AMERICAN JOURNAL OF OTOLARYNGOLOGY–HEAD AND NECK MEDICINE AND SURGERY 36 (2015) 647–656

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/amjoto

The improvement of internal consistency of the Acoustic Voice Quality Index

Ben Barsties [a,b,*], *Youri Maryn* [a,c,d]

[a] *Faculty of Medicine and Health Sciences, University of Antwerp, Belgium*
[b] *Medical School, Hochschule Fresenius University of Applied Sciences, Hamburg, Germany*
[c] *European Institute for ORL, Sint-Augustinus Hospital, Antwerp, Belgium*
[d] *Faculty of Education, Health & Social Work, University College Ghent, Belgium*

ARTICLE INFO

Article history:
Received 29 December 2014

ABSTRACT

Purpose: This investigation aims to explore the improvement of the relatively new hoarseness severity quantification method, called Acoustic Voice Quality Index (AVQI), which measures the concatenation of continuous speech (CS) and sustained phonation (SP) segments. Earlier investigations indicated that the proportion of the SP is more dominant in the final AVQI result than the CS.
Method: Sixty voice samples were selected with different voice pathologies and equal distribution of hoarseness severity ranged from normal to severe. Every voice sample varied in three different durations: voice duration-one (VD-1) with seventeen syllables text plus three seconds of SP, voice duration-two (VD-2) with customized length of CS plus three seconds of SP, and voice duration-three (VD-3) with a whole text plus three seconds SP. All voice samples were perceptually judged on overall voice quality by five experienced voice clinicians. AVQI's precision and concurrent validity were assessed in all three VDs. Finally, the internal consistency across all three VDs was analyzed.
Results: No significant differences were found in the perceptual evaluation of overall voice quality across all three VDs by acceptable rater reliability. The concurrent validity distinguished in all three VDs as a marked degree of correlation (i.e., ranged from r_s = 0.891 to r_s = 0.929) with no significant differences across all three VDs. The best precision was found in VD-2. Finally, the internal consistency showed in VD-2 a balanced out impact of the final AVQI score with no significant differences from both speech tasks.
Conclusion: Although AVQI currently uses the speech material of VD-1, the present study demonstrated the best results in VD-2 (i.e., precision and internal consistency). These features of VD-2 facilitate higher representativity and improve the validity of this objective diagnostic instrument.

1. Introduction

Voice quality is a feature of voice production that describes a perceptual phenomenon in the voice sound [1]. Generally, voice quality is not a clearly defined term in literature [1]. However, the overall voice quality is mostly compatible with the term hoarseness. Hoarseness is a voice symptom that perceptually deviates from normal voice quality recognized

* *Corresponding author at:* Universiteitsplein 1, 2610 WILRIJK, Antwerp, Belgium. Tel.: +32 3 265 26 39.
E-mail address: ben.barsties@t-online.de (B. Barsties).

http://dx.doi.org/10.1016/j.amjoto.2015.04.012
0196-0709/

by oneself or others [2]. Major subtypes of abnormal overall voice quality, which have received wide acceptance, are breathiness, roughness and strained [1].

Furthermore, variations in voice quality are the most frequent voice complaints in clinical practice [3]. Therefore, two broad approaches enable measuring voice quality [1]. First, a subjective method (i.e., auditory-perceptual judgment) is used to listen to a patient's voice and assign a score that reflects his/her judgment of the voice sound. Second, objective methods are used which apply specific algorithms to quantify certain aspects of a correlate of vocal production like vocal acoustic signal, the inverse-filtered oral airflow signal or its derivatives. All these tools may evaluate the presence, the degree, and the progression of abnormal voice quality in a sufficiently valid and reliable way. Traditionally, the auditory-perceptual judgment has been commonly used to determine all these three components in the evaluation of voice quality because of the simplicity and efficiency of this method. The use of auditory-perceptual judgment is not undisputed in literature. There are many factors which affect the auditory-perceptual judgment's reliability and accuracy [1]. The consideration of these factors is difficult to monitor in clinical practice and thus, objective tools may support examiners in their decision of rating voice quality. Acoustic-analysis of the voice signal is one possibility in the evaluation of voice quality and is the most used diagnostic instrument to identify voice disorders in research [4].

Recently, new methods in acoustic-analysis were used to analyze continuous speech and sustained phonation with sufficient accuracy and reliability, e.g., the Acoustic Voice Quality Index (AVQI) proposed by Maryn et al. [5]. This feature in acoustic-analysis facilitates higher ecological validity in the evaluation of voice quality.

AVQI is a six-factor acoustic model based on linear regression analysis used to measure overall voice quality in concatenated continuous speech and sustained phonation segments. In order to simplify clinical interpretation, the regression model was linearly rescaled in such a way that the outcome of the equation resulted in a score between 0 and 10 by calling this final model AVQI [4]. It is one of the first objective-acoustic models to judge continuous speech. To our knowledge the Cepstral Spectral Index of Dysphonia by Awan et al. [6,7] may also successfully evaluate overall voice quality in continuous speech and sustained phonation.

AVQI is an acoustic correlate of auditory-perceptual judgment because perceptual evaluation is considered the 'gold standard' in research and clinical practice.

The AVQI model uses a detection algorithm from Parsa and Jamieson [8] to separate voice and voiceless segments of the recording of continuous speech. This procedure allows acoustic measurements of continuous speech with many more meaningful acoustic markers based on the frequency-, time-, and amplitude domains in the evaluation of overall voice quality as shown in the meta-analysis from Maryn et al. [9].

Although AVQI was originally developed for Dutch speakers, this model has been validated and found reliable in different languages including native German [10], native English in pediatric population [11], and multilingual persons speaking Dutch, English, German and French [12]. Based on the results of these studies, AVQI seems to be cross-linguistically robust in Germanic languages. The performance of AVQI is relatively insulated from inter-language phonetic differences [12].

The relevance of acoustic measurements in clinical management by rating overall voice quality is to objectively monitor voice quality through the voice therapy process. Therefore, AVQI has thus far proven highly sensitive in voice changes through voice therapy (i.e., r = 0.80) [13].

Recent research about internal consistency and test–retest measurement of AVQI has shown a low level of AVQI score variability (i.e., an AVQI score = 0.54), but AVQI has been most strongly influenced by sustained phonation [14]. Furthermore, sustained phonation has revealed to have a significantly greater influence on AVQI score than continuous speech [14]. These findings suggest that more research is required for more representativity and ecological validity in AVQI to balance out the internal consistency through equal proportion of these two speech tasks.

This investigation aims to explore the equal proportion of the two speech tasks in AVQI by expanding the duration of continuous speech. Although the part of continuous speech in the current model of AVQI covers 17 to 22 syllables [12], the duration analysis material of continuous speech is significantly lower, after separating voice to voiceless segments of the detection algorithm, than the constant three seconds on sustained phonation [14].

Furthermore, the judgment validity of overall voice quality was verified anew between auditory-perceptual rating and AVQI scores by different durations of the analyzed segments.

Our research questions address the following:

1. Are there significant differences between different auditory-perceptual overall voice quality ratings (i.e., the judgment of concatenated continuous speech and sustained phonation) by varying the duration of continuous speech?
2. What is the impact of varying the voiced continuous speech duration on the correlation and perceptual diagnostic accuracy between perceptual ratings and AVQI values?
3. Does the internal consistency of AVQI improve when the proportions of sustained phonation and voiced continuous speech are adapted to reach higher ecological validity?

2. Methods

2.1. *Subjects*

The voice-disordered subjects were recruited retrospectively from the ENT caseload of the Sint-Jan General Hospital in Bruges, Belgium. Concatenated voice samples of continuous speech and sustained phonation were obtained from a database of 350 patients with various organic and non-organic etiologies [5,15]. These voice samples were chosen by selecting four groups with various degrees of hoarseness (i.e., absence/clear voice, slight, moderate, and severe). Firstly, the selection was based upon prior modal agreement across