Correlations between Automated Dysphonia Quantification and Perceptual Voice Evaluation

N. Ulozaitė[1*], T. Petrauskas[1], V. Šaferis[2], V.Uloza[1]

[1]*Department of Otorhinolaryngology, Lithuanian University of Health Sciences, Kaunas, Lithuania,*

[2]*Department of Physics, Mathematics and Biophysics, Lithuanian University of Health Sciences, Kaunas, Lithuania*

[*]*E-mail: nora.ulozaite@gmail.com*

Introduction. Voice disorders are relatively common affecting from 6 to 9 % of general population [1]. Automated analysis of voice signals is the most used diagnostic instrument to identify voice disorders in clinical practice and research. Acoustic markers computed from voice signal recordings are a convenient way for collection of objective non-invasive voice data, quantifying and documenting dysphonia severity and providing measurable evidence of the functional outcomes of therapeutic and/or phonosurgical treatment of voice problems [2, 3].

If an objective voice assessment is to be considered robust and ecologically valid, then the acoustic voice measure ideally should be calculated utilizing voice signals based on recordings of both speaking patterns - sustained phonation and running speech [4, 5].

Currently, Acoustic Voice Quality Index (AVQI) proposed by Maryn et al. [5] is one of two multiparametric models in the world, used to successfully evaluate voice quality on both speech types. AVQI has been recently validated to use in Lithuanian language (AVQI-LT) [6].

Aim of the study: The aim of the present study was to investigate the feasibility and robustness of AVQI-LT for dysphonia quantification and it's correlations with auditory-perceptual judgment.

Methods. Voice samples consisted of 153 individuals examined at the Department of Otolaryngology of Lithuanian University of Health Sciences, Kaunas, Lithuania. The mean age of the study group was 40 years.

The normal voice subgroup was composed of volunteers without vocal complains and no history chronic laryngeal diseases or other long-lasting voice disorders.

The pathological voice subgroup consisted of 107 patients who represented a rather common and clinically discriminative group of laryngeal diseases and voice disturbances including mass lesions and paralysis of vocal folds. The clinical diagnosis was based on clinical evaluation, videolaryngoscopy and direct microlaryngoscopy.

Voice recordings. The mixed gender database of voice recordings in this study contained 153 digital voice recordings of sustained phonation of the vowel sound /a/ and the Lithuanian sentence "Turėjo senelė žilą oželį" (The grandmother had a little grey goat) containing 13 syllables. Both of them were recorded with comfortable pitch and loudness. Voice samples from each subject were recorded in a T-series silent room for hearing testing with a cardioid AKG Perception 220 microphone placed at 10.0 cm distance from the mouth keeping at about 90° microphone-to-mouth angle.

The voice recordings were saved in ''wav'' file format using Audacity software (http://audacity.sourceforge.net/) at a sampling frequency of 44.1 kHz. The external sound card M-Audio (Cumberland, RI) was used for digitization of the voice recordings outside the computer.

Further formatting, selecting and segmenting of the recordings were completed manually using the program Praat version 5.3.57.[7] The voice samples were concatenated for auditory-perceptual judgment in the following order: text segment, a pause of one second, followed by the 3 second sustained vowel /a/ segment.

Acoustic Voice Analysis. For acoustic analysis, only voiced parts for the continuous speech were manually extracted and concatenated to the medial 3 seconds of sustained /a/ phonation. This chain of signals was used for acoustic analysis with the AVQI script version 02.02 developed for the program Praat.[8] The following multiple regression equation of AVQI includes acoustic markers from time, frequency, and quefrency domains was used:

$$AVQI = ((3.295 - (0.111 * CPPS) - (0.073 * HNR) - (0.213 * SHIM) + (2.789 * SHdB) - (0.032 * sLTAS) + (0.077 * ttLTAS)) * 2.208) + 1.797 \quad (1)$$

Abbreviations: AVQI - acoustic voice quality index; CPPS - smoothed cepstral peak prominence; HNR - harmonic to noise ratio; SHIM (shimmer) – amplitude perturbations as percentage; SHdB (shimmer) - amplitude perturbations expressed as dB; sLTAS - slope of the long-term average spectrum; ttLTAS - tilt of trendline through the long-term average spectrum.

The graphical presentation of AVQI-LT sample in case of laryngeal carcinoma is shown in Fig.1.

Auditory-perceptual evaluation. Five experienced physicians-laryngologists, who were all native Lithuanians, served as rater panel. They performed auditory-perceptual evaluations to quantify the vocal deviation. The experts judged the voice samples into four ordinal severity classes of Grade (G), Roughness (R), Breathiness (B) from the GRB scale (i.e. 0 = normal, 1 = slight, 2 = moderate, 3 = severe [9]. Additionally, overall severity of dysphonia was evaluated on the 10.0 cm long visual analogue scale (VAS) from the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) protocol as proposed the American Speech-Language and Hearing Association [10].

Results. Rater reliability. To assess the inter-rater reliability among the five judges for the ordinal rating, we computed the kappa coefficient according to Fleiss (Fk).[11] The estimation of the inter-rater reliability of the VAS-rating was assessed with the single measure of Interclass Correlation Coefficient (ICC). Guidelines for the strength of agreement of the kappa statistics indicate the following range: Fk<0.00 = poor, Fk=0.00-0.20 slight, Fk=0.21-0.40 fair, Fk=0.41-0.60 moderate, Fk=0.61-0.80 substantial, Fk=0.81-1.00 almost perfect agreement, respectively. The ICC ranges from 0.00 (i.e., total absence of reliability) to 1.00 (i.e., perfect reliability) and general guideline indicates that ICC>0.75 corresponds with good reliability.

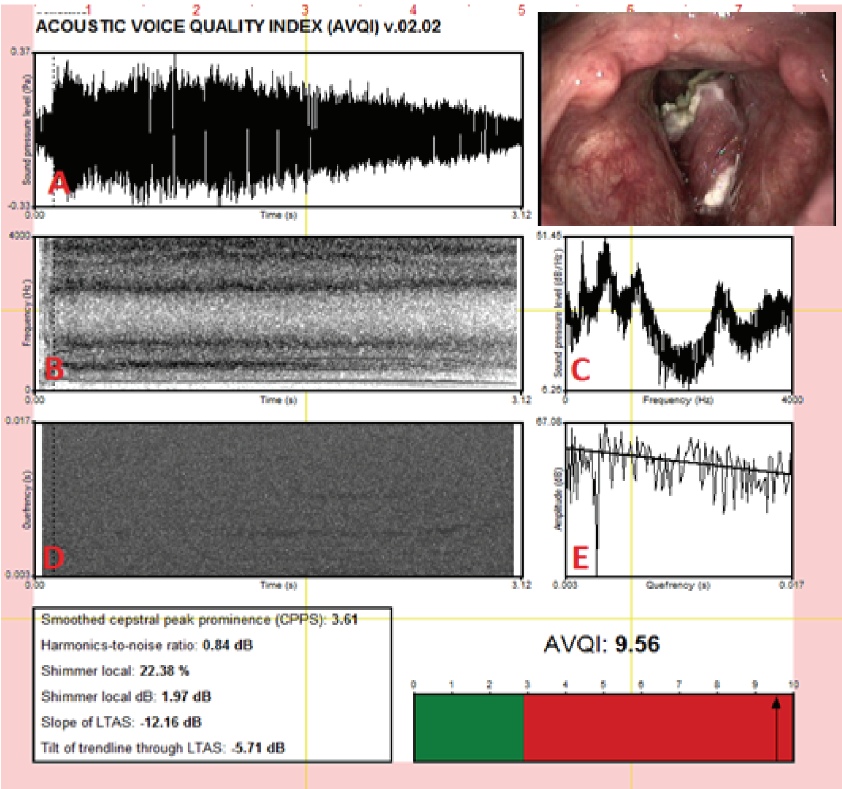

Fig. 1. Graphical presentation of AVQI-LT sample in case of laryngeal carcinoma:
A. Oscillogram: representation of the concatenated vowel and sentence recording air pressure fluctuations in the time domain.
B. Spectrogram: representation of the signal in terms of energy spread over its frequency over time.
C. Long term average spectrum (LTAS): logarithmic power spectral density which describes sound pressure level per frequency.
D. Power-Cepstrogram: shows cepstral splices as a function of time.
E. Power-Cepstrum: shows the tilt line of the power-cepstrum.

In the present study, the inter-rater reliability outcomes showed Fk= 0.90 for ordinal G ratings, Fk=0.82 for ordinal R ratings, Fk=0.81 for ordinal B ratings and ICC=0.86 for continuous VAS ratings. Thus, both outcomes of auditory-perceptual evaluation indicated almost perfect inter-rater reliability.

Concurrent validity. The concurrent validity of AVQI-LT and auditory-perceptual judgment, as expressed in Spearman's rank-order correlation coefficient and coefficient of determination, was highly significant and varied between rs = 0.87 ($p < 0.001$) for G, rs = 0.71 ($p < 0.001$) for R, rs = 0.86 ($p < 0.001$) for B and rs = 0.86 ($p < 0.001$) for VAS.

These results demonstrated significant and marked [12] concurrent validity between both auditory-perceptual judgment procedures and AVQI-LT.

Correlations between AVQI-LT and auditory-perceptual voice assessment. Results of correlation analysis between AVQI-LT and auditory-perceptual voice assessment are presented in Table 1.

Table 1. Correlation matrix between AVQI-LT and auditory-perceptual voice assessment

	AVQI	Gmean	Rmean	Bmean	VASmean
AVQI	1.00	0.87**	0.71**	0.86**	0.86**
Gmean	0.87**	1.00	0.86**	0.94**	0.98**
Rmean	0.71**	0.86**	1.00	0.71**	0.85**
Bmean	0.86**	0.94**	0.71**	1.00	0.94**
VASmean	0.86**	0.98**	0.85**	0.94**	1.00
N	153	153	153	153	153

***Correlation is significant at the 0.01 level (2-tailed).*

As shown in Table 1, very strong and statistically significant correlations between AVQI-LT and all factors of auditory-perceptual voice assessment (including G, R and B) were achieved in our series.

Discussion and conclusion. The importance of AVQI as an acoustical clinical marker of overall voice quality was already proven across multiple studies with various subjects groups and languages [4, 5, 15].

Generally, the results of the present study are in concordance with the data of literature, confirming that the performance of AVQI is relatively insulated from inter language phonetic differences and supports the cross-linguistic robustness of AVQI-LT as a valid and objective measure of overall voice quality.

The inter-rater reliability results showed Fk ranging from 0.81 to 0.90 for ordinar factorial rating and an ICC = 0.855 for VAS mean. Therefore, both outcome measures indicated moderate and thus sufficient strength of agreement in inter-rater reliability.

In the present study, the results of statistical analysis related to concurrent validity of the AVQI-LT are comparable with the outcomes reported for other languages, namely, Dutch, German, French, English, Finnish, Korean and Japanese [5, 13]. The results of the present study revealed significant and marked concurrent validity between two auditory-perceptual judgment procedures and AVQI-LT.

Very high results in correlation between AVQI-LT and auditory-perceptive voice assessment confirm robustness of this method of automated voice analysis. On the other hand, very strong and statistically significant correlations between AVQI-LT and all factors of auditory-perceptual voice assessment, i.e. G, R, and B, presume the G factor as qualificatory for overall dysphonia assessment, thus allowing reduction the number of auditory-perceptual voice assessment factors and simplifying subjective voice evaluation.

Therefore, AVQI may warrant an important step in making practical, reliable and objective voice assessment tool suitable to nonexperts and voice professionals feasible to support their decision in practice or research managing patients with voice disorders.

In conclusion, the AVQI-LT strongly correlates with auditory-perceptual voice assessment and is considered to be a valid and reliable tool for automated quantification of dysphonia in Lithuanian speaking population.Figures must be photographically reproducible: sharp, noise-free and good contrast. Figures will be published only in grey, not in colour.

Acknowledgment. This study was supported by project BN16-143 "Development of automated system for voice screening (BSAS)"

References

1. Roy N, Merrill RM, Thibeault S, Parsa RA, Gray SD, Smith EM. Prevalence of voice disorders in teachers and the general population. J Speech Lang Hear Res. 2004;47(2):281-293.

2. 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(2):187-196.

3. Geneid A, Rihkanen H, Kinnari TJ. Long-term outcome of endoscopic shortening and stiffening of the vocal folds to raise the pitch. Eur Arch Otorhinolaryngol. 2015;272(12):3751-3756.

4. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. J Voice. 2010;24(5):540-555.

5. 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(3):161-174.

6. Uloza V, Petrauskas T, Padervinskis E, Ulozaite N, Barsties B, Maryn Y. Validation of the Acoustic Voice Quality Index in the Lithuanian Language. J Voice. 2016.

7. Boersma P, Weenink D. Praat, a system for doing phonetics by computer. 2001.

8. Maryn Y, Weenink D. Objective Dysphonia Measures in the Program Praat : Smoothed Cepstral Peak Prominence and Acoustic Voice Quality Index. J Voice. 2015;29(1):35-43.

9. 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 Laryngolo. Eur Arch Otorhinolaryngol. 2001;258(2):77-82.

10. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. Am J Speech-Language Pathol. 2009;18(2):124-132.

11. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971;76(5):378-382.

12. Frey LR, Botan CH, Friedman PG KG. Investigating Communication: An Introduction to Research Methods. Englewood Cliffs: Prentice-Hall; 1991.

13. Maryn Y, De Bodt M, Barsties B, Roy N. The value of the acoustic voice quality index as a measure of dysphonia severity in subjects speaking different languages. Eur Arch Otorhinolaryngol. 2014;271(6):1609-1619.

14. Barsties B, Maryn Y. The Acoustic Voice Quality Index. Toward expanded measurement of dysphonia severity in German subjects]. HNO. 2012;60(8):715-720.

15. Hosokawa K, Barsties B, Iwahashi T, et al. Validation of the Acoustic Voice Quality Index in the Japanese Language. J Voice. 2016.

Correlations between Automated Dysphonia Quantification and Perceptual Voice Evaluation

N. Ulozaitė[1], T. Petrauskas[1], V. Šaferis[2], V.Uloza[1]

[1]Department of Otorhinolaryngology, Lithuanian University of Health Sciences, Lithuania

[2]Department of Physics, Mathematics and Biophysics, Lithuanian University of Health Sciences, Lithuania

The aim of the present study was to investigate the feasibility and robustness of Lithuanian version of Acoustic Voice Quality Index (AVQI-LT) for dysphonia quantification and it's correlations with auditory-perceptual judgment.

The AVQI-LT was calculated using the mixed gender database of voice recordings in this study containing 153 digital voice recordings of sustained phonation of the vowel sound /a/ and 13 syllables Lithuanian sentence. Five experienced physicians-laryngologists performed auditory-perceptual evaluations to quantify the vocal deviation. The results of the present study revealed significant and marked concurrent validity between two auditory-perceptual judgment procedures and AVQI-LT. Very strong and statistically significant correlations between AVQI-LT and all factors of auditory-perceptual voice assessment (including G, R and B) were achieved.

Conclusion. The AVQI-LT strongly correlates with auditory-perceptual voice assessment and is considered to be a valid and reliable tool for automated quantification of dysphonia in Lithuanian speaking population.