The Acoustic Voice Quality Index: Toward improved treatment outcomes assessment in voice disorders

Youri Maryna, Marc De Bodt, Nelson Roy

Abstract

Voice practitioners require an objective index of dysphonia severity as a means to reliably track treatment outcomes. To ensure ecological validity however, such a measure should survey both sustained vowels and continuous speech. In an earlier study, a multivariate acoustic model referred to as the Acoustic Voice Quality Index (AVQI), consisting of a weighted combination of 6 time-, frequency- and quefrency-domain metrics, was developed to measure dysphonia severity in both speaking tasks. In the current investigation, the generalizability and clinical utility of the AVQI are evaluated by first assessing its external cross-validity and then determining its sensitivity to change in dysphonia severity following surgical and/or behavioral voice treatment. The results, based upon a new set of normal and disordered voices compared favorably with outcomes reported earlier, indicating acceptable external validity. Furthermore, the AVQI was sensitive to treatment-related changes, validating its role as a potentially robust and objective voice treatment outcomes measure.

Learning outcomes: Readers will be able to: (1) explain methodological issues surrounding the development of voice treatment outcomes measures (such as external cross-validity and responsiveness to change), (2) appreciate the relevance of measuring dysphonia severity in both sustained vowels and connected speech, (3) describe the method of obtaining the Acoustic Voice Quality Index (AVQI), (4) appreciate differences among a variety of estimates of diagnostic accuracy, and (5) discuss the AVQI as a clinically valid treatment outcomes measure.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Measuring treatment outcomes is a fundamental component of evidence-based practice in speech-language pathology. An outcome is simply the result of an intervention. In the area of voice disorder management, many...
methods have been proposed as potential voice treatment outcomes measures (i.e., laryngoscopic, vibratory, aerodynamic, auditory-perceptual, etc.). However, acoustic measurement of voice has received substantial attention as a potential objective treatment outcomes measure due to its relatively low cost, non-invasiveness, ease of application, and numerical output (Awan & Roy, 2009; Maryn, Corthals, Van Cauwenberge, Roy, & De Bodt, 2009; Parsa & Jamieson, 2001). The general need for objective outcomes measures of voice treatment, has motivated clinicians and researchers to develop and investigate the clinical value of specific acoustic measures of dysphonia severity. As a consequence, there has been a proliferation of acoustic analysis algorithms sensitive to measures of F0 perturbation, amplitude perturbation, waveform perturbation, spectral configuration and cepstral characteristics in radiated and inverse filtered soundwaves (Buder, 2000). However, the validity and clinical utility of many of these acoustic measures, especially the more popular time-based, perturbation measures, has been strongly debated over the past two decades (e.g., De Bodt, 1997; Kreiman & Gerratt, 2005; Parsa & Jamieson, 2001; Titze, 1995). Moreover, Carding et al. (2004) confirmed inadequate sensitivity of any one of these time-domain perturbation measures, used in isolation, to treatment-related changes in voice and voice quality. To overcome the limited validity of single acoustic parameters, and also recognizing the multidimensionality of voice, several researchers have explored a multiparameter approach to measuring voice quality, and/or to distinguish among different voice types and levels of dysphonia severity (e.g., Awan & Roy, 2006; Ma & Yiu, 2006; Wuyts et al., 2000; Yu, Ouaknine, Revis, & Giovanni, 2001). One of these multivariable models, a 6-factor model referred to as the ‘Acoustic Voice Quality Index’ (i.e., ‘AVQI’), was developed by Maryn et al. (2009). Although the AVQI was constructed in a similar manner to other models, it is unique in that it permits objective assessment of dysphonia severity on sustained vowels as well as continuous speech.

In practice, acoustic measures are traditionally derived from sustained mid-vowel samples and not from continuous speech samples for several reasons. First, a sustained vowel represents relatively stable phonation whereas continuous speech involves fast and frequent glottal and supraglottal changes. Second, in contrast to continuous speech, sustained mid-vowel segments do not contain non-voiced phonemes, rapid voice on- and offsets and prosodic variations in F0 and amplitude. Third, sustained vowels are not influenced by speech rate and stress, vocal pauses, and phonetic context. Fourth, classic F0 or T0 perturbation and amplitude perturbation measures strongly rely on pitch detection and extraction algorithms; and consequently they become imprecise in continuous speech analyses, in which perturbation is significantly increased by intonational patterns, voice onsets and offsets, and unvoiced segments. Fifth, sustained vowels can be elicited and produced with less effort and in a more standardized manner than continuous speech. Sixth, there is no linguistic loading in a sustained vowel, resulting in relative immunity from influences related to language, dialect and region, etc. (Askenfelt & Hammarberg, 1986; Maryn et al., 2009; Parsa & Jamieson, 2001; Zraick, Wendel, & Smith-Olinde, 2005).

The inclusion of both speaking tasks or stimulus types (i.e., continuous speech and sustained vowel) in voice analysis is important for several reasons however. First, vocal inconstancies typically observed in continuous speech rather than in sustained vowels (e.g., voice onset/offset, prosodic modulations, voice breaks, etc.) can be decisive in auditory-perceptual voice quality evaluation (Hammarberg, Fritzell, Gauffin, Sundberg, & Wedin, 1980). Second, the two stimulus types can express different types/degrees of vocal dysfunction and, consequently, result in different perceptual ratings (Wolfe, Cornell, & Fitch, 1995; Zraick et al., 2005). Adductor spasmodic dysphonia, for example, can often be characterized by relatively normal voice during sustained vowels, whereas voice in continuous speech is often more severely disrupted (Roy, Gouse, Mauzycki, Merrill, & Smith, 2005). Third, dysphonia symptoms commonly emerge in conversational voice production instead of sustained vowels (except for singing voice) and they are usually revealed to patients in connected speech (Yiu, Worrall, Longland, & Mitchell, 2000). Therefore, if a voice treatment outcomes measure is to be considered “ecologically valid” (i.e., one that is truly representative of daily speech and voice use patterns), then the acoustic measure ideally should be calculated on recordings of both speaking tasks.

To our knowledge, the AVQI is the first measure to incorporate samples of continuous speech, in addition to the sustained vowel samples used in other measurement protocols. To calculate the AVQI, a weighted combination of the output of 6 acoustic time- (i.e., shimmer local, shimmer local dB and harmonics-to-noise ratio), frequency- (i.e., general slope of the spectrum and tilt of the regression line through the spectrum) and quefrency-domain (i.e., smoothed cepstral peak prominence) measures is modeled in a linear regression formula. The concurrent validity of the AVQI-model, i.e. its ability to measure overall severity of dysphonia (also known as Grade or G), is promising; and, a study of the diagnostic precision validated its ability to generally discriminate between normophonia and dysphonia (Maryn et al., 2009). The discriminatory power of the AVQI was considered to be at least equivalent to other
multivariate models (Eskenazi, Childers, & Hicks, 1990; Wolfe, Fitch, & Cornell, 1995; Wolfe, Fitch, & Martin, 1997; Wuyts et al., 2000), but additional evaluation was considered essential to confirm and extend these early reports.

The present investigation examines further the validity and clinical utility of the AVQI as a potential treatment outcomes measure. In this regard two experiments were conducted. The first experiment examined the external cross-validity of the AVQI. When evaluating a predictive instrument like the AVQI, data are initially collected on an experimental sample (i.e., the sample originally used in Maryn et al., 2009) and the multivariate algorithm associated with these data (e.g., predictive equation, cutoff score, etc.) is then applied later to a larger population. Data gathered on the initial experimental sample can sometimes differ substantially from data obtained from a subsequent sample containing different subjects and voices (Portney & Watkins, 2000). Consequently, it is unknown whether the AVQI’s performance (i.e., its validity and accuracy) will necessarily be the same as on the original sample. Therefore, cross-validation is necessary to establish generalizability, and valid treatment outcomes measures demand such validation (Frey, Botan, Friedman, & Kreps, 1991; Portney & Watkins, 2000). This generalizability or “external validity” of a measure can be investigated via two methods: internal and external cross-validity. Internal cross-validity works with subsets of subjects that are randomly drawn from the original sample. Maryn et al. (2009), have already explored the internal cross-validity of the AVQI. Via random selection from the original sample of 251 subjects, the authors formed 30 groups of 10, 50 and 100 subjects. For every group, a correlation coefficient between the AVQI and the auditory-perceptual ratings of dysphonia severity was then calculated. With mean correlations of 0.77 for the groups with 100 subjects, 0.75 for the groups with 50 subjects, and 0.80 for the groups with 10 subjects, it was concluded that the AVQI demonstrated stable validity across the different samples. External cross-validation on the other hand, is typically accomplished by assessing a measure’s performance on a totally new group of subjects, under the assumption that the original and new groups are both good representatives of the population under consideration (i.e., vocally normal and voice-disordered subjects with various degrees of dysphonia). To generalize the AVQI’s accuracy across samples, it is thus important to cross-validate it on a new sample (i.e., external cross-validity). The first experiment of this study represents the external cross-validity of the AVQI using a new sample of 39 subjects.

The second experiment focused on the AVQI’s responsiveness to change. The ability of a measure to detect change over time is a crucial issue if this measure is to be used to assess the potential effects (i.e., outcome) of an intervention, and to serve as a viable treatment outcomes measure (Portney & Watkins, 2000). An acoustic measure is considered to be “change-responsive” when a change in its score is proportional to a change in the score of its clinical criterion (traditionally the auditory-perceptual rating of dysphonia severity). As such, there are only a few reports exploring whether or not a particular acoustic parameter is a valid outcomes measure of voice therapy. However, the design of these studies does not always permit conclusions regarding the responsiveness to change. Plant, Hillel, and Waugh (1997), for example, examined the change in perceptual ratings and an acoustic measure called ‘pitch amplitude’ before and after medialization thyroplasty in 16 subjects with unilateral vocal fold immobility. ‘Pitch amplitude’ is the amplitude of the maximum correlation in the autocorrelation function of the inverse filtered voice signal (Davis, 1979). Although thyroplasty resulted in a statistically significant difference in both the perceptual rating and the pitch amplitude, no statistics regarding the proportionality between the differences/changes of the 2 measures were provided. Whether or not a change in pitch amplitude truly represented a proportionally equivalent change in perceived degree of dysphonia, i.e., the degree to which pitch amplitude is a valid outcomes measure of medialization thyroplasty, remains unclear. A similar study was recently reported by Jin et al. (2008). They monitored 40 patients with laryngopharyngeal reflux and investigated the changes in the reflux symptom index, the reflux finding score, and ‘jitter’ following medical therapy. ‘Jitter’ measures the mean difference in fundamental frequency or duration of adjacent periods, relative to the mean fundamental frequency or duration of all periods in the voice recording (Buder, 2000). Jin et al. (2008) found statistically significant differences for all measures before and after 1 month of treatment and concluded that “acoustic parameters can be used as indicators of treatment efficacy in the patients with laryngopharyngeal reflux” (p. 940). However, the proportional relationship between the magnitude in change of jitter and the magnitude of change in the reflux symptom index and the reflux finding score was not investigated, thus the use of jitter as a viable indicator of treatment (i.e., anti-reflux medication) efficacy also remains unknown. Awan and Roy (2009), on the other hand, investigated the relationship between the change scores associated with their 4-factor acoustic model and change scores observed on auditory-perceptual ratings of dysphonia severity before and after manual circumlaryngeal therapy in 88 patients with muscle tension dysphonia. They found a correlation (i.e., proportional relationship) of 0.75 between acoustically predicted change scores and perceived dysphonia severity change scores, and concluded that the acoustic model was a sensitive outcomes measure. Interestingly, although the acoustic model of Awan and Roy (2009) and the AVQI of Maryn et al.
were constructed independently, they share some similarities. For instance, except for the noise-to-harmonics ratio in the AVQI and the pitch sigma in the model of Awan and Roy (2009), both models contained measures of the cepstral peak, the ratio of low to high spectral energy, and the period-to-period amplitude perturbation. While the Awan and Roy (2009) acoustic model has shown respectable responsivity to change, the AVQI’s sensitivity to change and thus its potential as a treatment outcomes measure awaits experimental verification. Therefore, the second experiment of the present study investigated the AVQI’s responsiveness to change using a new set of 33 patients with various voice pathologies who underwent various kinds of intervention.

2. Methods

2.1. Subjects

Voice samples from 72 patients recruited from the ENT Department of the Sint-Jan General Hospital in Bruges, Belgium were employed in the 2 experiments. For the first experiment (i.e. the external cross-validation of AVQI), we used the recordings of 6 vocally normal and 33 voice-disordered subjects. Their voices were recorded at the beginning of a standard voice assessment. The group of subjects with a voice disorder consisted of 19 females and 14 males, ranging in age from 16 to 86 years (mean = 49.2 years, SD = 20.1 years). There also were 6 females without any voice disorder, ranging in age from 27 to 63 years (mean = 36.1 years, SD = 8.3 years).

For the second experiment (i.e. the AVQI’s responsiveness to change), the pre- and post-therapy recordings of another 33 subjects with organic and/or non-organic voice disorders were used (Table 1). This group consisted of 22 females and 11 males, with a mean age of 40.9 years (SD = 18.9 years, range 7–68 years). To examine the AVQI’s ability to track different degrees of change in dysphonia severity following treatment, recordings of subjects experiencing varying degrees of therapeutic progress and change in overall voice quality were included in this experiment. The type of voice treatment and the combination of behavioral techniques used to improve voice is described in Table 1.

2.2. Voice treatment

Table 1 provides an overview of the interventions employed for all the individual subjects included in the second experiment. All 33 subjects received an eclectic treatment program, with an individualized combination of behavioral voice therapy techniques and/or surgery. Behavioral voice therapy included indirect strategies (i.e., counseling and advice concerning vocal hygiene, and healthy voice use) as well as direct strategies (i.e., combined exercises on speech-breathing, resonance, pitch, loudness, phonatory facilitation and voice onset) to improve the voice and/or to decrease the number and severity of voice-related complaints. There were 6 patients (i.e., 18.2%) who were treated primarily with surgery, including phonosurgery (i.e., PS) in 3 subjects with a vocal fold cyst and 1 subject with severe polypoid degeneration, and laryngeal framework surgery (i.e., LFS) in 2 subjects with unilateral vocal fold paralysis who underwent type I thyroplasty.

The number of sessions of behavioral voice therapy ranged from 1 to 49, with an average of 7.4 sessions. Recordings of continuous speech and sustained vowels were made before and after the voice treatment, with a mean interval of 97.9 days (range 1–483 days).

2.3. Voice recordings

For all 105 recordings (i.e., 39 recordings in experiment 1, 33 pre-treatment recordings in experiment 2 and 33 post-treatment recordings in experiment 2), subjects were asked to sustain the vowel /a/ for at least 5 s and to read aloud a Dutch phonetically balanced text (Van de Weijer & Slis, 1991) 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 Corp., Lincoln Park, NJ, USA), and saved in WAV-format. For this study, a copy of every vowel sample was edited to include only the central 3 s and a copy of the read text samples was trimmed to contain only the first two sentences. In the program “Praat” (Paul Boersma, Institute of Phonetic Sciences, University of Amsterdam, The Netherlands), the text segment, a pause of 2 s and the vowel segment were concatenated for the perceptual evaluation of both speech types into a single rating. The resulting concatenated waveform is represented in Fig. 1.
Table 1
Overview of relevant information on the subjects (gender, age and voice disorders) and their voice treatment (number of sessions, interval between the 2 assessments and the type and subtype of therapy).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Gender</th>
<th>Age at pre-therapy recording (years)</th>
<th>Number of therapy sessions</th>
<th>Interval between pre- and post-therapy (days)</th>
<th>Voice disorder&lt;sup&gt;a,b&lt;/sup&gt;</th>
<th>Surgery&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Behavioral voice therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PS LFS</td>
<td>Indirect Breathing Resonance Pitch Loudness Facilitation Onset</td>
</tr>
<tr>
<td>1</td>
<td>M</td>
<td>7</td>
<td>9</td>
<td>104</td>
<td>Bilateral VF nodules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>65</td>
<td>21</td>
<td>221</td>
<td>Left VF paralysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>33</td>
<td>3</td>
<td>39</td>
<td>Polypoid mucosa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>24</td>
<td>1</td>
<td>3</td>
<td>Type 4 MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>46</td>
<td>15</td>
<td>151</td>
<td>Type 2b MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>49</td>
<td>5</td>
<td>154</td>
<td>Polypoid mucosa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>24</td>
<td>16</td>
<td>112</td>
<td>Type 6 MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>40</td>
<td>1</td>
<td>47</td>
<td>Left VF cyst</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>16</td>
<td>8</td>
<td>107</td>
<td>Type 6 MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>26</td>
<td>5</td>
<td>35</td>
<td>Polypoid mucosa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>11</td>
<td>15</td>
<td>252</td>
<td>Bilateral VF nodules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>51</td>
<td>7</td>
<td>133</td>
<td>Left VF paralysis, type 2b MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>63</td>
<td>1</td>
<td>119</td>
<td>Polypoid mucosa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>45</td>
<td>1</td>
<td>42</td>
<td>Right VF paralysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>68</td>
<td>1</td>
<td>1</td>
<td>Type 2b MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>51</td>
<td>2</td>
<td>483</td>
<td>Right VF cyst, type 1 MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>56</td>
<td>5</td>
<td>97</td>
<td>Type 3 MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>58</td>
<td>1</td>
<td>9</td>
<td>Type 3 MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>M</td>
<td>62</td>
<td>1</td>
<td>10</td>
<td>Type 2a MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>F</td>
<td>65</td>
<td>49</td>
<td>286</td>
<td>Left VF paralysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>F</td>
<td>57</td>
<td>1</td>
<td>11</td>
<td>Right VF cyst</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>F</td>
<td>31</td>
<td>1</td>
<td>1</td>
<td>Type 2b MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>F</td>
<td>23</td>
<td>7</td>
<td>54</td>
<td>Bilateral VF nodules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>M</td>
<td>14</td>
<td>7</td>
<td>91</td>
<td>Type 6 MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>F</td>
<td>49</td>
<td>2</td>
<td>9</td>
<td>Type 4 MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>M</td>
<td>10</td>
<td>12</td>
<td>186</td>
<td>Bilateral VF nodules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>F</td>
<td>26</td>
<td>8</td>
<td>113</td>
<td>Bilateral VF nodules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>F</td>
<td>35</td>
<td>1</td>
<td>126</td>
<td>Bilateral VF nodules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>F</td>
<td>49</td>
<td>7</td>
<td>66</td>
<td>Left VF paresis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>F</td>
<td>43</td>
<td>1</td>
<td>1</td>
<td>Type 2b MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>F</td>
<td>23</td>
<td>1</td>
<td>21</td>
<td>Type 4 MTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>M</td>
<td>65</td>
<td>14</td>
<td>38</td>
<td>Left VF paralysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>F</td>
<td>64</td>
<td>14</td>
<td>108</td>
<td>Left VF paralysis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> VF = vocal fold.  
<sup>b</sup> MTD = muscle tension dysphonia (types of MTD according to Rammage et al., 2001), type 1 MTD = laryngeal isometry with posterior incomplete adduction, type 2a MTD = glottal compression or hyperadduction, type 2b MTD = supraglottal hyperadduction, type 3 MTD = anteroposterior compression, type 4 MTD = longitudinal incomplete adduction, type 6 MTD = transitional falsetto register.  
<sup>c</sup> PS = phonosurgery, LFS = laryngeal framework surgery.
Since certain acoustic measures employed in this study are only valid for voiced segments of the continuous speech samples, an algorithm for detection, segmentation, and concatenation of these voiced segments was used. This algorithm was originally based on Parsa and Jamieson (2001, p. 332) and customized in Praat by Maryn et al. (2009). The resulting waveform is shown in Fig. 2.

2.4. Voice quality ratings

Five experienced speech-language pathologists were asked to rate each of the 105 concatenated voice samples. These raters were the same as in Maryn et al. (2009) for the 2 experiments. With exception of the first author (who collected all recordings), all raters were blinded regarding the identity, diagnosis and disposition of the 72 subjects (i.e. normal, pre-treatment, post-treatment, etc.). There were 2 rating sessions (i.e., 1 per experiment). The listening environment and procedures were comparable to those described in the previous study. All concatenated voice samples were presented in random order and judged on overall severity of dysphonia (Grade, $G$) with a 4-point equal-appearing interval scale, as suggested by the Japan Society of Logopedics and Phoniatrics (Hirano, 1981). Before judging the samples, $G$ was described using the definition provided by Kreiman and Gerratt (2000). Furthermore, as
recommended by Chan and Yiu (2002), an attempt was made to establish an external standard, ostensibly to increase the reliability of listener ratings. For the purpose of establishing an external standard, 12 samples were selected from the database from the previous study, i.e. three samples per level of $G$ (0 = normal, 1 = slight dysphonia, 2 = moderate dysphonia, 3 = severe dysphonia). These samples were selected based upon prior unanimous agreement across raters regarding the degree of dysphonia, thus these samples were considered to be highly representative of a specific level of $G$.

2.5. Acoustic measures

Objective measurement of overall voice quality consisted of determining the 6 acoustic parameters for calculating AVQI: smoothed cepstral peak prominence (CPPs) with the computer program “SpeechTool” (James Hillenbrand, Western Michigan University, Kalamazoo, MI, USA) and harmonics-to-noise ratio (HNR), shimmer local, shimmer local dB, general slope of the spectrum (slope) and tilt of the regression line through the spectrum (tilt) with Praat. The method for the determination of the six acoustic measures was identical to the method of Maryn et al. (2009). Consequently, the Acoustic Voice Quality Index (AVQI) was calculated according to the regression formula: $2.571 [3.295 – 0.111 (CPPs) – 0.073 (HNR) – 0.213 (shimmer local) + 2.789 (shimmer local dB) – 0.032 (slope) + 0.077 (tilt)]$. From an initial set of 13 acoustic measures, this combination of 6 weighted measures best predicted dysphonia severity; and this formula was constructed on the unstandardized coefficients of the statistical model after stepwise multiple linear regression analysis.

To assess the test–retest reliability of the acoustic analysis protocol, the first author later reanalyzed 20 samples (i.e., >25%) selected randomly. The samples were coded and deidentified prior to randomization and re-analysis, AVQI measures were recalculated and compared to the original analyses. Pearson’s correlation (i.e., $r_p$) was used to estimate acoustic remeasurement reliability and revealed an $r_p = 0.991$ (significant at the $p = .000$ level), indicating excellent test–retest reliability.

2.6. Statistics

All statistical analyses were completed using SPSS for Windows version 12.0 (SPSS Inc., Chicago, IL, USA). First, the inter-rater reliability of the five raters was investigated using the intraclass correlation coefficient (i.e., ICC). The ICC is a conventionally reliability index that measures the degree of consistency among ratings. Like other reliability coefficients, the ICC ranges from 0.00 (i.e., total absence of reliability) to 1.00 (i.e., perfect reliability). Although there are no standard values for the interpretation of the ICC, a general guideline suggests that values above 0.75 indicate good reliability, and values below 0.75 poor to moderate reliability. For many clinical measurements, however, reliability should exceed 0.90 to ensure reasonable reliability (Portney & Watkins, 2000). Both the single-measures ICC (i.e., reliability based on comparison of ratings from individual listeners) and the average-measures ICC (i.e., reliability based on averaged ratings) were calculated to estimate inter-rater reliability. Using averaged ratings has the effect of increasing reliability estimates, as means are considered better estimates of true scores, theoretically reducing error variance (Portney & Watkins, 2000, p. 562).

To assess the external cross-validity of the AVQI (i.e., how well can the AVQI measure the severity of dysphonia in a new set of clinical voice samples?), the Spearman rank-order correlation coefficient (i.e., $r_s$) and the coefficient of determination (i.e., $r^2_s$) between AVQI and mean $G$ (as averaged over the five raters) served to estimate the criterion-related concurrent validity of the AVQI. Furthermore, several estimates were calculated to appraise the diagnostic precision of the AVQI: how well can AVQI discriminate between normal and dysphonic voices? The diagnostic accuracy of a measure is commonly evaluated by its sensitivity and specificity. Sensitivity is a test’s ability to detect a condition or disease when present. In this case, sensitivity would be the percent of correctly identified dysphonics who test positive on the AVQI. Specificity is an estimate of a test’s ability to correctly identify non-cases (i.e., normophonics) when they test negative on the AVQI. However, depending on the cutoff point chosen to define a positive result, different sensitivity and specificity values can be found. This trade-off between sensitivity and specificity can be examined using a graphic representation called the receiver operating characteristic (ROC) curve. This ROC-curve is created by plotting a point for each cutoff score that represents the true positive score (i.e., sensitivity) on the ordinate and the false positive score (i.e., $1 – specificity$) on the abscissa. In the present study, a voice was considered normophonic only when all five judges rated it as normal (i.e., mean $G = 0.0$). On the other hand, a
voice was considered dysphonic at the moment one judge evaluated it as slightly dysphonic or G1 (i.e., 0.2 ≤ mean
G ≤ 3). The ability of the AVQI to discriminate between normal and dysphonic voices was represented by the “area
under ROC-curve” (i.e., \( A_{\text{ROC}} \)). An \( A_{\text{ROC}} = 1.0 \) is found for measures that perfectly distinguish between normal and
dysphonic voices. An \( A_{\text{ROC}} = 0.5 \) corresponds with chance-level diagnostic accuracy (Portney & Watkins, 2000). To
provide additional evidence regarding the value of a diagnostic measure and to help reduce problems with sensitivity/
specificity related to the base-rate differences in the samples (i.e., the uneven number of 6 normophonic and 33
dysphonic subjects in the sample), likelihood ratios should also be calculated (Dollaghan, 2007). The “likelihood ratio
for a positive result” (i.e., \( LR^+ \)) yields information regarding how the odds of the disease increase when the test is
positive. It is calculated by \( LR^+ = [(\text{sensitivity})/1 - \text{specificity}] \) and gives information regarding the likelihood that an
individual is dysphonic when testing positive. The “likelihood ratio for a negative result” (i.e., \( LR^- \)) is an estimate
that helps to determine if an individual does not have a particular disorder when they test negative on the diagnostic
test. It is calculated by \( LR^- = [(1 - \text{sensitivity})/\text{specificity}] \) and gives information regarding how much the odds of
having a disorder decreases when testing negative. As a general guideline, the diagnostic value of a measure is
considered to be high when \( LR^+ ≥ 10 \) and \( LR^- ≤ 0.1 \). Because the LR statistics consider sensitivity and specificity
simultaneously, they are less vulnerable to sample size characteristics and base-rate differences in the sample between
vocally normal and voice-disordered participants (Dollaghan, 2007). Based on the results of our previous study
(Maryn et al., 2009), the diagnostic statistics \( A_{\text{ROC}}, LR^+ \) and \( LR^- \) were calculated using an AVQI cutoff point of 2.95.

Third, AVQI’s responsiveness to change was investigated by means of the “standardized change score”. A change
score is obtained after subtracting the pre-therapy score from the post-therapy score. However, the auditory-perceptual
rating of dysphonia severity (i.e., \( G \)) and the acoustic measure (i.e., the AVQI) are based on distinct scales, and thus
their change scores cannot be directly compared. A unit free standardized score is necessary to make such a
comparison. In each case, for the auditory-perceptual rating and the acoustic measure alike, the standardized change
score (i.e., SCS) was obtained by dividing the change score by the standard deviation for the pre-therapy scores
(Portney & Watkins, 2000). Subsequently, the change-responsiveness of AVQI was analyzed by correlating the
standardized change score in mean \( G \) (i.e., SCS_{mean-G}) with the standardized change score in AVQI (i.e., SCS_{AVQI}).
The higher \( r_s \) observed between SCS_{mean-G} and SCS_{AVQI}, the more the AVQI is considered to be a responsive treatment
outcomes measure which is sensitive to changes in perceived dysphonia severity. Furthermore, the means of SCS_{mean-G}
and SCS_{AVQI} were compared and the difference between the values of these two paired variables was computed with the
Student t-test (\( \alpha \)-level = 0.05).

3. Results

3.1. Consistency of the auditory-perceptual ratings

The single-measures ICC of 0.698 (95% confidence interval: 0.609–0.779) indicated a moderate inter-rater
reliability between individual raters. As expected, however, this reliability estimate increased to a very acceptable
level with the average-measures ICC of 0.920 (95% confidence interval 0.886–0.946), because ratings of multiple
listeners have been averaged and error variance decreased before calculating the ICC. These ICC’s indicate that the 5
listeners consistently rated the overall severity of dysphonia, and thus the reliability of the \( G \)-ratings can be considered
acceptable for the purposes of the two experiments in this study.

3.2. Experiment 1: external cross-validity of AVQI

The first 2 columns of Table 2 list the descriptive data for AVQI in the group of 6 vocally normal cases and the 33
dysphonic subjects. The first item in the external cross-validation of AVQI was the criterion-related concurrent validity.
This is expressed as the bivariate correlation of \( r_s = 0.796 \) between mean \( G \) and the AVQI across the 39 subjects with or without various
voice disorders. This correlation corresponds with a \( r_s^2 = 0.634 \), designating that more than 60% of the variance of mean
\( G \) was accounted for by the AVQI, and confirms strong concurrent validity (Portney & Watkins, 2000). The proportional
relationship between mean \( G \) and the AVQI is illustrated by the scatterplot and regression line in Fig. 3.

The second issue surrounds the diagnostic accuracy of the AVQI. To evaluate the AVQI’s ability to distinguish
vocally normal from voice-disordered participants, a ROC-curve was constructed (see Fig. 4). The \( A_{\text{ROC}} \), with the
$G$-scores as the state variable (i.e., mean $G = 0$ indicating normal voice quality, mean $G > 0$ indicating disrupted voice quality) and the AVQI scores as the test variable, was 0.920. This result confirms excellent discriminatory power of the AVQI in distinguishing normal from pathological voices (with statistical significance at $p = 0.000$, under the assumption of a nonparametric distribution). Furthermore, an AVQI cutoff score of 2.95 was chosen based on our previous results as a demarcation point between normal and dysphonic voices. In the present study, this cutoff score was associated with a sensitivity = 0.85 and specificity = 1.00, and an extremely impressive $LR^+ > 2000$ and a respectable $LR^- = 0.16$.

### 3.3. Experiment 2: AVQI’s responsiveness to change

The descriptive data of AVQI in the group of 33 dysphonic subjects who received treatment are also listed in the third and fourth column of Table 2. To compare the outcomes before and after voice therapy, a standardized change score [i.e., $SCS = (\text{post-therapy score} - \text{pre-therapy score})/\text{standard deviation for pre-therapy score}$] was first calculated from both the mean $G$ data and the AVQI data. The size of the SCS indicated the degree of change in dysphonia severity: the higher the SCS, the greater the change in the voice quality following treatment. The sign of the SCS indicated the direction of the change in overall voice quality: a minus-sign corresponded with improvement or

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th></th>
<th>Experiment 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normophonic (N = 6)</td>
<td>Dysphonic (N = 33)</td>
<td>Pre-therapy (N = 33)</td>
<td>Post-therapy (N = 33)</td>
</tr>
<tr>
<td>Min</td>
<td>1.66</td>
<td>1.78</td>
<td>2.75</td>
<td>0.70</td>
</tr>
<tr>
<td>Max</td>
<td>2.89</td>
<td>7.33</td>
<td>10.16</td>
<td>4.07</td>
</tr>
<tr>
<td>Range</td>
<td>1.23</td>
<td>5.55</td>
<td>7.41</td>
<td>3.37</td>
</tr>
<tr>
<td>Mean</td>
<td>2.29</td>
<td>4.16</td>
<td>5.29</td>
<td>2.79</td>
</tr>
<tr>
<td>SD</td>
<td>0.45</td>
<td>1.48</td>
<td>2.09</td>
<td>0.65</td>
</tr>
</tbody>
</table>

G-scores as the state variable (i.e., mean $G = 0$ indicating normal voice quality, mean $G > 0$ indicating disrupted voice quality) and the AVQI scores as the test variable, was 0.920. This result confirms excellent discriminatory power of the AVQI in distinguishing normal from pathological voices (with statistical significance at $p = 0.000$, under the assumption of a nonparametric distribution). Furthermore, an AVQI cutoff score of 2.95 was chosen based on our previous results as a demarcation point between normal and dysphonic voices. In the present study, this cutoff score was associated with a sensitivity = 0.85 and specificity = 1.00, and an extremely impressive $LR^+ > 2000$ and a respectable $LR^- = 0.16$.

### 3.3. Experiment 2: AVQI’s responsiveness to change

The descriptive data of AVQI in the group of 33 dysphonic subjects who received treatment are also listed in the third and fourth column of Table 2. To compare the outcomes before and after voice therapy, a standardized change score [i.e., $SCS = (\text{post-therapy score} - \text{pre-therapy score})/\text{standard deviation for pre-therapy score}$] was first calculated from both the mean $G$ data and the AVQI data. The size of the SCS indicated the degree of change in dysphonia severity: the higher the SCS, the greater the change in the voice quality following treatment. The sign of the SCS indicated the direction of the change in overall voice quality: a minus-sign corresponded with improvement or
Fig. 4. ROC-curve illustrating the diagnostic accuracy of AVQI (the dashed line resembles a virtual chance-level discrimination between normophonia and dysphonia).

Fig. 5. Scatterplot, linear regression line \( y = 0.623x - 0.3425 \) and curvilinear regression line \( y = -0.1106x^2 + 0.3332x - 0.3612 \) illustrating the degree in which standardized changes in perceived overall voice quality are predicted by standardized changed in AVQI scores.
decrease in dysphonia severity, a plus-sign was associated with worsening or increase in dysphonia severity. For example, before voice therapy, subject no. 2 had a moderate dysphonia, as evidenced by a mean $G = 1.6$ and an AVQI = 6.31. However, following voice therapy, a normal voice quality was achieved which resulted in a mean $G = 0.0$ and an AVQI = 1.83. The SCSmean-$G = -2.48$ was reflected in a comparable SCSAVQI = -2.14. Subject no. 23 on the other hand, showed a slightly increased dysphonia severity after voice therapy, which was reflected in a positively valenced SCSmean-$G = 0.31$ (from mean $G$ of 0.8 to 1.0), and again a comparable SCSAVQI = 0.20.

The correlation observed between SCSmean-$G$ and SCSAVQI showed a strong proportional relationship ($r_s = 0.80$). Fig. 5 shows a scatterplot of the SCSmean-$G$ values and the SCSAVQI values for the 33 pre- and post-therapy voice samples. Higher SCSmean-$G$ values are mostly associated with proportionally higher SCSAVQI values, and vice versa. The linear regression line reflects a $r^2_s = 0.64$. However, inspection of the scatterplot in Fig. 5 raised suspicion that a nonlinear relationship might exist. Because the lowest and the highest SCSAVQI scores digressed somewhat to the lower side from what was expected based purely on the linear regression line, we also investigated a second-order polynomial relationship (with $r^2_s = 0.670$). However, there was no statistically significant difference (2-tailed $p = 0.842$) between the outcome of the linear model and the outcome of the curvilinear model.

Additionally, Fig. 5 demonstrates that the data range for the SCSmean-$G$ (from −3.71 to 0.93) was very similar to the data range for SCSAVQI (from −3.38 to 0.20). Statistical analysis of these data showed no significant difference ($t = 1.322$; 2-tailed $p = 0.195$), confirming the similarity in the data distributions of SCSmean-$G$ and SCSAVQI.

4. Discussion

To improve ecological validity, assessment of dysphonia severity and tracking of treatment outcomes should employ procedures which survey both sustained vowels and continuous speech (Hammarberg et al., 1980; Parsa & Jamieson, 2001; Yiu et al., 2000; Zraick et al., 2005). To our knowledge, Maryn et al. (2009) were the first to construct a multivariable model based on acoustic measures derived from both speaking tasks. This ‘Acoustic Voice Quality Index’ (a.k.a. ‘AVQI’) model incorporated the middle 3 s of a sustained /a/ and the first 2 sentences of a phonetically balanced Dutch text and concatenated them into 1 single sound file, upon which the following 6 acoustic measures were determined. Measures of harmonic versus noise energy (as ‘harmonics-to-noise ratio’ in the present study) and spectral tilt (as ‘tilt of the trend line through the long-term average spectrum’ and ‘slope of the long-term average spectrum’ in the present study) have traditionally been associated with insufficient glottal closure and breathiness (Awan & Roy, 2006; Sodersten & Lindestad, 1990). Measures of amplitude perturbation (as ‘shimmer local dB’ in the present study) have classically been related to irregular vocal fold vibration and roughness (Awan & Roy, 2006). Although cepstral measures (as ‘smoothed cepstral peak prominence’ in the present study) have mainly been associated with breathiness and less with roughness (Hillenbrand, Cleveland, & Erickson, 1994; Hillenbrand & Houde, 1996; Heman-Ackah, Michael, & Goding, 2002), it is assumed that all factors causing deviations in the voice signal decrease the prominence of the first raahmonic (Ferrer, de Bodt, Maryn, Van de Heyning, & Hernández-Díaz, 2007). Consequently, the combination of these 6 parameters subsumes several measures sensitive to potential vibratory distortions at the level of the glottis, and would serve as a metric of overall dysphonia severity.

Although Maryn et al. (2009) concluded that their results supported the feasibility, predictive validity and diagnostic accuracy of this multiparameter model, confirmation of these results through replication was needed to support the generalizability of AVQI. We therefore investigated the external cross-validity of AVQI using a new set of 39 normophonic and/or dysphonic subjects. While the initial study (Maryn et al., 2009) reported respectable concurrent validity (i.e., $r_s = 0.78$ between AVQI and mean $G$) and diagnostic precision (i.e., $A_{ROC} = 0.90$, sensitivity = 0.74, specificity = 0.96, $LR^+ = 19.98$, $LR^- = 0.27$), the present study provided the requisite external cross-validation of the AVQI by essentially replicating the results of the original study, but this time using a different set of voices. The $r_s = 0.796$ confirmed a strong relationship between the AVQI and mean $G$ (see Fig. 3), and substantiated the feasibility of AVQI as a measure sensitive to the continuum of dysphonia severity. Furthermore, as evidenced by impressive estimates of diagnostic precision (i.e., $A_{ROC} = 0.920$, sensitivity = 0.85, specificity = 1.00, $LR^+ > 2000$, $LR^- = 0.16$) the AVQI offers excellent discriminatory power, sufficient to distinguish normal from pathological voices. These statistics related to concurrent validity and diagnostic accuracy are comparable and in some cases superior to outcomes reported in the original study. This highlights the robustness of the AVQI, as a clinical measure of dysphonia severity and supports its external validity.
In order to serve as a valid treatment outcomes measure however, the AVQI needs to be sensitive to different degrees of change in dysphonia severity. Changes in perceived dysphonia severity (i.e., in ‘G’) should be reflected in proportionally equivalent changes in predicted voice quality (i.e., in AVQI scores). In a second experiment, we therefore evaluated the responsiveness to change of the AVQI using pre- and post-treatment voice samples from 33 voice-disordered patients who underwent surgical and/or individually customized behavioral voice treatment (Table 1). To compare change scores of 2 variables with different units/scales, such as mean G and AVQI, change scores were normalized over the standard deviation for the pre-therapy scores, resulting in unit free standardized change scores (i.e., SCS). Results indicated a strong correlation of $r_s = 0.80$, wherein 64% of the variance in SCS$_{mean-G}$ was accounted for by SCS$_{AVQI}$. These results support the sensitivity of the AVQI to treatment-related changes in dysphonia severity, and thus it should be considered as a valid treatment outcomes measure in voice.

In a very recent study, Awan and Roy (2009) also investigated the external cross-validity and the responsiveness to change of a 4-factor acoustic model for the prediction of dysphonia severity in sustained vowels, developed previously by the same authors in 2006. Interestingly, this model was constructed independently from our model, yet consisted of a similar set of acoustic parameters as AVQI. Both models contain measures of shimmer: “shimmer local” and “shimmer local dB” in AVQI and “shimmer (dB)” in Awan and Roy (2006, 2009). Correlations between shimmer measures and ratings of overall severity of dysphonia or hoarseness have indicated poor to moderately strong associations: $r = 0.54$ (Wolfe, Cornell, et al., 1995), $r = 0.41$ (De Bodt, 1997), and $r = 0.70$ Halberstam (2004). Both models also contain measures of spectral energy distribution: “slope of long-term average spectrum” and “tilt of trend line through long-term average spectrum” in AVQI and “discrete Fourier transform ratio” in the model of Awan and Roy (2006, 2009). Poor to moderate correlations have been found between similar spectral measures and scores of hoarseness or overall dysphonia severity: $r = 0.52$ (Dejonckere & Wieneke, 1996), and $r = −0.47$ (Eadie & Doyle, 2005). Most importantly however, both models also contained a measure of the cepstral peak as the most dominant feature: “smoothed cepstral peak prominence” in AVQI (Maryn et al., 2009) and “ratio of the actual amplitude of the cepstral peak prominence to the expected amplitude” in the model of Awan and Roy (2006, 2009). Several reports exist in the literature emphasizing that measures of the relative magnitude of the first rhamonic (i.e., the first cepstral peak) are promising correlates of dysphonia severity. Cepstral measures, in particular, have received much attention since Hillenbrand and his coworkers reported that the “cepstral peak prominence” (i.e., CPP) and the “smoothed cepstral peak prominence” (i.e., CPPs) were the strongest acoustic correlates of breathiness in sustained vowels ($r = −0.92$; Hillenbrand et al., 1994) as well as in continuous speech ($r = −0.88$; Hillenbrand & Houde, 1996). Their results have been confirmed by many other reports. Dejonckere and Wieneke (1996), for example, reported that the “cepstral magnitude” was the best predictor of hoarseness in sustained vowels ($r = −0.80$), compared to spectral (such as the “spectral noise above and under 6000 Hz”) and perturbation measures (such as the “jitter ratio”). Heman-Ackah et al. (2002) also found that the CPPs was strongly related to overall voice quality in sustained vowels ($r_p = −0.80$) and especially in continuous speech ($r_p = −0.86$), compared to the more traditional perturbation measures (for example “relative average perturbation” and “amplitude perturbation quotient”). Halberstam (2004) also reported very strong correlations of $r_p = −0.90$ and $r_p = −0.88$ between the degree of hoarseness and the respective CPP and CPPs measures on continuous speech. As in Heman-Ackah et al. (2002), these cepstral measures demonstrated superior validity in continuous speech, and consequently proved to be the most ecologically valid measures. Eadie and Baylor (2006) reported $r_s = 0.79$ between CPP and overall dysphonia severity in continuous speech and $r_s = 0.81$ between CPPs and overall dysphonia severity on sustained vowels. On the other hand, only Wolfe and Martin (1997) and Halberstam (2004) reported poor correlations between hoarseness and CPP ($r_p = −0.30$) and between hoarseness and CPPs ($r_p = −0.55$) on sustained vowel samples, respectively. Collectively, these findings together with the results from Maryn et al. (2009) and Awan and Roy (2006, 2009) confirm the power of the cepstrum-based parameters to predict the degree of overall dysphonia, and emphasize the superiority of these measures as compared to more traditional time-domain perturbation measures.

In some ways, it comes as no surprise that we found highly comparable outcomes for the 2 models – in terms of initial validity, internal cross-validity, external cross-validity and sensitivity to change. The initial validity of the model of Awan and Roy (2006) indicated a strong $r = 0.88$, whereas Maryn et al. (2009) found with $r = 0.78$ a slightly lower initial validity. The internal cross-validation based on repeated correlations in different randomized subgroups revealed a strong mean $r = 0.88$ in Awan and Roy (2006) and a less strong mean $r = 0.77$ in Maryn et al. (2009). The external cross-validation of the models in a new set of voice samples showed a very strong $r = 0.91$ in Awan and Roy (2009) and a strong $r_s = 0.80$ for the AVQI. These internal and external cross-validations, and the...
fact that acceptable results were found for the both models, confirm the robustness and stability of this multivariable approach to the measurement of overall voice quality. Finally, the responsiveness to change was investigated for both models. Awan and Roy (2009) found a strong $r = 0.75$ between the unstandardized change scores (i.e., post-therapy score minus pre-therapy score) of the perceived and the acoustically predicted dysphonia severity. The present study examined the standardized change scores of the perceptual ratings and the AVQI data and indicated a strong $r_s = 0.80$ (Fig. 5). Since both studies incorporated a continuum in voice quality change (i.e., from absence of change to dramatic change), it can be concluded that both models are acceptably responsive/sensitive to voice changes after voice treatment.

Based on these results, we conclude that the AVQI possesses respectable concurrent validity, concurrent cross-validity, diagnostic precision, and responsiveness to change. All of these are desirable attributes for an objective treatment outcomes measure. When one considers that the AVQI incorporates not only sustained vowels, but also continuous speech, it also possesses another desirable attribute “ecological validity”. Furthermore, because the AVQI can be computed using computer programs freely available as in Praat and SpeechTool, it has the additional appeal of easy access and affordability. Collectively, as clinicians, third party payers, and administrators demand objective evidence of positive or negative treatment outcomes, the results of this investigation confirm that the AVQI deserves further attention as a promising objective treatment outcomes measure that could be included as an important part of a multidimensional assessment of treatment effects.

Acknowledgments

The authors sincerely appreciate Dr. James Hillenbrand (from the Western Michigan University, Kalamazoo, MI, USA) for providing the SpeechTool software for the CPP and CPPs measures. The authors also thank Dr. Gwen Van Nuffelen, Dr. Bernadette Timmermans and Fons Mertens for their contributions in the perceptual rating of the many concatenated voice samples.

Appendix A. Continuing education

(1) To be ecologically valid, voice quality is ideally measured in both sustained vowels and continuous speech: (a) True, (b) False.

(2) The Acoustic Voice Quality Index consists of 6 spectrum- and cepstrum-based measures: (a) True, (b) False.

(3) Responsive to change is an important methodological issue, because it investigates the sensitivity of the AVQI to extra-experimental factors such as changes in the acoustic algorithms or changes in the data acquisition system: (a) True, (b) False.

(4) Cross-validation is also an important step in the development of the AVQI, because it investigates whether the results from the initial research can be generalized to other subject samples: (a) True, (b) False.

(5) The results of this study demonstrate that the multivariate approach of the AVQI is a reasonably valid and clinically viable method: (a) True, (b) False.

References


Halberstam, B. (2004). Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL*, 66, 70–73.


