their noninvasiveness, relative low cost, and ease of application.[2] They are able to yield a numerical output, and therefore they consistently permit tracking of treatment outcomes and communication of this information to voice clinicians, patients, third-party payers, physicians, and other stakeholders.[3,4] One specific and recently developed method to quantify the severity of overall dysphonia is the Acoustic Voice Quality Index (AVQI).

This index has the following attributes. First, the AVQI is developed to measure dysphonia in both continuous speech and sustained vowel recordings. Sustained vowels induce a comparatively steady action at subglottal, glottal, and supraglottal levels. Continuous speech, on the other hand, is characterized by temporal and spectral variations caused by voice onsets and offsets, interruptions, voiceless phonemes, phonetic context, prosodic modulations in fundamental frequency and intensity, speech tempo, and so on.[2,5,6] Because the vocal behavior differs considerably between these two voice/speech tasks, perception of type and severity of dysphonia can be hypothesized to vary across the kind of speech to be rated. Although two studies[7,8] found no statistically significant differences in the auditory perception of dysphonia severity between sustained vowels and continuous speech, four other studies[6,9–11] revealed that listeners rate dysphonia, and especially breathiness, more severely in sustained vowels than in continuous speech. These findings underline that for perceptual and instrumental methods in the clinical dysphonia assessment it is essential to record and analyze sustained vowels and continuous speech. Therefore, the AVQI requires recordings and measures overall dysphonia severity of both speech/voice contexts.

Second, the AVQI is a multivariate construct that combines multiple acoustic markers to yield a single number that correlates reasonably with overall dysphonia severity. This multiparametric approach was motivated by the multidimensional nature of voice quality and the fact that it is not related to a sole physical variable or a unique psychoacoustical determinant.[12] This is in contrast to pitch and loudness that many authors regard as synonymous to the distinctive features fundamental frequency and sound intensity, respectively. Kreiman and Gerratt[13] for example stated that vocal quality included all perceptual dimensions of the spectral envelope and its changes in time, and indicated a possibly major role of both time-domain and frequency-domain measures in the investigation of voice quality. Furthermore, bivariate correlational methods to estimate the proportional relationship between an auditory-perceptual rating and a single acoustic marker approach often lacked sufficient strength (see Maryn et al[14] for a meta-analysis on this topic). This prompted many researchers to apply multivariate statistics and to combine the merits of multiple measures for increased validity of objective/instrumental analysis of voice quality and/or more accurate discrimination among different perceptual categories/levels of dysphonia severity.[4,9,15–26] To construct a statistical model representing the best combination of acoustic predictors for the overall

Accepted for publication June 30, 2014.

From the *Department of Otorhinolaryngology and Head & Neck Surgery, Speech-Language Pathology, Sint-Jan General Hospital, Bruges, Belgium; †Department of Speech-Language Therapy and Audiology, Faculty of Education, Health and Social Work, University College Ghent, Ghent, Belgium; ‡Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; and the §Institute of Phonetic Sciences, Faculty of Humanities, University of Amsterdam, Amsterdam, The Netherlands.

Address correspondence and reprint requests to Youri Maryn, Sint-Jan General Hospital, Speech-Language Pathology and Audiology, Ruddershove 10, 8000 Bruges, Belgium. E-mail: yourimaryn@vvl.be

Journal of Voice, Vol. 29, No. 1, pp. 35-43
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2014.06.015

degree of disordered voice, only six from the thirteen acoustic measures that were initially applied in the study of Maryn et al[4] were retained after stepwise multiple linear regression analysis. The following multiple regression equation that was based on the unstandardized coefficients of this statistical model was called the AVQI = 2.571 (3.295 – 0.111 smoothed cepstral peak prominence − 0.073 harmonics-to-noise ratio − 0.213 shimmer local + 2.789 shimmer local dB − 0.032 slope of the long-term average spectrum + 0.077 tilt of the trendline through the long-term average spectrum). This equation thus includes acoustic markers from the time, frequency, and quefrency domains, and is a multidimensional representation of dysphonia severity.

Third, the main contributor to the AVQI model is the smoothed version of the cepstral peak prominence (CPPS). This measure represents the distance between the first rahmonic peak and the point with equal quefrency on the regression line through the smoothed cepstrum. The reasoning behind this acoustic marker is that the more periodic a voice signal, the more it displays a well-defined harmonic configuration in the spectrum (ie, the more harmonic the spectrum), and, consequently, the more the cepstral peak will be prominent. Since its introduction in the field of voice quality measurement by Hillenbrand et al[27] and Hillenbrand and Houde,[28] it has proven to be a reliable and valid measure of overall voice quality, and especially breathiness, across a multitude of studies.[4,29–37] Additionally, Maryn et al[14] performed a meta-analysis on correlation coefficients between auditory-perceptual ratings of overall dysphonia severity (ie, Grade or G) and 69 acoustic measures on sustained vowels/26 acoustic measures on continuous speech. Only four acoustic markers on sustained vowels (ie, Pearson *r* at autocorrelation peak, pitch amplitude, spectral flatness of residue signal, and smoothed cepstral peak prominence), and only three acoustic markers on continuous speech (ie, signal-to-noise ratio based on linear predictive coding and inverse filtering, cepstral peak prominence, and smoothed cepstral peak prominence) satisfied the meta-analytic criteria and were considered to be the most promising measures for the acoustic assessment of overall voice quality. The smoothed cepstral peak prominence (CPPS), however, was the only acoustic metric that yielded sufficient concurrent validity in both sustained vowel and continuous speech, and was therefore regarded as the superior acoustic measure of dysphonia severity.[14] In combination with the other five acoustic measures of the AVQI, the CPPS provided a solid basis to objectively/quantitatively approach dysphonia severity.

Fourth, because the original AVQI was developed with the help of the freely available and downloadable software packages "*Praat*" (Paul Boersma and David Weenink; Institute of Phonetic Sciences, University of Amsterdam, The Netherlands—http://www.praat.org/) and "*SpeechTool*" (James Hillenbrand; Western Michigan University, Kalamazoo, MI, USA—http://homepages.wmich.edu/~hillenbr/), it is within reach of most voice clinicians. Furthermore, with the *Praat* script from Maryn et al,[4] it was possible to automate and standardize most of the sound formatting and acoustic analyses. The program *Praat* has the additional advantage that it offers packages for multiple operating systems (ie, Windows, Macintosh, Linux, and so on.) and can be applied regardless the operating system that is used by the voice and speech clinician.

Fifth, the AVQI has been scrutinized on validity in several studies since its publication, for example the correlation coefficients (ie, r) were highly comparable across the different studies: $r = 0.780$ on two hundred fifty Dutch-speaking subjects,[4] $r = 0.796$ on thirty-nine Dutch-speaking subjects,[34] $r = 0.794$ on one hundred and seven English-speaking subjects,[38] $r = 0.790$ on sixty-one German-speaking subjects,[39] and an averaged $r = 0.829$ on fifty subjects speaking different languages.[40] Pooling these data of 507 subjects across five studies results in a homogeneous weighted correlation (ie, r_w, to credit the r of large sample studies more than the r of those with a smaller sample in the calculation of an average r across studies, the number of subjects is taken into account to weight the *r*-values) of $r_w = 0.790$. This finding indicates that the AVQI is a robust method, insulated from the differences across the languages in these studies, audio recording technology, reliability of the G-ratings of the experienced/professional listeners in these studies, or other factors that might affect its proportional relationship with G-ratings.

In conclusion, the AVQI is a multivariate, accessible, feasible, and reasonably valid method to clinically measure overall dysphonia severity in sustained vowel and continuous speech samples. Because the CPPS was only available in the program *SpeechTool* and not in the program *Praat*, the practical disadvantage of the AVQI was that it required a protocol combining both programs. This culminated in a relatively high number of steps in the AVQI procedure: (1) recording the continuous speech sample, (2) extracting and concatenating the voiced segments with a customized script in *Praat* (ie, part 1), (3) recording the medial 3 seconds of [a:] (ie, part 2), (4) chaining part 2 after part 1 in *Praat*, (5) determining CPPS in *SpeechTool*, (6) determining harmonics-to-noise ratio (HNR), shimmer local (SL), SLdB, Slope, and Tilt with a customized script in *Praat*, (7) completing the CPPS in a form in *Praat*, and (8) calculating the AVQI with a customized script in *Praat* (eg, Maryn et al[40]). The AVQI protocol was relatively laborious on that account. However, because the recent implementation of the cepstral peak prominence measures in *Praat* (since version 5.3.53)–with various cepstral applications via the "To PowerCepstrogram … " function and the possibility to obtain the cepstral metric via the "Get CPPS … " command—it became also possible to calculate the smoothed cepstral peak prominence in this program. This new development in *Praat* no longer necessitated a combination with *SpeechTool*, facilitating a single-program (ie, only in the program *Praat*) and even a single-script (ie, instead of working with multiple scripts as before, it became possible to merge all signal processing steps into only one *Praat*-macro) solution for the AVQI, minimizing the manual labor for the voice clinician, and therefore maximizing the AVQI's feasibility. A new script (ie, a second version or a beta version) for the AVQI was established to meet that purpose and will be examined in the present study.

The following two research aims were investigated in the present study. Because the construction of the power cepstrum