Validation of the Acoustic Voice Quality Index in the Lithuanian Language

***Virgilijus Uloza, *Tadas Petrauskas, *Evaldas Padervinskis, *Nora Ulozaitė, †·‡Ben Barsties, and †·§·‖Youri Maryn,** **Kaunas, Lithuania, †§Antwerp and ‖Ghent, Belgium, and ‡Hamburg, Germany*

Summary: Objectives. The aim of the present study was to validate the Acoustic Voice Quality Index in Lithuanian language (AVQI-LT) and investigate the feasibility and robustness of its diagnostic accuracy, differentiating normal and dysphonic voice.

Methods. A total of 184 native Lithuanian subjects with normal voices (n = 46) and with various voice disorders (n = 138) were asked to read aloud the Lithuanian text and to sustain the vowel /a/. A sentence with 13 syllables and a 3-second midvowel portion of the sustained vowel were edited. Both speech tasks were concatenated, and perceptually rated for dysphonia severity by five voice clinicians. They rated the Grade (G) from the Grade Roughness Breathiness Asthenia Strain (GRBAS) protocol and the overall severity from the Consensus Auditory-perceptual Evaluation of Voice protocol with a visual analog scale (VAS). The average scores (G_{mean} and VAS_{mean}) were taken as the perceptual dysphonia severity level for every voice sample. All concatenated voice samples were acoustically analyzed to receive an AVQI-LT score.

Results. Both auditory-perceptual judgment procedures showed sufficient strength of agreement between five raters. The results achieved significant and marked concurrent validity between both auditory-perceptual judgment procedures and AVQI-LT. The diagnostic accuracy of AVQI-LT showed for both auditory-perceptual judgment procedures comparable results with two different AVQI-LT thresholds. The AVQI-LT threshold of 2.97 for the G_{mean} rating obtained reasonable sensitivity = 0.838 and excellent specificity = 0.937. For the VAS rating, an AVQI-LT threshold of 3.48 was determined with sensitivity = 0.840 and specificity = 0.922.

Conclusions. The AVQI-LT is considered a valid and reliable tool for assessing the dysphonia severity level in Lithuanian-speaking population.

Key Words: Dysphonia–Acoustic Voice Quality Index–Acoustic voice analysis–Lithuanian language–Voice assessment.

INTRODUCTION

Voice disorders are relatively common, affecting 6–9% of the general population.[1,2] Dysphonia related to laryngeal dysfunction may be caused by functional, behavioral, neurologic, benign, or malignant factors.[3] Correct diagnosis of the laryngeal/voice disorder is an essential step toward its appropriate treatment and cost control. Clinical diagnostics of laryngeal/voice disorders is based on multidimensional approach including perception of voice quality, acoustic voice analysis, visualization of the larynx (videolaryngostroboscopy [VLS] and direct microlaryngoscopy), measurement of voice aerodynamics, and subjective rating of voice by the patient.[4] Acoustic voice analysis helps in the evaluation of voice quality, which is an important aspect in laryngeal diseases, because majority of these patients have an abnormal voice quality manifesting in various dysphonia severity types and levels.[5]

Automated analysis of voice signals is the most used diagnostic instrument to identify voice disorders in clinical practice and research. Acoustic markers computed from voice signal recordings are a convenient way for collection of objective noninvasive voice data, quantifying and documenting dysphonia severity and providing measurable evidence of the functional outcomes of therapeutic and/or phonosurgical treatment of voice problems.[6–10]

Traditionally, acoustic methods are used to yield objective data on only sustained vowels.[11] In most of earlier studies, sustained vowel /a/ has been chosen for analysis because the steady-state phonations are relatively simple to process and enable time effective analysis. Therefore, this simple acoustic structure manifests and provides reliable detection and computation of acoustic features at the lowest equal error rate in laryngeal pathology detection.[7,12,13] Furthermore, sustained vowels typically are not influenced by speech rate, and word or sentence stress, and do not contain fast voice onsets and terminations, voiceless phonemes, and prosodic perturbations in fundamental frequency and amplitude.[14] Finally, sustained vowel phonation is rather insulated from aspects related to different languages. Therefore, sustained vowels are generally used as universal and suitable for voice diagnostics and otolaryngology clinical routine.

On the other hand, acoustic voice analysis using vowel phonation and based on one acoustic parameter (eg, jitter, shimmer, etc) revealed in different studies quite poor reliability and correlation to auditory-perceptual judgment.[8,15] However, multivariate approaches in voice assessment models showed higher correlation with auditory-perceptual judgment and higher reliability and validity in voice pathology detection.[16–19]

Furthermore, in contrast to sustained phonation, analysis of connected speech presents a great interest for research and clinical practice because symptoms of disordered voice quality are

Accepted for publication June 8, 2016.

From the *Department of Otolaryngology, Lithuanian University of Health Sciences, Kaunas, Lithuania; †Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; ‡Medical School, Hochschule Fresenius University of Applied Sciences, Hamburg, Germany; §European Institute for ORL-HNS, Sint-Augustinus Hospital, Antwerp, Belgium; and the ‖Faculty of Education, Health and Social Work, University College Ghent, Ghent, Belgium.

Address correspondence and reprint requests to Nora Ulozaitė, Department of Otolaryngology, Lithuanian University of Health Sciences, Kaunas, Lithuania. E-mail: nora.ulozaite@gmail.com

Journal of Voice, Vol. 31, No. 2, pp. 257.e1–257.e11
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.06.002

more typically revealed in continuous speech.[20] Moreover, the sustained vowel phonation is an artificial type of phonation, and that feature determines the major limitation of the sustained vowel signal analysis. On the other hand, connected speech is highly representative of daily voicing and can be considered more "ecologically valid" (ie, more representative of daily speech and voice use patterns). Therefore, if an objective voice assessment is to be considered robust and ecologically valid, then the acoustic measure ideally should be calculated using voice signals based on recordings of both speaking patterns—sustained phonation and running speech.[19,21] Currently, only two multiparametric models successfully evaluate voice quality on these two speech types: the Acoustic Voice Quality Index (AVQI) proposed by Maryn et al,[21] and Cepstral Spectral Index of Dysphonia published by Awan et al.[22] In further investigations, both models performed highly comparable and demonstrated acceptable accuracy and reliability in detecting voice quality abnormality.[23]

AVQI's stability and validity across different linguistic and phonetic structures have been approved for the following western European languages: Dutch,[11,21,23] German,[23,24] English (Germanic group),[23,25] and French (Roman group).[23] The results of all these studies revealed that the performance of AVQI is relatively insulated from interlanguage phonetic differences. Additionally, a preliminary study of AVQI in the Finnish-speaking population also showed favorable results.[26] A further study with Korean language (ie, an Altaic language) supported the cross-linguistic robustness of AVQI as a valid and objective marker of overall dysphonia severity across phonetical differences.[27]

The remarkable results on AVQI's performance from the studies mentioned above confirm that AVQI is a promising objective and quantitative measurement of the dysphonia severity level in several languages. Thus, new perspectives and widening across linguistic condition and generalization of this method of voice assessment are allowed. Nevertheless, continuous speech can vary substantially in its phonetic and phonatory composition depending on the differences of language groups or even dialects. Therefore, vocal physiology and phonatory output can presumably be different across languages influencing vocal behavior, and possibly the quality of voice.[23] Consequently, the demand for further research and validation of AVQI for other languages still exists.

Lithuanian language is one of the two living Baltic languages. Lithuanian inherited the old Indo-European word inflexion and phonological system, considering some marked differences from other European languages.[28] Therefore, introduction of a new voice assessment method, which employs continuous speech segments, ie, the AVQI in the Lithuanian language, requires an appropriate validation procedure.

The aim of the present study was to validate AVQI in the Lithuanian language and investigate the feasibility and robustness of its diagnostic accuracy, differentiating between normophonic and dysphonic voices.

MATERIALS AND METHODS

This study was approved both by Kaunas Regional Ethics Committee for Biomedical Research (No. P2-24/2013) and by Lithuanian State Data Protection Inspectorate for Working with Personal Patient Data (No. 2R-648 [2.6-1]).

Voice sample consisted of 184 individuals examined at the Department of Otolaryngology of the Lithuanian University of Health Sciences, Kaunas, Lithuania. The mean age of the study group was 44.1 (standard deviation [SD] = 17.5) years.

The normal voice subgroup was composed of 46 volunteers without vocal complaints and no history of chronic laryngeal diseases or other long-lasting voice disorders. All of them were free from any known hearing problems and free from common cold or upper respiratory infections at the time of voice recording. These voice samples were also evaluated as healthy voices by clinical voice specialists. No pathological alterations in the larynx of the healthy subjects were found during VLS. Digital high quality VLS recordings were performed with a XION Endo-STROB DX device (XION GmbH, Berlin, Germany) using a 70° rigid endoscope.

The pathological voice subgroup consisted of 138 patients who represented a rather common and clinically discriminative group of laryngeal diseases and voice disturbances, ie, mass lesions of the vocal folds, paralysis, reflux laryngitis, Parkinson disease, functional dysphonia, presbylaryngis, and Huntington chorea. In the present study, the included mass lesions of the vocal folds were nodules, polyps, cysts, Reinke's hyperplasia, papillomata, chronic hyperplastic laryngitis, keratosis, and carcinoma. The clinical diagnosis was based on clinical evaluation during VLS and direct microlaryngoscopy. All the patients with mass lesions of the vocal folds underwent endolaryngeal microsurgical interventions. The voice recordings were made before the surgical intervention. The final diagnosis was proven by the results of histological examination of the removed tissue.

Demographic data of the total study group and diagnoses of the pathological voice subgroup are presented in Table 1. All selected patients were serially enrolled (except those with Parkinson disease and Huntington chorea). Therefore, the real incidence of voice pathologies in our series might be clinically representative of the population of patients with voice disorders, reflecting different ages and gender groups, and different types and severity degrees of voice quality. Because the aim of the present study was to validate AVQI in the Lithuanian language and investigate the feasibility and robustness of its diagnostic accuracy differentiating between normophonic and dysphonic voices, some patients with neurological disorders, which manifest with abnormal voice characteristics, were included in the studied group.

Voice recordings

The mixed gender database of voice recordings in this study contained 184 digital voice recordings of sustained phonation of the vowel sound /a/ (as in the English word "large") and the first Lithuanian sentence "Turėjo senelė žilą oželį" (The grandmother had a little gray goat) containing 13 syllables from the phonetically balanced text "The Wolf and Little Goats." Both of them were recorded with comfortable pitch and loudness. According to the International Phonetic Alphabet, this Lithuanian sentence is pronounced as [tʊreːjɔ sɛnɛːle ʒɪlaː ɔʒɛːliː].