European Archives of Oto-Rhino-Laryngology (2019) 276:1737–1745
https://doi.org/10.1007/s00405-019-05433-5

LARYNGOLOGY

Exploring the feasibility of the combination of acoustic voice quality index and glottal function index for voice pathology screening

Nora Ulozaite-Staniene[1] · Tadas Petrauskas[1] · Viktoras Šaferis[2] · Virgilijus Uloza[1]

Received: 10 December 2018 / Accepted: 12 April 2019 / Published online: 23 April 2019

Abstract

Purpose The aim of this study was to explore the diagnostic value of the combination of Acoustic Voice Quality Index (AVQI) and Glottal Function Index (GFI) as a screening tool for voice disorders, and to compare the AVQI measurements obtained using oral and smartphone (SP) microphones.

Methods A study group consisted of 183 adult individuals including 86 subjects with normal voice and 97 patients with pathological voice. Voice recordings were performed simultaneously with an oral AKG Perception 220 and SP iPhone 6s microphones. To evaluate the diagnostic accuracy differentiating normal and pathological voice, the receiver-operating characteristic statistics [area under curve (AUC), positive and negative likelihood ratios (LR+ and LR−)], and correct classification rate (CCR) were used.

Results The AVQI cut-off scores of 3.31 for oral and 3.32 for SP microphones were associated with very good test accuracy (AUC = 0.857 and AUC = 0.818), resulting in balance between sensitivity and specificity (70.0% vs 86.0% and 70% vs 87.0%). The CCR reached 78%. The combined AVQI and GFI cut-off scores of 6.65 for oral and 7.1 for SP microphones corresponded with excellent test accuracy (AUC = 0.976 and AUC = 0.965) and sensitivity and specificity (93.0% vs 93.0% and 91.0% vs 94%). Very respectable levels of LR+ and LR− both for oral microphone (13.3 and 0.08) and for SP microphone (15.6 and 0.10) voice recordings were achieved. CCRs of 93% and 92% confirmed the results of ROC statistics.

Conclusions Combination of AVQI and GFI measurements significantly improved diagnostic accuracy in differentiating normal vs pathological voice.

Keywords AVQI · GFI · Automated voice screening · Voice analysis

Introduction

Diagnostics of laryngeal/voice disorders is based on a multidimensional approach including perception of voice changes, visualization of the larynx [video laryngostroboscopy (VLS)], objective measurement of voice quality, measurement of voice aerodynamics, and subjective rating of voice by both the patient and physician. Timely, qualified, and complex otolaryngology examination involving measurements mentioned above may increase diagnostic and treatment effectiveness and reduce health care expenditures in the evaluation and management of patients with diverse laryngeal/voice disorders including laryngeal carcinoma.

Ensuring an early visit to an otolaryngologist, screening healthy or at-risk populations for early signs of a voice abnormality, seems to be useful and promising, because many laryngeal/voice disorders may be most successfully treated if detected early. In general, the aim of a screening

✉ Nora Ulozaite-Staniene
nora.ulozaite@lsmuni.lt

Tadas Petrauskas
tadas@petrauskas.co.uk

Viktoras Šaferis
viktoras.saferis@lsmuni.lt

Virgilijus Uloza
virgilijus.ulozas@kaunoklinikos.lt

[1] Department of Otorhinolaryngology, Lithuanian University of Health Sciences, Eiveniu 2, LT-5009 Kaunas, Lithuania

[2] Department of Physics, Mathematics and Biophysics, Lithuanian University of Health Sciences, Eiveniu 2, LT-5009 Kaunas, Lithuania

tool can be defined as providing identification of individuals who should undergo further diagnostic procedures to more precisely define the presence and nature of the disorder [1].

An automated acoustic analysis-based voice-screening tool could serve as one of the potential approaches to help primary care physicians and other public health care services identify the patients who require otolaryngological referral, thereby improving diagnostics and management of laryngeal/voice disorder patients. The main goal of automated pathological voice detection systems is to categorize any input voice as either normal or pathological [2].

Multiparametric models for evaluation of voice quality have been shown to be more suitable than single measures, because use of multiple acoustic parameters in voice assessment models demonstrated higher correlations to auditory-perceptual judgment and higher reliability and validity in voice pathology detection [3, 4]. Multiparametric acoustic measures present an acceptable capacity to discriminate between the presence and the absence of laryngeal alteration and to differentiate several laryngeal diagnoses [5].

Nowadays, two multiparametric models based on sustained vowels and on continuous speech have been used for evaluation of voice quality: the Cepstral Spectral Index of Dysphonia (CSID) introduced by Awan et al. in 2009 and the Acoustic Voice Quality Index (AVQI) proposed by Maryn et al in 2010 [6, 7].

The value of AVQI as a clinical marker of overall voice quality has already been proven across multiple studies revealing adequate diagnostic accuracy, high and consistent concurrent validity and test–retest variability, high sensitivity to voice quality changes through voice therapy, and robust interlanguage validity and insulation from interlanguage phonetic differences [8–15]. Moreover, a recent study showed that the AVQI values do not depend on gender and age, thus widening the perspectives for further generalization of this objective, and quantitative measurement for dysphonia severity level and possible voice-screening purposes [16].

The recent studies demonstrated that CSID can provide a valid estimate of dysphonia severity and a strong level of accuracy for the classification of voice-disordered cases versus controls, particularly when auditory-perceptual judgment is used as the reference standard. Thus use of CSID measures represents a potentially significant advance in voice disorder identification [1, 17].

Another and additional source of information suitable for screening purposes could be special voice-related questionnaires, because questionnaire data may contain important information that is significant for voice classification into normal/pathological classes. The results of solitary studies indicated that voice-related questionnaire data can also be used for voice-screening purposes; however, only a few attempts in this field have been made [18, 19].

Several patient-reported questionnaires have been proposed to assess the impact of voice disorders on patient's quality of life, and to evaluate treatment outcomes—Voice Handicap Index (VHI), Voice-Related Quality of Life (V-RQOL), Voice Outcome Survey, Voice Activity and Participation Profile, and Voice Symptom Scale [20–24].

In 2005, Bach et al. developed and validated the 4-item Glottal Function Index (GFI) questionnaire [25]. This brief, easily understood and self-administered, reproducible and reliable symptom-based battery was designed to assess the presence and degree of vocal dysfunction in adults. Cohen et al. in 2007 presented results of the validation of the Hebrew version of the GFI and proved the utility of this tool to improve the evaluation of children with voice disorders [26]. The cut-off score of >3.0 points distinguishing patients and healthy controls with sensitivity of 70% and specificity of 72% was revealed. A culturally adapted and validated Lithuanian version of GFI (GFI-LT) showed a statistically significant high reliability and internal consistency (Cronbach $\alpha=0.8$, $r=0.50$), and moderate item-total correlation ($r=0.41$–0.55). The cut-off score of above 3.0 points in the adult population demonstrated a high diagnostic sensitivity and specificity (88% and 84%, respectively) and was also confirmed to be a valid and reliable tool for self-assessment of the severity of voice disorders in Lithuanian-speaking patients [27].

The questionnaires most widely used in the assessment of voice mainly reflect voice impact on quality of life and the impact of the disorder in everyday and professional activities. They are also useful in the evaluation of treatment outcomes. Despite their high value in research and clinical voice assessment, questionnaires have not been elaborated specifically for screening purposes, i.e., to identify individuals who are at risk of having voice or laryngeal related problem, even in its early stages [28].

This voice questionnaire-based screening demand was partially decreased by the development of a few specialized voice-screening questionnaires. In 2005, van Gogh et al. designed a 5-item screening questionnaire for voice problems after Treatment of Early Glottic Cancer [29]. This questionnaire proved to be a reliable, valid, and feasible method to detect voice impairment in daily life and was useful for both radiation oncologists and otorhinolaryngologists in their follow-up of patients treated for early glottic cancer. Later, Ghirardi et al. developed a Screening Index for Voice Disorder questionnaire comprising 12 voice-related symptoms and validated it as an effective tool providing high levels of sensitivity for screening teachers for voice disorder [28]. On the other hand, Morawska et al. in 2018 found that the V-RQOL may prove useful in screening procedures for occupational voice disorders and, therefore, may be used as a primary source of referral in clinical practice, particularly in occupational medicine [30].

Several later investigations revealed acceptable the screening potential of GFI: in the task of distinguishing between the normal and dysphonic voice classes, the GFI data even outperformed the acoustic voice data when using Random Forest Classifier. Furthermore, fusing audio data with responses to GFI items consistently improved detection further [31, 32] However, the discriminatory power of questionnaire data and questionnaire features are still seldom used for screening of laryngeal disorders.

The possibility of the combination of AVQI and GFI data has not been investigated before. Therefore, the aim of this study was to explore the diagnostic value of the combination of AVQI and GFI as a screening tool for voice disorders, and to compare the AVQI measurements obtained using oral and smartphone microphones.

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).

183 adult individuals were included in the study, 94 men and 89 women. The mean age of the study group was 43.03 (SD = 15.38) years. They were all examined at the Department of Otolaryngology of the Lithuanian University of Health Sciences, Kaunas, Lithuania.

The normal voice subgroup (noncase group) consisted of 86 selected healthy volunteer individuals, 43 men and 43 women; mean age 34.73 (SD = 11.83) years. This subgroup was collected following three criteria to define a vocally healthy subject: (a) all selected subjects considered their voice as normal and had no actual voice complaints, no history of chronic laryngeal diseases or voice disorders; (b) no pathological alterations in the larynx of the healthy subjects were found during video laryngostroboscopy (VLS) using an XION Endo-STROB DX device (XION GmbH, Berlin, Germany) 70° rigid endoscope (this subgroup of subjects was considered as laryngoscopic-negative); and (c) all these voice samples were evaluated as normal voices by clinical voice specialists.

The pathological voice subgroup (case group) consisted of 97 patients, 50 men and 47 women; mean age 50.46 years (SD = 14.36). They presented with a large variety of laryngeal diseases and voice disturbances, i.e., mass lesions of the vocal folds, unilateral and bilateral vocal fold paralysis, reflux laryngitis, sulcus glottidis, and functional dysphonia. The clinical diagnosis was based on clinical examination (complaints, history) and results of VLS and/or direct microlaryngoscopy. Laryngoscopic findings included a wide range of signs representing vocal fold hypertrophy/atrophy, paresis/paralysis, tremor, and benign and malignant mass lesions of vocal folds. Thus, this subgroup of subjects was considered as laryngoscopic positive. All the patients with mass lesions of the vocal folds underwent endolaryngeal microsurgical interventions. In those cases, the final diagnosis was proven by the results of histological examination of the removed tissue.

The clinical diagnosis served as the criterion to sort individuals into pathological voice or normal voice groups and was considered as the "gold" standard to evaluate the diagnostic accuracy of the combination of AVQI and GFI, discriminating normal voice (noncase) and pathological voice subjects (case). Demographic data of the study group and diagnoses of the pathological voice subgroup are presented in Table 1. All selected patients were recruited consecutively during the time period of 2014–2017 years. Therefore, the incidence of voice pathologies in our series is considered to be clinically representative for the population of voice-disordered patients reflecting different age and gender groups.

Voice recordings

Voice samples from each subject were recorded in a T-series silent room for hearing testing (T-room, CATegnerAB, Bromma, Sweden) simultaneously using two different microphones, a cardioid AKG Perception 220 microphone (AKG Acoustics, Vienna, Austria) and an internal (bottom) microphone of smartphone (SP) iPhone 6s. Both microphones were placed alongside at a 10.0 cm distance from the mouth, keeping at about 90° microphone-to-mouth angle. Each participant was required to complete two vocal tasks which were digitally recorded. The tasks consisted of (a) sustaining

Table 1 Demographic data of the study group

Diagnosis	Total	Age (years)	
		Mean	SD
Normal voice	86	34.73	11.83
Vocal fold noduli	5	28.2	8.76
Vocal fold polyp	22	42.09	12.56
Vocal fold cyst	1	56	–
Vocal fold carcinoma	14	68.07	6.91
Vocal fold polypoid hyperplasia	11	51.55	7.62
Vocal fold keratosis	5	48.2	12.51
Vocal fold papilloma	8	47.5	12.3
Unilateral vocal fold paralysis	5	49	12.88
Bilateral vocal fold paralysis	7	54.43	17.11
Chronic hyperplastic laryngitis	14	52.46	8.67
Vestibular cyst	1	62	–
Sulcus glottidis	4	48.33	10.97
Grand total	183	43.03	15.38

SD standard deviation

phonation of the vowel sound [a:] for at least 5 s duration and (b) reading a phonetically balanced text segment in Lithuanian "Turėjo senelė žilą oželį" (the grandmother had a little grey goat). The participants completed both vocal tasks at personally comfortable loudness and pitch level. Only one sample of the vowel [a] and one sample of the text were used for acoustic voice analysis.

All voice recordings were captured at a sampling frequency of 44.1 kHz and exported in 16 bit depth lossless "wav" audio file format.

Acoustic voice analysis

For AVQI calculations, the signal processing of the voice samples was done in the Praat software (version 5.3.57; https://www.fon.hum.uva.nl/praat/). Only voiced parts for the continuous speech were manually extracted and concatenated to the medial 3 s of sustained [a] phonation. This chain of signals was used for acoustic analysis with the AVQI script version 02.02 developed for the program Praat (https://www.vvl.be/documenten-en-paginas/praat-script-avqi-v0202). Six acoustic measures (i.e., the smoothed cepstral peak prominence, harmonics-to-noise ratio, shimmer local, shimmer local dB, slope of the long-term average spectrum, and tilt of the trendline through the average long-term average spectrum) were used in multiple regression equation of AVQI [33].

Glottal function index questionnaire

Each participant of the study (normal and pathological voice subgroups) filled in the GFI_LT questionnaire at the baseline along with voice recordings. Based on the results of our previous study, the GFI-LT scores higher than 3.0 were considered to be the limiting value distinguishing normal and pathological voice subgroups [27]. In this study, 84 individuals (86.6%) from the pathological voice subgroup resulted with a GFI score higher than 3.0. The healthy voice subgroup subjects filled in the GFI questionnaires with the score 3.0 and lower in 91.9% (79 individuals).

Statistical analysis

All statistical analyses were completed using IBM SPSS Statistics 25.0 (IBM Corp., Armonk, NY). First, to evaluate the level of agreement between AVQI measurements obtained from recordings done with oral and SP microphones, Bland–Altman plots were used [34]. The Bland–Altman plot represents a scatterplot of two variables: the average between the two measurements on the horizontal axis and the difference between the two measurements on the vertical axis. Thus, the plot shows the amount of disagreement between the two measures (via differences) and indicates how this disagreement relates to the magnitude of the measurements [35].

Second, the intraclass correlation coefficient (ICC) was employed to measure reliability of two different voice recording options—oral microphone and SP microphone. The ICC assesses the reliability of measurement tools by comparing the variability of different measures of the same subject with the total variation across all measures and all subjects [36]. One-way random-effects ANOVA model was used. The ICC ranges from 0.00 (i.e., total absence of reliability) to 1.00 (i.e., perfect reliability; to ensure reasonable reliability in ICC score, the value should exceed 0.90) [36].

Third**,** the receiver-operating characteristics (ROC) and several estimates of ROC, namely sensitivity and specificity, the "area under ROC curve" (i.e., AUC) were used to calculate the accuracy of AVQI in discriminating between normal and pathological voices. Thus, AUC was used to estimate the accuracy of AVQI in discriminating between normal and pathological voices. In general, AUC of 0.5–0.6 is not useful, 0.7–0.8 is considered good, 0.8–0.9 very good and >0.9 excellent. The applicability of the best cut-off score for a clinical decision was assessed by the balance between the "likelihood ratio for a positive result" (LR+) and the "likelihood ratio for a negative result" (LR−). As a general guideline, the diagnostic value of this measure is considered to be high when $LR+ \geq 10$ and $LR- \leq 0.1$ [37]. To summarize accuracy measures and feasibility of AVQI using different microphones, correct classification rate (CCR) was calculated.

Fourth, to evaluate diagnostic value of combination of AVQI and GFI measurements discriminating normal and pathological voices, the results of both tests were summated, and ROCs' analysis was performed.

Fifth, the feasibility of AVQI using different microphones and combining AVQI and GFI, and the Correct Classification Rate (CCR) were calculated.

Sixth, to test the statistical significances of the difference between the AUCs of dependent ROC curves, the method of DeLong et al. was used [38].

Results

Agreement between AVQI measurement results obtained with different microphones

Bland–Altman analysis available in Fig. 1 showed no systemic difference in AVQI measurements between voice recordings made with oral and SP microphones, although some divergences were revealed only in the pathological voice group. Besides these outliers, remaining differences were in the smaller range and the mean of differences was on zero for both groups of subjects. Moreover, ICC of 0.903

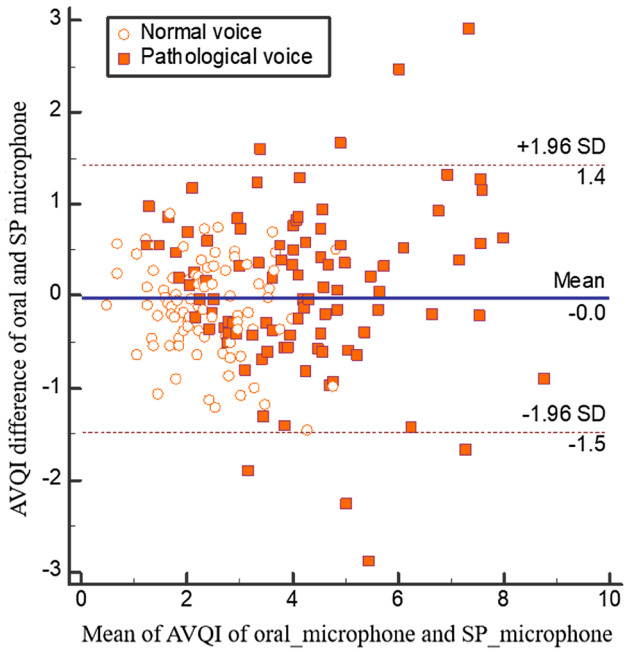

Fig. 1 Bland–Altman plot for measurements of AVQI. The horizontal solid line denotes the mean of differences and the horizontal dashed lines correspond to ± 2 standard deviations from the mean of differences. Dark color and squares correspond to pathological voice subjects, and bright color and circles correspond to normal voice subjects

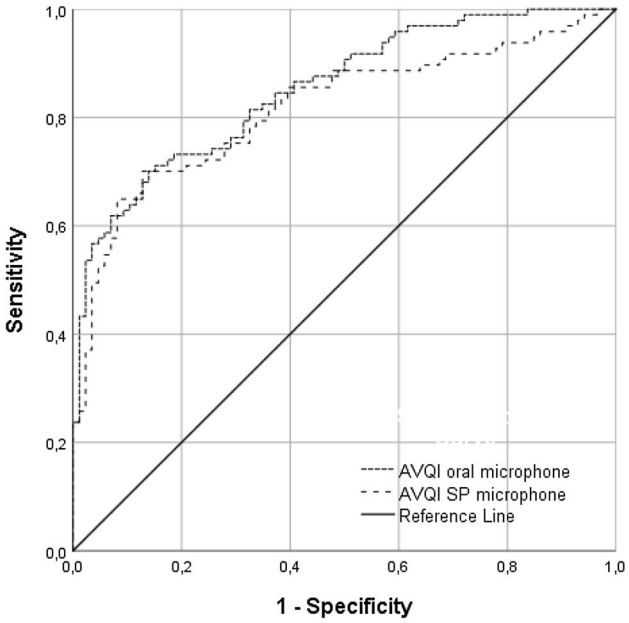

Fig. 2 Receiver-operating characteristics (ROC) curves illustrating the diagnostic accuracy of AVQI obtained from oral and smart phone microphones

revealed almost perfect agreement and reliability between AVQIs obtained from two voice recording tools (oral and SP microphones).

Diagnostic accuracy

ROC analysis was applied to estimate diagnostic accuracy of AVQI in differentiating between normal and pathological voice. (Fig. 2). The ROC curves for oral and SP voice recordings were inspected visually to identify optimum cut-off scores. Both ROC curves occupied the largest part of the graph, clearly revealing respectable power to discriminate between normal and pathological voices.

The cut-off scores of AVQI related to discrimination between normal and pathological voice groups were calculated with the best balance between sensitivity and specificity and resulted with 3.31 for oral microphone and 3.32 for SP microphone. The diagnostic accuracy was marked as very good in both cases, with the AUC = 0.857 for oral microphone and AUC = 0.818 for SP microphone. Statistics related to AVQI potential to accurately differentiate between normal and pathological voice qualities are presented in Table 2.

The AVQI cut-off score (i.e., AVQI = 3.31 pertaining to oral microphone voice recording revealed reasonable sensitivity of 70% and specificity of 86%. For SP voice recordings, the almost identical AVQI cut-off score of 3.32 was determined with also reasonable sensitivity of 70% and specificity of 87%. However, the LR statistics did not reach the recommended levels (LR + ≥ 10 and LR − ≤ 0.1) both for oral (5.02 and 0.35) microphone and (5.48 and 0.34) for SP microphone voice recordings, respectively. Furthermore, the identical CCR of AVQI indicated 78% diagnostic accuracy between normal and pathological voice groups.

To summarize, across the AUC, sensitivity, specificity, LR+ and LR− statistics, and CCR, the data demonstrated almost identical results of AVQI performance between two voice recording instruments, i.e., oral and SP microphones.

Combining AVQI and GFI

The ROC analysis was applied to estimate the diagnostic accuracy of the combination of AVQI and GFI in discriminating between normal and pathological voices. A new cut-off score of combined AVQI and GFI reached 6.65 for oral microphone with the best balance between sensitivity of 93% and specificity of 93% and 7.1 for SP microphone with sensitivity of 91% and specificity of 94%, respectively. (Fig. 3).

Thus, combining these two measurements together significantly improved diagnostic accuracy in discriminating normal and pathological voice. Consequently, the mean AUC increased to excellent with AUC = 0.976 for oral microphone and AUC = 0.965 for SP microphone. (Fig. 4).

Moreover, data complied with acceptable results above the threshold criteria of LR + ≥ 10 for both types

Table 2 Statistics illustrating AVQI ability to differentiate normal and pathological voice

Microphone	Parameter	Cut-off score	AUC	Sensitivity (%)	Specificity (%)	LR+	LR−	CCR (%)
Oral	AVQI	3.31	0.857	70	86	5.02	0.35	78
Smartphone	AVQI	3.32	0.818	70	87	5.48	0.34	78

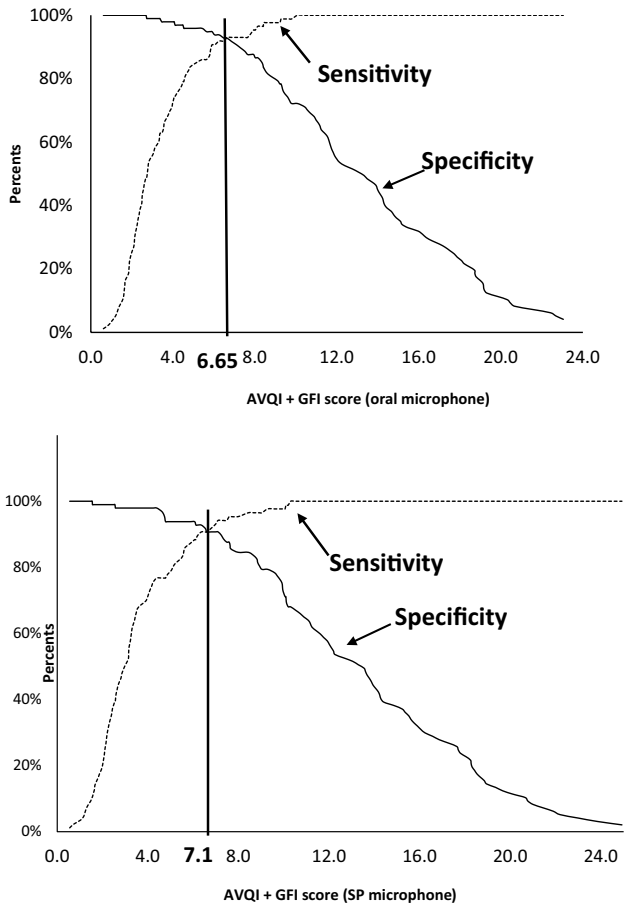

Fig. 3 Sensitivity versus specificity chart for the classification of normal versus pathological voices. Above —combined AVQI and GFI cut-off score of oral microphone voice recordings; below—combined AVQI and GFI cut-off score of smartphone microphone voice recordings

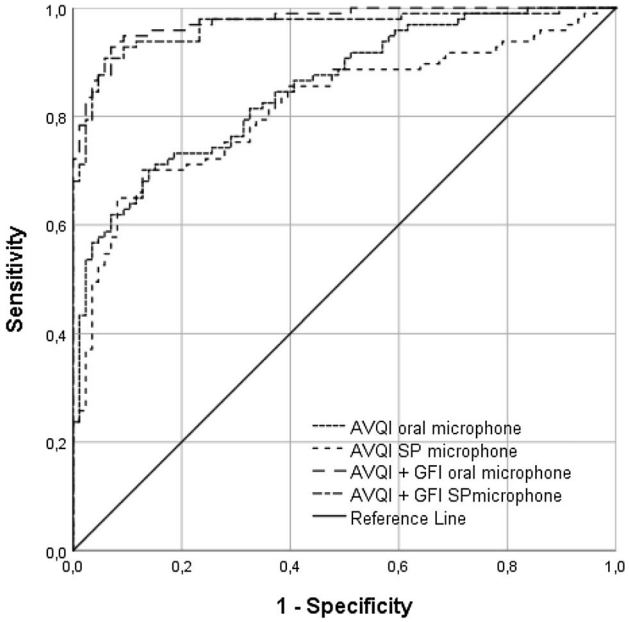

Fig. 4 Receiver-operating characteristic (ROC) curves illustrating the diagnostic accuracy of combination of AVQI and GFI obtained from oral and smart phone microphones and AVQI alone

of recording and the AVQI + GFI ability to differentiate between normal and abnormal voice quality. In addition, for LR, a high diagnostic value of the measure was reached in both cases with LR− ≤ 0.1, as proposed in general guidelines [37]. CCR of this combined method increased to 92% and 93%, respectively. (Table 3).

Finally, a comparison of the AUCs-dependent ROC curves (i.e., single AVQI measurements against combined AVQI + GFI measurements) according to the test of DeLong et al., confirmed a statistically significant increase ($p < 0.001$) of the AUCs of combined measurements [38]. To summarize, the combination of AVQI and GFI measurements significantly improved diagnostic accuracy in differentiating normal vs. pathological voice.

Discussion

The purpose of this study was to explore the diagnostic value of the AVQI and the combination of AVQI and GFI as a screening tool for voice disorders. Analysis of the results revealed that the AVQI showed a strong ability to discriminate between normal voice vs. pathological voice individuals as determined via clinical diagnosis of laryngeal disorder.

The AVQI cut-off scores of 3.31 for oral microphone and 3.32 for SP microphone were associated with very good test accuracy (AUC = 0.857 and AUC = 0.818, respectively), resulting in a balance between sensitivity and specificity (70.0% vs 86.0% and 70% vs 87.0%, respectively). However, the recommended levels of positive and negative likelihood ratios both for oral microphone (5.02 and 0.35) and for SP microphone (5.48 and 0.34) voice recordings were not achieved. The CCR reached 78% for both types of voice recordings.

Table 3 Statistics illustrating ability of combined AVQI and GFI score to differentiate normal and pathological voice

Microphone	Parameters	Cut-off score	AUC	Sensitivity (%)	Specificity (%)	LR+	LR−	CCR (%)
Oral	AVQI + GFI	6.65	0.976	93	93	13.3	0.08	93
Smartphone	AVQI + GFI	7.10	0.965	91	94	15.6	0.10	92

It is obvious that acoustic measures (particularly those that have been validated using auditory-perceptual ratings) are most appropriate for use when disordered versus nondisordered groups are classified on the basis of auditory-perceptual judgments [1]. On the other hand, from clinical practice it is well known that subjects with laryngoscopically atypical signs may also produce perceptually ''normal'' voice and vice versa. This may be determined by the phenomenon that the presence of a mass lesion or other structural variation of vocal folds does not necessarily lead to perceived or acoustically measured dysphonia, especially if the location of the lesion does not have a substantial effect on the vibratory characteristics of the vocal folds [1]. For instance, some conditions such as edema, erythema, and even small vocal fold lesions such as vocal nodules and/or polyps located on the upper surface of the vocal fold may have insignificant effect on the periodicity of vocal fold vibration. On the other hand, the clinically severely disordered voices (for example, Parkinson disease) typically are associated with structurally normal larynges. Therefore, it is possible that acoustic measures that have been validated as estimates of perceived dysphonia severity may not be strong classifiers of subjects with or without laryngoscopic abnormality.

The AVQI model, based on acoustic voice parameters, with smoothed cepstral peak prominence and shimmer as part of the components, is primarily related to periodicity prominence and, therefore, presumed to solely measure vocal sound quality [8]. Therefore, AVQI has been proved as the most effective in differentiating normal versus pathological voice based on perceptual assessment. Despite this fact, the results of our previous study and data of this investigation demonstrated that AVQI is also highly effective in differentiating normal and pathological voice based on the "gold standard", i.e., voice disorder diagnosis [39].

It is evident from clinical practice that dysphonia may manifest with varying features, having voice quality distortion as only one of the many other features along with physical manifestations of a voice disorder, which include throat discomfort, vocal fatigue, limited voice range, extra effort to speak, voice cracks, etc. [40, 41]. For the patient, the presence of these glottal-functions-related symptoms sometimes may be even more motivating to visit an otolaryngologist as compared to voice disturbance. The results of several studies showed that patients with voice problems detected in the self-assessment by VoiSS questionnaire have more deviated acoustic parameters; however, the relationship between acoustic measurements and vocal self-assessment, in general, is not linear. These findings confirm that these different procedures (VoiSS and acoustic analysis) provide complementary rather than redundant information [40].

In their study, D'haeseleer et al. 2017 found that objective vocal quality of theater actors, measured by the AVQI, did not change immediately after a theater performance. Nevertheless, about 50% of the theater actors had vocal complaints after performing was revealed with questionnaires. The authors suggest that it is possible that vocal complaints related to the presence of vocally violent behavior do not have an impact on the objective vocal quality immediately after the performance [42]. Consequently, use of two different information modalities while evaluating voice quality in theater actors may be considered as complimentary.

It is well known from clinical practice that some individuals may have no complaints regarding voice quality and consider their voice normal; however, other clinical symptoms related to glottal function may manifest. Most of physical manifestations of a voice disorder can be evaluated with GFI. In such case, the abnormal result of GFI would reflect distorted glottal function. The combined result of AVQI and GFI assessment would classify those subjects to the voice-disturbed group with large probability of a correct result.

Therefore, the rationale of using a combination of AVQI and GFI was based on presumption that these two modalities provide related, but rather different forms of discriminatory information valuable for distinguishing normal/pathological or voice-disordered cases vs. non cases. Results of the present study confirmed these considerations: by combining information from two different modalities (AVQI and GFI), a considerable improvement in classification accuracy normal (noncase) versus pathological voice (voice-disordered case) was obtained in our series.

The combined AVQI and GFI cut-off scores of 6.65 for oral microphone and 7.1 for SP microphone corresponded with excellent test accuracy (AUC = 0.976 and AUC = 0.965, respectively) and sensitivity and specificity (70.0% vs. 86.0%, respectively). Very respectable levels of positive and negative likelihood ratios both for oral microphone (13.3 and 0.08) and for SP microphone (15.6 and 0.10) voice recordings were achieved, respectively. Finally, the very high CCR of 93% and 92% confirmed the results of ROC statistics.

It is important to notice that, for the patient, the GFI is brief, easily understood, and readily completed [25]. Therefore, this short, pragmatic, and user-friendly 4-item

tool should be well suitable for a voice disorder screening purpose. However, some limitations of the GFI have to be considered. Due to the shortness of the GFI, this symptom-oriented survey is feasible to cover only a rather narrow spectrum of voice-related symptoms and complaints. As with other vocal impairment batteries, the GFI is not specific for any single clinical entity and shows only utility for assessment of severity of abnormalities of glottal function.

Results of the present study are in some contradiction with Awan et al. 2016 data. In their study, the authors observed that addition of another reference standard (namely the VHI score) did not add much to the discriminatory performance of the CSID to define voice-disordered cases versus noncases [1]. This is not surprising, because subjects with self-perceived vocal disability may actually produce perceptually and acoustically ''normal'' voice and patients with dysphonic voice may not consider voice-related disability in their everyday life. Therefore, these two information modalities (CSID and VHI) reflect rather different aspects of the problem and do not necessarily supplement each other regarding evaluation of both perceived and acoustically distorted voice quality. The VHI does not objectively reflect physical or physiological symptoms, but, instead, quantifies the psychosocial consequences of voice disorders, which are more influenced by contextual factors (personal and environmental) than the voice acoustic parameters [40]. On the other hand, the GFI is aimed specifically at identifying the presence and degree of symptoms of glottal dysfunction in evaluating patients presenting with dysphonia resulting from a variety of clinical entities [25].

Another important issue of this study was proving of suitability of SP microphone recordings for voice analysis and estimation of AVQI. In general, this is in concordance with novel data of the literature, confirming that smartphones have proven to be comparable to standard external microphones in recording and acoustic analysis of normal and dysphonic voice signals [32, 43–45]. Moreover, results of the present study validate the suitability of the SP microphone signal for the task of voice screening.

Conclusions

In summary, the combination of two different information sources (i.e., acoustic analysis such as the AVQI and GFI as a glottal function symptom-based questionnaire) used in this study significantly improved diagnostic accuracy in differentiating normal vs. pathological voice and could be clearly best used as indicators of dysphonia. In clinical practice, the combination of AVQI and GFI would represent a potentially valuable and robust screening tool.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest\

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. 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]).

References

1. Awan SN, Roy N, Zhang D, Cohen SM (2016) Validation of the Cepstral Spectral Index of dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. J Voice 30(2):130–144. https://doi.org/10.1016/j.jvoice.2015.04.009
2. Kaleem MF, Ghoraani B, Guergachi A, Krishnan S (2011) Telephone-quality pathological speech classification using empirical mode decomposition. Conf Proc IEEE Eng Med Biol Soc 2011:7095–7098. https://doi.org/10.1109/IEMBS.2011.6091793
3. Linder R, Albers AE, Hess M, Pöppl SJ, Schönweiler R (2008) Artificial neural network-based classification to screen for dysphonia using psychoacoustic scaling of acoustic voice features. J Voice 22(2):155–163. https://doi.org/10.1016/j.jvoice.2006.09.003
4. Uloza V, Verikas A, Bacauskiene M et al (2011) Categorizing normal and pathological voices: automated and perceptual categorization. J Voice 25(6):700–708. https://doi.org/10.1016/j.jvoice.2010.04.009
5. Lopes LW, Simões LB, da Silva JD et al (2017) Accuracy of acoustic analysis measurements in the evaluation of patients with different laryngeal diagnoses. J Voice 31(3):382.e15–382.e26. https://doi.org/10.1016/j.jvoice.2016.08.015
6. Awan SN, Roy N, Dromey C (2009) Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model. Clin Linguist Phon 23(11):825–841. https://doi.org/10.3109/02699200903242988
7. Maryn Y, De Bodt M, Roy N (2010) The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. J Commun Disord 43(3):161–174. https://doi.org/10.1016/j.jcomdis.2009.12.004
8. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M (2010) Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. J Voice 24(5):540–555. https://doi.org/10.1016/j.jvoice.2008.12.014
9. Barsties B, Maryn Y (2012) The Acoustic Voice Quality Index toward expanded measurement of dysphonia severity in German subjects. HNO 60(8):715–720. https://doi.org/10.1007/s00106-012-2499-9
10. Reynolds V, Buckland A, Bailey J et al (2012) Objective assessment of pediatric voice disorders with the Acoustic Voice Quality Index. J Voice 26(5):672.e1–672.e7. https://doi.org/10.1016/j.jvoice.2012.02.002

11. Maryn Y, De Bodt M, Barsties B, Roy N (2014) The value of the Acoustic Voice Quality Index as a measure of dysphonia severity in subjects speaking different languages. Eur Arch Oto Rhino Laryngol 271(6):1609–1619. https://doi.org/10.1007/s00405-013-2730-7
12. Barsties B, Maryn Y (2015) The improvement of internal consistency of the Acoustic Voice Quality Index. Am J Otolaryngol Head Neck Med Surg 36(5):647–656. https://doi.org/10.1016/j.amjoto.2015.04.012
13. Maryn Y, Kim H-T, Kim J (2016) Auditory-perceptual and acoustic methods in measuring dysphonia severity of Korean speech. J Voice 30(5):587–594. https://doi.org/10.1016/j.jvoice.2015.06.011
14. Hosokawa K, Barsties B, Iwahashi T et al (2017) Validation of the Acoustic Voice Quality Index in the Japanese Language. J Voice 31(2):260.e1–260.e9. https://doi.org/10.1016/j.jvoice.2016.05.010
15. Nunez-Batalla F, Diaz-Fresno E, Alvarez-Fernandez A, Munoz Cordero G, Llorente Pendas JL (2017) Application of the Acoustic Voice Quality Index for objective measurement of dysphonia severity. Acta Otorrinolaringol Esp 68(4):204–211. https://doi.org/10.1016/j.otorri.2016.11.010
16. Latoszek BBV, Ulozaitė-Stanienė N, Maryn Y, Petrauskas T, Uloza V (2017) The influence of gender and age on the Acoustic Voice Quality Index and dysphonia severity index: a normative study. J Voice. https://doi.org/10.1016/j.jvoice.2017.11.011
17. Lee JM, Roy N, Peterson E, Merrill RM (2017) Comparison of two multiparameter acoustic indices of dysphonia severity: the Acoustic Voice Quality Index and cepstral spectral index of dysphonia. J Voice 2:515
18. Verikas A, Gelzinis A, Bacauskiene M, Uloza V, Kaseta M (2009) Using the patient's questionnaire data to screen laryngeal disorders. Comput Biol Med 39(2):148–155. https://doi.org/10.1016/j.compbiomed.2008.11.008
19. Verikas A, Bacauskiene M, Gelzinis A, Vaiciukynas E, Uloza V (2012) Questionnaire—versus voice-based screening for laryngeal disorders. Expert Syst Appl 39(6):6254–6262. https://doi.org/10.1016/j.eswa.2011.12.037
20. Jacobson BH, Johnson A, Grywalski C et al (1997) The Voice Handicap Index (VHI). Am J Speech Lang Pathol 6(3):66. https://doi.org/10.1044/1058-0360.0603.66
21. Hogikyan ND, Sethuraman G (1999) Validation of an instrument to measure voice-related quality of life (V- RQOL). J Voice 13(4):557–569. https://doi.org/10.1016/S0892-1997(99)80010-1
22. Gliklich RE, Glovsky RM, Montgomery WW (1999) Validation of a voice outcome survey for unilateral vocal cord paralysis. Otolaryngol Head Neck Surg 120(2):153–158. https://doi.org/10.1016/S0194-5998(99)70399-2
23. Ma EPM, Yiu EML (2001) Voice activity and participation profile: assessing the impact of voice disorders on daily activities. J Speech Lang Hear Res 44(3):511–524. https://doi.org/10.1044/1092-4388(2001/040)
24. Deary IJ, Wilson JA, Carding PN, MacKenzie K (2003) VoiSS: a patient-derived voice symptom scale. J Psychosom Res 54(5):483–489. https://doi.org/10.1016/S0022-3999(02)00469-5
25. Bach KK, Belafsky PC, Wasylik K, Postma GN, Koufman JA (2005) Validity and reliability of the Glottal Function Index. Arch Otolaryngol Head Neck Surg 131(11):961–964. https://doi.org/10.1001/archotol.131.11.961
26. Cohen JT, Oestreicher-Kedem Y, Fliss DM, DeRowe A (2007) Glottal Function Index: a predictor of glottal disorders in children. Ann Otol Rhinol Laryngol 116(2):81–84. https://doi.org/10.1177/000348940711600201
27. Pribuišiene R, Baceviciene M, Uloza V, Vegiene A, Antuseva J (2012) Validation of the Lithuanian version of the glottal function index. J Voice 26(2):73–78. https://doi.org/10.1016/j.jvoice.2011.01.012
28. Ghirardi ACDAM, Ferreira LP, Giannini SPP, de Oliveira MDRD (2013) Screening index for voice disorder (SIVD): development and validation. J Voice 27(2):195–200. https://doi.org/10.1016/j.jvoice.2012.11.004
29. van Gogh CD, Verdonck-de Leeuw IM, Boon-Kamma BA, Langendijk JA, Kuik DJ, Mahieu HF (2005) A screening questionnaire for voice problems after treatment of early glottic cancer. Int J Radiat Oncol Biol Phys 62(3):700–705. https://doi.org/10.1016/j.ijrobp.2004.10.027
30. Morawska J, Niebudek-Bogusz E, Wiktorowicz J, Sliwinska-Kowalska M (2018) Screening value of v-rqol in the evaluation of occupational voice disorders. Med Pr 69(2):119–128. https://doi.org/10.13075/mp.5893.00649
31. Vaiciukynas E, Verikas A, Gelzinis A et al (2015) Fusing voice and query data for non-invasive detection of laryngeal disorders. Expert Syst Appl 42(22):8445–8453. https://doi.org/10.1016/j.eswa.2015.07.001
32. Uloza V, Padervinskis E, Vegiene A et al (2015) Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening. Eur Arch Otorhinolaryngol 272(11):3391–3399. https://doi.org/10.1007/s00405-015-3708-4
33. Uloza V, Petrauskas T, Padervinskis E, Ulozaitė N, Barsties B, Maryn Y (2017) Validation of the Acoustic Voice Quality Index in the Lithuanian language. J Voice 31(2):257.e1–257.e11. https://doi.org/10.1016/j.jvoice.2016.06.002
34. Bland J, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327(8476):307–310. https://doi.org/10.1016/j.jsv.2006.09.024
35. Elliott AC, Woodward WA. *Statistical Analysis Quick Reference Guidebook: With SPSS Example.*; 2007. doi:10.4135/9781412985949
36. Portney, LG, Watkins M. *Foundations of Clinical Research: Applications to Practice (Vol. 2)*. Vol 47.; 2000. https://books.google.com.au/books/about/Foundations_of_Clinical_Research.html?id=apNJPgAACAAJ&pgis=1
37. Dollaghan CA (2007) The handbook for evidence-based practice in communication disorders. MD Brooks, Baltimore
38. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44(3):837–845. https://doi.org/10.2307/2531595
39. Latoszek BBV, Ulozaitė-Stanienė N, Petrauskas T, Uloza V, Maryn Y (2018) Diagnostic accuracy of dysphonia classification of DSI and AVQI. Laryngoscope 129(3):692–698. https://doi.org/10.1002/lary.27350
40. Lopes LW, da Silva JD, Simões LB et al (2017) Relationship between acoustic measurements and self-evaluation in patients with voice disorders. J Voice 31(1):119.e1–119.e10. https://doi.org/10.1016/j.jvoice.2016.02.021
41. Maryn Y, Roy N, De Bodt M, Van Cauwenberge P, Corthals P (2009) Acoustic measurement of overall voice quality: a meta-analysis. J Acoust Soc Am 126(5):2619–2634. https://doi.org/10.1121/1.3224706
42. D'haeseleer E, Meerschman I, Claeys S, Leyns C, Daelman J, Van Lierde K. Vocal quality in theatre actors. J Voice. 2017;31(4):510–510. doi:10.1016/j.jvoice.2016.11.008
43. Lin E, Hornibrook J, Ormond T (2012) Evaluating iPhone recordings for acoustic voice assessment. Folia Phoniatr Logop 64(3):122–130. https://doi.org/10.1159/000335874
44. Grillo EU, Brosious JN, Sorrell SL, Anand S (2016) Influence of smartphones and software on acoustic voice measures. Int J Telerehabilitation 8(2):9–14. https://doi.org/10.5195/IJT.2016.6202
45. Manfredi C, Lebacq J, Cantarella G et al (2017) Smartphones offer new opportunities in clinical voice research. J Voice 31(1):111.e1–111.e7. https://doi.org/10.1016/j.jvoice.2015.12.020