Smartphone-Based Voice Wellness Index Application for Dysphonia Screening and Assessment: Development and Reliability

[*]Virgilijus Uloza, [*]Nora Ulozaitė-Stanienė, [*]Tadas Petrauskas, [*]Kipras Pribuišis, [*]Ingrida Ulozienė, [†]Tomas Blažauskas, [†]Robertas Damaševičius, and [†]Rytis Maskeliūnas, *†*Kaunas, Lithuania*

Summary: Objective. This study aimed to develop a Voice Wellness Index (VWI) application combining the acoustic voice quality index (AVQI) and glottal function index (GFI) data and to evaluate its reliability in quantitative voice assessment and normal versus pathological voice differentiation.

Study design. Cross-sectional study.

Methods. A total of 135 adult participants (86 patients with voice disorders and 49 patients with normal voices) were included in this study. Five iOS and Android smartphones with the "Voice Wellness Index" app installed were used to estimate VWI. The VWI data obtained using smartphones were compared with VWI measurements computed from voice recordings collected from a reference studio microphone. The diagnostic efficacy of VWI in differentiating between normal and disordered voices was assessed using receiver operating characteristics (ROC).

Results. With a Cronbach's alpha of 0.972 and an ICC of 0.972 (0.964–0.979), the VWI scores of the individual smartphones demonstrated remarkable inter-smartphone agreement and reliability. The VWI data obtained from different smartphones and a studio microphone showed nearly perfect direct linear correlations (r = 0. 993–0.998). Depending on the individual smartphone device used, the cutoff scores of VWI related to differentiating between normal and pathological voice groups were calculated as 5.6–6.0 with the best balance between sensitivity (94.10–95.15%) and specificity (93.68–95.72%), The diagnostic accuracy was excellent in all cases, with an area under the curve (AUC) of 0.970–0.974.

Conclusion. The *"Voice Wellness Index"* application is an accurate and reliable tool for voice quality measurement and normal versus pathological voice screening and has considerable potential to be used by healthcare professionals and patients for voice assessment.

Key Words: Voice screening app–Dysphonia screening–AVQI–GFI–VWI.

INTRODUCTION

Voice disorders manifesting as dysphonia are common, affecting approximately 3–9% of the population.[1,2] These include pathologies ranging from functional voice disorders and benign mass lesions to disabling chronic disorders and fatal malignant tumors.

Voice acoustic data represent a non-invasive, relatively easy-to-capture, and accurate biomarker, suggesting feasible and reliable possibilities for the screening and monitoring of dysphonia. Over the past several decades, researchers have developed several multiparametric acoustic voice indices to address the multidimensionality of voice signals and overcome the limited validity of a single acoustic parameter compared to the multidimensionality of the voice. These indices analyze and combine several acoustic voice parameters, considering both sustained phonation and connected speech, to provide a single score for measuring voice quality.[3,4] Currently, two multiparametric models based on sustained vowels and continuous speech have been used to evaluate voice quality: the cepstral peak prominence (CPP) and acoustic voice quality index (AVQI).

CPP, an acoustic measurement obtained from the cepstrum of a sound wave, is a potential acoustic marker of dysphonia.[5] CPP can be extracted from connected speech and sustained vowels and does not require direct computation of the fundamental frequency. CPP values fell within a continuous range, with lower values typically correlated with greater levels of dysphonia.[6] Cepstral spectral index of dysphonia (CSID) and smoothed CPP, derived from CPP, can provide a valid estimate of dysphonia severity and a high level of accuracy for the classification of voice-disordered cases versus controls, particularly when auditory-perceptual judgment is used as the reference standard. Therefore, the use of CPP measures represents a considerable advancement in voice disorder identification, assessment of voice deviation severity, and evaluation of speech therapy outcomes.[7–12]

The acoustic voice quality index (AVQI), developed by Maryn et al in 2010[13] is an exemplary multiparametric acoustic voice index. AVQI is a six-variable acoustic model for the multiparametric measurement of voice quality,

Accepted for publication October 12, 2023.

This project received funding from the European Regional Development Fund (project No. 13.1.1-LMT-K-718-05-0027) under a grant agreement with the Research Council of Lithuania (LMTLT). Funded as the European Union's measures in response to the COVID-19 pandemic.

From the *Department of Otorhinolaryngology, Lithuanian University of Health Sciences, Kaunas, Lithuania; and the †Faculty of Informatics, Kaunas University of Technology, Kaunas, Lithuania.

Address correspondence and reprint requests to Kipras Pribuišis, Department of Otorhinolaryngology, Lithuanian University of Health Sciences, 50061 Kaunas, Lithuania. E-mail: Kipras.pribuisis@lsmu.lt

Journal of Voice, Vol xx, No xx, pp. xxx–xxx

0892-1997

https://doi.org/10.1016/j.jvoice.2023.10.021

concatenating both the sustained vowel [a:] and the voiced parts of a continuous speech fragment. Several studies have confirmed the following remarkable features of AVQI: high internal consistency and inter-rater reliability, concurrent validity, test-retest reliability, high sensitivity to voice quality changes through voice therapy, utility in discriminating across perceptual levels of dysphonia severity, and adequate diagnostic accuracy with good discriminatory power in differentiating between normal and abnormal voice qualities.[14–19] Consequently, AVQI is considered a globally recognized multiparametric tool for voice quality assessment in both clinical and research applications.[20–22]

In the context of the multidimensionality of voice, studies investigating the integration of multidimensional information obtained in voice clinics have gained considerable attention.[23] However, existing data on the association between acoustic voice analysis, auditory-perceptual measures, and patients' voice self-assessment (self-reported measurements) of voice problems are scarce and controversial.

The reliability and correlation of the auditory-perceptual assessment of acoustic voice analysis based on vowel phonation and one acoustic parameter were poor.[24,25] However, multivariate techniques in voice assessment models demonstrated a stronger association with auditory-perceptual judgment as well as higher validity and reliability in identifying voice diseases.[15,26] Listeners' ratings and changes observed before and after treatment are strongly correlated and significantly associated with the AVQI- and CSID-estimated dysphonia severity for sustained vowels, connected speech, and a combined context.[9,10] Lopez et al (2017) investigated the relationship between acoustic measurements and self-evaluations in patients with voice disorders. The Voice Symptoms Scale (VoiSS) and the acoustic measures were correlated. Greater acoustic deviations were observed in patients with self-reported voice issues on the VoiSS. However, the voice handicap index (VHI) scores and acoustic measurements were not correlated, and no differences in the averages of these measurements were observed between patients with and without voice disorders identified from the VHI cutoffs.[23] A recent study by Saeedi et al (2023) revealed that the means of the CPP analysis did not differ significantly between all scores of the nonstandard hoarseness self-assessment questionnaire in the dysphonic and normal voice groups. Nevertheless, the observed relationship between cepstral analysis and self-assessment is promising for the development of multiparametric tests in the future.[27]

A study by Edie et al found that when acoustic and auditory-perceptual measures were combined, the accuracy in detecting the presence or absence of a voice disorder, along with gauging the severity of this deviation, increased to 100%.[28] Furthermore, Lopes et al revealed that although the relationship between acoustic voice measurements and vocal self-assessment is not linear, these different methods provide complementary rather than redundant information and support the utility of both voice assessment procedures for investigating different aspects of a voice disorder.[23] In a recent study, Saeedi et al demonstrated the relationship between cepstral analysis and voice self-assessment tools and concluded that this is promising for the development of robust multiparametric tests.[27]

In this context, the glottal function index (GFI) questionnaire developed by Bach et al in 2005 should be carefully considered. GFI is a concise (4-item) self-administered tool that is easily comprehensible, reliable, symptom-based and focuses on functional aspects. GFI evaluates the extent of vocal dysfunction in adults, making it a noteworthy inclusion.[29] A GFI cutoff score > 3.0 points helps distinguish dysphonic patients from healthy, normal voice controls with a high level of sensitivity and specificity.[30,31] Moreover, moderate correlations between the GFI and GRBAS scale ratings, except for roughness[32] and GFI, and the voice handicap index (VHI) questionnaire total scores were reported.[33]

Several studies have utilized GFI for examining patients with dysphonia caused by paralyzing glottal insufficiency and vocal fold atrophy and to evaluate the response to treatment.[34–36] In the task of differentiating between normal and dysphonic voice classes, the GFI results outperformed the classification results based on acoustic voice parameters.[37] In contrast, Spellman et al showed that GFI exhibited higher precision in diagnosing dysphonia than did VHI; however, it was less precise than that of acoustic voice measures.[34] Furthermore, combining acoustic voice parameters with responses to GFI items enhanced normal and dysphonic voice categorization.[38] A subsequent study revealed that the combination of AVQI and GFI measurements significantly improved the diagnostic accuracy in differentiating between normal and pathological voices.[39]

Recently, mobile communication devices such as smartphones and tablets have become widely available to both clinicians and patients. Current advances in smartphone technology and microphone quality propose a cost-effective and easily available alternative to studio microphones, contributing to a clinically feasible and effective tool for the detection, assessment, and care of voice disorders, indicating that a variety of commonly available modern smartphones can be used to collect high-quality voice recordings suitable for acoustic analysis.[8,40] The feasibility of smartphone voice recordings for estimating AVQI has already been demonstrated.[18,41–43] Furthermore, data on the AVQI performance using different applications for mobile communication devices are available.[43–45]

Previous study findings enabled us to presume the feasibility of integrating voice-related data from two different information sources (ie, acoustic voice analysis, such as the AVQI and GFI, as patient-reported outcome measures) to estimate the combined Voice Wellness Index (VWI). Consequently, the current research was designed to answer the following questions regarding the possibility of a smartphone-based application for VWI estimation: (1) Is the diagnostic accuracy of VWI relevant to differentiating normal and pathological voices? and (2) Are the different

smartphone-estimated average VWI values consistent and comparable? We hypothesized that using different smartphones for voice recording and VWI estimation would enable quantitative voice assessment and voice screening.

Therefore, the present study aimed to develop a universal platform-based application combining AVQI and GFI data and evaluate its reliability in quantitative voice evaluation and normal vs. pathological voice differentiation.

MATERIALS AND METHODS

Methods

This study was approved by the Kaunas Regional Ethics Committee for Biomedical Research (2022-04-20; No. BE-2-49).

The sample size was determined using a power analysis. The chosen power (1-beta) for this analysis was 0.9 presuming an equal sample size and a significance level of 0.05. Considering equal sample sizes, the Fleiss analysis resulted in a required sample size of at least 42 participants in the normal and 42 participants in pathological voice groups to detect statistically significant differences.[46]

All study participants were examined, and measures were taken at standard-of-care clinical appointments at the Department of Otolaryngology of the Lithuanian University of Health Sciences, Kaunas, Lithuania.

The inclusion criteria to define a vocally healthy participant are as follows: (1) all selected participants considered their voice as normal and had no actual voice complaints or history of chronic laryngeal diseases or voice disorders; (2) the participants had no hearing problems and were free from common cold or upper respiratory infections at the time of voice recording, (3) no pathological alterations in the larynx of the healthy participants were found during video laryngostroboscopy (VLS); (4) voice samples were evaluated as normal voices by a laryngologist; (5) age ≥18 years; (6) signed informed consent. The exclusion criteria are as follows: (1) age ≤18 years, (2) complaints about voice disorder, (3) pathological alterations in the larynx, (4) refusal to sign informed consent.

The criteria for inclusion in the pathological voice subgroup are as follows: (1) participants aged ≥18 years, (2) those with complaints of voice disorder, (3) voice evaluated as pathological by a laryngologist, (4) presence of laryngoscopically positive signs, (5) histologically verified diagnosis in cases of mass lesions of vocal folds; (6) signed informed consent. The exclusion criteria are as follows: (1) age ≤18 years, (2) no complaints about voice disorder, (3) normal voice, (4) absence of laryngoscopically positive signs, (5) refusal to sign informed consent.

Pretreatment data from patients with voice disorders were collected at baseline. Voice samples were classified into four ordinal severity classes based on the GRBAS scale (G0 = normal voice, G1 = slight, G2 = moderate, and G3 = severe dysphonia).[47] In the pathological voice group, a threshold for perceptual dysphonia severity of G ≥ 1.0 was considered.

The diagnosis of voice disorders was based on clinical examination (complaints and pertinent history) along with the results of videolaryngostroboscopy (VLS) and/or direct microlaryngoscopy. High-quality digital VLS recordings were performed using a XION Endo-STROB DX device (XION GmbH, Berlin, Germany) with a 70° rigid endoscope. Positive laryngoscopic findings include a wide range of signs representing vocal fold hypertrophy/atrophy, paresis/paralysis, tremors, and benign and malignant mass lesions of the vocal fold. All patients with mass lesions of the vocal folds (nodules, polyps, cysts, Reinke's hyperplasia, papillomata, chronic hyperplastic laryngitis with keratosis, or carcinoma) underwent endolaryngeal microphonosurgery. The final diagnosis was confirmed by histological examination of the surgical specimen and was used in the present study. The final diagnosis served as the gold standard for evaluating the diagnostic accuracy of VWI in discriminating between participants with normal and pathological voices.

Although the AVQI was originally developed and presumed to be the most appropriate for differentiating between disordered and non-disordered voice quality groups based on auditory–perceptual judgment, a previous study demonstrated that this multivariate index could also recognize vocally healthy individuals and patients with voice disorders based on the diagnosis of laryngeal disorder.[48]

Glottal function index questionnaire

Each participant (normal and pathological voice subgroups) completed the GFI questionnaire at baseline, along with voice recordings. The GFI scores range from 0 to 20 points, and scores higher than 3.0 points are considered the limiting value distinguishing normal and pathological voice subgroups.[31] The GFI questionnaire is presented in the Appendix.

Voice recordings

Voice samples from the study participants were recorded in a T-series sound-proof room for hearing testing (T-room, CATegner AB, Bromma, Sweden) using a studio oral cardioid AKG Perception 220 microphone (AKG Acoustics, Vienna, Austria). The microphone was placed 10.0 cm distance from the mouth while maintaining a microphone-to-mouth angle of 90°. Each participant completed two digitally recorded vocal tasks. The tasks consisted of sustaining phonation of the vowel sound [a:] for at least 4 seconds duration (to ensure voice recording of the patients with glottal closure insufficiency and shortened phonation time) and reading a phonetically balanced text segment in Lithuanian "Turėjo senelė žilą oželį" ("The granny had a small grey goat"). Participants were instructed to complete both vocal tasks at a comfortable level of loudness and pitch.

Smartphone microphone frequency response curves were evaluated to ensure that all voice recordings coincided between genuine smartphone microphones and studio microphone speech recordings while preserving consistent environmental conditions. A signal analysis was performed to reduce the influence of non-smartphone factors, including room acoustics,

reflections, user-to-microphone distances, directionality, and user loudness, which would naturally arise when dealing with human participants. Ableton digital audio workstation (DAW) was used as the audio modeling environment. The MFreeform Equalizer VST (Virtual Studio plugin) by MeldaProduction (available at https://www.meldaproduction.com/MFreeformEqualizer) was used to import and filter audio to match the frequency response datasets derived from each smartphone. MFreeformEqualizer filter quality was set to the highest available setting with a 0% curve smoothing configuration. Subsequently, all audio files were processed as 44,100 Hz 16-bit WAV files. This methodology enabled the flat AKG Perception 220 recording to be filtered exactly as a signal recorded through a specific smartphone while circumventing any other possible non-smartphone variable influence. The following microphone frequency-response data of the following smartphones were employed in the model: iPhone SE, iPhone 13 Max Pro, Huawei P50 Pro, Samsung S22 Ultra, and OnePlus 9 Pro. The AKG Perception 220 studio microphone flat frequency response recording was compared against five different voice recording groups derived for specific smartphone models using the described microphone matching method and further analyzed on smartphones with the VWI application.

AVQI estimation

For AVQI calculations, signal processing of the voice samples was performed using Praat software (version 5.3.57; https://www.fon.hum.uva.nl/praat/). The voice samples were concatenated in the following order: text segment, 2-second pause, followed by a 3-second sustained vowel /a/ segment. 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-v0203.download=AcousticVoiceQualityIndexv.02.03.txt.[3] The following multiple regression equation for the AVQI was used[15]:

AVQI = 9.072–0.245× smoothed cepstral peak prominence −0.161× harmonics to noise ratio −0.470× shimmer local +6.158× shimmer local dB −0.071× slope of the long-term average spectrum −0.170× tilt of the trendline through the average long-term average spectrum.

This equation includes acoustic markers from the time, frequency, and quefrency domains and is a multidimensional representation of dysphonia severity. AVQI scores ranged between 0 and 10 points, with a higher score indicating more severe dysphonia. The mean cutoff score for differentiating normal and pathological voices is approximately 3.05 and may range between 2.80 and 3.46 depending on the software and language used.[48]

Development of a universal-platform-based "*Voice Wellness Index*" application

The Voice Wellness Index (VWI) is the proportion summation of the AVQI and GFI scores. To equalize the AVQI and GFI inputs into the final VWI score, the following equation was used: VWI = AVQI + GFI/ 2. The rationale for this approach was as follows: VWImax (20 points) = AVQImax (10 points) + GFImax 20/2 points).

The *"Voice Wellness Index"* application for use both with iOS and Android operating devices was developed based on the "*VoiceScreen"* application[44] and its further development, that is, the universal-platform-based (UPB) "*Voice Screen"* application.[49] In this case, all measures were estimated on the server; therefore, the computationally costly sound processing was not dependent on the computational capabilities of the user device. Background noise monitoring, voice recording, automated AVQI calculations, GFI estimation, and VWI processing are implemented in this application. Consequently, the "*Voice Wellness Index"* application allows voice recording, automatically extracting acoustic voice features and displaying the VWI result alongside a recommendation to the user Figure 1.

In this study, the UPB "*Voice Wellness Index"* application was installed in five different smartphones (iPhone Pro Max 13, iPhone SE [iOS operating system], OnePlus 9 PRO, Samsung S22 Ultra, and Huawei P50 pro [Android operating system]). The VWI measurements estimated with the *"Voice Wellness Index"* application from voice recordings obtained from a flat-frequency response studio microphone, AKG Perception 220, were compared with the VWI results obtained using these smartphone devices.

STATISTICAL ANALYSIS

Statistical analyses were performed using IBM SPSS Statistics for Windows (version 27.0; (IBM Corp., Armonk, NY, USA) and MedCalc Version 20.118 (Ostend, BE: MedCalc Software Ltd.). A *P*-value of 0.05 indicated statistical significance.

The data distribution was determined according to the normality law by applying the Shapiro–Wilk normality test and calculating the coefficients of skewness and kurtosis. Student's *t*-test was used to test the equality of means for normally distributed data. Analysis of variance (ANOVA) was used to determine significant differences between the multiple means of the independent groups.[50] Cronbach's alpha was used to measure the internal consistency. The Pearson's correlation coefficient was used to assess the linear relationship between variables obtained from continuous scales.

Receiver operating characteristic (ROC) curves were used to obtain optimal sensitivity and specificity at optimal scale cutoff points. The area under the ROC curve (AUC) served to calculate the accuracy of the discriminatory scale. A pairwise comparison of the ROC curves, as described by De Long et al, was used to determine whether there was a statistically significant difference between two or more variables when categorizing normal/pathological voices.[51]

RESULTS

Study group

The study group comprised 135 adults (58 men and 77 women) with a mean age of 42.92 years (SD 15.26). The

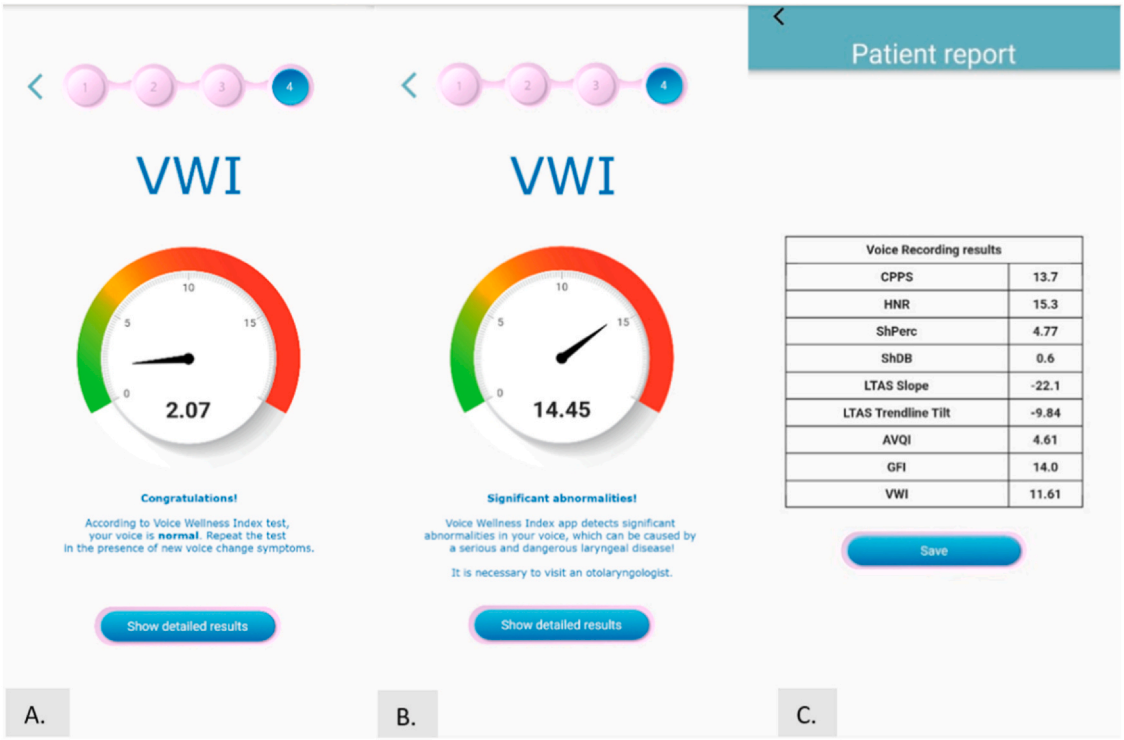

FIGURE 1. Screenshots of the "***Voice Wellness Index***" mobile application. **A**. Example of normal voice VWI result and recommendation; **B.** Example of pathological voice VWI result and recommendation: **C.** Display of AVQI, GFI and VWI results. (For details see the Supplementary Material).

normal voice subgroup included 49 selected healthy volunteers (16 men and 33 women; mean age 31.69 years [SD 9.89]). The pathological voice subgroup consisted of 86 consecutive patients (42 men and 44 women; mean age 50.8 years [SD, 14.3]). They present with a relatively common and clinically discriminative group of laryngeal diseases and related voice disturbances. Table 1 shows the demographic data of the study group and the diagnoses of the pathological voice subgroups.

In the present study, the control and patient groups were matched by sex ($P = 0.07$). Despite not being matched by age ($P = 0.01$), they were considered suitable for AVQI-related data analysis because previous studies did not demonstrate correlations between the participant's age, sex, and AVQI measurements, indicating that these variables do not affect the overall AVQI value.[45,52–54]

VWI evaluation outcomes

The distribution of the mean VWI scores in the study group according to clinical diagnosis is presented in Table 2.

ANOVA revealed statistically significant differences between the mean scores of the VWI ($P = 0.01$; $F = 10.9$), AVQI ($P = 0.01$; $F = 34.9$), and GFI ($P = 0.01$; $F = 4.8$)

TABLE 1.
Demographic Data of the Study Group

		Age		Sex		G	
Diagnosis	n	Mean	SD	Male (n)	Female (n)	Mean	SD
Chronic hyperplastic laryngitis	10	55.9	7.34	8 (80%)	2 (20%)	1.70	0.68
Mass lesions of vocal folds	49	44.39	12.40	18 (36.7%)	31 (63.3%)	2.10	0.77
Vocal fold cancer	11	65.09	7.710	10 (90.1%)	1 (9.9%)	2.55	0.69
Reflux laryngitis	2	57.0	15.56	1 (50%)	1 (50%)	3.0	0.0
Laryngeal paralysis	10	45.60	13.47	3 (30%)	7 (70%)	2.30	0.48
Parkinson's disease	2	71.50	9.19	0 (0%)	2 (100%)	1.50	0.71
Functional dysphonia	2	39.0	2.04	2 (100%)	0 (0%)	1.50	2.10
Normal voice	49	31.69	9.89	16 (32.7%)	33 (67.3%)	0.29	0.46
Total	135	42.92	15.26	58 (43%)	77 (57%)		

Abbreviation: G, grade of dysphonia.

TABLE 2.
Distribution of VWI Scores Along With Compounding Mean AVQI and GFI Scores in the Study Group

		VWI			AVQI			GFI/2		
Diagnosis	n	Mean	SD	p	Mean	SD	p	Mean	SD	*P*
Normal voice	49	2.48	1.14	0.01	2.04	0.79	0.01	0.44	0.74	0.01
Mass lesions of vocal folds	49	9.53	2.85		4.22	1.74		5.31	2.36	
Vocal fold cancer	11	7.98	2.76		5.02	2.38		2.95	1.25	
Chronic hyperplastic laryngitis	10	7.96	2.91		3.46	1.66		4.50	2.06	
Laryngeal paralysis	10	10.87	3.30		4.42	1.55		6.45	2.81	
Functional dysphonia	2	7.12	6.05		3.87	2.67		3.25	4.60	
Reflux laryngitis	2	9.39	0.29		4.64	0.46		4.75	0.37	
Parkinson's disease	2	3.20	2.05		1.70	1.01		1.50	1.41	

Abbreviations: VWI, voice wellness index; AVQI, acoustic voice quality index; GFI, glottal function index; SD, standard deviation.

TABLE 3.
Correlations of VWI Scores Obtained With a Studio Microphone and Different Smartphones

Microphones		iPhone SE	iPhone Pro Max 13	Huawei P50 pro	Samsung S22 Ultra	OnePlus 9 PRO
AKG Perception 220	r	0.998	0.997	0.993	0.996	0.998
	p	0.001	0.001	0.001	0.001	0.001
	n	135				

Abbreviations: VWI, voice wellness index; r, Pearson's correlation coefficient; p, statistical significance.

between the normal and pathological voice subgroups. As shown in Table 2, the mean VWI scores were significantly higher in the pathological voice subgroup than in the normal voice subgroup.

The evaluation of the VWI scores of individual smartphones displayed excellent inter-smartphone agreement with a Cronbach's alpha of 0.972. Moreover, the inter-smartphone VWI measurement reliability was excellent, with an average intraclass correlation coefficient of 0.972 (range: 0.964–0.979). A one-way ANOVA did not detect statistically significant differences between the mean VWI scores obtained using different smartphones ($F = 0.155$; $P = 0.978$). Furthermore, Bonferroni analysis reaffirmed the insignificance of the differences among the VWI scores obtained from smartphones ($P = 1.0$, estimated Bonferroni p for statistically significant difference, $P = 0.01$). The largest observed mean VWI difference between different smartphones was 0.4.

Furthermore, almost perfect direct linear correlations were observed between the VWI results obtained from voice recordings using a studio microphone and different smartphones. The Pearson's correlation coefficients ranged between 0.993 and 0.998 (Table 3).

The relationships between the VWI scores obtained using the studio microphone and different smartphones are graphically presented in Figure 2.

As demonstrated in Figure 2, the VWI results obtained using different smartphones closely resembled the VWI results obtained using a studio microphone with a single data point outside of the 95% confidence interval ($R^2 = 0.993$). These results confirmed that the VWI score obtained with the smartphone was directly compatible with that obtained with the reference studio microphone.

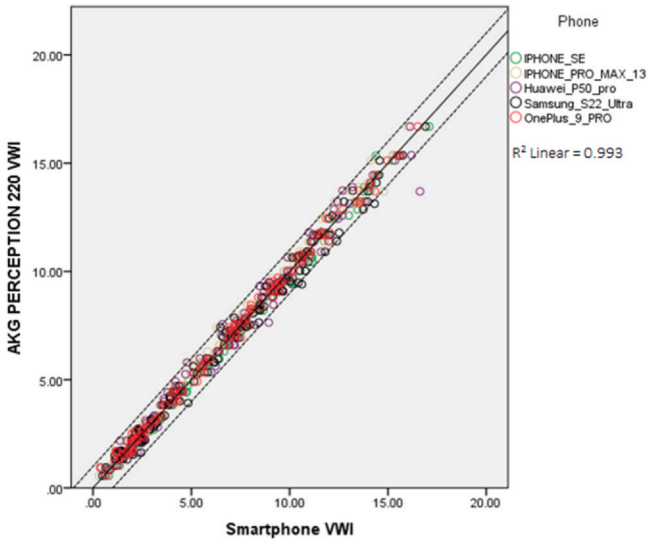

FIGURE 2. Scatterplot illustrating the correlation between the VWI results obtained from the studio microphone and different smartphones with a 95% confidence interval.

Normal versus pathological voice diagnostic accuracy of the VWI using different smartphones

The ROC analysis was applied to estimate the diagnostic accuracy of VWI obtained from a studio microphone and different smartphones for differentiating between normal

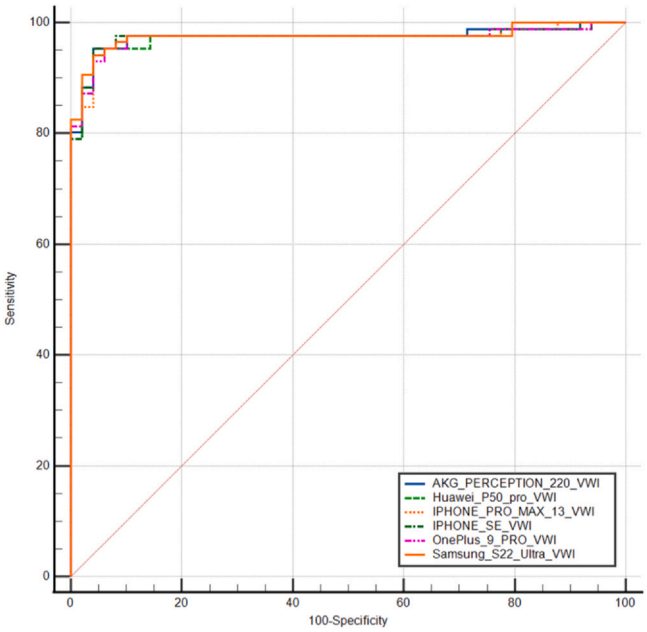

FIGURE 3. ROC curves illustrating the diagnostic accuracy of VWI obtained from voice recordings with studio microphones and different smartphones in discriminating between normal versus pathological voices.

and pathological voices. The ROC curves of the VWI were visually inspected to identify the optimum cutoff scores according to the general interpretation guidelines.[55] The ROC curves can be observed in Figure 3.

As shown in Figure 3, all ROC curves were almost identical and occupied the greater part of the graph, clearly revealing the highly respectable power of VWI obtained from voice recordings using a studio microphone and different smartphones in differentiating between normal and pathological voices.

The AUC analysis revealed a high level of precision of VWI in discriminating between normal and pathological voices, with a suggested AUC threshold of 0.800. Table 4 shows the results of the ROC statistical analyses.

As presented in Table 4, the ROC analysis determined the optimal VWI cutoff values for distinguishing between normal and pathological voices for each smartphone used. All employed devices passed the proposed 0.8 AUC threshold and revealed an acceptable Youden-index value. Depending on the individual smartphone device used, the cutoff scores of VWI related to discrimination between normal and pathological voice groups were calculated as 5.6–6.0 with the best balance between sensitivity (94.10–95.15%) and specificity (93.68–95.72%). The diagnostic accuracy was excellent in all cases, with an area under the curve of 0.970–0.974.

A pairwise comparison of the AUC-dependent ROC curves (VWI measurements obtained from studio microphones and different smartphones) according to a test by DeLong et al confirmed no statistically significant differences between the AUCs ($P > 0.05$). The largest observed difference between the AUCs was 0.004.

DISCUSSION

This study was the first to use a novel universal platform-based "*Voice Wellness Index*" application combining the AVQI and GFI data for quantitative voice assessment and normal versus pathological voice differentiation in patients with various voice disorders and normal voice controls. The results of the present study, to some extent, fulfill the need reported in the literature for studies that investigate the integration of multidimensional information obtained in the vocal clinic.[23]

The management of vocal problems using voice assessment software for mobile communication devices has been the subject of several studies. According to a 2015 study by Mat Baki et al, speech recordings processed and analyzed using the OperaVoxTM program were statistically equivalent to the "gold standard".[56] In 2018, Cesari et al developed the android application Vox4Health to calculate primary acoustic voice measures obtained from the vocalization of the vowel /a/ to determine the potential presence of a voice disorder[57] in real-time. Furthermore, the Dysphonia Detection Index, developed in conjunction with the Vox4Heath, represents a multiparametric acoustic marker that may assess voice quality and identify potential vocal problems.[57] In 2018, Kojima et al proposed the Voice Analyzer or "VA" smartphone application for quantitative analysis of voice quality.[58] A highly accessible real-time acoustic voice analysis device (VArt) expressing a new hoarseness index, real-time Ra (Rart), was created by Fujimura et al in 2019, after further system development.[58,59]

TABLE 4.
Statistics Illustrating the VWI Accuracy Differentiating Normal and Pathological Voices Using Studio Microphones and Different Smartphones

Microphone	AUC	Cut-off score	Sensitivity %	Specificity %	Youden-index J
AKG Perception 220	0.972	6.0	95.15	95.72	0.91
iPhone SE	0.972	6.0	95.15	95.72	0.91
iPhone Pro Max 13	0.972	5.6	95.15	95.72	0.91
Huawei P50 pro	0.970	5.6	94.10	95.72	0.9
Samsung S22 Ultra	0.974	5.6	95.15	93.68	0.87
OnePlus 9 PRO	0.971	5.6	95.15	93.68	0.87

Abbreviations: VWI, voice wellness index; AUC, area under the curve.

The application (VoiceEvalU8), which uses the Praat source code and algorithms to provide an automatic option for the accurate computation of several acoustic voice measurements and AVQI on iOS and Android smartphones, was presented by Grillo et al. in 2023.[43] In 2022, Shabnam et al reported the creation of a user-friendly program that offers a condensed output of AVQI cutoff values that can be understood by people with voice disorders, non-experts, and medical professionals.[18] Another contemporary development, VOXplot, is a freeware program for acoustic speech quality analysis based on Praat signal processing methods. The VOXplot is specifically designed to examine voice quality and was built to satisfy the needs of physicians and researchers for standardized and simple use. Therefore, the complete acoustic voice quality assessment workflow is covered by a VOXplot that includes recordings and recording quality evaluation.[11]

However, each of the aforementioned speech analysis applications determines the required acoustic voice characteristics using only the acoustic information from the voice signal. In contrast, the VWI application combines data from two different sources (AVQI and GFI) and outputs the resulting results. Moreover, it provides recommendations to users based on the test results. To our knowledge, no similar devices have been reported in the literature.

The concept of VWI development is based on the assumption that the voice assessment process should consider the multidimensionality involved in the manifestation of voice disorders. However, from a clinical perspective, the relationship between morphological alterations of the vocal folds and voice disturbances is not always direct and linear. For instance, individuals with laryngoscopically atypical signs may produce a 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 and/or acoustically measured dysphonia. 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 minor effects on the periodicity of vocal fold vibrations and perceived voice quality. In contrast, severely disordered voices (Parkinson's disease and functional voice disorders) are typically associated with structurally normal larynges. Simultaneously, dysphonia may manifest clinically with different features, resolving voice quality distortion as one of many other complaints. The physical manifestations of a voice disorder, such as throat discomfort, vocal fatigue, limited voice range, extra effort to speak, and voice cracks, may have an even greater impact on a patient's comprehension of the disease. Therefore, the assessment of glottal function-related symptoms using the GFI questionnaire is considered clinically substantial and relevant. Furthermore, the integration of multidimensional information related to vocal function should increase the accuracy of determining the presence or absence of a voice disorder and the severity of this deviation.

The results of the present study confirmed that combining two different procedures (acoustic analysis and glottal function questionnaire data) provided complementary information that ensured the robustness of VWI estimation across the reference studio microphone and different platforms' smartphones and significantly improved diagnostic accuracy in differentiating between normal and pathological voices.

The VWI scores of individual smartphones displayed excellent inter-smartphone agreement and reliability, achieving a Cronbach's alpha of 0.972 and an ICC of 0.964–0.979. Almost perfect direct linear correlations ($r = 0.993–0.998$) were observed between the VWI results obtained using a studio microphone and those obtained using different smartphones, confirming the direct compatibility among the devices used in the study. Correlation analysis showed that all VWI measurements were highly correlated (Pearson's $r = 0.993–0.998$) across the devices used in the present study. Excellent agreement and reliability, combined with no statistical differences in VWI scores obtained with different smartphones, further suggest that VWI evaluation results are compatible and reproducible on different smartphone platforms.

Analysis of the study results revealed that the VWI yielded a remarkable ability to discriminate between normal and pathological voices, as determined by clinical diagnosis ($AUC = 0.970–0.974$, depending on the smartphone used), resulting in an appropriate balance between very high sensitivity and specificity (94.10–95.15%) and (93.68–95.72%), respectively. These findings suggest that VWI is a reliable tool for differentiating normal and pathological voices independently of voice recordings from tested studio microphones and different smartphones and presents remarkable importance from a practical point of view.

Furthermore, the high dysphonia screening potential of another component of the VWI, the GFI questionnaire, has already been reported.[34,37,38] Consequently, fusion data from the two information sources previously mentioned allows VWI to be the most appropriate for discriminating between disordered and non-disordered voice quality groups. Finally, the relatively high discrimination power of the GFI data is another important feature of VWI, because such a sensor-independent data source with high discrimination strength reduces possible acoustic parameter-dependent variances related to the differences in smartphone microphones and equalizes the impact of both compounding parts (AVQI and GFI) on the VWI score. This feature is of great importance when different voice recording devices, such as different smartphones or other mobile communication devices, are used.

The present study had several limitations. Voice recordings for the current investigation were made in a soundproof room. However, the omnidirectional built-in microphones of smartphones may produce different results in actual clinical settings when surrounding noise is present. Further studies are necessary to assess both the effect of the voice recording environment and the unique characteristics of the microphones on the application of different smartphones in real clinical settings. This can be accomplished

by simultaneously conducting voice recordings on various smartphones. The results of the present study were based on a group that presented with clinically discriminative laryngeal and voice disorders. To achieve the greatest comparability of acoustic voice features derived from voice recordings collected using mobile communication devices and reference studio microphones, further studies including a wide variety of voice diseases, including functional voice disorders, are required. The results of further research may enable outcomes and advancements to be applied in healthcare applications.

In summary, the VWI is an easy-to-use tool that can be used by patients, healthy individuals, and medical professionals. Filling the 4-item GFI questionnaire and recording the sustained vowel [a] phonation and reading of a phonetically balanced standard sentence takes approximately 1–2 minutes. For the individual, the VWI application immediately displays the calculated VWI score along with a recommendation to the user. A separate window of the application displays the compounding parts of the WVI: the AVQI and six acoustic voice parameters, as well as the results of the GFI evaluation. This information could be of great importance for medical voice professionals to provide the necessary data for diagnosing voice disorders and monitoring treatment results. In addition, displaying AVQI and GFI results for medical professionals predisposes the evaluation of the "weight" of acoustic voice disturbances and/or glottal function symptoms on the final VWI score.

CONCLUSION

The *"Voice Wellness Index"* application represents an accurate and reliable tool for voice quality measurement and normal versus pathological voice screening, resulting in the cutoff scores of VWI ranging between 5.6 and 6.0 obtained using smartphones with excellent diagnostic accuracy (AUC = 0.97–0.974) and the best balance between sensitivity (94.10–95.15%) and specificity (93.68–95.72%). Therefore, this application has considerable potential for use by healthcare professionals and patients for voice assessment.

Institutional Review Board Statement

This study was conducted in accordance with the tenets of the Declaration of Helsinki (1975), and the protocol was approved by the Kaunas Regional Ethics Committee for Biomedical Research (2022-04-20 No. BE-2-49).

Informed Consent Statement

Informed consent was obtained from all participants involved in the study.

Data Availability Statement

The data presented in this study are available from the corresponding author upon reasonable request.

Declaration of Competing Interest

None.

Appendix

Appendix 1. Original GFI questionnaire

Within the last MONTH, How Did the Following Problems Affect You?	0 = No Problem 5 = Severe Problem
1. Speaking Took Extra Effort	0 1 2 3 4 5
2. Throat Discomfort or Pain After Using Your Voice	0 1 2 3 4 5
3. Vocal Fatigue (Voice Weakened as You Talked)	0 1 2 3 4 5
4. Voice Cracks or Sounds Different	0 1 2 3 4 5
	Total

.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jvoice.2023.10.021.

References

1. Roy N, Merrill RM, Thibeault S, et al. Prevalence of voice disorders in teachers and the general population. *J Speech Lang Hear Res.* 2004;47:281–293. https://doi.org/10.1044/1092-4388(2004/023).
2. Bhattacharyya N. The prevalence of voice problems among adults in the United States. *Laryngoscope.* 2014;124:2359–2362. https://doi.org/10.1002/lary.24740.
3. Maryn Y, Weenink D. Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index. *J Voice.* 2015;29:35–43. https://doi.org/10.1016/j.jvoice.2014.06.015.
4. Awan SN, Roy N, Zhang D, et al. Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. *J Voice.* 2016;30:130–144. https://doi.org/10.1016/j.jvoice.2015.04.009.
5. Watts CR, Awan SN, Maryn Y. A comparison of cepstral peak prominence measures from two acoustic analysis programs. *J Voice.* 2017;31:387.e1–387.e10. https://doi.org/10.1016/j.jvoice.2016.09.012.
6. Murton O, Hillman R, Mehta D. Cepstral peak prominence values for clinical voice evaluation. *Am J Speech Lang Pathol.* 2020;29:1596–1607. https://doi.org/10.1044/2020_AJSLP-20-00001.
7. Awan SN, Roy N, Dromey C. Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model. *Clin Linguist Phon.* 2009;23:825–841. https://doi.org/10.3109/02699200903242988.
8. Awan SN, Shaikh MA, Awan JA, et al. Smartphone recordings are comparable to "Gold standard" recordings for acoustic measurements of voice. *J Voice.* 2023. https://doi.org/10.1016/j.jvoice.2023.01.031.S0892-1997(23)00031-0.
9. Lee JM, Roy N, Peterson E, et al. Comparison of two multiparameter acoustic indices of dysphonia severity: the acoustic voice quality index and cepstral spectral index of dysphonia. *J Voice.* 2018;32:515.e1–515.e13. https://doi.org/10.1016/j.jvoice.2017.06.012.
10. Barsties V, Latoszek B, Mathmann P, Neumann K. The cepstral spectral index of dysphonia, the acoustic voice quality index and the acoustic breathiness index as novel multiparametric indices for acoustic assessment of voice quality. *Curr Opin Otolaryngol Head Neck Surg.* 2021;29:451–457. https://doi.org/10.1097/MOO.0000000000000743.
11. Barsties V, Latoszek B, Mayer J, et al. Advances in clinical voice quality analysis with VOXplot. *J Clin Med.* 2023;12:4644. https://doi.org/10.3390/jcm12144644.
12. Hofman EC, Dassie-Leite AP, Martins PDN, Pereira EC. Acoustic measurements of CPPS and AVQI pre and post speech therapy. *CoDAS.* 2023;35:e20220136. https://doi.org/10.1590/2317-1782/20232022136pt.
13. Maryn Y, De Bodt M, Roy N. The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord.* 2010;43:161–174. https://doi.org/10.1016/j.jcomdis.2009.12.004.
14. Barsties B, Maryn Y. The acoustic voice quality index. toward expanded measurement of dysphonia severity in german subjects. *Head Neck Oncol.* 2012;60:715–720. https://doi.org/10.1007/s00106-012-2499-9.
15. Uloza V, Petrauskas T, Padervinskis E, et al. Validation of the acoustic voice quality index in the Lithuanian language. *J Voice.* 2017;31:257.e1–257.e11. https://doi.org/10.1016/j.jvoice.2016.06.002.
16. Kankare E, Barsties V, Latoszek B, et al. The acoustic voice quality index version 02.02 in the Finnish-speaking population. *Logoped Phoniatr Vocol.* 2020;45:49–56. https://doi.org/10.1080/14015439.2018.1556332.
17. Lehnert B, Herold J, Blaurock M, et al. Reliability of the acoustic voice quality index AVQI and the acoustic breathiness index (ABI) when wearing CoViD-19 protective masks. *Eur Arch Otorhinolaryngol.* 2022;279:4617–4621. https://doi.org/10.1007/s00405-022-07417-4.
18. Shabnam S, Pushpavathi M, Gopi Sankar R, et al. A comprehensive application for grading severity of voice based on acoustic voice quality index v.02.03. *J Voice.* 2022. https://doi.org/10.1016/j.jvoice.2022.08.013.S0892-1997(22)00245-4.
19. Penido FA, Gama ACC. Accuracy analysis of the multiparametric acoustic indices AVQI, ABI, and DSI for speech-language pathologist decision-making. *J Voice.* 2023. https://doi.org/10.1016/j.jvoice.2022.11.027. S0892-1997(22)00380-0.
20. Jayakumar T, Benoy JJ. Acoustic voice quality index (AVQI) in the measurement of voice quality: a systematic review and meta-analysis. *J Voice.* 2022. https://doi.org/10.1016/J.JVOICE.2022.03.018. S0892-1997(22)00084-4.
21. Batthyany C, Latoszek BBV, Maryn Y. Meta-analysis on the validity of the acoustic voice quality index. *J Voice.* 2022. https://doi.org/10.1016/j.jvoice.2022.04.022. S0892-1997(22)00132-1.
22. Saeedi S, Aghajanzade M, Khatoonabadi AR. A literature review of voice indices available for voice assessment. *J Rehabil Sci Res.* 2022;9:151–155. https://doi.org/10.30476/jrsr.2022.93362.1235.
23. Lopes LW, da Silva JD, Simões LB, et al. Relationship between acoustic measurements and self-evaluation in patients with voice disorders. *J Voice.* 2017;31:119.e1–119.e10. https://doi.org/10.1016/j.jvoice.2016.02.021.
24. Maryn Y, Roy N, De Bodt M, et al. Acoustic measurement of overall voice quality: a meta-analysis. *J Acoust Soc Am.* 2009;126:2619–2634. https://doi.org/10.1121/1.3224706.
25. Barsties B, Maryn Y. Test-retest variability and internal consistency of the acoustic voice quality index. *Head Neck Oncol.* 2013;61:399–403. https://doi.org/10.1007/s00106-012-2649-0.
26. Maryn Y, Corthals P, Van Cauwenberge P, et al. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J Voice.* 2010;24:540–555. https://doi.org/10.1016/j.jvoice.2008.12.014.
27. Saeedi S, Aghajanzadeh M, Khoddami SM, et al. Relationship of cepstral analysis with voice self-assessments in dysphonic and normal speakers. *Eur Arch Otorhinolaryngol.* 2023;280:1803–1813. https://doi.org/10.1007/s00405-022-07690-3.
28. Eadie TL, Doyle PC. Classification of dysphonic voice: acoustic and auditory-perceptual measures. *J Voice.* 2005;19:1–14. https://doi.org/10.1016/j.jvoice.2004.02.002.
29. Bach KK, Belafsky PC, Wasylik K, et al. Validity and reliability of the glottal function index. *Arch Otolaryngol Head Neck Surg.* 2005;131:961–964. https://doi.org/10.1001/archotol.131.11.961.
30. Cohen JT, Oestreicher-Kedem Y, Fliss DM, et al. Glottal function index: a predictor of glottal disorders in children. *Ann Otol Rhinol Laryngol.* 2007;116:81–84. https://doi.org/10.1177/000348940711600201.
31. Pribuisiene R, Pribuisis K, Liutkevicius V, et al. Glottal function index questionnaire for screening of pediatric dysphonia. *Int J Pediatr Otorhinolaryngol.* 2019;123:97–101. https://doi.org/10.1016/j.ijporl.2019.04.045.
32. Fujiki RB, Thibeault SL. Examining relationships between GRBAS ratings and acoustic, aerodynamic and patient-reported voice measures in adults with voice disorders. *J Voice.* 2023;37:390–397. https://doi.org/10.1016/j.jvoice.2021.02.007.
33. Torabi H, Ansari NN, Zamani F, et al. Glottal function index: validity and reliability of the persian language version in patients with voice disorders. *J Voice.* 2023;37:140.e1–140.e6. https://doi.org/10.1016/j.jvoice.2020.10.013.
34. Spellman J, Coulter M, Roth C, et al. Prevalence, characteristics and impact of dysphonia in US marine corps drill instructors. *J Voice.* 2020;34:694–701. https://doi.org/10.1016/j.jvoice.2019.02.015.
35. Hoffman MR, Vandiver B, Derise N, et al. Effect of medialization on dyspnea index in unilateral vocal fold paralysis. *Otolaryngol Head Neck Surg.* 2022;167:327–333. https://doi.org/10.1177/01945998211056515.
36. Bick E, Dumberger LD, Farquhar DR, et al. Does voice therapy improve vocal outcomes in vocal fold atrophy? *Ann Otol Rhinol Laryngol.* 2021;130:602–608. https://doi.org/10.1177/0003489420952464.
37. Vaiciukynas E, Verikas A, Gelzinis A, et al. Fusing voice and query data for non-invasive detection of laryngeal disorders. *Expert Syst Appl.* 2015;42:8445–8453. https://doi.org/10.1016/j.eswa.2015.07.001.
38. Uloza V, Padervinskis E, Vegiene A, et al. Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening. *Eur Arch Otorhinolaryngol.* 2015;272:3391–3399. https://doi.org/10.1007/s00405-015-3708-4.

39. Ulozaite-Staniene N, Petrauskas T, Šaferis V, et al. Exploring the feasibility of the combination of acoustic voice quality index and glottal function index for voice pathology screening. *Eur Arch Otorhinolaryngol.* 2019;276:1737–1745. https://doi.org/10.1007/S00405-019-05433-5.
40. Fahed VS, Doheny EP, Busse M, et al. Comparison of acoustic voice features derived from mobile devices and studio microphone recordings. *J Voice.* 2022. https://doi.org/10.1016/j.jvoice.2022.10.006.S0892-1997(22)00312-5.
41. Maryn Y, Ysenbaert F, Zarowski A, et al. Mobile communication devices, ambient noise, and acoustic voice measures. *J Voice.* 2017;31:248.e11–248.e23. https://doi.org/10.1016/j.jvoice.2016.07.023.
42. Grillo EU, Brosious JN, Sorrell SL, et al. Influence of smartphones and software on acoustic voice measures. *Int J Telerehabil.* 2016;8:9–14. https://doi.org/10.5195/ijt.2016.6202.
43. Grillo EU, Wolfberg J. An assessment of different praat versions for acoustic measures analyzed automatically by VoiceEvalU8 and manually by two raters. *J Voice.* 2023;37:17–25. https://doi.org/10.1016/j.jvoice.2020.12.003.
44. Uloza V, Ulozaite-Staniene N, Petrauskas T. An iOS-based VoiceScreen application: feasibility for use in clinical settings-a pilot study. *Eur Arch Otorhinolaryngol.* 2023;280:277–284. https://doi.org/10.1007/s00405-022-07546-w.
45. Shabman S, Pushpavathi M. Effect of gender on acoustic voice quality index 02.03 and dysphonia severity index in Indian normophonic adults. *Indian J Otolaryngol Head Neck Surg.* 2022;74:5052–5059. https://doi.org/10.1007/s12070-021-02712-8.
46. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions.* 3rd ed. John Wiley & Sons; 2013.
47. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. guideline elaborated by the committee on phoniatrics of the european laryngological society (ELS). *Eur Arch Otorhinolaryngol.* 2001;258:77–82. https://doi.org/10.1007/S004050000299.
48. Barsties V, Latoszek B, Ulozaitė-Stanienė N, et al. Diagnostic accuracy of dysphonia classification of DSI and AVQI. *Laryngoscope.* 2019;129:692–698. https://doi.org/10.1002/lary.27350.
49. Uloza V, Ulozaitė-Stanienė N, Petrauskas T, et al. Reliability of universal-platform-based voice screen application in AVQI measurements captured with different smartphones. *J Clin Med.* 2023;12:4119. https://doi.org/10.3390/jcm12124119.
50. McHugh ML. Multiple comparison analysis testing in ANOVA. *Biochem Med.* 2011;21:203–209. https://doi.org/10.11613/bm.2011.029.
51. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143:29–36. https://doi.org/10.1148/radiology.143.1.7063747.
52. Barsties v. Latoszek B, Ulozaitė-Stanienė N, Maryn Y, et al. The influence of gender and age on the acoustic voice quality index and dysphonia severity index: a normative study. *J Voice.* 2019;33:340–345. https://doi.org/10.1016/j.jvoice.2017.11.011.
53. Batthyany C, Maryn Y, Trauwaen I, et al. A case of specificity: How does the acoustic voice quality index perform in normophonic subjects? *Appl Sci.* 2019;9:2527. https://doi.org/10.3390/APP9122527.
54. Jayakumar T, Benoy JJ, Yasin HM. Effect of age and gender on acoustic voice quality index across lifespan: a cross-sectional study in Indian population. *J Voice.* 2022;36:436.e1–436.e8. https://doi.org/10.1016/j.jvoice.2020.05.025.
55. Dollaghan CA. *The Handbook for Evidence-Based Practice in Communication Disorders.* Baltimore, MD, US: Paul H. Brookes Publishing Co; 2007.
56. Mat Baki M, Wood G, Alston M, et al. Reliability of OperaVOX against multidimensional voice program (MDVP). *Clin Otolaryngol.* 2015;40:22–28. https://doi.org/10.1111/coa.12313.
57. Cesari U, De Pietro G, Marciano E, et al. Voice disorder detection via an m-health system: Design and results of a clinical study to evaluate Vox4Health. *Biomed Res Int.* 2018;2018:8193694. https://doi.org/10.1155/2018/8193694.
58. Kojima T, Fujimura S, Hori R, et al. An innovative voice analyzer "VA" smart phone program for quantitative analysis of voice quality. *J Voice.* 2019;33:642–648. https://doi.org/10.1016/j.jvoice.2018.01.026.
59. Fujimura S, Kojima T, Okanoue Y, et al. Real-time acoustic voice analysis using a handheld device running android operating system. *J Voice.* 2020;34:823–829. https://doi.org/10.1016/j.jvoice.2019.05.013.