Validation of the Acoustic Voice Quality Index in the Korean Language

*Geun-Hyo Kim, *Yeon-Woo Lee, †In-Ho Bae, ‡Hee-June Park, *Soo-Geun Wang, and §Soon-Bok Kwon, *Busan, †Yangsan, and ‡§South Korea

Summary: Objective. The purpose of this study was to explore the concurrent validity of the Acoustic Voice Quality Index (AVQI) in a Korean population. We investigated the feasibility of its cutoff values and diagnostic accuracy in discriminating between normal and dysphonic voices.

Materials and Methods. A total of 1,524 native Korean subjects with normal voices (n = 113) and with voice disorders (n = 1,411) were asked to speak and sustain the vowel "a" and to read the Korean text "Walk" aloud. A 2-second mid-speech portion of the sustained vowel sound and two sentences (with 26 syllables) were edited and concatenated, and the AVQI was calculated. Additionally, two auditory-perceptual assessments, namely the Grade, Roughness, Breathiness, Asthenia, Strain scale, and the Consensus Auditory-Perceptual Evaluation of Voice were used by five speech language pathologists to rate the severity of dysphonia.

Results. Both auditory-perceptual assessments showed high confidence levels among the five speech language pathologists. The AVQI correlated with grade (G) and overall severity (OS). There were statistically significant differences in AVQI, G, and OS between normal and pathological speech groups. In the receiver operating characteristic curve analysis, AVQI, G, and OS cutoff values were <3.33, <0.00, and <22.00, respectively. The receiver operating characteristic curve analysis indicated that AVQI had excellent diagnostic accuracy for discriminating between normal and dysphonic voices (area under the curve: 0.970-0.977).

Conclusions. We demonstrated the concurrent validity of AVQI as a promising tool for assessing overall voice quality and quantifying dysphonia in a Korean-speaking population.

Key Words: Acoustic Voice Quality Index–Voice–Dysphonia–GRBAS–CAPE-V.

INTRODUCTION

Dysphonia evaluations often rely on the integration of information from acoustic and auditory-perceptual evaluations.[1] These evaluations should comprise both sustained vowel sounds and verbalized sentences in clinical voice evaluations.[2–5] The results of vowel phonation and sentence reading may differ depending on the presence of and causes of voice disorders.[6–8] Quantifications of dysphonia are performed by MDVP, Dr. Speech, ADSV, SpeechTool, and other tools.[9–12] The voices of the patients are analyzed to confirm the effectiveness of laryngeal microsurgery, injection laryngoplasty, botulinum toxin treatment, and voice therapy.[13–16] The voice samples obtained during vowel phonation and sentence reading are analyzed to judge whether the patients' voices have recovered after treatment. Laryngeal examinations such as laryngeal endoscopy and laryngeal videostroboscopy can confirm the possibility of voice impairment, but are invasive and can be uncomfortable to the patient.[17,18] Alternatively, acoustic voice analysis is noninvasive and easy to use because there is no inconvenience to the patient.[19,20]

Auditory-perceptual assessment is considered to be the gold standard to verify the validity of other methods. Various auditory-perceptual ratings, such as the Grade, Roughness, Breathiness, Asthenia, Strain (GRBAS) scale, or the Consensus Auditory-Perceptual Assessment of Voice (CAPE-V), have been suggested to standardize auditory-perceptual assessments, improve their reliability, and to quantify dysphonia severity.[21–24] However, the reliability of the evaluation was based on the experience of the examiner with voice disorders in clinical practice, since auditory-perceptual assessments are subjective examinations.

To overcome these limitations, Maryn et al introduced the Acoustic Voice Quality Index (AVQI), which can quantify the severity of dysphonia.[25–29] Because judgments of the severity of voice disorders may differ remarkably between voice tasks, the AVQI was designed to quantify concatenated voice samples in vowel and sentence reading. The robustness of the AVQI in different languages and phonemic environments has been reported. Despite the different language and phonemic situations, there was a high correlation between AVQI and auditory-perceptual assessments.[30–35] A previous Korean study reported a similarly high correlation, but the number of subjects was relatively low compared to other studies.

In clinical practice, voice specialists should be able to distinguish between normal and pathological voices. Previous studies suggested cutoff values of AVQI for dysphonia populations using the receiver operating characteristic (ROC).[28,30,33,34] ROC curves are simple analytical methods and helpful for assessing the diagnostic capability of two or

Accepted for publication June 20, 2018.
Financial Disclosure: The authors have no funding or financial relationship to disclose
Declarations of interest: none.
From the *Department of Otorhinolaryngology-Head and Neck Surgery and Biomedical Research Institute, Pusan National University Hospital, Busan, South Korea; †Department of Otorhinolaryngology-Head and Neck Surgery, Pusan National University Yangsan Hospital, Yangsan, Gyeongsangnam-do, South Korea; ‡Department of Speech Rehabilitation, Choonhae College of Health Sciences, Ulsan, South Korea; and the §Department of Humanities, Language and Information, Pusan National University, Pusan, South Korea.
Address correspondence and reprint requests to Soon-Bok Kwon, Department of Humanities, Language and Information, Pusan National University, Pusan, South Korea. E-mail: voicesbkwon@gmail.com
Journal of Voice, Vol. 33, No. 6, pp. 948.e1–948.e9
0892-1997

https://doi.org/10.1016/j.jvoice.2018.06.007

more examination or for comparing the predictive power of two or more variables.[36–40] The cutoff values can be measured based on the two indices of "sensitivity" and "specificity," both of which are used in clinical examinations. The accuracy of the test depends on how well the test can distinguish the normal group from the group with the disease. Accuracy is calculated by the area under the ROC curve (AUC). An area of one represents a perfect test while an area of 0.5 represents an inadequate test (AUC < 0.5 = fail; 0.7–0.8 = acceptable; 0.8–0.9 = excellent; 0.9–1 = outstanding).[41]

In this study, we investigated the correlation between the AVQI and auditory-perceptual assessments using a large number of voice samples, and reported the cutoff and accuracy of the AVQI, grade (G), and overall severity (OS) of dysphonia between normal and pathological groups.

This study addressed the following research questions.

(1) Can AVQI be used as an objective measurement of the degree of dysphonia in Korean speakers?

(2) How closely does AVQI relate to the auditory-perception scale (G, OS)?

(3) How do G and OS correlate in the two auditory-perception scales?

(4) What are the cutoff values and the accuracy of the AVQI, G, and OS between normal and pathological speech status?

The purpose of this study was to investigate the concurrent validity of the AVQI compared to two standardized auditory-perceptual assessment tools in measuring the severity of dysphonia in a Korean population.

MATERIALS AND METHODS

Subjects

The initial study group included 1,604 subjects' voice samples. After calculating the signal-to-noise ratio (SNR) of the recordings, the sound files of 80 subjects were excluded from the study. Consequently, voice data of 1,524 native-speaking Korean subjects from the Pusan National University Hospital were used in this study. The characteristics of the study subjects are shown in Table 1. The clinical diagnoses were based on clinical evaluations via direct laryngoscopy and laryngeal videostroboscopy.

The study subjects were considered to be representative samples of the population with voice disorders, reflecting different ages and sexes, voice quality, and organic, functional, neurologic laryngeal pathologies. The focus of this study was to assess the concurrent validity of the AVQI in a Korean population in distinguishing between normal and dysphonic voices. The normal group was defined as subjects with no laryngeal disease or dysphonia and with an auditory-perceptual assessment of Grade 0.

TABLE 1.
The Primary Laryngological Diagnoses Included in the Present Study

Laryngeal status	Absolute number	% of total
Vocal nodules	330	21.7%
Presbylaryngis	261	17.1%
Vocal folds paralysis	198	13.0%
Vocal polyp	190	12.5%
Normal	113	7.4%
Papilloma	85	5.6%
PTC	62	4.1%
Glottic cancer	57	3.7%
Spasmodic dysphonia	53	3.5%
Vocal cyst	49	3.2%
Leukoplakia	40	2.6%
Functional dysphonia	33	2.2%
Sulcus vocalis	25	1.6%
LPRD	17	1.1%
Edema	10	0.7%
Total	1,524	100.0%

The subjects' voices were recorded during the regular voice evaluation, prior to surgical intervention, behavior voice therapy, or medication. The group consisted of 783 men and 741 women, ranging in age from 9 to 89 years (mean± standard deviation [SD]: 51.9± 15.7 years). The Institutional Review Board of Pusan National University Hospital approved this study (H-1801-012-063).

Voice recordings

We complied with the basic protocol of the European Laryngological Society for voice evaluation proposed by Dejonckere et al.[1] All voice samples were recorded using a Shure SM48 (Shure Inc), digitized at a sampling rate of 44.1 kHz and a quantization of 16 bits using the Computerized Speech Lab (CSL, model4500, KayPENTAX, Lincoln Park, NJ), and saved in WAV format.

Each voice sample was recorded in a quiet room with the ambient noise level kept below 40 dB. Subjects were instructed to vocalize the vowel /a/ at a comfortable pitch and loudness three times. Relatively more stable vowel recordings were selected for each voice sample concatenation. An active recording window was implemented to comprise only a mid-vowel portion of 2 seconds. Next, the subjects were asked to read two sentences from the Korean passage "Walk," which contains 26 syllables.[35,42,43] Audio recording, editing, and SNR measurement methods were similar to those presented in a previous study.[25] Sustained vowel and sentence files were concatenated into one single voice sample for the AVQI calculation.

Measurement of SNR

To measure the SNR and evaluate the voice quality, parts of pause and nonspeech components were subtracted from the concatenated voice samples. Previous studies have reported that the recommended, acceptable, and unacceptable levels of SNR are above 42 dB, above 30 dB, and below 30 dB, respectively.[44,45] The SNR ranged between 17.3 dB and 75.5 dB, with a mean of 43.4 dB. The recordings of 80 subjects were excluded from this study because their SNRs were below 30 dB. The voice samples of the 1,524 subjects were recorded with sufficient SNR and therefore, were determined acceptable for further analysis.

Auditory-perceptual assessments

We asked five speech-language pathologists (SLP), all native Koreans with more than 7 years of experience in voice quality assessment, to rate our voice samples (three were SLPs working in ENT clinics, and two were professors specializing in voice disorders). All the raters were blinded to the 1,524 subjects' information. In the auditory-perceptual assessment of dysphonia severity, concatenated voice samples were randomly presented. The samples were judged separately as to overall voice quality following GRBAS and CAPE-V. The first parameter, "Grade" (G), was based on a four-point Likert scale (normal: 0; mild: 1; moderate: 2; severe: 3), and was evaluated by the methods of the Japan Society of Logopedics and Phoniatrics.[46] The second parameter, "Overall severity" (OS), was based on the CAPE-V method which uses a visual analog scale of 100 mm, with anchoring points, as suggested by the Special Interest Division 3 Voice and Voice Disorders of the American Speech-Language-Hearing Association.[24]

It has been suggested to evaluate OS by CAPE-V when the following tasks are completed: sustained vowel /a/ and /i/ for 3–5 seconds each, saying six sentences aloud, and spontaneous speech.[22–24] This study focused on the analysis of AVQI and therefore did not include various utterance tasks as suggested by CAPE-V. The standardized CAPE-V sentences in Korean do not yet exist. The Korean sentences from 'Walk' used in this study are phonetically and phonologically balanced sentences, as shown in previous studies.[35,42,43]

All ratings were conducted in a single session. The recordings of 300 randomly selected subjects (20% of rated voice samples taken from the first session) were evaluated a second time (after two weeks) to assess intrarater variability. Scores of auditory-perceptual assessments between raters were compared to calculate inter-rater variability.

The Google questionnaire (for GRBAS) and a customized computer program (Digital CAPE-V) were used to facilitate the data analysis and management, and to save the ratings of the voice samples. The graphic examples for auditory-perceptual assessments are displayed in Figure 1. After listening to the audio files, raters provided scores for the auditory-perceptual assessments, which were then stored for statistical analysis.

Theory and calculations

Acoustic measure for AVQI measurement

We applied the Praat (Version 5.4.19) script of Maryn et al[25] for the automated detection, separation, and synthesis of the voiced components since the specific acoustic methods used in the present study are only valid for voiced components of the concatenated voice files. For the AVQI measurement, a 2-second mid-vowel component was attached to the chain of voiced text components resulting in a single sound file. Voice quality was quantified by analyzing the same utterance. Objective measurements of overall voice quality included calculating the AVQI on the synthesized acoustic signals.[29]

AVQI was designed as a multivariable correlate of dysphonia severity comprising various acoustic parameters. The AVQI was calculated by the following regression equation:

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 × general

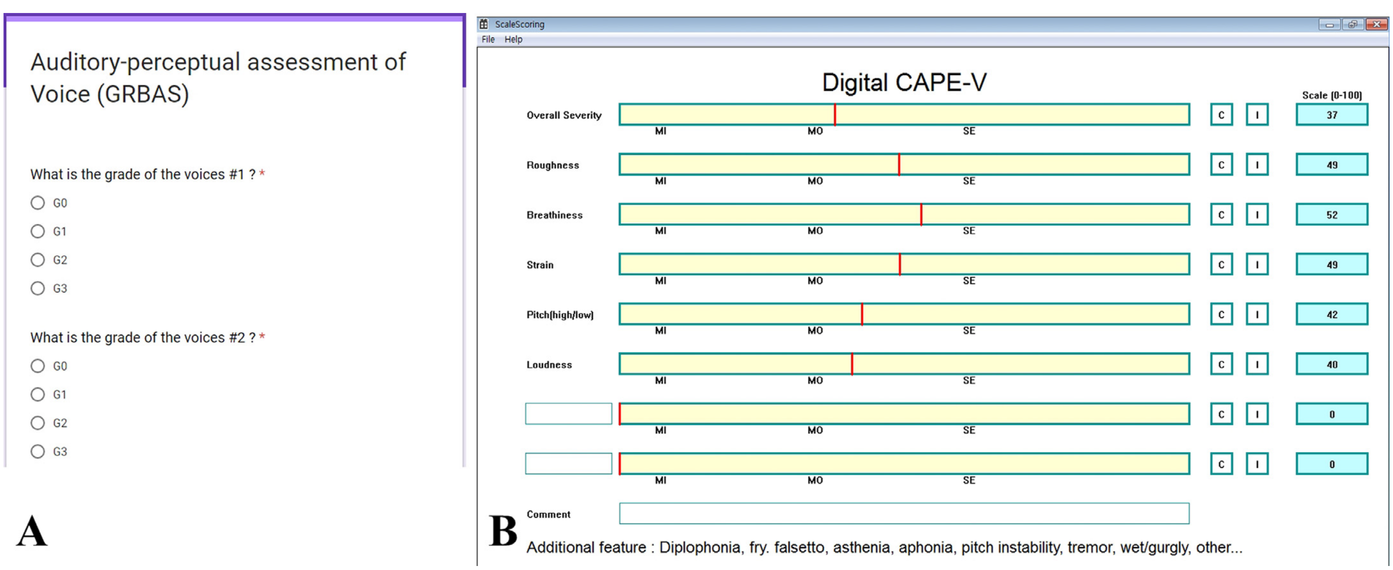

FIGURE 1. Example of the Google questionnaire for GRBAS (A) and a customized computer program for CAPE-V (B). GRBAS, grade, roughness, breathiness, asthenia, strain; CAPE-V, consensus auditory-perceptual evaluation of voice.

TABLE 2.
Inter-rater Reliability of the Five SLPs Who Judged G and OS in Concatenated Voice Samples

	Rater 2		Rater 3		Rater 4		Rater 5	
	ICC on G	ICC on OS	ICC on G	ICC on OS	ICC on G	ICC on OS	ICC on G	ICC on OS
Rater 1	0.693**	0.756**	0.687**	0.700**	0.878**	0.955**	0.849**	0.877**
Rater 2			0.580**	0.741**	0.712**	0.756**	0.702**	0.797**
Rater 3					0.661**	0.762**	0.926**	0.943**
Rater 4							0.895**	0.934**

** $P < 0.01$.
ICC, interrater correlation coefficient; G, grade; OS, overall severity; SLP, speech and language pathologists.

slope of the long-term average spectrum) + (0.170 × tilt of the regression line through the long-term average spectrum).

Statistical analysis

All statistical analyses were completed using R version 3.4.1 (The R Foundation for Statistical Computing, Vienna, Austria) and RStudio 1.0.143 (RStudio Inc., Boston, MA). The inter-rater variability of the Likert G scores and the visual analog scale OS scores for the five raters were examined using the intraclass correlation coefficient (ICC). The intra-rater reliability of the five raters for the 300 voice samples was calculated. For each concatenated voice sample, the scores of the five raters were averaged, resulting in mean G and OS scores for the 1,524 subjects.

The correlation coefficient and coefficient of determination (r^2) was used to investigate the concurrent validity of the AVQI in Korean. The Pearson correlation coefficient (r_p) was measured between the auditory-perceptual assessments (G and OS) and AVQI. In addition, the Pearson correlation coefficient (r_p) was measured to investigate the correlation between G and OS.

An independent sample *t* test was used to assess the differences between the two groups. To obtain the cutoff values and accuracy (area under curve: AUC) of the AVQI, G, and OS between normal (G0) and pathological status, a ROC curve analysis was performed. For analyzing and outputting data, R packages such as ggplot2, ggsignif, corrplot, GGally, plyr, reshape2, and hexbin were used.

RESULTS

Inter-rater variability

Table 2 summarizes the variability statistics for the auditory-perceptual assessments between the five raters. The inter-rater reliability ranged from low (ICC = 0.580 between rater 2 and 3 for G) to very high (ICC = 0.955 between rater 1 and 4 for OS), with a mean inter-rater ICC of 0.790. The overall reliability between the raters was felt to be sufficient for this study.

Intra-rater variability

Table 3 summarizes the data on variability in the auditory-perceptual assessments among the five raters. The intra-rater reliability for G varied from moderate (ICC = 0.746, rater 2) to high (ICC = 0.938, rater 1), and the intra-rater reliability for OS varied from moderate (ICC = 0.742, rater 1) to high (ICC = 0.932, rater 5), with a mean intra-rater ICC of 0.842 (G) and 0.825 (OS). The general reliability within the raters was considered acceptable for the purpose of the present study.

AVQI, G, and OS values according to laryngeal status

The AVQI, G, and OS values according to laryngeal pathology are shown in Figure 2. The mean values of AVQI, G, and OS ranged from normal (AVQI = 2.4 ± 0.7, G = 0.00, OS = 12.1 ± 5.4) to values consistent with glottic cancer (AVQI = 8.7 ± 0.8, G =2.4 ± 0.4, OS = 69.9 ± 18.7). The measured values of AVQI varied depending on the laryngeal pathology.

AVQI and auditory-perceptual assessments (G and OS)

The correlation between AVQI data and G score was $r = 0.863$, $r^2 = 0.745$ ($P < 0.01$). The correlation between AVQI data and OS score was $r = 0.860$, $r^2 = 0.740$ ($P < 0.01$). The strong proportional relationships to the G and OS scores were accounted for by the variance (74.5% and 74%) in the AVQI data, respectively. This result indicates that voices with higher G and OS score are expected to correlate with an equally higher AVQI, and the opposite is also the case (Figure 3).

G and OS

The correlation between G and OS ranged from $r = 0.704$ ($P < 0.01$, rater 2) to $r = 0.878$ ($P < 0.01$, rater 5). The correlation between the general (average of the five raters'

TABLE 3.
Intrarater Reliability of the Five SLPs Who Evaluated G and OS in Concatenated Voice Samples

ICC	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
G	0.938**	0.746**	0.791**	0.854**	0.884**
OS	0.742**	0.784**	0.778**	0.891**	0.932**

** $P < 0.01$.
ICC, intrarater correlation coefficient; G, grade; OS, overall severity; SLP, speech and language pathologists.

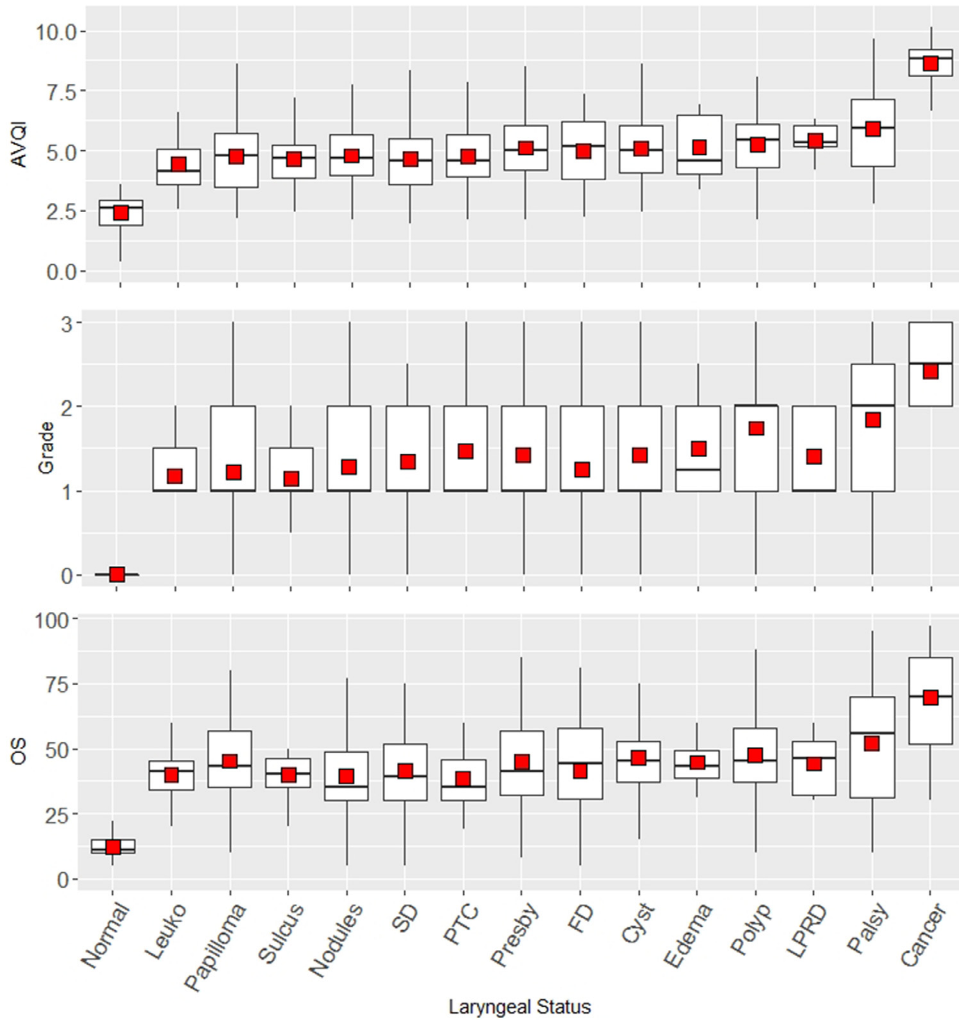

FIGURE 2. Box plots of AVQI, G, and OS according to laryngeal status. AVQI, acoustic voice quality index; G, grade; OS, overall severity.

score) G and OS corresponded with r = 0.848, $r^2 = 0.719$ ($P < 0.01$) on concatenated voice samples (Figure 3).

Cutoff values and accuracy of the AVQI, G and OS between normal and pathological status

The AVQI, G, and OS of the normal group was significantly different from those of the pathological group (Figure 4). The normal group had relatively low values of AVQI, G, and OS, whereas the pathological group had relatively high values (Figure 5). The AVQI threshold of 3.33 for the detection of dysphonic voice exhibited reasonable sensitivity (0.900) and excellent specificity (0.965) (AUC = 0.970). For the G scores, a G threshold of 0.00 was measured with a sensitivity of 0.950 and a specificity of 1.000 (AUC = 0.975). For the OS scores, an OS threshold of 22.0 was measured, with a sensitivity of 0.934 and a specificity of 0.973 (AUC = 0.977) (Figure 6).

DISCUSSION

The present study explored the concurrent validity of the AVQI in a Korean population, and investigated the practicality of its diagnostic accuracy and cutoff values in discriminating between normal and dysphonic voices. It was important to evaluate not only sustained vowel phonation but also sentence reading in the assessment of natural voice and speech. Acoustic quantification of concatenated voice samples has not been widely reported. The reliability of the auditory-perceptual assessments was verified, and the differences in AVQI, G, and OS between normal and pathological voices were confirmed in this study. In addition, we determined that AVQI is effective in evaluating voice disorders by confirming the correlation between acoustic and auditory-perceptual evaluations.

The AVQI is an acoustic analysis method designed to quantify the severity of the disordered voice using both the sustained spoken vowel and sentence reading. Although the validity of the AVQI in a Korean population has been verified, the results were based on a small number of subjects. For this reason, the present study was conducted on a group of subjects with more diverse voice diseases and a greater total number of subjects for strong validity. In contrast to the correlation among AVQI, G, and OS, the cutoff values of AVQI, G, and OS between normal (G0) and pathological

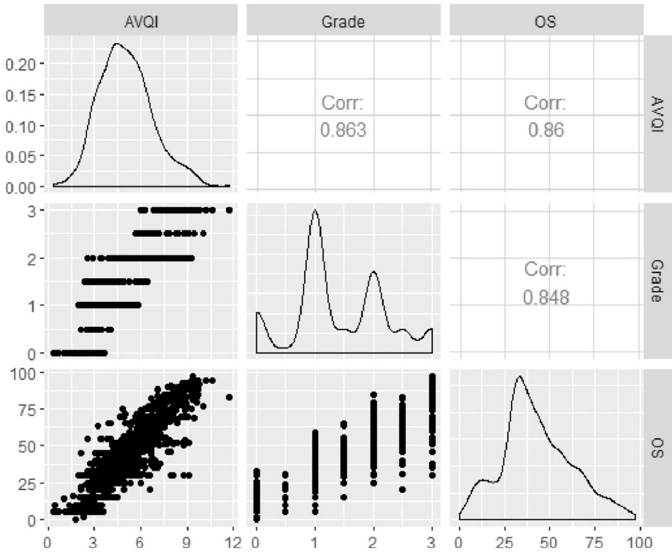

FIGURE 3. Correlation plots and correlation coefficients among AVQI, G, and OS. AVQI, acoustic voice quality index; G, grade; OS, overall severity.

voice status has not previously been analyzed. Therefore, this study explored the relationship between, and cutoff values for two widely used auditory-perceptual assessments (GRBAS and CAPE-V) and the AVQI in a Korean population.

Experienced evaluators were able to achieve results with high reliability using a variety of evaluation tools. In the case of the GRBAS, it is easily evaluated on a four-point scale and the test-retest reliability is high. With the CAPE-V, the degree of severity is evaluated based on a scale from 0 to 100, so the test-retest reliability is low; the advantage is that the severity of voice impairments can be evaluated in more detail. The reason for the high correlation between the evaluation tools with different characteristics is that the intensive auditory-perceptual assessment training was conducted using various voice samples.

The ability of the skilled clinician to evaluate dysphonia plays a crucial role in the voice evaluation, and it has been reported that voice impairments can be accurately identified. We found that the AVQI did correlate with G and OS. This result is consistent with previous studies that reported a high correlation between acoustic measurements and auditory-perceptual assessments. Shin et al[47] and Uloza et al[48] reported that GRBAS, jitter, shimmer, and noise to harmonic ratio were improved significantly after laryngeal microsurgery. In addition, the acoustic measurements (maximum phonation time, jitter, shimmer, and noise to harmonic ratio) were highly correlated with the auditory-perceptual assessment (GRBAS), and improvements after treatment were confirmed by jitter and shimmer. Research on the correlation between the AVQI and auditory evaluations has consistently been reported. There were many reports on GRBAS at the beginning of the study, and CAPE-V was also reported frequently thereafter. The correlation between AVQI and G was reported to be 0.794–0.868 in the English, 0.780–0.858 in Germans, 0.781 in the French, 0.828 in Japanese, 0.911 in Koreans, and 0.680 in Spaniards. The correlation between AVQI and OS was reported to be 0.852–0.876 in Germans and 0.924 in Koreans.[30–35,49,50] In the present study, the correlation between G and OS was 0.836, and in the previous study was reported to be 0.950 in the English, 0.969 in Brazilians-Portuguese, and 0.955–0.965 in Koreans.[22,35,51,52] The AVQI has been studied in many languages and has been reported to show very high confidence in the quantification of voice impairment. As in previous studies, this study revealed a high correlation among AVQI and G and OS, which is consistent with the results of previous Korean studies.

Cutoff values of AVQI, G, and OS and its AUC were calculated for measuring the probability for dysphonic population. Normal and pathological groups showed significant differences in AVQI, G, and OS. The normal group consisted of Grade 0 subjects who did not complain of laryngeal disease or voice disorders. The cutoffs for the three variables

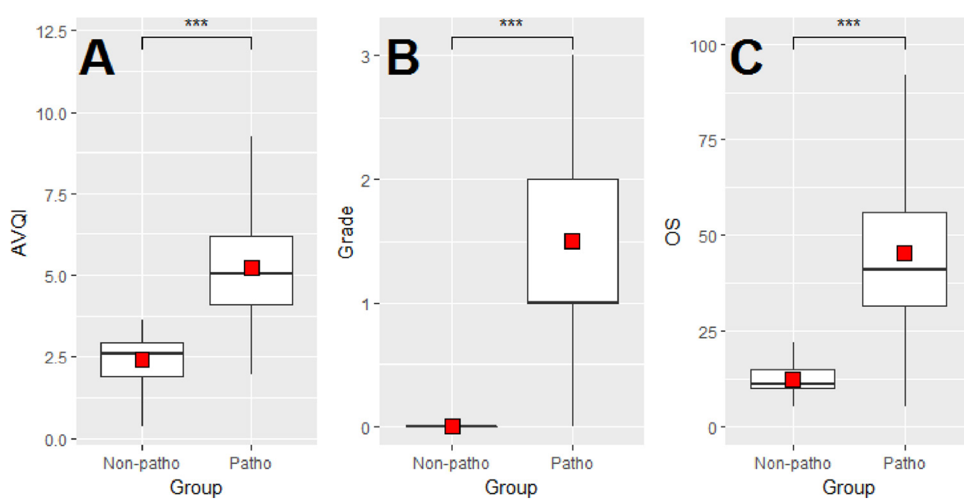

FIGURE 4. Comparison of AVQI, G, and OS between normal and pathologic groups. (A: AVQI, B: Grade, C: Overall severity). AVQI, acoustic voice quality index; G, grade; OS, overall severity.

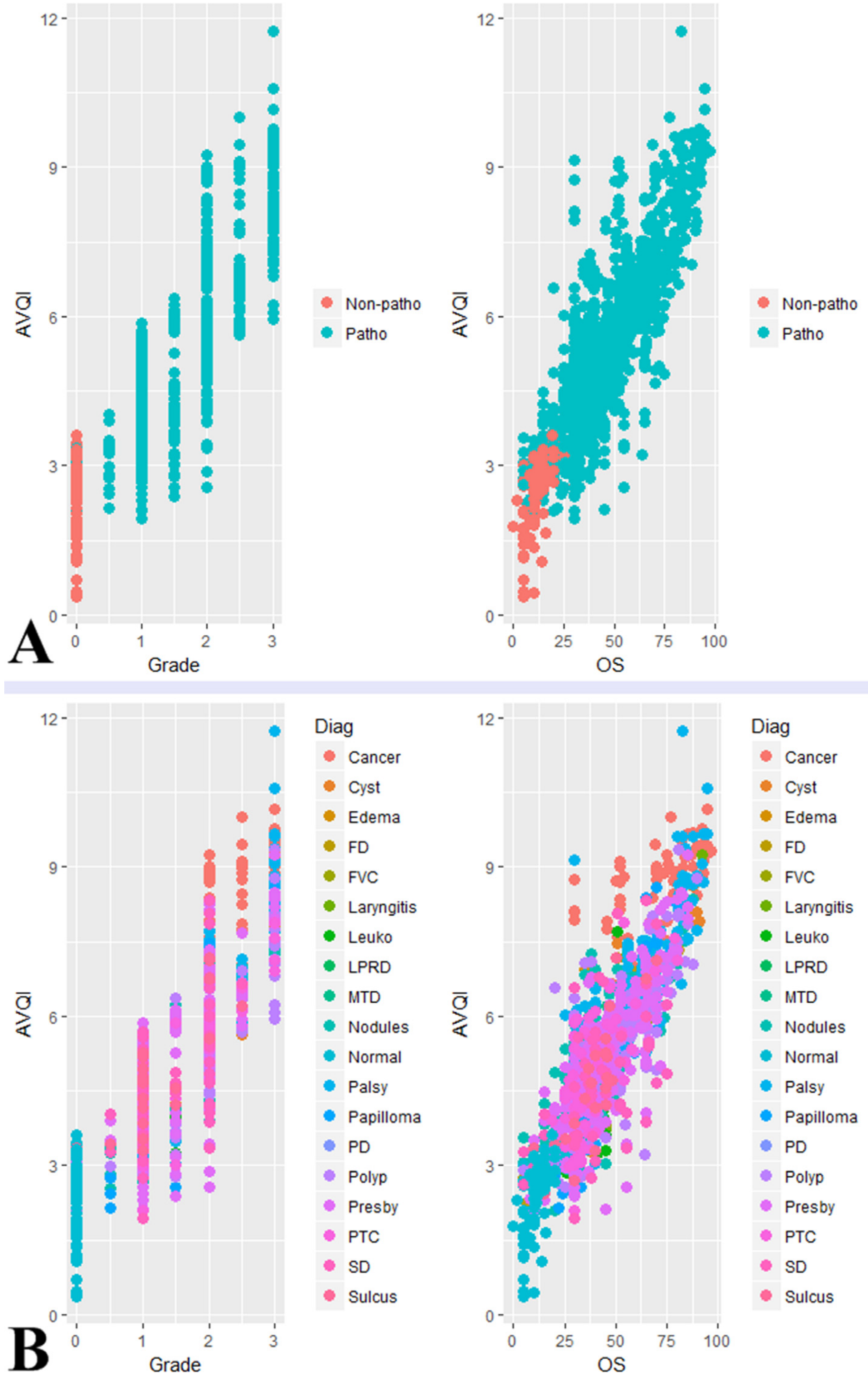

FIGURE 5. Scatter plot of AVQI, G, and OS. (A: normal versus pathologic group; B: according to diagnosis). AVQI, acoustic voice quality index; G: grade; OS, overall severity.

were 3.33, 0, and 22, respectively. When the pathological groups were compared with the normal group, significantly higher values in all AVQI, G, and OS scores were found for the pathological groups. The current results are consistent with previous studies reporting significant differences between normal and dysphonic groups in subjects speaking different languages. Previous studies have reported cutoffs of 3.19−3.66 in the Dutch, 3.25−3.29 in Americans, 3.05 in Germans, 3.07 in the French, 2.97 in Lithuanians, and 3.15 in the Japanese.[28,30,33,34] The differences in cutoffs are attributed to the application of other versions of the AVQI regression formula and the relatively low AVQI in the pathological group, which included some Grade 0 subjects with laryngeal disease. In addition, a statistically significant difference was reported between the normal group, with a cutoff of 6.23, and the pathological group with a cutoff of 7.70,

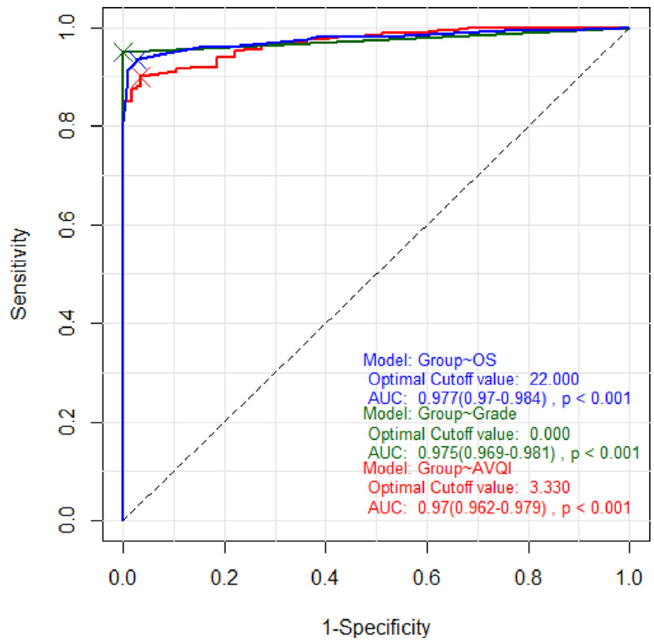

FIGURE 6. ROC curve analyses for distinguishing normal and dysphonic voices. ROC, receiver operating characteristic.

in Spaniards, but the value of the AVQI was much greater than those of previous studies.[32] The results of this study support the results of previous studies showing that AVQI can discriminate between normal and dysphonic voices well.

The perturbation analysis is not accurate in the case of highly aperiodic speech signals, and the cepstral analysis is performed to compensate for this limitation. However, the CPP value has a limitation depending on the pause and the unvoiced interval in the sentence. From this point of view, AVQI analyzes the same vowel phonation length and sentence reading, and additionally removes the pause and unvoiced interval using the Praat script. It is possible to quantify voice impairment by analyzing voice samples obtained under the same conditions. If the same vowel phonation section and sentence reading are united, it will be comparable within the same language. The probability of classifying dysphonic voice through acoustic and auditory-perceptual assessments is 97.0–97.7%. It is possible to improve the prediction of voice recovery by analyzing the visual, acoustic, auditory-perceptual, and patient self-reports altogether. In addition, this study suggests that the newly proposed AVQI method, including the acoustic and auditory-perceptual assessments, could be used as a promising tool in screening for voice disorders.

The limitations of this study are that the regression equations for the AVQI measurements are constantly being developed and updated. In this study, the version of AVQI applied is 2.02, but now version 3.01 has also been introduced.[31,53,54] In order to quantify and differentiate the pathological voice, the newly proposed regression equation should be continuously applied to improve the discrimination accuracy. Most recently published studies have applied version 2.02 and are also testing for version 3.01.

CONCLUSIONS

It has been found that AVQI is highly correlated with auditory-perceptual assessments and has high reliability in quantifying the degree of voice disorders in the dysphonic population. This means that the AVQI and auditory-perceptual evaluations are complementary. This study confirmed the concurrent validity of the AVQI and the Korean standard auditory-perceptual assessment tool for quantifying dysphonia.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jvoice.2018.06.007.

REFERENCES

1. 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 Oto-Rhino-L*. 2001;258:77–82.
2. Sakata T, Kubota N, Yonekawa H, et al. Grbas evaluation of running speech and sustained phonation. *Folia Phoniatr Logo*. 1995;47. 101-101.
3. Carding P, Carlson E, Epstein R, et al. Formal perceptual evaluation of voice quality in the United Kingdom. *Logoped Phoniatr Vocol*. 2000;25:133–138.
4. Awan SN, Lawson LL. The effect of anchor modality on the reliability of vocal severity ratings. *J Voice*. 2009;23:341–352.
5. Barsties B, Beers M, Ten Cate L, et al. The effect of visual feedback and training in auditory-perceptual judgment of voice quality. *Logoped Phoniatr Vocol*. 2017;42:1–8.
6. Maryn Y, Dick C, Vandenbruaene C, et al. Spectral, cepstral, and multivariate exploration of tracheoesophageal voice quality in continuous speech and sustained vowels. *Laryngoscope*. 2009;119:2384–2394.
7. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res*. 2001;44:327–339.
8. McGuirt WF, Blalock D, Koufman JA, et al. Comparative voice results after laser resection or irradiation of T1 vocal cord carcinoma. *Arch Otolaryngol Head Neck Surg*. 1994;120:951–955.
9. Putzer M. Multiparametric description of voice quality for normal male and female voices. *Folia Phoniatr Logo*. 2001;53:73–84.
10. Smits I, Ceuppens P, De Bodt MS. A comparative study of acoustic voice measurements by means of Dr. Speech and Computerized Speech Lab. *J voice*. 2005;19:187–196.
11. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Lang Hear Res*. 1996;39:311–321.
12. Deliyski DD, Shaw HS, Evans MK. Influence of sampling rate on accuracy and reliability of acoustic voice analysis. *Logoped Phoniatr Vocol*. 2005;30:55–62.
13. Lee YS, Lee DH, Jeong GE, et al. Treatment efficacy of voice therapy for vocal fold polyps and factors predictive of its efficacy. *J Voice*. 2017;31:120–e9.
14. Siu J, Tam S, Fung K. A comparison of outcomes in interventions for unilateral vocal fold paralysis: a systematic review. *Laryngoscope*. 2016;126:1616–1624.
15. Cannito MP, Woodson GE, Murry T, et al. Perceptual analyses of spasmodic dysphonia before and after treatment. *Arch Otolaryngol Head Neck Surg*. 2004;130:1393–1399.

16. Ma Y, Green R, McCabe D, et al. Long-term voice outcome following radiation versus laser microsurgery in early glottic cancer. *J Voice* 2017. [In Press]
17. Berkow LC, Morey TE, Urdaneta F. The technology of video laryngoscopy. *Anesth Analg*. 2018;126:1527–1534.
18. John Doyle D, Dahaba A, LeManach Y. Advances in anesthesia technology are improving patient care, but many challenges remain. *BMC Anesthesiol*. 2018;18:39.
19. Niebudek-Bogusz E, Kotyło P, Śliwińska-Kowalska M. Evaluation of voice acoustic parameters related to the vocal-loading test in professionally active teachers with dysphonia. *Int J Occup Med Environ Health*. 2007;20:25–30.
20. Batra K, Bhasin S, Singh A. Acoustic analysis of voice samples to differentiate healthy and asthmatic persons. *IJECS*. 2015;4:13161–13164.
21. De Bodt MS, Wuyts FL, Van de Heyning PH, et al. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice*. 1997;11:74–80.
22. Karnell MP, Melton SD, Childes JM, et al. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice*. 2007;21:576–590.
23. Zraick RI, Kempster GB, Connor NP, et al. Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Am J Speech-Lang Pat*. 2011;20:14–22.
24. Kempster GB, Gerratt BR, Abbott KV, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech-Lang Pat*. 2009;18:124–132.
25. 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.
26. 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.
27. Reynolds V, Buckland A, Bailey J, et al. Objective assessment of pediatric voice disorders with the acoustic voice quality index. *J Voice*. 2012;26. 672-e671.
28. Maryn Y, De Bodt M, Barsties B, et al. The value of the Acoustic Voice Quality Index as a measure of dysphonia severity in subjects speaking different languages. *Eur Arch Oto-Rhino-L*. 2014;271:1609–1619.
29. 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.
30. Barsties B, Maryn Y. The acoustic voice quality index. Toward expanded measurement of dysphonia severity in German subjects. *HNO*. 2012;60:715–720.
31. 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* 2018,515–e1.
32. Núñez-Batalla F, Díaz-Fresno E, Álvarez-Fernández A, et al. Application of the acoustic voice quality index for objective measurement of dysphonia severity. *Acta Otorrinolaringol*. 2017;68:204–211. (English Edition).
33. Uloza V, Petrauskas T, Padervinskis E, et al. Validation of the acoustic voice quality index in the lithuanian language. *J Voice*. 2017;31. 257-e251.
34. Hosokawa K, Barsties B, Iwahashi T, et al. Validation of the Acoustic Voice Quality Index in the Japanese language. *J Voice*. 2017;31. 260-e261.
35. Maryn Y, Kim HT, Kim J. Auditory-perceptual and acoustic methods in measuring dysphonia severity of Korean speech. *J Voice*. 2016;30:587–594.
36. Behlau M, Madazio G, Moreti F, et al. Efficiency and cutoff values of self-assessment instruments on the impact of a voice problem. *J Voice*. 2016;30. 506.e509-506.e518.
37. Moradi N, Pourshahbaz A, Soltani M, et al. Cutoff point at voice handicap index used to screen voice disorders among persian speakers. *J Voice*. 2013;27. 130.e131-130.e135.
38. Tafiadis D, Kosma EI, Chronopoulos SK, et al. Acoustic and perceived measurements certifying tango as voice treatment method. *J Voice*. 2018;32:256–e13.
39. Tafiadis D, Chronopoulos SK, Kosma EI, et al. Using receiver operating characteristic curve to define the cutoff points of voice handicap index applied to young adult male smokers. *J Voice*. 2018;32:443–448.
40. Tafiadis D, Kosma EI, Chronopoulos SK, et al. Voice handicap index and interpretation of the cutoff points using receiver operating characteristic curve as screening for young adult female smokers. *J Voice*. 2018;32:64–69.
41. Hanley J, McNeal B. A simple generalization of the area under the ROC curve to multiple class classification problems. *Radiology*. 1982;143:29–36.
42. Moon BJ, Kim SY, Choi S-h, et al. Clinical analysis of the laryngeal premalignant lesion. *Korean J Otorhinolaryngol-Head Neck Surg*. 2007;50:525–528.
43. Seo IH, Moon J-H, Chung P-S, et al. Role and effectiveness of postoperative voice therapy in laryngeal microsurgery. *Korean J Otorhinolaryngol-Head Neck Surg*. 2018;61:200–207.
44. Ingrisano DRS, Perry CK, Jepson KR. Environmental noise: a threat to automatic voice analysis. *Am J Speech-Lang Pat*. 1998;7:91–96.
45. Deliyski DD, Shaw HS, Evans MK. Adverse effects of environmental noise on acoustic voice quality measurements. *J Voice*. 2005;19:15–28.
46. Hirano M. Psyco-acoustic evaluation of voice. Clinical Examination of Voice (Disorders of Human Communication) 1981:81−84.
47. Shin YS, Chang JW, Yang SM, et al. Persistent dysphonia after laryngomicrosurgery for benign vocal fold disease. *Clin Exp Otorhinolar*. 2013;6:166–170.
48. Uloza V, Saferis V, Uloziene I. Perceptual and acoustic assessment of voice pathology and the efficacy of endolaryngeal phonomicrosurgery. *J Voice*. 2005;19:138–145.
49. Barsties VLB, Ulozaite-Staniene 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* 2017. [In Press]
50. Maryn Y, Debo K. Is perceived dysphonia related to perceived healthiness? *Logoped Phoniatr Vocol*. 2015;40:122–128.
51. Wuyts FL, De Bodt MS, Van de Heyning PH. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice*. 1999;13:508–517.
52. Nemr K, Simoes-Zenari M, Cordeiro GF, et al. GRBAS and cape-V scales: high reliability and consensus when applied at different times. *J Voice*. 2012;26. 812-e817.
53. Hosokawa K, Barsties VLB, Iwahashi T, et al. The acoustic voice quality index version 03.01 for the japanese-speaking population. *J Voice* 2017. [In Press]
54. Barsties B, Maryn Y. External validation of the acoustic voice quality index version 03.01 with extended representativity. *Ann Otol Rhinol Laryngol*. 2016;125:571–583.