ACOUSTIC VOICE QUALITY INDEX VERSION 03.01 FOR THE JAPANESE-SPEAKING POPULATION

K. Hosokawa[1], B. Barsties von Latoszek, T. Iwahashi, M. Iwahashi, C. Kato, S. Iwaki,
M. Yoshida, H. Sasai, A. Miyauchi, H. Inohara, M. Ogawa, Y. Maryn

[1]Department of Otorhinolaryngology, JCHO Osaka Hospital, Osaka, Japan
khosokawa@wg8.so-net.ne.jp

Objective: The acoustic voice quality index (AVQI) was developed as an ecologically valid tool to estimate overall voice quality, including that of both continuous speech (CS) and sustained vowel (SV) tasks. The recent revision to AVQI version 03.01 (AVQIv3) in the Dutch language equally balanced the time length of CS and SV analyses by using 34 syllables for the CS task. However, to correctly apply this tool in other languages, the required syllable number should be adjusted for the different articulatory and phonatory physiology characteristics of different languages. Therefore, to utilize AVQIv3 in the Japanese-speaking population, we first investigated how many syllables should be included in the CS reading text. We then confirmed the concurrent validity and diagnostic accuracy of AVQIv3 for the Japanese-speaking population.

Methods: First, we collected 336 recordings of CS samples with 58 syllables from the recordings used in our previous study on AVQIv2 for the Japanese language. After removing silent or consonant segments using the AVQI script, we visually counted how many syllables should be included in each CS sample. From this procedure, 30 Japanese syllables was judged to be the most appropriate number for achieving time balance between CS and SV samples. Next, we calculated AVQIv3 for 455 voice recordings (including samples from 55 control participants) using CS samples with 30 syllables, then conducted auditory-perceptual judgments of overall voice quality. Subsequently, we estimated the concurrent validity and diagnostic accuracy using correlation and receiver operating characteristic (ROC) analyses, respectively.

Results: Concurrent validity was confirmed by a strong correlation (Spearman's rho = 0.873) between the AVQIv3 and overall voice quality. Furthermore, the area under ROC was 0.915, demonstrating the excellent discriminative capability of AVQIv3. The AVQIv3 threshold of 2.43 in the original Dutch language corresponded with sensitivity and specificity percentages of 64.6% and 97.3%, respectively, whereas a threshold of 1.41 achieved the best accuracy for Japanese, with balanced sensitivity and specificity percentages of 84.4% and 85.6%, respectively. Furthermore, the 95th percentile of the control participants produced a threshold of 2.06, showing sensitivity and specificity percentages of 72.1% and 93.8%, respectively, as well as reasonable positive and negative likelihood ratios of 11.7 and 0.30, respectively.

Conclusions: The AVQIv3 using 30 Japanese syllables is an ecologically valid and reliable measurement for estimating overall voice quality and detecting hoarse voices.