

Article

Not peer-reviewed version

---

# Reliability of Universal-Platform-Based VoiceScreen Application in AVQI Measurements Captured With Different Smartphones

---

[Virgilijus Uloza](#) , [Nora Ulozaite Staniene](#) <sup>\*</sup> , Tadas Petrauskas , [Kipras Pribuišis](#) , [Tomas Blažauskas](#) , [Robertas Damaševičius](#) , [Rytis Maskeliūnas](#)

Posted Date: 7 June 2023

doi: 10.20944/preprints202306.0540.v1

Keywords: VoiceScreen app; dysphonia screening; AVQI; smartphones



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Reliability of Universal-Platform-Based *VoiceScreen* Application in AVQI Measurements Captured with Different Smartphones

Virgilijus Uloza <sup>1</sup>, Nora Ulozaite-Stanienė <sup>1,\*</sup>, Tadas Petrauskas <sup>1</sup>, Kipras Pribuišis <sup>1</sup>,  
Tomas Blažauskas <sup>2</sup>, Robertas Damaševičius <sup>2</sup> and Rytis Maskeliūnas <sup>2</sup>

<sup>1</sup> Department of Otorhinolaryngology, Lithuanian University of Health Sciences, 50061 Kaunas, Lithuania

<sup>2</sup> Faculty of Informatics, Kaunas University of Technology, 51368 Kaunas, Lithuania

\* Correspondence: nora.ulozaite@ismuni.lt

**Abstract:** The aim of the study was to develop the universal-platform-based (UPB) application suitable for different smartphones for estimation of the Acoustic Voice Quality Index (AVQI) and evaluate its reliability in AVQI measurements and normal and pathological voice differentiation. Our study group consisted of 135 adult individuals, including 49 with normal voices and 86 patients with pathological voices. The developed UPB “*VoiceScreen*” application installed on five iOS and Android smartphones was used for AVQI estimation. The AVQI measures calculated from voice recordings obtained from a reference studio microphone were compared with AVQI results obtained using smartphones. The diagnostic accuracy of differentiating normal and pathological voice was evaluated by applying receiver-operating characteristics. The one-way ANOVA analysis did not detect statistically significant differences between mean AVQI scores revealed using a studio microphone and different smartphones ( $F=0.759$ ;  $p=0.58$ ). Almost perfect direct linear correlations ( $r=0.991-0.987$ ) were observed between the AVQI results obtained with a studio microphone and different smartphones. An acceptable level of precision of AVQI in discriminating between normal and pathological voices was yielded, with areas under the curve (AUC) displaying  $0.834-0.862$ . There were no statistically significant differences between the AUCs ( $p>0.05$ ) obtained from studio and smartphone microphones. The significant difference revealed between the AUCs was only  $0.028$ . The UPB “*VoiceScreen*” application represents an accurate and robust tool for voice quality measurement and normal vs. pathological voice screening purposes, demonstrating the potential to be used by patients and clinicians for voice assessment employing both iOS and Android smartphones.

**Keywords:** VoiceScreen app; dysphonia screening; AVQI; smartphones

## 1. Introduction

Mobile communication devices such as smartphones or tablets are widely available to most of the global population, with an expected number of smartphone subscriptions to reach about 7,145 billion by 2024 [1]. The increasing number of validated applications for smartphones in the field of general otorhinolaryngology and especially in a field related to voice assessment and management of voice disorders is permanently monitored in the literature [2–6]. Advances in smartphone technology and microphone quality offer an affordable and accessible alternative to studio microphones traditionally used for speech analysis, thus providing an effective tool for assessing, detecting, and caring for voice disorders [7–9].

The combination of variables in smartphone hardware and software may lead to differences between voice quality measures. Whether acoustic voice features recorded using smartphones sufficiently match a current gold standard for remote monitoring and clinical assessment with a studio microphone remains uncertain [7,10,11]. Some controversies on this matter in the literature

still exist. Several studies found that smartphones' provided voice recordings and derived acoustic voice quality parameters are comparable to those derived using standard studio microphones [8,12–14]. On the other hand, using some other acoustic voice quality parameters can be discouraging [15]. Two recent studies found none of the studied smartphones could replace the professional microphone in a voice recording to evaluate the six parameters analyzed, except for  $f_0$  and jitter. Moreover, passing a voice signal through a telecom channel induced filter and noise effects that significantly impacted common acoustic voice quality measures [16,17].

Nowadays, multiparametric models for voice quality assessment are generally accepted to be more reliable and valid than single-parameter measures because they demonstrate stronger correlations with auditory-perceptual voice evaluation and are more representative of daily use patterns. For example, the Acoustic Voice Quality Index (AVQI) is a six-variable acoustic model for the multiparametric measurement evaluating both the voiced parts of a continuous speech fragment and sustained vowel [a:] developed by Maryn et al. in 2010 [18,19].

Multiple studies across different languages have attested to the reliability of AVQI as a clinical voice quality evaluation tool. High consistency, concurrent validity, test-retest reliability, high sensitivity to voice quality changes through voice therapy, utility in discriminating across the perceptual levels of dysphonia severity, and adequate diagnostic accuracy with good discriminatory power of the AVQI in differentiating between normal and abnormal voice qualities were observed [19–26]. It is noteworthy that several studies have reported that gender and age do not affect the overall AVQI value, thus proving the perspectives for further generalization of this objective and quantitative voice quality measurement [26–29]. Therefore, nowadays, AVQI is considered a recognized-around-the-globe multiparametric construct of voice quality assessment for its clinical and research applications [30–32].

Several previous studies have proved the suitability of using smartphone voice recordings performed both in acoustically treated sound-proof rooms or in ordinary users' environments to estimate the AVQI [9,11,14,26,33,34]. However, just a few studies in the literature provide data about AVQI realization using different applications for mobile communication devices [9,26,35].

The study by Grillo et al. in 2020 presented the application (*VoiceEvalU8*) which provides an automatic option for the reliable calculation of several acoustic voice measures and AVQI on iOS and Android smartphones using Praat source code and algorithms [35]. A user-friendly application/graphical user interface for the Kannada-speaking population was proposed by Shabnam et al. in 2022. That application provides a simplified output for AVQI cut-off values to depict the AVQI-based severity of dysphonia which can be comprehensible by patients with voice disorder and health professionals [26]. The multilingual "*VoiceScreen*" application developed by Uloza et al. allows voice recording in clinical settings, automatically extracting acoustic voice features, estimating the AVQI result and displaying it alongside a recommendation to the user [9]. However, the "*VoiceScreen*" application runs the iOS operating system, and that feature limits usability only to iPhones, tablets, etc.

The results of the studies mentioned above enabled us to presume the feasibility of voice recordings captured with different smartphones for the estimation of AVQI. Consequently, the current research was designed to answer the following questions regarding the possibility of a smartphone-based "*VoiceScreen*" application for AVQI estimation: (1) are the average AVQI values estimated by different smartphones consistent and comparable; (2) is the diagnostic accuracy of different smartphones estimated AVQIs relevant to differentiate normal and pathological voices. We hypothesize that using different smartphones for voice recordings and estimation of AVQI will be feasible for the quantitative voice assessment.

Therefore, the present study aimed to develop the universal-platform-based (UPB) application suitable for different smartphones for estimation of AVQI and evaluate its reliability in AVQI measurements and normal/pathological voice differentiation.

2. Materials and Methods

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki of 1975, and the protocol was approved by the Kaunas Regional Ethics Committee for Biomedical Research (2022-04-20 No. BE-2-49).-

The study group consisted of 135 adult individuals, 58 men and 77 women. The mean age of the study group was 42.9 (SD 15.26) years. They were all examined at the Department of Otolaryngology of the Lithuanian University of Health Sciences, Kaunas, Lithuania.

The pathological voice subgroup consisted of 86 patients, 42 men and 44 women: with a mean age of 50.8 years (SD 14.3). They presented with a relatively common and clinically discriminative group of laryngeal diseases and related voice disturbances, i.e., benign and malignant mass lesions of the vocal folds and unilateral paralysis of the vocal fold. The normal voice subgroup consisted of 49 selected healthy volunteer individuals, 16 men and 33 women: mean age 31.69 (SD 9.89) years. This subgroup was collected following three criteria to define a vocally healthy subject: (1) all selected subjects considered their voice as normal and had no actual voice complaints, no history of chronic laryngeal diseases or voice disorders; (2) no pathological alterations in the larynx of the healthy subjects were found during video laryngoscopy; and (3) all these voice samples were evaluated as normal voices by otolaryngologists working in the field of voice. Demographic data of the study group and diagnoses of the pathological voice subgroup are presented in Table 1.

Table 1. Demographic data of the study group.

Diagnosis	n	Age	
		Mean	SD
Normal voice	49	31.69	9.89
Mass lesions of vocal folds	49	44.39	12.4
Vocal fold cancer	11	65.09	7.71
Chronic hyperplastic laryngitis	10	55.9	7.34
Laryngeal paralysis	10	45.6	13.47
Functional dysphonia	2	39	24.04
Reflux laryngitis	2	57	15.56
Parkinson's disease	2	71.5	9.19
Total	135	42.92	15.26

Abbreviation: SD – standard deviation.

No correlations between the subject's age, sex, and AVQI measurements were found in the previous study [27] . Therefore, in the present study, the control and patient groups were considered suitable for AVQI-related data analysis despite these groups not being matched by sex and age.

Original Voice Recordings

Voice samples from each subject 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 at a 10.0 cm distance from the mouth, keeping a 90° microphone-to-mouth angle. Each participant was asked to complete two vocal tasks, which were digitally recorded. The tasks consisted of (1) sustaining phonation of the vowel sound [a:] for at least 4 s duration and (2) reading a phonetically balanced text segment in Lithuanian “Turėjo senelė žilą oželį” (“The granny had a small grey goat”). The participants completed both vocal tasks at a personally comfortable loudness and pitch. All voice recordings were

captured with Audacity recording software (<https://www.audacityteam.org/>) at a sampling frequency of 44.1 kHz and exported in a 16-bit depth lossless “wav” audio file format into the computer’s hard disk drive (HDD).

#### *Auditory-Perceptual Evaluation*

Five experienced physicians-laryngologists, who were all native Lithuanians, served as the rater panel. Blind to all relevant information regarding the subject (i.e., identity, age, gender, diagnosis, and disposition of the voice samples), they performed auditory-perceptual evaluations to quantify the vocal deviation, judging the voice samples into four ordinal severity classes of Grade from the GRBAS scale (i.e., 0 = normal, 1 = slight, 2 = moderate, 3 = severe dysphonia [36]. A detailed description of the auditory-perceptual evaluation is presented elsewhere [21].

#### *Transmitting Studio Microphone Voice Recordings to Smartphones*

The impact on voice recordings caused by technical differences in studio and smartphone microphones was averted by applying the filtration (equalization) of the already recorded flat frequency audio using the data from the smartphone frequency response curves. The filtered result would represent audio recorded with the selected smartphone. Using this method, the only variable affected is the frequency response, keeping other variables, i.e., room reflections, distance to the microphone, directionality, user loudness, and other variables, constant. Ableton DAW (digital audio workstation) was implemented as an editing environment, and the VST (virtual studio plugin) plugin MFreeformEqualizer by MeldaProduction (<https://www.meldaproduction.com/MFreeformEqualizer/features>) was used to import the frequency response datasets and equalize the frequencies according to the required frequency response. The MFreeformEqualizer filter quality was set to the extreme (highest available), with 0% curve smoothing. All the audio files were then re-exported as 44100 Hz 16-bit wav files.

With this method, the digital voice recordings obtained with a studio microphone were directly transmitted to different smartphones, avoiding the surrounding environment’s impact.

#### *AVQI Estimation*

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 of the continuous speech were manually extracted and concatenated to the medial 3 s of sustained [a] phonation. The voice samples were concatenated for auditory-perceptual judgment in the following order: text segment, a 2-second pause, followed by the 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>.

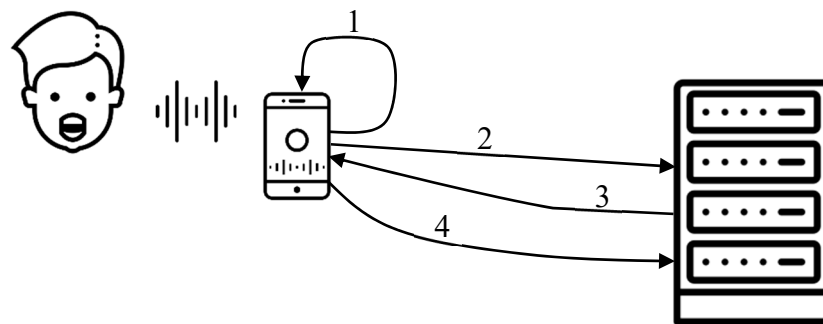
#### *Development of a Universal-Platform-Based “VoiceScreen” Application for Automated AVQI Estimation*

The “VoiceScreen” application for use with iOS operating devices was developed in the initial stage. Background noise monitoring, voice recording, and developed automated AVQI calculation were implemented in the application. Consequently, the “VoiceScreen” application allows voice recording, automatically extracts acoustic voice features and displays the AVQI result alongside a recommendation to the user [9].

The upgraded UPB version of the “VoiceScreen” application, suitable for iOS and Android devices, was elaborated in the next stage. In this case, the calculation of the AVQI and its characteristics is performed on the server; therefore, the computationally costly sound processing is not dependent on the user’s device’s computational capabilities. We use the Flutter framework (<https://flutter.dev/>) to create our client application. It allows compiling applications for different platforms (devices and their operating systems) from a single code base. The framework ensures that the same algorithms run on different devices and that no new software errors are introduced while



porting the application. Currently, our application works with both iOS and Android devices. Figure 1 shows the structure of the system. The numbers in the picture depict the flow of the operations.



**Figure 1.** Structure of the system and flow of the operations.

In the first step, the given smartphone (iOS or Android) records sound waves acquired while saying given phrases aloud. The sound waves are preprocessed (see Step 1 in Figure 1) in real-time. The preprocessing aims to clean the sound waves from pauses and ensure the minimum amount of sound suitable for further analysis. Step 2 sends the preprocessed sound wave to the server for further analysis. The server runs a Linux operating system and provides web services for software in Python. That software is based on Praat (<https://www.fon.hum.uva.nl/praat/>) application ported into a Python library by the Parselmouth project (<https://parselmouth.readthedocs.io/>). We use this library to calculate AVQI and other sound characteristics used in AVQI calculation. In step 3, the AVQI index and the related data are returned to the smartphone and displayed to the user. Step 4 is optional. If the user chooses to save the results, the sound waves and calculated characteristics are saved into the server database. No personal data relating to a specific person with the calculated AVQI and its parameters is saved on a server.

In the present study, the UPB “VoiceScreen” application was installed in five different smartphones (namely, iPhone Pro Max 13, iPhone SE (iOS operating system), OnePlus 9 PRO, Samsung S22 Ultra, Huawei P50 pro (Android operating system)) used for AVQI estimation. The AVQI measures estimated with the “VoiceScreen” application from voice recordings obtained from a flat frequency response studio microphone AKG Perception 220, were compared with AVQI results obtained using these smartphone devices.

### Statistical Analysis

Statistical analysis was performed using IBM SPSS Statistics for Windows, version 20.0 (Armonk, NY: IBM Corp.) and MedCalc Version 20.118 (Ostend, BE: MedCalc Software Ltd.). The chosen level of statistical significance was 0.05.

The data distribution was determined according to the normality law by applying the Shapiro-Wilk test of normality and calculating the coefficients of skewness and kurtosis. Student’s *t*-test was used to test the equality of means in normally distributed data [37]. An analysis of variance (ANOVA) was employed to determine if there were significant differences between the multiple means of the independent groups [38]. Cronbach’s alpha was used to measure the internal consistency of measures [39]. Pearson’s correlation coefficient was applied to assess the linear relationship between variables obtained from continuous scales. Spearman’s correlation coefficient was used to determine the relationship in ordinal results. Receiver operating characteristic (ROC) curves were used to obtain the optimal sensitivity and specificity at optimal AVQI cut-off points. The “area under the ROC curve” (AUC) served to calculate the possible discriminatory accuracy of AVQI performed with a studio microphone and different smartphones. A pairwise comparison of ROC curves, as described by De Long et al., was used to determine if there was a statistically significant difference between two or more variables when categorizing normal/pathological voice [40].

3. Results

*Raters’ Perceptual Evaluation Outcomes*

The rater panel demonstrated excellent inter-rater agreement (Cronbach’s  $\alpha = 0.967$ ) with a mean intra-class correlation coefficient of 0.967 between five raters (from 0.961 to 0.973).

*AVQI Evaluation Outcomes*

An individual smartphone AVQI evaluation displayed excellent agreement by achieving a Cronbach’s alpha of 0.984. The inter-smartphone AVQI measurements’ reliability was excellent, with an average Intra-class Correlation Coefficient (ICC) of 0.983 (ranging from 0.979 to 0.987).

The mean AVQI scores provided by different smartphones and a studio microphone can be observed in Table 2.

**Table 2.** Comparison of the mean AVQI results obtained with different smartphones and studio microphone.

Microphone	n	Mean AVQI	Std. Deviation	F	p
AKG Perception 220	135	3.43	1.83	0.759	0.58
iPhone SE		3.56	1.86		
iPhone Pro Max 13		3.16	1.83		
Huawei P50 pro		3.37	1.96		
Samsung S22 Ultra		3.52	1.93		
OnePlus 9 PRO		3.42	1.86		

Abbreviation: AVQI – acoustic voice quality index.

As shown in Table 2, the one-way ANOVA analysis did not detect statistically significant differences between mean AVQI scores revealed using different smartphones ( $F = 0.759$ ;  $p = 0.58$ ). Further Bonferroni analysis reaffirmed the lack of difference between the AVQI scores obtained from different smartphones ( $p = 1.0$ , estimated Bonferroni’s  $p$  for statistically significant difference  $p = 0.01$ ). The mean AVQI differences ranged from 0.01 to 0.4 points when comparing different smartphones.

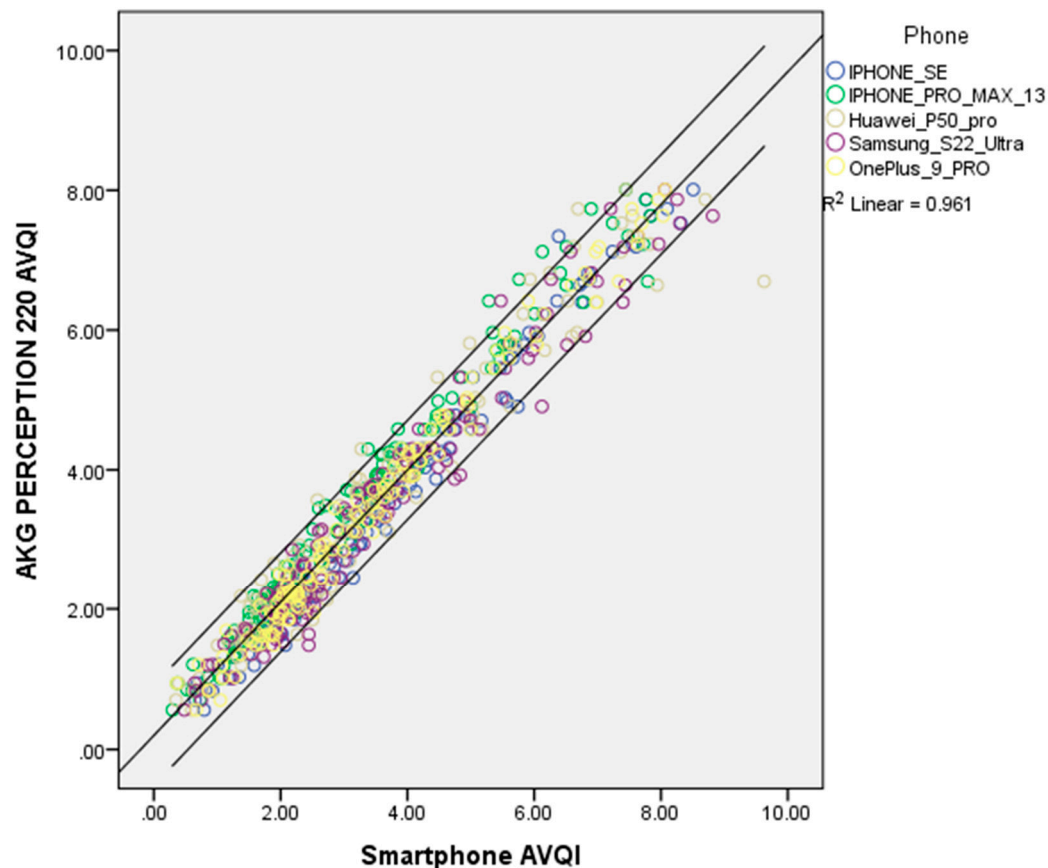
Almost perfect direct linear correlations were observed between the AVQI results obtained with a studio microphone and different smartphones. Pearson’s correlation coefficients ranged from 0.991 to 0.987 and can be observed in Table 3.

**Table 3.** Correlations of AVQI scores obtained with 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	0.991	0.987	0.970	0.979	0.992
	p	.001	.001	.001	.001	.001
	n	135	135	135	135	135

Abbreviations: r- Pearsons’s correlation coefficient; p – statistical significance.

The relationships between the AVQI scores obtained with a studio microphone and different smartphones are graphically presented in Figure 2.



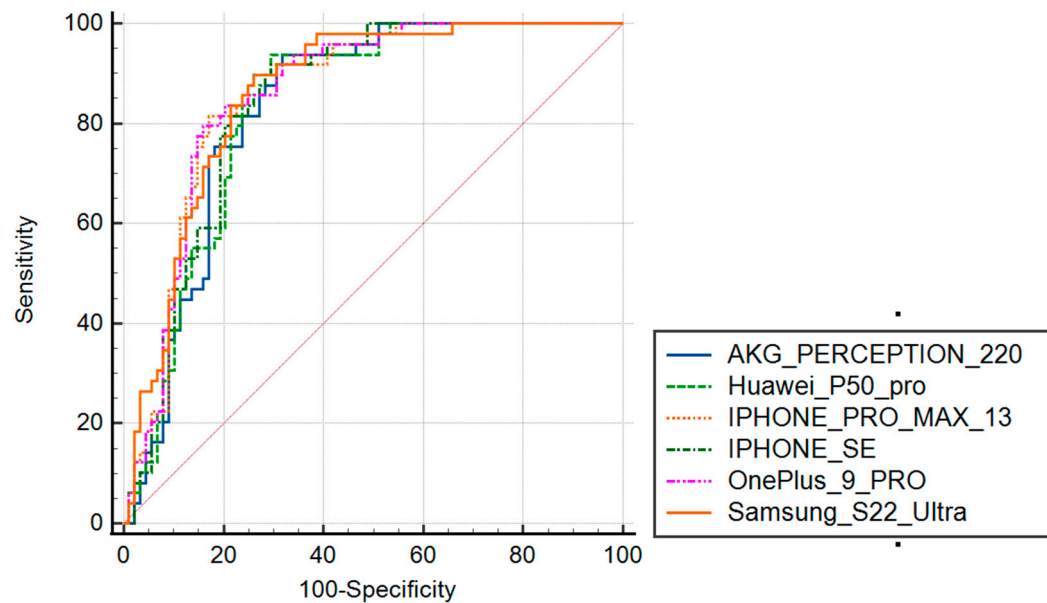
**Figure 2.** Scatterplot illustrating the correlation between the AVQI results obtained from the studio and different smartphones with a 95% confidence interval.

As demonstrated in Figure 2, it is evident that AVQI results obtained with different smartphones closely resemble the AVQI results obtained with a studio microphone, with very few data points outside of the 95% confidence interval ( $R^2 = 0.961$ ). Therefore, it is safe to conclude that the AVQI scores obtained with smartphones are directly compatible with the ones obtained with the reference studio microphone.

#### *The Normal vs. Pathological Voice Diagnostic Accuracy of the AVQI Using Different Smartphones*

First, the ROC curves of AVQI obtained from a studio microphone and different smartphone voice recordings were inspected visually to identify optimum cut-off scores according to general interpretation guidelines [41]. All of the ROC curves were visually almost identical and occupied the largest part of the graph, clearly revealing their respectable power to discriminate between normal and pathological voices. Figure 3.





**Figure 3.** ROC curves illustrating the diagnostic accuracy of studio and different smartphone microphones in discriminating normal/pathological voice.

Second, as revealed by the AUC statistics analysis, a high level of precision of the AVQI in discriminating between normal and pathological voices was yielded with the suggested AUC = 0.800 threshold. The results of the ROC statistical analysis are presented in Table 4.

**Table 4.** Statistics illustrating the accuracy the AVQI differentiating normal and pathological voices recorded using studio and different smartphone microphones.

AVQI	AUC	Cut-off	Sensitivity %	Specificity %	Youden-index J
AKG Perception 220	0.834	3.27	93.88	68.18	0.62
iPhone SE	0.844	3.23	91.84	70.45	0.62
iPhone Pro Max 13	0.858	2.14	81.63	82.95	0.65
Huawei P50 pro	0.835	3.08	93.88	70.45	0.64
Samsung S22 Ultra	0.862	2.93	89.8	73.86	0.64
OnePlus 9 PRO	0.86	2.3	79.59	84.09	0.64

Abbreviations: AVQI acoustic voice quality index, AUC area under the curve.

As demonstrated in Table 4, the ROC analysis determined the optimal AVQI cut-off values for distinguishing between normal and pathological voices for each smartphone. All employed microphones passed the proposed 0.8 AUC threshold and revealed an acceptable Youden-index value.

Third, a pairwise comparison of the significance of the differences between the AUCs revealed in the present study is presented in Table 5.

**Table 5.** A pairwise comparison of the significance of differences between the AUCs.

p	AKG Perception 220	iPhone SE	iPhone Pro Max 13	Huawei P50 pro	Samsung S22 Ultra	OnePlus 9 PRO
AKG Perception 220	-	0.163	0.099	0.966	0.11	0.086

iPhone SE	0.163	-	0.367	0.579	0.282	0.863
iPhone Pro Max 13	0.099	0.367	-	0.268	0.718	0.863
Huawei P50 pro	0.966	0.579	0.268	-	0.223	0.256
Samsung S22 Ultra	0.11	0.282	0.718	0.223	-	0.863
OnePlus 9 PRO	0.086	0.863	0.863	0.256	0.863	-

Abbreviations: AUC area under the curve.

As shown in Table 5, a comparison of the AUCs-dependent ROC curves (AVQI measurements obtained from studio microphone and different smartphones), according to the test of DeLong et al., confirmed no statistically significant differences between the AUCs ( $p > 0.05$ ). The most considerable observed difference between the AUCs was only 0.028. These results confirm compatible results of the AVQI's diagnostic accuracy in differentiating normal vs. pathological voice when using voice recordings from a studio microphone and different smartphones.

#### 4. Discussion

In the present study, the novel UPB “VoiceScreen” application for the estimation of AVQI and detection of voice deteriorations in patients with various voice disorders and healthy controls was tested for the first time simultaneously with different smartphones. The AVQI was chosen for voice quality assessment because of some essential favorable features of this multiparametric measurement: the less vulnerability of the AVQI to environmental noise compared to other complex acoustic markers, the robustness of the AVQI regarding the interaction between acoustic voice quality measurements and room acoustics; no significant differences within-subject for both women and men when comparing the AVQI across different voice analysis programs [11,14,42]. Another essential attribute of the AVQI is that Praat is the only freely available program that estimates the AVQI. That eliminates the impact of possible software differences on AVQI computation.

In the present study, the results of the ANOVA analysis did not detect statistically significant differences between mean AVQI scores revealed using different smartphones ( $F = 0.759$ ;  $p = 0.58$ ). Moreover, the mean AVQI differences ranged from 0.01 to 0.4 points when comparing AVQI estimated with different smartphones, thus establishing a low level of variability. This corresponds with a value of 0.54 for the absolute retest difference of AVQI values proposed by Barsties and Maryn in 2013 [19,43]. Consequently, these outcomes of AVQI measurements with different smartphones were considered neither statistically nor clinically significant, justifying the possibility of practical use of the UPB “VoiceScreen” app.

The correlation analysis showed that all AVQI measurements were highly correlated (Pearson's  $r$  ranged from 0.991 to 0.987) across the devices used in the present study. This concurred with the literature data on the high correlation between acoustic voice features derived from studio microphones and smartphones and examined both for control participants and synthesized voice data [7,12–14].

Furthermore, analysis of the results revealed that the AVQI showed a remarkable ability to discriminate between normal and pathological voices as determined by auditory-perceptual judgment. The ROC analysis determined the optimal AVQI cut-off values for distinguishing between normal and pathological voices for each smartphone used. A remarkable precision of AVQI in discriminating between normal and pathological voices was yielded (AUC 0.834 – 0.862), resulting in an acceptable balance between sensitivity and specificity. These findings suggest that AVQI is a reliable tool in differentiating normal/pathological voice independently of the voice recordings from tested studio microphone and different smartphones. The comparison of the AUCs-dependent ROC curves (AVQI measurements obtained from studio microphone and different smartphones) demonstrated no statistically significant differences between the AUCs ( $p > 0.05$ ), with the largest

revealed difference between the AUCs of only 0.028. These results confirm compatible results of the AVQI diagnostic accuracy in differentiating normal vs. pathological voice when using voice recordings from studio microphone and different smartphones and present remarkable importance from a practical point of view.

Several limitations of the present study have to be considered. Despite the encouraging results of the AVQI measurements, some individual discrepancies between AVQI results revealed with different smartphones still exist. Therefore, further research in a wide diversity of voice pathologies, including functional voice disorders, is needed to ensure the maximum comparability of acoustic voice features derived from voice recordings obtained with mobile communication devices and reference studio microphones. Moreover, it is also essential to evaluate the impact of the voice recording environment on the AVQI estimation in real clinical settings by performing simultaneous voice recordings with different smartphones, making possible the results and improvements to be employed in healthcare applications.

Summarizing the results of the previous and present studies allows for the presumption that the performance of the novel UPB “VoiceScreen” app using different smartphones represents an adequate and compatible performance of AVQI estimation.

However, it is important to note that due to existing differences in recording conditions, microphones, hardware, and software, the results of acoustic voice quality measures may differ between recording systems [11]. Therefore, using the UPB “VoiceScreen” app with some caution is advisable. For voice screening purposes, it is more reliable to perform AVQI measurements using the same device, especially when performing repeated measurements. Moreover, these bits of advice should be considered when comparing data of acoustic voice analysis between different voice recording systems, i.e., different smartphones or other mobile communication devices, and when using them for diagnostic purposes or monitoring voice treatment outcomes.

## 5. Conclusions

The UPB “VoiceScreen” app represents an accurate and robust tool for voice quality measurement and normal vs. pathological voice screening purposes, demonstrating the potential to be used by patients and clinicians for voice assessment employing both iOS and Android smartphones.

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

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki of 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 subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mobile network subscriptions worldwide 2028. Available online: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> (Accessed on Apr 3, 2023).
2. Casale, M.; Costantino, A.; Rinaldi, V.; Forte, A.; Grimaldi, M.; Sabatino, L.; Oliveto, G.; Aloise, F.; Pontari, D.; Salvinelli, F. Mobile applications in otolaryngology for patients: An update. *Laryngoscope Investigative Otolaryngology* **2018**, *3*, 434.
3. Eleonora, M.C.T.; Lonigro, A.; Gelardi, M.; Kim, B.; Cassano, M. Mobile Applications in Otolaryngology: A Systematic Review of the Literature, Apple App Store and the Google Play Store. *Ann Otol Rhinol Laryngol* **2021**, *130*, 78–91, DOI 10.1177/0003489420940350. Available online: <https://doi.org/10.1177/0003489420940350>.

4. Grillo, E.U.; Wolfberg, J. An Assessment of Different Praat Versions for Acoustic Measures Analyzed Automatically by VoiceEvalU8 and Manually by Two Raters. *Journal of Voice* **2023** Jan;37(1):17-25. doi: 10.1016/j.jvoice.2020.12.003.
5. Boogers, L.S.; Chen, B.S.J.; Coerts, M.J.; Rinkel, R.N.P.M.; Hannema, S.E. Mobile Phone Applications Voice Tools and Voice Pitch Analyzer Validated With LingWAVES to Measure Voice Frequency. *Journal of Voice* **2022**, DOI 10.1016/j.jvoice.2022.10.013. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199722003186>.
6. Kojima, T.; Hasebe, K.; Fujimura, S.; Okanou, Y.; Kagoshima, H.; Taguchi, A.; Yamamoto, H.; Shoji, K.; Hori, R. A New iPhone Application for Voice Quality Assessment Based on the GRBAS Scale. *Laryngoscope* **2021**, 131, 580-582, DOI 10.1002/lary.28796.
7. Fahed, V.S.; Doheny, E.P.; Busse, M.; Hoblyn, J.; Lowery, M.M. Comparison of Acoustic Voice Features Derived From Mobile Devices and Studio Microphone Recordings. *Journal of Voice* **2022**, DOI 10.1016/j.jvoice.2022.10.006. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199722003125>.
8. Awan, S.N.; Shaikh, M.A.; Awan, J.A.; Abdalla, I.; Lim, K.O.; Misono, S. Smartphone Recordings are Comparable to “Gold Standard” Recordings for Acoustic Measurements of Voice. *Journal of Voice* **2023**, DOI 10.1016/j.jvoice.2023.01.031. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199723000310>.
9. 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, DOI 10.1007/s00405-022-07546-w.
10. Munnings, A.J. The Current State and Future Possibilities of Mobile Phone “Voice Analyser” Applications, in Relation to Otorhinolaryngology. *Journal of Voice* **2020**, 34, 527-532, DOI 10.1016/j.jvoice.2018.12.018. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199718302595>.
11. Maryn, Y.; Ysenbaert, F.; Zarowski, A.; Vanspauwen, R. Mobile Communication Devices, Ambient Noise, and Acoustic Voice Measures. *Journal of Voice* **2017**, 31, 248.e11-248.e23, DOI 10.1016/j.jvoice.2016.07.023. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199716301965>.
12. Kardous, C.A.; Shaw, P.B. Evaluation of smartphone sound measurement applications. *J Acoust Soc Am* **2014**, 135, EL186-192, DOI 10.1121/1.4865269.
13. Manfredi, C.; Lebacqz, J.; Cantarella, G.; Schoentgen, J.; Orlandi, S.; Bandini, A.; DeJonckere, P.H. Smartphones Offer New Opportunities in Clinical Voice Research. *Journal of Voice* **2017**, 31, 111.e1-111.e7, DOI 10.1016/j.jvoice.2015.12.020. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199716000059>.
14. Grillo, E.U.; Brosious, J.N.; Sorrell, S.L.; Anand, S. Influence of Smartphones and Software on Acoustic Voice Measures. *Int J Telerehabil* **2016**, 8, 9-14, DOI 10.5195/ijt.2016.6202.
15. Schaeffler, F.; Jannetts, S.; Beck, J.M. Reliability of clinical voice parameters captured with smartphones – measurements of added noise and spectral tilt. *Proceedings of the 20th Annual Conference of the International Speech Communication Association INTERSPEECH*, Graz, Austria **2019**, 2523-2527, DOI 10.21437/Interspeech.2019-2910. Available online: <https://ereseach.qmu.ac.uk/handle/20.500.12289/10013> (accessed on Apr 3, 2023).
16. Marsano-Cornejo, M.; Roco-Videla, Á. Comparison of the Acoustic Parameters Obtained With Different Smartphones and a Professional Microphone. *Acta Otorrinolaringol Esp (Engl Ed)* **2022**, 73, 51-55, DOI 10.1016/j.otoeng.2020.08.009.
17. Pommée, T.; Morsomme, D. Voice Quality in Telephone Interviews: A preliminary Acoustic Investigation. *Journal of Voice* **2022**, DOI 10.1016/j.jvoice.2022.08.027. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199722002685>.
18. 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, DOI 10.1016/j.jcomdis.2009.12.004. Available online: <https://www.sciencedirect.com/science/article/pii/S0021992409000884>.
19. Barsties, B.; Maryn, Y. [The Acoustic Voice Quality Index. Toward expanded measurement of dysphonia severity in German subjects]. *HNO* **2012**, 60, 715-720, DOI 10.1007/s00106-012-2499-9.
20. Hosokawa, K.; Barsties, B.; Iwahashi, T.; Iwahashi, M.; Kato, C.; Iwaki, S.; Sasai, H.; Miyauchi, A.; Matsushiro, N.; Inohara, H.; Ogawa, M.; Maryn, Y. Validation of the Acoustic Voice Quality Index in the Japanese Language. *Journal of Voice* **2017**, 31, 260.e1-260.e9, DOI 10.1016/j.jvoice.2016.05.010. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199716300789>.
21. Uloza, V.; Petrauskas, T.; Padervinskis, E.; Ulozaite, N.; Barsties, B.; Maryn, Y. Validation of the Acoustic Voice Quality Index in the Lithuanian Language. *Journal of Voice* **2017**, 31, 257.e1-257.e11, DOI 10.1016/j.jvoice.2016.06.002. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199716300716>.

22. Kankare, E.; Barsties V Latoszek, B.; Maryn, Y.; Asikainen, M.; Rorarius, E.; Vilpas, S.; Ilomäki, I.; Tyrmä, J.; Rantala, L.; Laukkanen, A. The acoustic voice quality index version 02.02 in the Finnish-speaking population. *Logoped Phoniatr Vocol* **2020**, *45*, 49-56, DOI 10.1080/14015439.2018.1556332.
23. Englert, M.; Lopes, L.; Vieira, V.; Behlau, M. Accuracy of Acoustic Voice Quality Index and Its Isolated Acoustic Measures to Discriminate the Severity of Voice Disorders. *Journal of Voice* **2022**, *36*, 582.e1-582.e10, DOI 10.1016/j.jvoice.2020.08.010. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199720302939>.
24. Yeşilli-Puzella, G.; Tadihan-Özkan, E.; Maryn, Y. Validation and Test-Retest Reliability of Acoustic Voice Quality Index Version 02.06 in the Turkish Language. *Journal of Voice* **2022**, *36*, 736.e25-736.e32, DOI 10.1016/j.jvoice.2020.08.021. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199720303222>.
25. Englert, M.; Latoszek, B.B.v.; Behlau, M. Exploring The Validity of Acoustic Measurements and Other Voice Assessments. *Journal of Voice* **2022**, DOI 10.1016/j.jvoice.2021.12.014. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199721004392>.
26. Shabnam, S.; Pushpavathi, M.; Gopi Sankar, R.; Sridharan, K.V.; Vasanthalakshmi, M.S. A Comprehensive Application for Grading Severity of Voice Based on Acoustic Voice Quality Index v.02.03. *Journal of Voice* **2022**, DOI 10.1016/j.jvoice.2022.08.013. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199722002454>.
27. Barsties v. Latoszek, B.; Ulozaitė-Stanienė, N.; Maryn, Y.; Petrauskas, T.; Uloza, V. The Influence of Gender and Age on the Acoustic Voice Quality Index and Dysphonia Severity Index: A Normative Study. *Journal of Voice* **2019**, *33*, 340-345, DOI 10.1016/j.jvoice.2017.11.011. Available online: <https://www.sciencedirect.com/science/article/pii/S089219971730468X>.
28. Batthyany, C.; Maryn, Y.; Trauwaen, I.; Caelenberghe, E.; van Dinther, J.; Zarowski, A.; Wuyts, F. A case of specificity: How does the acoustic voice quality index perform in normophonic subjects? *Applied Sciences (Switzerland)* **2019** Appl Sci <http://doi.org/10.3390/app9>
29. Jayakumar, T.; Benoy, J.J.; Yasin, H.M. Effect of Age and Gender on Acoustic Voice Quality Index Across Lifespan: A Cross-sectional Study in Indian Population. *Journal of Voice* **2022**, *36*, 436.e1-436.e8, DOI 10.1016/j.jvoice.2020.05.025. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199720301995>.
30. Jayakumar, T.; Benoy, J.J. Acoustic Voice Quality Index (AVQI) in the Measurement of Voice Quality: A Systematic Review and Meta-Analysis. *Journal of Voice* **2022**, DOI 10.1016/j.jvoice.2022.03.018. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199722000844>.
31. Batthyany, C.; Latoszek, B.B.V.; Maryn, Y. Meta-Analysis on the Validity of the Acoustic Voice Quality Index. *Journal of voice : official journal of the Voice Foundation* **2022**. <https://doi.org/10.1016/J.JVOICE.2022.04.022>
32. Saeedi, S.; aghajanzadeh, M.; Khatoonabadi, a.r. A Literature Review of Voice Indices Available for Voice Assessment. *JRSR* **2022**, *9*, 151-155, DOI 10.30476/jrsr.2022.93362.1235.
33. Grillo, E.U.; Wolfberg, J. An Assessment of Different Praat Versions for Acoustic Measures Analyzed Automatically by VoiceEvalU8 and Manually by Two Raters. *Journal of Voice* **2020** Jan;37(1):17-25. doi: 10.1016/j.jvoice.2020.12.003. Epub 2020 Dec 29.
34. Uloza, V.; Ulozaitė-Stanienė, N.; Petrauskas, T.; Kregždytė, R. Accuracy of Acoustic Voice Quality Index Captured With a Smartphone - Measurements With Added Ambient Noise. *J Voice* **2021**, S0892-4, DOI 10.1016/j.jvoice.2021.01.025.
35. Grillo, E.U.; 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, DOI 10.1016/j.jvoice.2020.12.003.
36. Dejonckere, P.H.; Bradley, P.; Clemente, P.; Cornut, G.; Crevier-Buchman, L.; Friedrich, G.; Van De Heyning, P.; Remacle, M.; Woisard, V. 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, DOI 10.1007/s004050000299.
37. Senn, S.; Richardson, W. The first t-test. *Statist Med* **1994**, *13*, 785-803, DOI 10.1002/sim.4780130802. Available online: <https://doi.org/10.1002/sim.4780130802>.
38. McHugh, M.L. Multiple comparison analysis testing in ANOVA. *Biochem Med (Zagreb)* **2011**, *21*, 203-209, DOI 10.11613/bm.2011.029.
39. Cho, E. Making Reliability Reliable: A Systematic Approach to Reliability Coefficients. *Organ Res Methods* **2016**, *19*, 651-682, DOI 10.1177/1094428116656239. Available online: <https://doi.org/10.1177/1094428116656239>.
40. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29-36, DOI 10.1148/radiology.143.1.7063747.



41. Dollaghan, C.A. *The handbook for evidence-based practice in communication disorders*, Paul H. Brookes Publishing Co: Baltimore, MD, US, 2007; pp. xi, 169.
42. Bottalico, P.; Codino, J.; Cantor-Cutiva, L.C.; Marks, K.; Nudelman, C.J.; Skeffington, J.; Shrivastav, R.; Jackson-Menaldi, M.C.; Hunter, E.J.; Rubin, A.D. Reproducibility of Voice Parameters: The Effect of Room Acoustics and Microphones. *Journal of Voice* **2020**, *34*, 320-334, DOI 10.1016/j.jvoice.2018.10.016. Available online: <https://www.sciencedirect.com/science/article/pii/S0892199718304338>.
43. Lehnert, B.; Herold, J.; Blaurock, M.; Busch, C. 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, DOI 10.1007/s00405-022-07417-4.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.