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 differs between the two programs, it was first estimated how well the recently implemented CPPS in *Praat* corresponded with the original CPPS in *SpeechTool* (ie, are the CPPS data in *Praat* interchangeable with the CPPS data in *SpeechTool*?). Consequently, given that the sound signal processing and the cepstral metric differ between the two programs, this study also investigated how well the second version of the AVQI corresponded with the original version of the AVQI (ie, are data of the second AVQI version transposable with the initial AVQI data?).

METHODS

The methods of the present study are highly comparable to the methods of two earlier studies[4,34] in terms of subjects. Only the acoustic cepstral methods and the program to elicit them were different from the previous studies.

Subjects

This research is based on one pool of voice recordings collected in two sets of subjects from the ENT caseload of the Sint-Jan General Hospital in Bruges, Belgium. First, the voice samples of 22 vocally normal and 228 voice-disordered participants from the study of Maryn et al[4] were used for further analysis. Second, we used the recordings of six vocally normal and 33 voice-disordered subjects of Maryn et al[34] This resulted in a total group of 289 (ie, 28 normophonic and 261 dysphonic) subjects, with 193 females between 10 and 86 years (mean = 36.8 years, Standard Deviation = 18.4 years) and 96 males between eight and 85 years (mean = 43.0 years, Standard Deviation = 21.5 years). Table 1 summarizes the variety of laryngoscopic/laryngostroboscopic diagnoses included in the sample. This study group is regarded as an adequate representation of a voice clinic population, reflecting different ages and genders and different types and degrees of voice quality disruption, and including nonorganic as well as organic laryngeal pathologies. Their voices were recorded at the beginning of a standard voice/larynx assessment.

TABLE 1.
List of Laryngeal Diagnoses of the 289 Subjects, With Their Absolute and Relative Occurrence

Laryngeal Diagnosis	Absolute Number	Relative Number
Normal	28	9.7
Abnormal	261	90.3
Functional dysphonia/MTD	88	30.4
Nodules	46	15.9
Polypoid mucosa/edema	31	10.7
Paralysis/paresis	24	8.3
Polyp	11	3.8
Cyst	9	3.1
Acute laryngitis	7	2.4
Hemorrhage	6	2.1
Leukoplakia	5	1.7
Mutational falsetto	5	1.7
Granuloma	4	1.4
Presbylarynx	4	1.4
Interarythenoidal pachyderm	4	1.4
Tumor	3	1.0
Sulcus glottidis	3	1.0
Postradiotherapy	3	1.0
Ventricular hypertrophy	2	0.7
Web	2	0.7
Postphonosurgery	1	0.3
Larynx trauma	1	0.3
Spasmodic dysphonia	1	0.3
Hyperkeratosis	1	0.3

MTD, muscle tension dysphonia.

Voice recordings

All 289 participants were asked to sustain the vowel [a:] for at least 5 seconds and to read aloud a Dutch phonetically balanced text[41] at comfortable pitch and loudness. Both voice samples were recorded using an AKG C420 head-mounted condenser microphone (AKG Acoustics, München, Germany), digitized at a sampling rate of 44.1 kHz and a resolution of 16 bits using the Computerized Speech Lab model 4500 (Kay Elemetrics Corporation, currently known as KayPENTAX, Lincoln Park, NJ, USA), and saved in WAV-format. A copy of every vowel sample was edited to include only the medial 3 seconds and a copy of the read text samples was trimmed to contain only the first two sentences in the program *Praat*.

Acoustic measures

Because certain acoustic measures used in this study are only valid for voiced segments of the continuous speech samples, the *Praat* script of Maryn et al[4] for automated detection, segmentation, and concatenation of these voiced segments was applied. The three-second midvowel segment was appended to this chain of voiced text segments, resulting into a single sound wave for acoustic analyses.

Seven acoustic markers were determined on these concatenated sound files. First, the HNR was administered in the program *Praat* as the base-10-logarithm of the ratio between the periodic energy and the noise energy multiplied by 10. Second, the SL was determined in the program *Praat* as the absolute mean difference between the amplitudes of successive periods divided by the average amplitude. Third, the "shimmer local dB" (ie, SLdB) was specified in the program *Praat* as the base-10-logarithm of the differences between the amplitudes of successive periods multiplied by 20. Fourth, the "general spectral slope" (ie, slope) was measured in the program *Praat* as the difference between the energy in the 0–1 kHz range and the energy in the 1–10 kHz range of the long-term average spectrum. Fifth, the "spectral trendline inclination" (ie, tilt) was computed in the program *Praat* as the difference between the energy in the 0–1 kHz range and the energy in the 1–10 kHz range of the trendline through the long-term average spectrum. Sixth, the "smoothed cepstral peak prominence" (ie, $CPPS_{SpeechTool}$) was determined in the program *SpeechTool* as the distance between the first rahmonic's peak and the point with equal quefrency on the regression line through the smoothed cepstrum. Seventh, the "smoothed cepstral peak prominence" (ie, $CPPS_{Praat}$) was also determined in the program *Praat* as the distance between the first rahmonic's peak and the point with equal quefrency on the regression line through the smoothed cepstrum. The $CPPS_{Praat}$ has just recently been implemented in the program *Praat*.

Smoothed cepstral peak prominence: *SpeechTool* versus *Praat*

The methods to yield the CPPS in *SpeechTool* and *Praat* are similar but not equal. Both programs construct a power cepstrum by first taking the logarithm of the amplitude spectrum. However, instead of taking the inverse Fourier transform of the log spectrum to obtain the cepstrum as in the *Praat* application (Weenink, in preparation), the *SpeechTool* application[27,28] takes the forward Fourier transform to obtain the cepstrum from the log spectrum. Furthermore, the background estimation in $CPPS_{Praat}$ is based on Theil robust fitting method, whereas the $CPPS_{SpeechTool}$ uses a plain linear regression for that purpose. The latter method is therefore more sensitive to outliers. Finally, some minor differences are the choice of windowing function (ie, a Gaussian window in the $CPPS_{Praat}$ algorithm and a Hamming window in the $CPPS_{SpeechTool}$ application), a parabolic interpolation to find the peak value in *Praat* and a sampling frequency independent preemphasis in *Praat*.[42]

Acoustic voice quality index: alfa version versus beta version

Based on the data of these acoustic markers, two versions of the AVQI were calculated. In the first version—ie, the original or alfa version, as developed by Maryn et al[4]—the AVQI was determined according to the following formula: $AVQI_{alfa} = 8.471 - 0.285 \times CPPS_{SpeechTool} - 0.188 \times HNR - 0.548 \times SL + 7.171 \times SLdB - 0.082 \times Slope - 0.198 \times Tilt$. So, the $AVQI_{alfa}$ is a combination of five acoustic markers in the program *Praat* plus one acoustic marker in the program *SpeechTool*. However, having to combine two computer programs to come to one single index of dysphonia severity decreased the userfriendliness/feasibility of the $AVQI_{alfa}$ and induced a relatively laborious procedure with at least eight steps.

To maximize its feasibility and to lower the number of processing steps and commands required from the clinician, a second or beta version of the AVQI was developed. Instead of inserting *SpeechTool*'s CPPS data in the $AVQI_{alfa}$ formula, *Praat*'s recently introduced CPPS data were inserted. Stepwise multiple linear regression analysis and linear rescaling of the outcomes of the equation to scores between 0 and 10 resulted in the following formula: $AVQI_{beta} = 9.072 - 0.245 \times CPPS_{Praat} - 0.161 \times HNR - 0.470 \times SL + 6.158 \times SLdB - 0.071 \times Slope - 0.170 \times Tilt$.

With this beta version, only the program *Praat* is involved and the number of commands is minimized to only two: (1) making the necessary recordings, and (2) activating the recently customized *Praat* script that automatically yields an $AVQI_{beta}$ score. The different parts and settings/parameters of this script can be consulted in Appendix, and examples of the output of this *Praat* script (ie, the relevant data and graphs) in a subject with a normal voice and a subject with a pathological voice are shown in Figures 1 and 2, respectively.

Statistical analyses

All statistical analyses in the present study were completed using *SPSS for Windows version 12.0* (SPSS Inc., Chicago, IL, USA). To answer the first question of the present study, the proportional relationship between $CPPS_{SpeechTool}$ and $CPPS_{Praat}$ was investigated with the parametric Pearson product-moment correlation coefficient (ie, r_P). The variance in $CPPS_{SpeechTool}$ that is accounted for by the variance in $CPPS_{Praat}$ was investigated with the coefficient of determination (ie, r_P^2). Similarly, to respond to the second research question, the degree of correspondence between $AVQI_{alfa}$ and $AVQI_{beta}$ was examined with r_P and their shared variance with r_P^2.

RESULTS

Correspondence between $CPPS_{SpeechTool}$ and $CPPS_{Praat}$

The descriptive statistics in Table 2 show that the $CPPS_{Praat}$ data are systematically higher than the $CPPS_{SpeechTool}$ data because of the dissimilarities in signal processing between the two measures. However, the scatterplot in Figure 3 illustrates the proportional association between the $CPPS_{SpeechTool}$ and $CPPS_{Praat}$ data and represents a quasi-perfect correlation of $r_P = 0.961$ ($P < 0.0001$, df = 287). With $r_P^2 = 0.924$, this means that 92.4% of the variance in $CPPS_{SpeechTool}$ is accounted for by the variance in $CPPS_{Praat}$.

Correspondence between $AVQI_{alfa}$ and $AVQI_{beta}$

The descriptive data in Table 2 highlight that the $AVQI_{beta}$-values very much resemble the $AVQI_{alfa}$-values. The proportional relationship between the $AVQI_{alfa}$ and the $AVQI_{beta}$ is clarified by the scatterplot in Figure 4. Based on the $r_P = 0.980$ ($P < 0.0001$, df = 287) between these data and the corresponding $r_P^2 = 0.960$, it can be stated that the $AVQI_{beta}$ almost perfectly (ie, for 96.0%) represents the $AVQI_{alfa}$.

DISCUSSION

The present study investigated the interconnection between versions of two modern methods to objectively-acoustically assess general dysphonia severity: CPPS and AVQI. The CPPS is determined in the smoothed power cepstrum by the difference in amplitude between the peak of first rahmonic and the corresponding value on the fit line exactly below the peak. This measure merits special attention as a tool in the clinical measurement of overall voice quality. It was introduced in the field of voice pathology by Hillenbrand et al[27] and Hillenbrand and Houde,[28] and although initially intended to estimate the degree of breathiness in sustained vowels and continuous speech, respectively, it also surfaced to be a valid and reliable measure of overall dysphonia severity.[24] The AVQI[4] is a multivariate, accessible/feasible, and reasonably valid method to clinically measure overall dysphonia severity in chained recordings of both a sustained vowel and continuous speech.

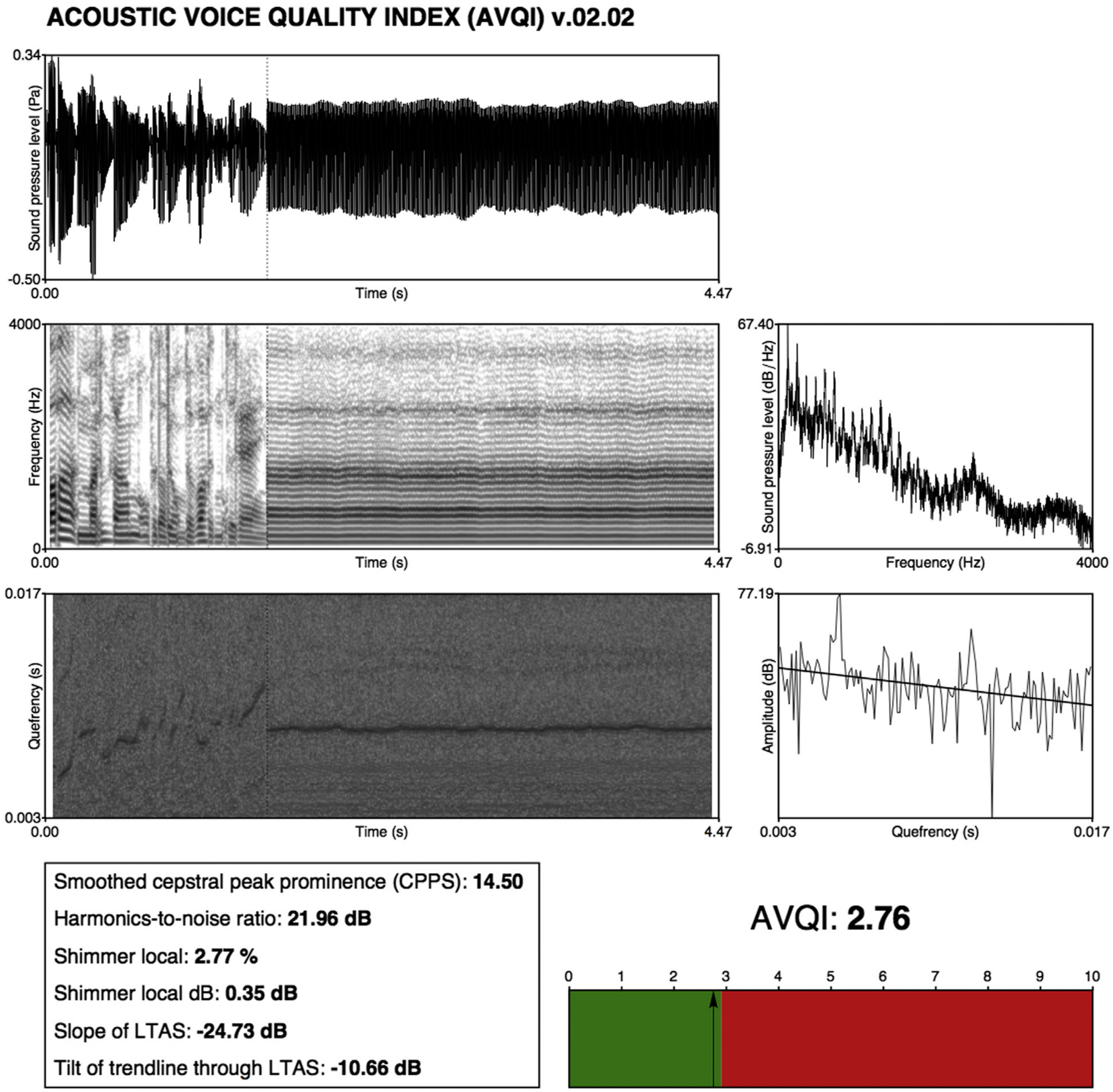

FIGURE 1. Example of the graphical output of the *Praat* script for the $AVQI_{beta}$. Subject 3 was a woman of 27 years with a normal larynx. The $AVQI_{beta} = 2.76$ confirms the $G_{mean} = 0.0$ or unanimously rated normophonia. Top graph: oscillogram. Center left graph: narrowband-spectrogram with window length = 0.03 s, time step = 0.002 s, and frequency step = 20 Hz. This spectrogram clearly shows harmonic structures from the bottom up to at least 4000 Hz in the continuous speech and the sustained vowel. There is neither subharmonic nor interharmonic noise, resulting in relatively low shimmer. Center right graph: long-term average spectrum with frequency step = 1 Hz. This spectrum shows a relatively unsteep inclination of the spectral energy (ie, less negative slope and tilt) and on average 21 harmonics emerge in the concatenated sound, resulting in an increased HNR. Bottom left and right graphs: respectively power cepstrogram and power cepstrum with time step = 0.02 s and ranging between 0.00303 s and 0.01667 s (ie, 330 Hz and 60 Hz, respectively). Both the power cepstrogram and the power cepstrum show clear emergence of a sharp first rahmonic, leading to increased CPPS.

Because the originally published AVQI-method combines two computer programs (ie, *Praat* for the acoustic measures HNR, SL, SLdB, Slope, and Tilt, and *SpeechTool* for the acoustic measure CPPS), it is considered relatively labor-intensive in terms of number of computer actions and therefore relatively time-consuming. To overcome this barrier (ie, to enable a single-program strategy) and to increase feasibility, a second version of the AVQI had to be developed in which the program *SpeechTool* no longer needed to be combined with the program *Praat* to yield all six acoustic measures of the AVQI model. Because all six acoustic measures of the AVQI are available now in *Praat* a simplification of the AVQI protocol became possible. However, the recently implemented calculation of the CPPS in *Praat*[42] deviates somewhat from the original calculation of the CPPS in *SpeechTool*,[27,28] precipitating the following two research questions. First, is there an acceptable correlation/

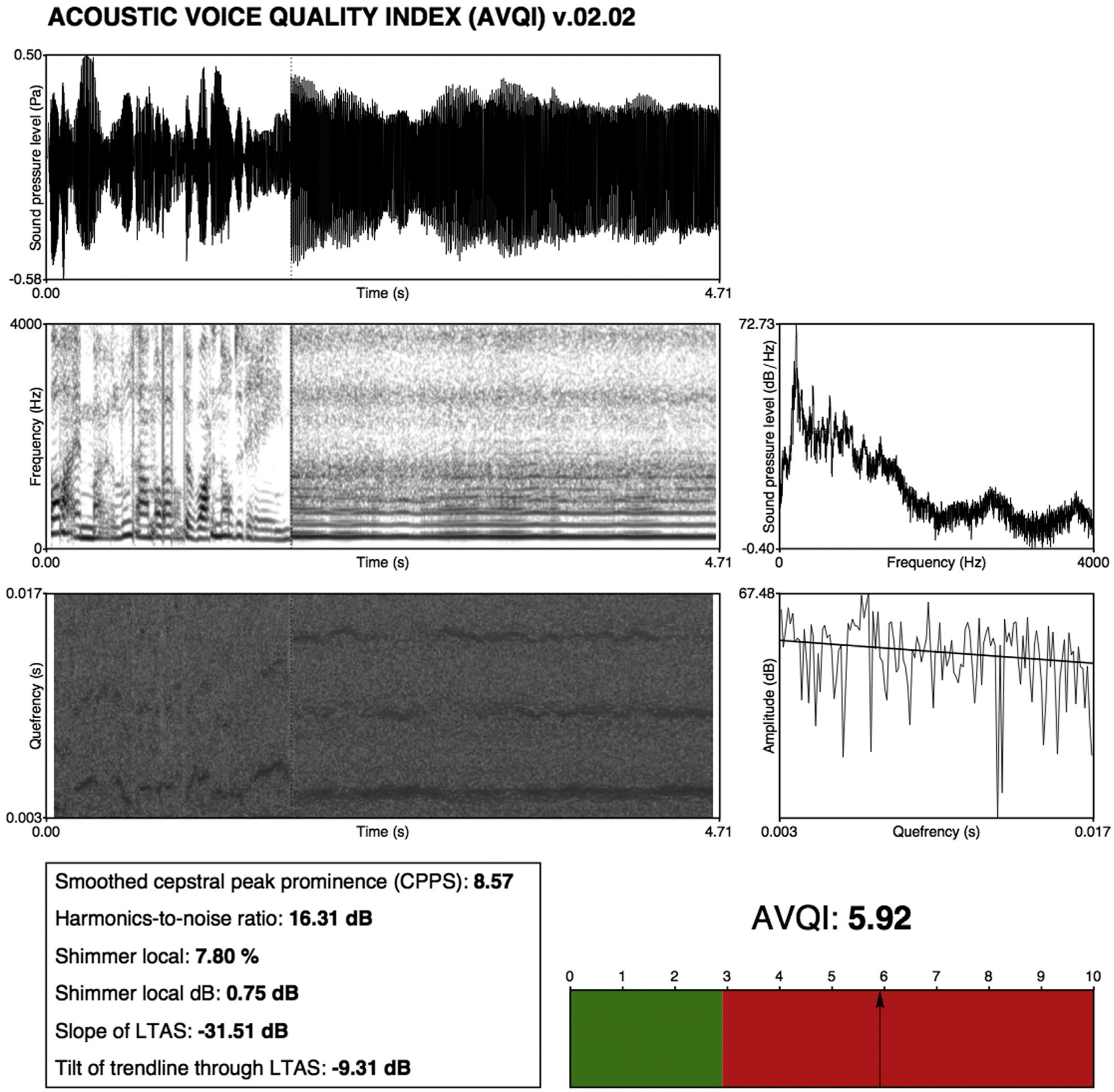

FIGURE 2. Example of the graphical output of the *Praat* script for the $AVQI_{beta}$. Subject 51 was a woman of 43 years with bilateral sulcus glottidis. The $AVQI_{beta} = 5.92$ confirms the $G_{mean} = 2.0$ or unanimously rated moderate dysphonia. Top graph: oscillogram. Center left graph: narrowband-spectrogram with window length = 0.03 s, time step = 0.002 s and frequency step = 20 Hz. This spectrogram clearly shows noise energy between 700 Hz and at least 4000 Hz in both the sustained vowel and the continuous speech. Furthermore it displays intermittent segments with triplophonia or f_0/3-phonation. The noise and the subharmonics result in increased shimmer. Center right graph: long-term average spectrum with frequency step = 1 Hz. This spectrum shows a steep inclination of the spectral energy (ie, more negative slope and tilt) and on average only six harmonics emerge in the concatenated sound, resulting in a decreased HNR. Bottom left and right graphs: respectively power cepstrogram and power cepstrum with time step = 0.02 s and ranging between 0.00303 s and 0.01667 s (ie, 330 Hz and 60 Hz, respectively). Both the power cepstrogram and the power cepstrum show insufficient emergence and broadening of the first rahmonic, leading to decreased CPPS.

correspondence between the two versions of the CPPS? Second, is there an acceptable interconnection between the two versions of the AVQI? Based on analyses of clinical recordings of a sustained [a:] and continuous speech provided by 28 normophonic and 261 dysphonic subjects, the present study can give a positive answer to both questions. Up to 92.4% of the variance in the original $CPPS_{SpeechTool}$ is explained by the variance in the derived $CPPS_{Praat}$. This indicates that the $CPPS_{Praat}$ is a highly acceptable approximation of the $CPPS_{SpeechTool}$. Furthermore, the variance of the $AVQI_{beta}$ accounted for 96.0% the variance of the $AVQI_{alfa}$. This means that there is an almost perfect correspondence between the first and the second version of the AVQI. It can thus be concluded that it is very reasonable to apply the beta version of the AVQI

TABLE 2.
Descriptive Statistics of the Outcomes of the Four Variables in This Study

Acoustic Measure	N	Average	Standard Deviation	Minimum	Maximum
$CPPS_{SpeechTool}$	289	6.61	1.79	0.89	10.94
$CPPS_{Praat}$	289	11.66	2.68	2.65	17.68
$AVQI_{alfa}$	289	3.83	1.63	0.00	10.00
$AVQI_{beta}$	289	3.97	1.60	0.63	10.10

(ie, the *Praat* script in Appendix) as a more feasible, faster, and userfriendly surrogate of the original AVQI.

Caveats and limitatons

Although *Praat* is freeware, runs on most of the popular computer operating platforms, and is therefore open to most of the voice clinicians, caution is required regarding the data acquisition system in which it is applied. Many aspects related to digital data acquisition and sound recording are known to affect the outcome of algorithms for the acoustic measurement of voice quality and their influence may not be underestimated. First, external/environmental noise (ie, all sounds and signals that are not meant to be recorded and analyzed during the clinical voice assessment) such as ventilation/fan noise of the computer, turbulence due to air conditioning, sounds outside the recording room (eg, traffic, heavy wind, and so on.), mobile phone interference, and so on may add erroneous signals to the direct voice/speech signal before it is captured by the microphone.[43] It is therefore crucial to minimize noise and to strive for a recording environment with an optimal signal-to-noise ratio (ie, SNR). Studies of Deliyski et al[44,45] show that an SNR = 30 dB is the minimum for voice measures to be valid and reliable, and that a SNR = 42 dB is required to obtain 99% accuracy in acoustic voice quality measures. This can be achieved in an anechoic chamber, a special sound treated room, or another room that is known to have sufficiently low noise levels. For the voice to be recorded reliably at eg, 70 dB on average, the voice clinician is advised to avoid interference so the noise does not exceed 40 dB. Second, microphone type and placement may also influence the sound recordings. In general, the following recommendations can lead the choice of a microphone[46–48]: a head-mounted condenser microphone with an XLR connection, a low impedance, a cardioid directionality pattern, a frequency response curve that is flat (ie, within 2 dB variation) between circa 50–8000 Hz, and a dynamic measuring range between circa 40–130 dB without distortion of the voice signal. Additionally, such a microphone should be placed off-axis (ie, at 45° or 90° from the direct air stream from the mouth) at approximately 5–10 cm from the mouth.[49] Third, it is generally advised to build-in an external mixer soundcard between the microphone and the computer,[47] and to make sure that the technical characteristics are similar to those of the microphone in terms of dynamic range, frequency response, and so on.[48] Fourth, aspects related to analog-to-digital conversion are also important and it is advised to record the voice sounds with a sampling frequency of at least 26 kHz[50] and a resolution of at least 24 bits.[48] Many studies have examined the effects of these

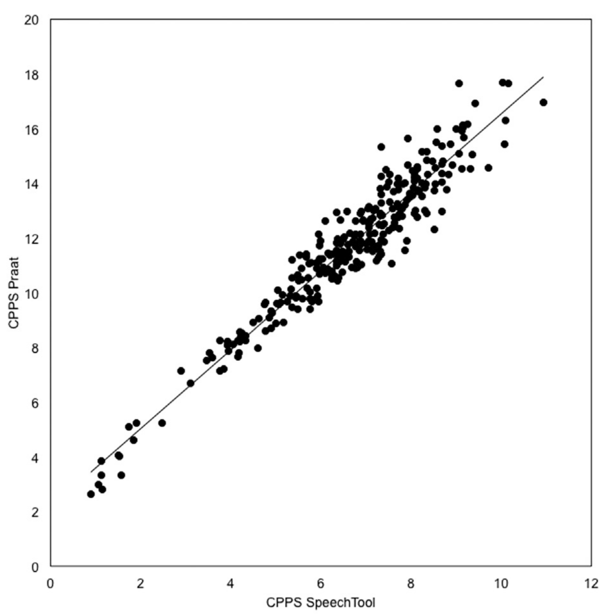

FIGURE 3. Scatterplot illustrating the proportional relationship between the $CPPS_{SpeechTool}$ and the $CPPS_{Praat}$. The linear regression line is defined by $y = 1.4387x + 2.1543$ and corresponds with $r_P^2 = 0.924$.

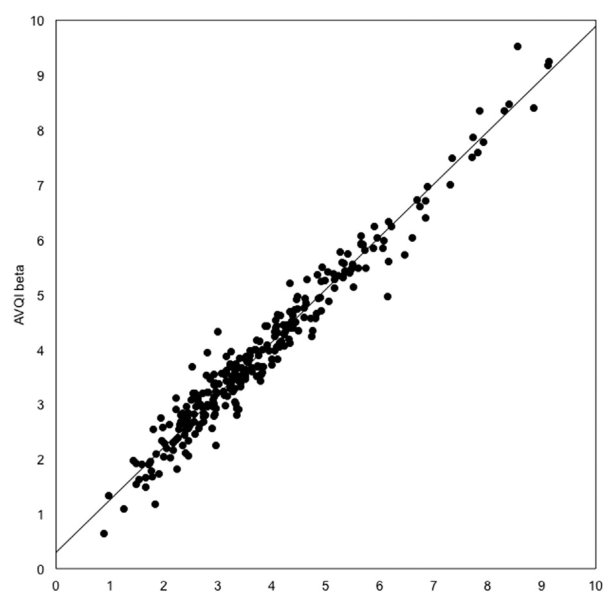

FIGURE 4. Scatterplot illustrating the proportional relationship between the $AVQI_{alfa}$ and the $AVQI_{beta}$. The linear regression line is defined by $y = 0.9597x + 0.2942$ and corresponds with $r_P^2 = 0.960$.

digital sound recording-related items on voice perturbation measures (ie, jitter, shimmer, and/or HNR) and found significant results. To our knowledge, however, the impact of microphone type and placement, environmental acoustics, digitization characteristics, and so forth on the power cepstrum, the prominence of the cepstral peak, and the outcome of the AVQI have not been investigated before. Therefore, to promote standardized and reliable methods for CPPS, AVQI, and other measures, future research should explore the influence of recording-related variables on the outcome of these measures. Meanwhile voice clinicians are advised to use the highest quality recording equipment in an environment with sufficient SNR.

Acknowledgments

The authors wish to thank the Flemish Association for Speech-Language Pathologists (Belsele, Belgium) for the financial support to implement the power cepstrum, the power cepstrogram, and the CPPS in the program *Praat*.

Supplementary data

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.jvoice.2014.06.015.

REFERENCES

1. Buder EH. Acoustic analysis of voice quality: a tabulation of algorithms 1902–1990. In: Kent RD, Ball MJ, eds. *Voice Quality Measurement*. San Diego, CA: Singular Publishing Group; 2000:119–244.
2. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res*. 2001;44: 327–339.
3. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall; 2000.
4. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J Voice*. 2010; 24:540–555.
5. Askenfelt AG, Hammarberg B. Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. *J Speech Hear Res*. 1986;29:50–64.
6. Zraick RI, Wendel K, Smith-Olinde L. The effect of speaking task on perceptual judgment of the severity of dysphonic voice. *J Voice*. 2005;19: 574–581.
7. de Krom G. Consistency and reliability of voice quality ratings for different types of speech fragments. *J Speech Hear Res*. 1994;37:985–1000.
8. Revis J, Giovanni A, Wuyts F, Triglia J. Comparison of different voice samples for perceptual analysis. *Folia Phoniatr Logop*. 1999;51:108–116.
9. Wolfe V, Cornell R, Fitch J. Sentence/vowel correlation in the evaluation of dysphonia. *J Voice*. 1995;9:297–303.
10. De Bodt M. *A framework for voice assessment: the relation between subjective and objective parameters in the judgement of normal and pathological voice* [doctoral thesis]. Antwerp, Belgium: University of Antwerp, Faculty of Medicine; 1997.
11. Maryn Y, Roy N. Sustained vowels and running speech in the auditory-perceptual evaluation of dysphonia severity. *J Soc Bras Fonoaudiol*. 2012;24:107–112.
12. Kreiman J, Vanlancker-Sidtis D, Gerratt BR. Perception of voice quality. In: Pisoni DB, Remez RE, eds. *The Handbook of Speech Perception*. Malden, MA: Blackwell Publishing; 2005:338–362.
13. Kreiman J, Gerratt BR. Validity of rating scale measures of voice quality. *J Acoust Soc Am*. 1998;104:1598–1608.
14. Maryn Y, Roy N, De Bodt M, Van Cauwenberge P, Corthals P. Acoustic measurement of overall voice quality: a meta-analysis. *J Acoust Soc Am*. 2009;126:2619–2634.
15. Eskenazi L, Childers DG, Hicks DM. Acoustic correlates of vocal quality. *J Speech Hear Res*. 1990;33:298–306.
16. Giovanni A, Robert D, Estublier N, Teston B, Zanaret M, Cannoni M. Objective evaluation of dysphonia: preliminary results of a device allowing simultaneous acoustic and aerodynamic measurements. *Folia Phoniatr Logop*. 1996;48:175–185.
17. Wolfe V, Fitch J, Martin D. Acoustic measures of dysphonic severity across and within voice types. *Folia Phoniatr Logop*. 1997;49:292–299.
18. Piccirillo JF, Painter C, Fuller D, Haiduk A, Fredrickson JM. Assessment of two objective voice function indices. *Ann Otol Rhinol Laryngol*. 1998;107: 396–400.
19. Wuyts FL, De Bodt MS, Molenberghs G, et al. The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach. *J Speech Lang Hear Res*. 2000;43:796–809.
20. Yu P, Ouaknine M, Revis J, Giovanni A. Objective voice analysis for dysphonic patients: a multiparametric protocol including acoustic and aerodynamic measurements. *J Voice*. 2001;15:529–542.
21. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurement. *J Voice*. 2004;18:299–304.
22. Awan SN, Roy N. Toward the development of an objective index of dysphonia severity: a four-factor acoustic model. *Clin Linguist Phon*. 2006;20:35–49.
23. Ma E, Yiu E. Multiparametric evaluation of dysphonic severity. *J Voice*. 2006;20:380–390.
24. Maryn Y, Dick C, Vandenbruaene C, Vauterin T, Jacobs T. Spectral, cepstral, and multivariate exploration of tracheoesophageal voice quality in continuous speech and sustained vowels. *Laryngoscope*. 2009;119: 2384–2394.
25. Awan SN, Solomon NP, Helou LB, Stojadinovic A. Spectral-cepstral estimation of dysphonia severity: external validation. *Ann Otol Rhinol Laryngol*. 2013;122:40–48.
26. Lowell SY, Colton RH, Kelley RT, Mizia SA. Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech. *J Voice*. 2013;27:393–400.
27. Hillenbrand J, Cleveland RA, Erickson R. Acoustic correlates of breathy vocal quality. *J Speech Hear Res*. 1994;37:769–778.
28. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Hear Res*. 1996;39: 311–321.
29. Wolfe V, Martin D. Acoustic correlates of dysphonia: type and severity. *J Commun Disord*. 1997;30:403–415.
30. Heman-Ackah YD, Michael DD, Goding GS. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice*. 2002;16:20–27.
31. Heman-Ackah YD, Heuer RJ, Michael DD, et al. Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol*. 2003;112:324–333.
32. Halberstam B. Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL J Otorhinolaryngol Relat Spec*. 2004;66: 70–73.
33. Eadie TL, Baylor CR. The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *J Voice*. 2006;20:527–544.
34. Maryn Y, De Bodt M, Roy N. The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord*. 2010;43:161–174.
35. Lowell SY, Colton RH, Kelley RT, Hahn YC. Spectral- and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker. *J Voice*. 2011;25:e223–e232.
36. Moers C, Möbius B, Rosanowski F, Nöth E, Eysholdt U, Haderlein T. Vowel- and text-based cepstral analysis of chronic hoarseness. *J Voice*. 2012;26:416–424.
37. Samlan RA, Story BH, Bunton K. Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. *J Speech Lang Hear Res*. 2013;56:1209–1223.

38. Reynolds V, Buckland A, Bailey J, et al. Objective assessment of pediatric voice disorders with the acoustic voice quality index. *J Voice*. 2012;26:672.e1–672.e7.
39. Barsties B, Maryn Y. The acoustic voice quality index. Toward expanded measurement of dysphonia severity in German subjects. *HNO*. 2012;60:715–720.
40. Maryn Y, De Bodt M, Barsties B, Roy N. The value of the acoustic voice quality index as a measure of dysphonia severity in subjects speaking different languages. *Eur Arch Otorhinolaryngol*. 2014;271:1609–1619.
41. Van de Weijer JC, Slis IH. Nasaliteitsmeting met de nasometer. [Nasality measurement with the nasometer]. *Tijdschrift voor Logopedie en Foniatrie*. 1991;63:97–101.
42. Weenink D. Implementation of the cepstral peak prominence measure in Praat. In preparation.
43. Howard DM, Murphy DT. *Voice Science, Acoustics and Recording*. San Diego, CA: Plural Publishing; 2008.
44. Deliyski DD, Shaw HS, Evans MK. Adverse effects of environmental noise on acoustic voice quality measurements. *J Voice*. 2005;19:15–28.
45. Deliyski DD, Shaw HS, Evans MK, Vesselinov R. Regression tree approach to studying factors influencing acoustic voice analysis. *Folia Phoniatr Logop*. 2006;58:274–288.
46. Winholtz WS, Titze IR. Miniature head-mounted microphone for voice perturbation analysis. *J Speech Lang Hear Res*. 1997;40:894–899.
47. Vogel AP, Morgan AT. Factors affecting the quality of sound recording for speech and voice analysis. *Int J Speech Lang Pathol*. 2009;11:431–437.
48. Švec JG, Granqvist S. Guidelines for selecting microphones for human voice production research. *Am J Speech Lang Pathol*. 2010;19:356–368.
49. Titze IR, Winholtz WS. Effect of microphone type and placement on voice perturbation measurements. *J Speech Hear Res*. 1993;36:1177–1190.
50. Deliyski DD, Shaw HS, Evans MK. Influence of sampling rate on accuracy and reliability of acoustic voice measures. *Logoped Phoniatr Vocol*. 2005;30:55–62.