

# Application of the New Version of the Acoustic Voice Quality Index with Korean Speakers

Geun-Hyo Kim<sup>a</sup>, Yeon-Yoo Lee<sup>a</sup>, In-Ho Bae<sup>b</sup>, Hee-June Park<sup>c</sup>, Soon-Bok Kwon<sup>d</sup>

<sup>a</sup>Department of Otorhinolaryngology-Head and Neck Surgery and Biomedical Research Institute, Pusan National University Hospital, Busan, Korea

<sup>b</sup>Department of Otorhinolaryngology-Head and Neck Surgery, Pusan National University Yangsan Hospital, Yangsan, Korea

<sup>c</sup>Department of Speech-Language Therapy, Choonhae College of Health Sciences, Ulsan, Korea

<sup>d</sup>Department of Humanities, Language and Information, Pusan National University, Busan, Korea

Correspondence: Soon-Bok Kwon, PhD

Department of Language and Information, Pusan National University, 2 Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan 46241, Korea

Tel: +82-51-510-2003

Fax: +82-51-582-0572

E-mail: sbkwon@pusan.ac.kr

Received: October 2, 2018

Revised: November 9, 2018

Accepted: November 9, 2018

**Objectives:** The purpose of the present study was to compute the new version of Acoustic Voice Quality Index (AVQI) and compare with existing AVQI versions. Additionally, we also investigated the correlation between AVQI measurements and auditory-perceptual ratings, and confirmed the optimal cutoff value and diagnostic predictive power to distinguish between normal and pathological voice. **Methods:** Voice samples of 2,369 native Korean subjects with normal voices (N = 143) and with dysphonia (N = 2,226) were recorded through both 5 seconds of sustained vowel /a/ and connected speech (Korean text "Walk"). We used the Praat script to extract the voiced segments and measure the AVQI. Additionally, two auditory-perceptual ratings (grade in GRBAS and overall severity in CAPE-V) were used by five speech language pathologists to assess the voice disorder severity. **Results:** Overall AVQI measurements were higher in order of AVQIv2, AVQIv4-K and AVQIv3; and AVQIv3 was significantly lower than other AVQI measures. In each version, there was a significant difference between normal and pathological voice groups. The correlation between AVQI and auditory evaluation was above .82, and the correlation with AVQIv4-K was the highest. All versions of the AVQI showed an above .94 for AUC, above 9.5 for LR(+), and below .2 for LR(-). **Conclusion:** AVQIv4-K is considered to have a high level of diagnostic power comparable to the existing AVQI versions.

**Keywords:** Acoustic Voice Quality Index (AVQI), Acoustic analysis, Dysphonia, Auditory-perceptual assessment, Praat

음향학적 음성 질 지수(Acoustic Voice Quality Index, AVQI)란 모음연장발성과 연속 말 과제(읽기, 자발화 등)를 연결하여 분석함으로써 음성장애의 중증도를 정량화하는 방법으로 소개되었다 (Maryn, De Bodt, & Roy, 2010b). 섹션분석과 캡스트럼 분석 변수들에 가중치를 둔 회귀식을 통해서 AVQI의 측정이 이루어진다. AVQI는 0-10점 사이의 점수로 계산되며 중증도가 높아질수록 AVQI 점수가 높아지도록 설계되었다. 음성장애를 정확하게 평가하기 위해서는 다양한 말 과제를 수행해야 한다. 주로 사용하는 것이 모음연장발성과 문장읽기, 자발화 등과 같은 연속발화가 포함된다. 대상자에 따라서 모음연장과 연속발화에서 다른 수행력을 보이는 경우가 있다. 일반적으로 모음과 연속발화의 분석이 각각 이루어

어지며 두 과제의 결과를 통합해서 판단해야 한다. AVQI를 사용함으로써 이러한 제한점을 극복할 수 있게 되었다. AVQI의 중요한 특징은 모음과 연속 말 과제를 각각 분석하는 것이 아니라 하나의 음성샘플로 연결한 후 무성음(voiceless segment)과 쉼(pause) 구간을 제외한 유성음 분절(voiced segment)만을 추출하여 자동으로 AVQI 측정치를 계산하고 결과값을 제시해 준다는 것이다. 모음과 연속발화를 연결하여 분석하기 때문에 음성평가 대상자의 전체적인 음질을 객관적으로 정량화할 수 있다. 이러한 방식은 Praat (Boersma, 2002)을 이용함으로써 가능하며 대량의 음성샘플로부터 연구자가 원하는 측정 변수들을 신속하고 일관된 방식으로 도출해낼 수 있다. 또한 이러한 AVQI 측정치는 청지각적 평가들과 상

관관계가 높다는 연구들이 많이 발표되어 공인타당도(concurrent validity)도 검증되었다(Barsties, Lehnert, & Janotte, 2018a; Maryn, De Bodt, Barsties, & Roy, 2014; Maryn, Kim, & Kim, 2016; Uloza et al., 2017).

이러한 AVQI 회귀식은 다양한 언어권에서 검증되고 있으며 초기 버전 alpha부터 버전 2.02 (v2), 버전 3.01 (v3)까지 보고된 상태이다(Barsties et al., 2018a; Barsties & Maryn, 2015, 2016; Delgado Hernández et al., 2018; Hosokawa et al., 2017a, 2017b; Kim et al., 2018a; Maryn, Corthals, Van Cauwenberge, Roy, & De Bodt, 2010a; Maryn et al., 2014; Maryn et al., 2016; Maryn & Weenink, 2015). 초기 AVQI<sub>alpha</sub> 버전에서는 SpeechTool (Hillenbrand, 2008)로 측정된 평활화된 켈스트럴 피크 현저성(smoothed cepstral peak prominence, CPPS)과 Praat에서 산출된 5개 음향 변수들—배음 대 소음비(harmonics-to-noise ratio, HNR), 강도 변동률(shimmer local, SL), 강도 변동률dB (shimmer local dB, SLdB), 장구간 평균스펙트럼의 기울기(general slope of the spectrum, Slope), 장구간 평균스펙트럼의 추세선 기울기(tilt of the regression line through the spectrum, Tilt)—의 값을 이용하였고 두 프로그램(SpeechTool, Praat)의 사용으로 분석 시 시간과 노력이 많이 요구되었다. 이후 AVQI<sub>beta</sub>에서는 Praat을 이용한 CPPS 측정이 이루어지고 단일 프로그램 전략이 가능해져 쉽고 빠르게 분석하는 것이 가능해졌다(Maryn & Weenink, 2015). AVQI<sub>beta</sub> 버전에서 Praat 스크립트 단계는 (1) 모음연장 샘플 및 연속 말 과제 샘플의 연결(concatenation), (2) 연결음성 샘플(concatenated voice samples)의 무성음 및 숨 구간 제거, (3) 신호 대 소음비(signal to noise ratio, SNR), CPPS를 포함한 7개 음향 변수의 측정 및 AVQI를 계산하는 스크립트의 활성화로 구성되었다. 추후 AVQI 버전 2.02 (AVQIv2)로 명명되었다.

최근에는 AVQI 버전 3.01 (AVQIv3)이 소개되었고 이 회귀식은 유성음 추출의 구간이 모음의 길이 3초와 비슷하도록 조정(유성음 추출 구간 3초+모음 3초)하여 도출되었다. 선행연구들에서 제시한 음절수를 확인하면 독일어 27음절, 일본어 30음절, 스페인어 33음절, 네덜란드어 34음절 등으로 다양하였다(Barsties et al., 2018a; Barsties & Maryn, 2015; Delgado Hernández et al., 2018; Hosokawa et al., 2017b). 하지만 이것은 모든 음성샘플들을 일괄적으로 스크립트를 적용하는 것이 아니라 음성샘플에 따라 개별적으로 적용하여 분석하였다. 연구자가 수작업으로 유성음 분절 구간 길이를 확인하고 음성샘플을 편집하게 된다. 이러한 방식은 연구자의 주관 이 개입될 수 있고 숙련도에 따라 다양한 측정치가 산출될 수 있다고 하였다(Barsties & Maryn, 2015).

기존의 한국어를 대상으로 한 연구는 AVQIv2의 적용에 관한 것

만 보고되고, AVQIv3의 적용은 전무하다(Kim et al., 2018a; Kim, Lee, Lee, & Kwon, 2018b; Maryn et al., 2016). 세 연구를 통해서 한국어 화자에 대한 음성장애 정량화 및 높은 수준의 진단 예측력이 측정되었고 음성장애의 감별에도 도움이 될 것으로 확인되었다. 세 연구에서는 ‘산책’ 문단의 일부를 사용하였다(예: 넓게 펼쳐있는 바다를 바라보면 내 마음 역시 넓어지는 것 같다). 다양한 언어권에서 검증받은 AVQI이지만 더 높은 감별능력 및 진단 예측력을 위해서 지속적으로 개발, 수정되고 있다. AVQI 회귀식은 각기 다른 말 과제 구성, 언어로부터 도출되었기 때문에 한국어를 기반으로 한 회귀식이 개발된다면 음성장애의 감별에 있어 더 높은 신뢰도를 확보할 수 있을 것으로 생각되었다. 기존 연구(Kim et al., 2018a)에서 분석된 음향학적 측정들을 바탕으로 한국어인 화자 및 음성환경 등에 특화된 회귀식을 개발하고 음성장애 감별에 대한 향상된 진단 예측력을 제시하기 위해서 본 연구를 수행하였다.

따라서, 본 연구에서는 음향학적 측정 및 청지각적 평가를 이용하여 새로운 버전의 회귀식을 AVQIv4-K로 명명하기로 결정하였고, 기존의 AVQIv2, AVQIv3와 비교하려고 하였다. 본 연구 문제는 다음과 같다.

- 1) 새로운 버전의 AVQI의 회귀모형은 무엇이며, 기존의 AVQI 버전들과 측정치에 차이가 있는가?
- 2) 3개의 AVQI 버전과 청지각적 평가(G, OS)와의 상관관계는 어떠한가?
- 3) 정상과 병리적 음성 집단을 구분할 수 있는 최적 절사값 및 진단 예측력은 어떠한가?

## 연구방법

### 연구대상

연구 초기의 집단은 2,457명이었으며 음성샘플의 SNR 분석 이후 88명이 제외되었다. 모음연장 및 연속 말 과제의 연결음성샘플은 부산·울산·경남 지역의 대학병원 이비인후과에서 수집한 2,369명의 데이터베이스로부터 얻었다. 정상음성 대상자 143명, 병리적 음성장애 대상자 2,226명이 포함되었고 남성 1,084명(9-89세, 평균 53.8 ± 17.6세), 여성 1,285명(10-88세, 평균 53.3 ± 15.1세)였다. 연구에 포함된 대상들의 연령, 성별, SNR, 음성장애 진단 범주는 Table 1에 제시하였다. 본 연구 집단은 다양한 연령대, 성별, 음질, 기질적, 기능적, 신경학적 후두병리를 포함하며 음성장애 인구들을 적절히 대표하는 것으로 고려하였고, 중재(음성치료, 후두미세수술, 성대내 주입술 등) 전의 음성샘플만을 대상으로 하였다. 임상적 진단은 후두경(laryngoscopy), 후두 스트로보스코피(laryngeal videostro-

**Table 1.** The primary diagnostic categories included in the present study

Category	Age (yr)	Sex		SNR	Absolute number	% of total
		Male	Female			
Normal	65.9±10.1	115 (80.4)	28 (19.6)	40.6±7.5	143	6.0
Cancer	55.7±14.4	167 (60.1)	111 (39.9)	43.5±9.7	278	11.7
Cyst	44.1±14.6	3 (5.0)	57 (95.0)	43.7±7.7	60	2.5
Edema	45.3±17.0	7 (13.0)	47 (87.0)	43.1±8.9	54	2.3
FD	48.3±16.8	55 (41.0)	79 (59.0)	45.5±8.0	134	5.7
Leuko	38.3±19.2	123 (80.9)	29 (19.1)	45.0±8.0	152	6.4
LPRD	58.8±11.3	55 (56.7)	42 (43.3)	41.4±12.0	97	4.1
Nodules	52.6±14.2	179 (60.5)	117 (39.5)	44.9±8.3	296	12.5
Palsy	65.9±11.2	102 (73.9)	36 (26.1)	44.9±8.7	138	5.8
Papilloma	49.2±15.9	12 (18.5)	53 (81.5)	47.5±10.2	65	2.7
Polyp	36.0±16.9	43 (35.5)	78 (64.5)	48.6±7.0	121	5.1
Presby	50.8±14.1	133 (44.6)	165 (55.4)	39.2±9.1	298	12.6
PTC	61.2±12.8	17 (7.1)	222 (92.9)	41.1±7.9	239	10.1
SD	58.4±13.6	12 (6.1)	184 (93.9)	43.4±8.2	196	8.3
Sulcus	55.1±10.5	61 (62.2)	37 (37.8)	46.2±8.9	98	4.1
Total	53.5±16.3	1,084 (45.8)	1,285 (54.2)	43.4±9.1	2,369	100.0

Values are presented as mean ± SD or number (%).

SNR=signal to noise ratio; Normal=vocally healthy subjects; Cancer=glottic cancer; Cyst=vocal fold cyst; Edema=vocal fold edema; FD=functional dysphonia; Leuko=laryngeal leukoplakia; LPRD=laryngopharyngeal reflux disease; Nodules=vocal fold nodules; Palsy=vocal fold paralysis; Papilloma=laryngeal papillomatosis; Polyp=vocal fold polyp; PTC=papillary thyroid carcinoma; SD=spasmodic dysphonia; Sulcus=sulcus vocalis.

boscopy), 차트 리뷰를 통해서 확인하였다. 본 연구에서 정상음성 집단은 후두질한 및 음성장애가 없으며 청지각적으로 음성을 들었을 때 GBRAS 평가에서 G척도가 0 (normal voice)으로 판단된 음성샘플들로 분류하였다. 본 연구는 음성샘플에 대한 후향적 분석 연구로서 부산대학교병원 연구윤리위원회의 심의를 면제받았다 (No. H-1801-012-063).

### 음성샘플

/a/ 모음 연장발성을 3회 반복하여 가장 안정적으로 녹음된 샘플을 선택하고 Praat을 사용하여 발성 시작과 끝 부분을 제외한 중간 2초 구간을 편집하였다. 연속 말 과제는 '산책' 문단 중 다음 문장(예: 넓게 펼쳐있는 바다를 바라보면 내 마음 역시 넓어지는 것 같다)을 편안한 속도와 강도로 읽었다(Maryn et al., 2016) 모든 음성 녹음은 이비인후과 음성검사실 내 방음부스에서 Computerized Speech Lab 4500 (Kay Electronic, USA)과 단일지향성 다이내믹 마이크인 SM48 (Shure Inc.)을 사용하여 샘플링속도(sampling rate) 44.1 kHz, 양자화 16 bit, WAV포맷으로 저장되었다. 정확하고 타당한 음성평가를 위해서 관련 임상평가지침을 근거로 음향학적 녹음 및 분석, 청지각적 평가를 수행하였다(Choi, 2013; Patel et al., 2018).

음성분석의 높은 신뢰도를 확보하기 위해서 최소 30 dB 이상의 SNR을 확보하였다(Deliyski, Shaw, & Evans, 2005; Ingrisano, Perry, & Jepson, 1998). 전체 음성샘플들의 SNR 범위는 16.3-75.7 dB이었으며, 평균 43.4 dB이었다. SNR이 30 dB 미만인 88명은 본 연구에서 제외하였다.

### 청지각적 평가

전반적인 음질에 대해서 5명의 1급 언어재활사가 청지각적 평가를 수행하였다. 평가자는 5명의 언어재활사로 구성되어 있으며, 음성평가, 음성치료를 전문으로 하며 청지각적 평가에 대한 경험이 5년에서 20년까지(평균 11.4 ± 5.5년) 다양하였다. 각 평가자는 연결된 음성샘플의 전반적인 음질을 평가했다. 청지각적 평가 절차는 이전 연구에서 수행한 것과 동일하였다(Kim et al., 2018a). 평가자들은 GRBAS 척도 중 Grade (G), Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)에서 전반적인 음성 중증도(overall severity, OS)를 사용하였다. GRBAS는 4점 척도를 사용하며, G점수는 비정상적인 음성 중증도(0=normal voice, 1=mild dysphonia, 2=moderate dysphonia, 3=severe dysphonia)를 표현해준다(Hirano, 1981). CAPE-V의 전반적인 음성의 중증도(OS)는 100 mm 길이의 막대선에 증상의 정도에 해당하는 위치에 표시를 한 후 표시된 위치까지의 길이를 측정하여 수치를 기록하는 시각적 아날로그 척도(visual analogue scale)를 사용하며, 음성장애가 심할수록 큰 수치로 기록하였다(Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2009). 모든 음성샘플은 소음이 40 dB 아래로 통제된 조용한 방에서 제시되었고 헤드셋을 사용하여 청지각적 평가를 수행하였다. 평가자들은 최종 판정을 내리는 데 필요한 만큼 반복하여 음성샘플을 들을 수 있었다.

학습효과를 최소화하기 위하여 한 세션 내에서 음성샘플들은 무작위로 제시되고 평가되었다. 또한 모든 평가자들은 음성샘플에 대한 사전 정보 없이 청지각적 평가를 수행하였다. 반복 평가로 인한 피로도, 집중력 저하와 같은 요소들을 통제하기 위해서 30개의 음성샘플을 평가하고 짧은 휴식시간을 가졌다(Kreiman, Gerratt, Kempster, Erman, & Berke, 1993).

### 음향학적 측정

Maryn 등(2010a)이 제시한 Praat 스크립트에 의해서 모음과 연속 말 과제를 연결하는 샘플을 만들고, 무성음, 쉼 구간 등의 분리가 이루어졌다. AVQI를 계산하기 위해서 Praat을 이용하여 6개의 음향 변수들(CPPS, HNR, SL, SLdB, Slope, Tilt)을 측정하였다. CPP는 캡스트럼의 첫 번째 라모닉(rahmonic) 위치에서의 피크가



뚜렷해지는데 그 주변의 값들에 비해 얼마만큼 두드러지는지를 통해 주기성의 정도를 표현한다. CPPS는 일정 범위의 인접 프레임 및 인접 쿼프렌시 영역 모두에 대해 평활화 작업을 거친 로그 캡스טר럼으로부터 구한 CPP를 의미한다. HNR은 배음에너지와 소음에너지 간의 비율이며, SL, SLdB는 진폭에 대한 변수이다. Slope는 0-1,000 Hz 사이의 장구간 평균스펙트럼 에너지와 1,000-10,000 Hz 사이의 장구간 평균스펙트럼 에너지의 차이이며, Tilt는 slope에 대한 회귀선 기울기를 의미한다. 각 버전에 따른 AVQI 회귀식은 다음과 같다(Table 2).

**통계분석**

청지각적 평가의 평가자 내 신뢰도를 평가하기 위해서 전체 음성 샘플 중 약 10%에 해당하는 240명의 음성샘플을 무작위로 선정하여 첫 평가 후 2주 뒤 재평가를 실시하였다. 또한 평가자 간 신뢰도를 측정하기 위해서 평가자 간 청지각적 평가 점수들이 비교되었다. 평가자 간 신뢰도는 이원배치 혼합효과 모형(two-way mixed effects model)을, 평가자 내에는 일원배치 임의효과 모형(one-way random effect model)을 활용하여 각 G와 OS별로 단일 측도(single measures)의 급내상관계수(intraclass correlation coefficient, ICC)를 측정하였다. 그 결과 평가자 간 신뢰도(G .779-.938 [mean = .866]; OS .744-.893 [mean = .808]), 평가자 내 신뢰도(G .786-.917 [mean = .875]; OS

.757-.866 [mean = .826])는 높은 수준이었다.

새로운 버전의 AVQI 모형을 도출하기 위해서 SPSS 22.0 프로그램(IBM-SPSS Inc., Armonk, NY, USA)을 사용하였다. 선행연구에서 제시한 방법으로, 선형회귀방정식의 결과(G척도)가 0-10 사이의 점수로 출력되도록 선형재조정(linear rescaling)한 후 단계적 다중선형회귀분석(stepwise multiple linear regression analysis)을 이용하여  $G_{rescaling}$  척도와 6개의 변수들(PraatCPPS, HNR, SL, SLdB, Slope, Tilt)과의 관계를 분석하였다(Barsties & Maryn, 2015; Barsties, Maryn, Gerrits, & De Bodt, 2017). 3개의 AVQI 버전을 비교하기 위해서 일원배치 분산분석과 Bonferroni 사후분석을 실시하였고 유의수준은  $p < .017$ 로 설정하였다. 3개의 AVQI 버전과 청지각적 평가(G, OS) 간에 상관관계를 확인하기 위하여 피어슨 상관분석(Pearson correlation)을 실시하였고 유의수준은  $p < .05$ 로 설정하였다. 정상과 병리적 음성 집단을 감별하기 위한 receiver operating characteristic curve (ROC) 분석을 실시하여 민감도(sensitivity)와 특이도(specificity), 곡선 하 영역(area under the curve, AUC), 양성우도비(likelihood ratio for a positive result, LR+), 음성우도비(likelihood ratio for a negative result, LR-)를 산출하였고, 각 버전의 AUC 간에 쌍대 비교(pairwise comparison)를 수행하였다. 피어슨 상관분석 및 ROC 분석은 R version 3.5.1 (The R Foundation for Statistical Computing, Vienna, Austria)과 RStudio 1.1.456 (RStu-

**Table 2.** Description of AVQI versions

AVQI version	Regression formula	Contents	Note
AVQI <sub>alpha</sub>	$8.471 - (.285 \times CPPS_{SpeechTool}) - (.188 \times HNR) - (.548 \times SL) + (7.171 \times SLdB) - (.082 \times Slope) - (.198 \times Tilt)$	3 seconds vowel+ 17 syllables	Use SpeechTool for CPPS and Praat for AVQI measurement
AVQI <sub>beta</sub> version 2.02 (v2)	$9.072 - (.245 \times CPPS_{Praat}) - (.161 \times HNR) - (.470 \times SL) + (6.158 \times SLdB) - (.071 \times Slope) - (.170 \times Tilt)$	3 seconds vowel+ 17 syllables	Use Praat for CPPS and AVQI measurement
AVQI version 3.01 (v3)	$(4.152 - (.177 \times CPPS_{Praat}) - (.006 \times HNR) - (.037 \times SL) + (.941 \times SLdB) + (.01 \times Slope) + (.093 \times Tilt)) \times 2.8902$	3 seconds vowel+ 27-34 syllables	Use Praat for CPPS and AVQI measurement

AVQI=Acoustic Voice Quality Index; CPPS=smoothed cepstral peak prominence; HNR=harmonics-to-noise ratio; SL=shimmer local; SLdB=shimmer local dB.

**Table 3.** Multiple linear regressions analysis predicting AVQIv4-K using PraatCPPS, HNR, SL, SLdB, Slope, and Tilt

Model		Unstandardized coefficients		Standardized coefficient B	t	Sig.	R	R <sup>2</sup>	Adj. R <sup>2</sup>	VIF
		B	SE							
Step 6	(Constant)	12.877	.399		32.288	.000***	.869	.755	.755	
	PraatCPPS	-.404	.016	-.462	-25.371	.000				3.195
	Tilt	.320	.013	.290	25.481	.000				1.249
	HNR	-.186	.011	-.294	-16.991	.000				2.893
	Slope	-.067	.007	-.124	-9.434	.000				1.672
	SLdB	3.300	.649	.344	5.081	.000				4.311
	SL	-.223	.056	-.270	-3.961	.000				4.875

AVQI=Acoustic Voice Quality Index; CPPS=smoothed cepstral peak prominence; HNR=harmonics-to-noise ratio; SL=shimmer local; SLdB=shimmer local dB.

\*\*\* $p < .001$ .

dio Inc., Boston, MA, USA)을 이용하여 분석하였다.

## 연구결과

### 새로운 버전의 AVQI 모형

단계적 다중선형회귀분석을 사용하여  $G_{\text{rescaling}}$  척도와 6개의 음향 변수들과의 관계를 분석하여 Table 3에 제시하였다. 최종 모형 6 단계에서 결정되었고 유의미한 것으로 나타났다( $R = .869, R^2 = .755, p < .001$ ). 다중공선성(Multicollinearity)에 있어서 공차한계가 .1보다 크고 분산팽창요인(variance inflation factor, VIF)이 10보다 작은 경우 다중공선성이 없는 것으로 판단한다(Graham, 2003). Praat-CPPS, HNR, SL, SLdB, Slope, Tilt 등 6개 요인 모두가 이러한 기준을 충족시키며 다중공선성의 문제는 없는 것으로 나타났다. 새로운 버전의 AVQIv4-K는 다음과 같다.

$$AVQIv4-K = 12.877 - (.404 \times CPPS) - (.186 \times HNR) - (.223 \times SL) + (3.3 \times SLdB) - (.067 \times Slope) + (.32 \times Tilt).$$

**Table 4.** Comparison of AVQI according to AVQI versions

	Normal group	Pathological group	p-value
AVQIv2	2.6±.7	5.2±1.4	<.001
AVQIv3	1.1±.9	4.0±1.9	<.001
AVQIv4-K	1.6±.9	5.1±2.0	<.001

Values are presented as mean ± SD.  
AVQI = Acoustic Voice Quality Index.

### 3가지 버전의 AVQI 비교

3가지 버전(AVQIv2, AVQIv3, AVQIv4-K)에 따른 AVQI 측정치의 기술통계와 박스플롯은 Table 4와 Figure 1에 제시하였다. 각 버전에 따른 정상과 병리적 음성 집단의 평균은 각각 AVQIv2에서  $2.6 \pm .7$  vs.  $5.2 \pm 1.4$ , AVQIv3에서  $1.1 \pm .9$  vs.  $4.0 \pm 1.9$ , AVQIv4-K에서  $1.6 \pm .9$  vs.  $5.1 \pm 2.0$ 으로 집단 간 차이가 유의미한 것( $p < .001$ )으로 확인되었다. 전체 대상자에 대한 AVQI 측정치는 AVQIv2 ( $5.04 \pm 1.46$ ), AVQIv4-K ( $4.92 \pm 2.10$ ), AVQIv3 ( $3.87 \pm 1.98$ ) 순으로 크게 나타났다. Bonferroni 사후분석을 실시한 결과, AVQIv3는 AVQIv2와 AVQIv4-K에 비해 유의미하게 낮은 반면( $p < .001$ ), AVQIv2와 AVQIv4-K 간에는 유의미한 차이가 없는 것( $p = .091$ )으로 확인되었다.

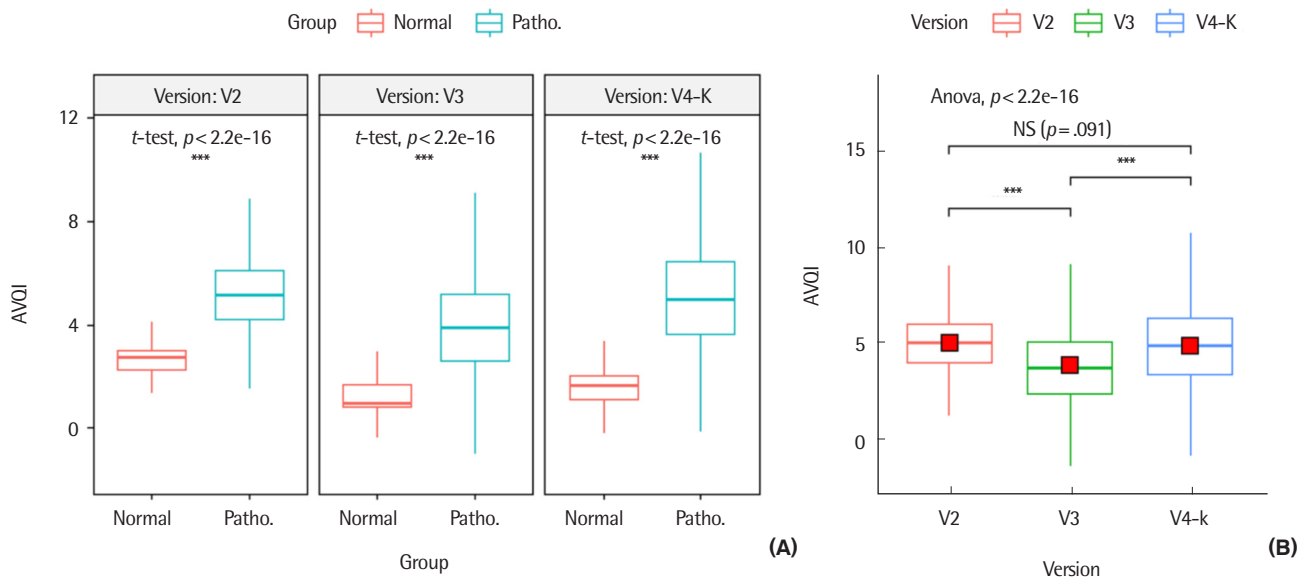
### 3가지 버전의 AVQI와 청지각적 평가(G, OS) 간의 상관관계

AVQI와 청지각적 평가 간의 상관관계를 분석한 결과, 3가지 버전의 AVQI는 G, OS와 모두 유의미한 상관관계를 보였으며 Table 5

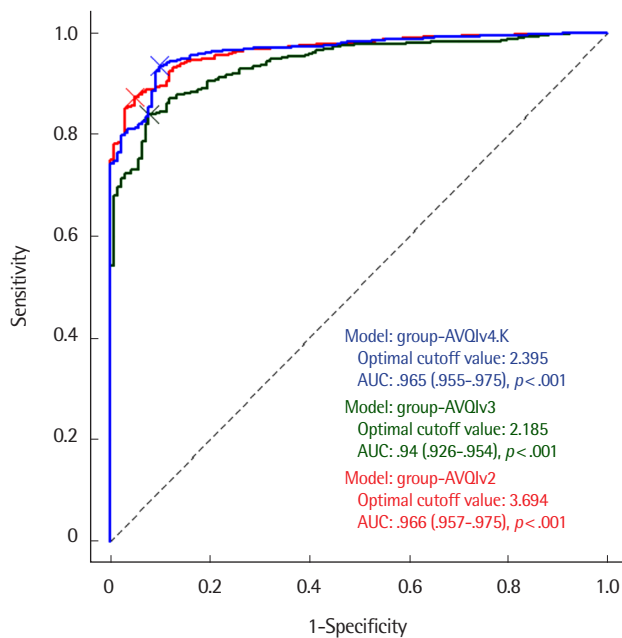
**Table 5.** Correlation coefficient between acoustic measures and A-P ratings

	AVQIv2	AVQIv3	AVQIv4-K	Grade	OS
AVQIv2	1	.950**	.987**	.858**	.826**
AVQIv3	-	1	.978**	.850**	.815**
AVQIv4-K	-	-	1	.869**	.835**
Grade	-	-	-	1	.859**

AVQI = Acoustic Voice Quality Index; OS = overall severity.  
\*\* $p < .01$ .



**Figure 1.** Comparison of Acoustic Voice Quality Index (AVQI) between normal and pathological group across AVQI versions (A) and Bonferroni's post-hoc analysis of AVQI value (B).



**Figure 2.** Receiver operator characteristic curve of AVQI versions. AVQI=Acoustic Voice Quality Index; AUC=area under the curve.

에 제시하였다. 청지각적 평가와의 상관관계는 AVQIv4-K가 가장 높았다.

### 정상과 병리적 음성 집단을 감별하기 위한 ROC 분석

병리적 음성 집단을 선별하기 위한 ROC 분석을 실시한 결과를 Figure 2, Table 6에 제시하였다. 3가지 버전의 AVQI에 대한 최적 절사값은 각각 3.694, 2.185, 2.395이었고 AUC는 AVQIv2 (.966), AVQIv4-K (.965), AVQIv3 (.940) 순으로 높았다. 각 버전의 AUC는 .94 이상으로 모두 뛰어난 수준(excellent)이었다. 각 버전의 LR(+)는 17.82, 10.90, 9.54, LR(-)는 .13, .17, .07로 나타났다. AUC를 쌍대 비교한 결과, 사후 분석과 마찬가지로 AVQIv3는 AVQIv2, AVQIv4-K와 모두 유의미한 차이를 보였지만 AVQIv2, AVQIv4-K 간에는 차이가 없었다(Table 7).

### 논의 및 결론

본 연구에서는 새로운 버전의 AVQI 모형을 제시하고, 음성장애의 정량화 측정치를 기존 버전들과 비교하고, 청지각적 평가와의 상관관계도 파악하고, 병리적 음성을 감별할 수 있는 최적 절사값 및 진단 예측값도 제시하였다. 본 연구의 질문에 대한 논의는 다음과 같다.

첫째, 새로운 버전의 AVQI의 회귀모형은 무엇이며 기존 AVQI

**Table 6.** Descriptive statistics of ROC curve according to AVQI versions

	Sensitivity	Specificity	Cutoff	AUC	95% CI	LR(+)	LR(-)
AVQIv2	.873	.951	3.694	.966	.955-.975	17.82	.13
AVQIv3	.839	.923	2.185	.940	.926-.954	10.90	.17
AVQIv4-K	.935	.902	2.395	.965	.957-.975	9.54	.07

ROC=receiver operator characteristic; AVQI=Acoustic Voice Quality Index; AUC=area under the curve; CI=confidence interval; LR(+)=likelihood ratio for a positive result; LR(-)=likelihood ratio for a negative result.

**Table 7.** Pairwise comparison of ROC curve

	Difference between areas	SE	95% CI	Z statistics	p-value
AVQIv2-AVQIv3	.0261	.00557	.0152-.0370	4.684	< .0001
AVQIv2-AVQIv4-K	.000993	.00260	-.00411-.00609	.381	.7029
AVQIv3-AVQIv4-K	.0251	.00415	.0169-.0332	6.040	< .0001

ROC=receiver operator characteristic; AVQI=Acoustic Voice Quality Index; SE=standard error; CI=confidence interval.

버전의 측정치에 차이가 있는가?

본 연구에서 새로운 버전의 AVQI 회귀모형이 제시되었다. 6개의 음향 변수들의 가중치는 이전 버전에 비하여 변경되었다. 3개 버전의 AVQI 중에서 AVQIv3는 AVQIv2와 AVQIv4-K보다 유의미하게 작은 값으로 측정되었고, AVQIv2와 AVQIv4-K 간에는 유의미한 차이는 없었다. 이러한 결과는 선행연구의 결과와 유사하였다 (Barsties & Maryn, 2015). 이는 각 버전의 회귀모형이 만들어진 음성샘플의 구성과 연관이 있다고 여겨진다. 3가지 버전의 AVQI 회귀식 중에서 연속 말 과제와 구간의 길이가 가장 길게 설정된 것은 AVQIv3이다. AVQIv2는 17음절과 3초의 모음 안정구간을 연결한 음성샘플을, AVQIv3은 연속 말 과제의 음성샘플을 무성음, 쉼 구간 제거 스크립트를 적용한 후(유성음 분절만 포함)의 최종 길이가 3초 정도로 조정된 후 3초의 모음을 연결하여 분석하였다. AVQIv3에서 제시한 방법은 음성샘플에서 유성음 분절을 추출한 길이가 3초에 가깝도록 조절하는 것이며 대상자에 따라 24-61음절로 다양하였다. 연속 말 과제와 모음의 길이를 동등하게 하여 AVQI에 미치는 영향 또한 비슷하게 맞추려고 하였지만 자동화된 작업을 할 수 없고 연구자가 수작업으로 구간을 조절해야 하는 제한점이 있다. 그렇기 때문에 재현성 및 신뢰도가 떨어지고 Praat 스크립트를 사용하는 의미가 약해진다고 하였다. 본 연구의 AVQIv4-K는 25음절과 2초의 모음을 연결하여 회귀모형을 도출하였다. 한국어 화자들이 널리 사용하는 '산책' 문단을 분석하여 음성장애 중증도를 정량화할 수 있는 회귀모형을 제시하려고 하였다.

25음절의 유성음 추출에 대한 예비조사(pilot study)에서 2.1 ±

.8초로 나타나 모음구간을 2초로 선정하였다. 앞으로 유성음 추출에 대한 연구를 후속 진행하여 모음 2초, 3초에 대응하는 문장 음절수를 분석한다면 AVQI에 미치는 모음과 연속발화의 영향을 확인할 수 있을 것으로 생각된다.

기존 연구들은 전체 대상자의 AVQI 측정치가 3.07-7.13으로 다양하였으며(Hosokawa et al., 2017a; Kim et al., 2018a; Maryn et al., 2010a; Maryn et al., 2010b; Maryn et al., 2014; Maryn & Weenink, 2015; Núñez-Batalla, Díaz-Fresno, Álvarez-Fernández, Cordero, & Pendás, 2017), 정상과 병리적 음성 집단으로 나누어 분석해 보면, 초기 AVQI 버전에서 정상집단은 2.29, 병리적 음성 집단은 4.16으로 보고되었고(Maryn et al., 2010b), AVQIv2 버전에서는 정상집단은 1.56-6.50, 병리적 음성 집단은 4.79-7.76으로 보고되었다(Hosokawa et al., 2017a; Kim et al., 2018a; Núñez-Batalla et al., 2017). AVQIv3 버전에서 정상집단은 2.33, 병리적 음성 집단은 4.46으로 보고되었고, 각 집단 간 비교보다는 진단 예측력에 관한 연구가 주를 이루었다(Barsties et al., 2018a; Barsties & Maryn, 2016; Barsties, Ulozaite-Staniene, Petrauskas, Uloza, & Maryn, 2018b; Delgado Hernández et al., 2018; Hosokawa et al., 2017b; Lee, Roy, Peterson, & Merrill, 2018).

AVQI 회귀식에는 PraatCPPS 값이 포함되기 때문에 연속 말 과제의 길이가 AVQI 측정치에 미치는 영향이 있다는 것으로 판단할 수 있다. 연속 말 과제에는 숨이나 무성음 구간 등이 포함되어 상대적으로 모음 과제보다 CPPS 값이 상대적으로 낮다는 사실에 근거하여 추정할 수 있다(Awan, Giovinco, & Owens, 2012; Awan, Helou, Stojadinovic, & Solomon, 2011; Heman-Ackah et al., 2003; Hillenbrand & Houde, 1996). 하지만 읽기 구간이 무조건 길다고 많은 영향을 미치는 것은 아니기 때문에 음성의 분석 및 신뢰도 확보를 위한 적절한 구간, 음절의 선택은 필요한 것으로 생각한다(Barsties & Maryn, 2015). 본 연구에서 AVQIv3 분석을 위해서 개별적인 구간 조정을 하지 않았지만 추후 연구에서는 독일 27음절, 네덜란드어 34음절, 일본 30음절, 스페인 33음절 등과 같이 AVQIv3에 맞는 읽기 음절수를 결정하는 것도 필요할 것으로 생각한다(Barsties et al., 2018a; Barsties & Maryn, 2015, 2016; Delgado Hernández et al., 2018; Hosokawa et al., 2017b).

둘째, 3개의 AVQI 버전과 청지각적 평가(G, OS)와의 상관관계는 어떠한가?

우선 각 버전 간의 AVQI 측정치들은 .950 이상의 높은 상관관계를 보였으며 AVQI 측정치들은 G와는 .850 이상, OS와는 .815 이상의 높은 상관관계를 나타내었다. AVQI와 청지각적 평가 간의 높은 상관관계는 지속적으로 보고되어 왔다. 선행연구에서 초기 버전의

AVQI는 GBRAS 척도의 G와 .781-.905, AVQIv2는 .828-.911 (G), .680-.924 (OS), .680 (breathy, B), AVQIv3는 G와 .815-.860, 기식음(B)과는 .835-.850의 상관관계를 보고하였다(Barsties et al., 2018a; Barsties & Maryn, 2012, 2015; Delgado Hernández et al., 2018; Kim et al., 2018a; Lee et al., 2018; Maryn et al., 2010b; Maryn et al., 2014; Maryn et al., 2016; Maryn & Weenink, 2015; Núñez-Batalla et al., 2017; Reynolds et al., 2012; Uloza et al., 2017). 선행연구에서 성대용종 환자의 후두미세수술(laryngeal microsurgery, LMS) 전후의 음성을 비교한 결과, 수술 전후 모두 AVQI를 포함한 음향학적 측정과 청지각적 평가 간의 상관관계(G .703-.749; OS .800-.884)가 ‘중간-강함’ 수준으로 확인되어, 음성회복을 예측하는 데 도움이 된다고 보고하였다(Kim et al., 2018b). AVQI는 중증도가 높은 음성을 신뢰성 있게 분석이 가능하기 때문에 성대마비 환자의 성대 내 주입술 전후의 비교나 성문틈(glottal gap)이 큰 환자의 음성 치료 전후의 음성 비교도 가능할 것으로 생각된다.

각 버전의 AVQI 중 새롭게 제시된 AVQIv4-K와 가장 높은 상관관계를 보였다. 한국어 화자들이 수행한 말 과제들을 통해서 회귀식을 도출해 내었기 때문으로 사료된다. 본 연구에서는 임상현장에서 주로 사용되는 ‘산책’ 문단과 ‘가을’ 문단 중에서 ‘산책’ 문단만을 사용하여 회귀식을 도출하였다. 문장 내 포함되어 있는 자음, 모음의 음소 비율에 따라서 AVQI 측정치가 달라질 수 있으며 후속 연구에서는 다양한 문단 및 위치의 문장들을 사용하여 비교하는 것도 필요할 것으로 생각된다(Lee, Lim, & Choi, 2017).

셋째, 정상과 병리적 음성 집단을 구분할 수 있는 최적 절사값 및 진단 예측력(AUC)은 어떠한가?

초기 버전에서는 최적 절사값이 2.70-3.66이었고 AVQIv2에서는 2.30-3.96, AVQIv3는 1.41-4.00의 범위로 나타났고 본 연구의 결과는 비슷하였다(Barsties & Maryn, 2012, 2015, 2016; Delgado Hernández et al., 2018; Hosokawa et al., 2017a; Kankare et al., 2015; Maryn et al., 2010b; Maryn et al., 2014; Reynolds et al., 2012; Uloza et al., 2017). 최적 절사값은 적용된 AVQI 버전과 분석에 포함된 모음의 길이, 연속 말 과제의 내용에 따라 달라질 수 있다. 본 연구에서 각 버전의 AVQI를 통해서 병리적 음성을 감별할 수 있는 AUC는 .940 이상, 10 이상의 LR(+), .2 이하의 LR(-)로 매우 높은 수준이었다. 기존 연구에서도 초기 버전 AVQI에서 .876-.940의 AUC, LR(+) 10.25-19.98, LR(-) .14-.27, AVQIv2에서는 .905-.962의 AUC, 10.81-15.10의 LR(+), .17-.29의 LR(-), AVQIv3 연구에서는 .842-.915의 AUC, 5.79-13.85의 LR(+), .26-.32의 LR(-)를 보고하였다(Barsties & Maryn, 2015; Barsties et al., 2018b; Delgado Hernández et al., 2018; Hosokawa et al., 2017a, 2017b; Maryn et al., 2010b; Reynolds



et al., 2012; Uloza et al., 2017; Uloza et al., 2018). LR은 전체 경우의 수 중 특정 사건이 나타나는 경우의 수를 확률로 제시한 것이다 (Attia, 2003). LR(+)의 측정은 실제로 장애가 없지만 특정 검사에서 양성(positive)으로 판정되는 경우가 어느 정도인지를 알아야 한다. 민감도를 (1-특이도)로 나눔으로써 확인할 수 있다(Hwang, 2012). 또한 LR(-)를 구하기 위해서 특정 검사에서 음성(negative) 판정 시 실제로 해당 질병이 존재할 확률이 얼마나 낮은지를 알아야 한다. LR(+)는 1 이상, LR(-)는 0-1 사이의 소수로 나타나며, LR(+)가 클수록, LR(-)가 작을수록 검사의 정확도가 증가한다고 알려져 있다. 일반적으로 LR(+)가 10 이상, LR(-)가 .2 이하 정도면 진단 예측력이 높은 검사라도 알려져 있으며, LR(+)가 20 이상, LR(-)가 .1 이하이면 높은 수준의 진단 정확도라고 하였다(Deeks & Altman, 2004; Dollaghan, 2004). 선행연구와 같은 기준을 적용하더라도, 본 연구에서 측정된 각 버전의 AVQI는 음성장애 환자들에 있어서 음성의 비정상성에 대한 강력한 예측 지표로 보인다. 이러한 AVQI 이외에도 Acoustic Breathiness Index (ABI), Dysphonia severity index (DSI) 등과 같은 다양한 지수(index)들이 소개되고 있다(Barsties et al., 2018a; Barsties et al., 2018b; Hakkesteeft, Brocaar, Wieringa, & Feenstra, 2006; Delgado Hernández et al., 2018; Uloza et al., 2017; Wuyts et al., 2000). 단일 변수가 아닌 다중 변수를 사용하여 음성장애의 중증도를 확인하고 감별하기 위한 시도는 지속되어야 한다.

종합해 보면, 본 연구의 결과들은 AVQI를 통해서 다양한 유형의 음성장애를 수치화하고 주관적인 청지각적 평가와 상관관계를 확인하여 AVQI의 임상적 유용성을 뒷받침하였다는 점에서 의의가 있다. 동일한 음성자료를 분석하더라도 AVQI의 회귀식에 따라 측정치가 다양해지며 최적 절사값 및 진단 예측력이 달라졌다. 25음절과 2초의 연장모음을 이용한 AVQIv4-K는 기존의 AVQI 버전과 비교하여 동등한 진단 예측력을 갖는 것으로 확인되었다. 후속 연구에서는 25음절에 대해서 유성음 분절 추출 후 구간 길이가 얼마나 되는지, 2초 혹은 3초의 유성음 분절 길이를 확보하기 위한 최적 음절수를 확인하는 것도 중요한 작업일 것이다. 또한 다양한 읽기 문장의 종류, 음절수를 대상으로 AVQI 측정을 실시한다면 음성장애 정량화 및 진단 도구로서의 유용성에 대한 더욱 세밀하고 정확한 정보를 제공할 수 있을 것이다.

## REFERENCES

- Attia, J. (2003). Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests. *Australian Prescriber*, 26, 111-113.
- Awan, S. N., Giovinco, A., & Owens, J. (2012). Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of Voice*, 26, 670-e15.
- Awan, S. N., Helou, L. B., Stojadinovic, A., & Solomon, N. P. (2011). Tracking voice change after thyroidectomy: application of spectral/cepstral analyses. *Clinical Linguistics & Phonetics*, 25, 302-320.
- Barsties, B., Lehnert, B., & Janotte, B. (2018a). Validation of the Acoustic Voice Quality Index version 03.01 and Acoustic Breathiness Index in German. *Journal of Voice*. Advanced online publication. DOI: 10.1016/j.jvoice.2018.07.026.
- Barsties, B., & Maryn, Y. (2012). Der Acoustic Voice Quality Index in Deutsch. *HNO*, 60, 715-712.
- Barsties, B., & Maryn, Y. (2015). The improvement of internal consistency of the Acoustic Voice Quality Index. *American Journal of Otolaryngology*, 36, 647-656.
- Barsties, B., & Maryn, Y. (2016). External validation of the Acoustic Voice Quality Index version 03.01 with extended representativity. *Annals of Otolaryngology, Rhinology & Laryngology*, 125, 571-583.
- Barsties, B., Maryn, Y., Gerrits, E., & De Bodt, M. (2017). The acoustic breathiness index (ABI): a multivariate acoustic model for breathiness. *Journal of Voice*, 31, 511-e11.
- Barsties, B., Ulozaitė-Stanienė, N., Petrauskas, T., Uloza, V., & Maryn, Y. (2018b). Diagnostic Accuracy of Dysphonia Classification of DSI and AVQI. *The Laryngoscope*. Advanced online publication. DOI: 10.1002/lary.27350.
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341-345.
- Choi, S. H. (2013). Speech-language pathologists' voice assessment and voice therapy practices: a survey for standard clinical guideline and evidence-based practice. *Communication Sciences & Disorders*, 18, 473-485.
- Deeks, J. J., & Altman, D. G. (2004). Diagnostic tests 4: likelihood ratios. *BMJ*, 329, 168-169.
- Delgado Hernández, J., León Gómez, N. M., Jiménez, A., Izquierdo, L. M., & Barsties v. Latoszek, B. (2018). Validation of the Acoustic Voice Quality Index version 03.01 and the Acoustic Breathiness Index in the Spanish language. *Annals of Otolaryngology, Rhinology & Laryngology*, 127, 317-326.
- Deliyski, D. D., Shaw, H. S., & Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice*, 19, 15-28.
- Dollaghan, C. A. (2004). Evidence-based practice in communication disorders: what do we know, and when do we know it? *Journal of Communication Disorders*, 37, 391-400.
- Graham, M. H. (2003). Confronting multicollinearity in ecological multiple



- regression. *Ecology*, 84, 2809-2815.
- Hakkesteegt, M. M., Brocaar, M. P., Wieringa, M. H., & Feenstra, L. (2006). Influence of age and gender on the Dysphonia Severity Index. *Folia Phoniatrica et Logopaedica*, 58, 264-273.
- Heman-Ackah, Y. D., Michael, D. D., Baroody, M. M., Ostrowski, R., Hillenbrand, J., Heuer, R. J., ... & Sataloff, R. T. (2003). Cepstral peak prominence: a more reliable measure of dysphonia. *Annals of Otolaryngology & Laryngology*, 112, 324-333.
- Hillenbrand, J. (2008). *SpeechTool with CPP scripts*. <https://homepages.wmich.edu/~hillenbr/>.
- Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39, 311-321.
- Hirano, M. (1981). Psycho-acoustic evaluation of voice. In *Clinical examination of voice* (pp. 81-84). Wien: Springer.
- Hosokawa, K., Barsties, B., Iwahashi, T., Iwahashi, M., Iwaki, S., Kato, C., ... & Inohara, H. (2017b). The Acoustic Voice Quality Index version 03.01 for the Japanese-speaking population. *Journal of Voice*. Advanced online publication. DOI: 10.1016/j.jvoice.2017.10.003.
- Hosokawa, K., Barsties, B., Iwahashi, T., Iwahashi, M., Kato, C., Iwaki, S., ... & Ogawa, M. (2017a). Validation of the acoustic voice quality index in the Japanese language. *Journal of Voice*, 31, 260-e1.
- Hwang, M. (2012). Sentence repetition as a clinical marker of specific language impairment in Korean-speaking preschool children. *Korean Journal of Communication & Disorders*, 17, 1-14.
- Ingrisano, D. R., Perry, C. K., & Jepson, K. R. (1998). Environmental noise: a threat to automatic voice analysis. *American Journal of Speech-Language Pathology*, 7, 91-96.
- Kankare, E., Barsties, B., Maryn, Y., Ilomäki, I., Laukkanen, A. M., Tyrmi, J., ... & Vilpas, S. (2015). A preliminary study of the Acoustic Voice Quality Index in Finnish speaking population. *Proceedings of the 11th Pan European Voice Conference*, Florence, Italy.
- Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, 18, 124-132.
- Kim, G. H., Lee, Y. W., Bae, I. H., Park, H. J., Wang, S. G., & Kwon, S. B. (2018a). Validation of the acoustic voice quality index in the Korean language. *Journal of Voice*. Advanced online publication. DOI: 10.1016/j.jvoice.2018.06.007.
- Kim, G. H., Lee, Y. W., Lee, B. J., & Kwon, S. B. (2018b). Acoustic and auditory-perceptual evaluation as predictor of voice recovery after laryngeal microsurgery in patients with vocal polyp. *Korean Journal of Otorhinolaryngology-Head and Neck Surgery*, 61, 361-369.
- Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech, Language, and Hearing Research*, 36, 21-40.
- Lee, J. M., Roy, N., Peterson, E., & Merrill, R. M. (2018). Comparison of two multiparameter acoustic indices of dysphonia severity: the Acoustic Voice Quality Index and Cepstral Spectral Index of dysphonia. *Journal of Voice*, 32, 515-e1.
- Lee, S. J., Lim, S. E., & Choi, H. S. (2017). A comparison of cepstral and spectral measures according to measurement position in a reading passage. *Communication Sciences & Disorders*, 22, 818-826.
- Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010a). Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *Journal of Voice*, 24, 540-555.
- Maryn, Y., De Bodt, M., Barsties, B., & Roy, N. (2014). The value of the Acoustic Voice Quality Index as a measure of dysphonia severity in subjects speaking different languages. *European Archives of Oto-Rhino-Laryngology*, 271, 1609-1619.
- Maryn, Y., De Bodt, M., & Roy, N. (2010b). The Acoustic Voice Quality Index: toward improved treatment outcomes assessment in voice disorders. *Journal of Communication Disorders*, 43, 161-174.
- Maryn, Y., Kim, H. T., & Kim, J. (2016). Auditory-perceptual and acoustic methods in measuring dysphonia severity of Korean speech. *Journal of Voice*, 30, 587-594.
- Maryn, Y., & Weenink, D. (2015). Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index. *Journal of Voice*, 29, 35-43.
- Núñez-Batalla, F., Díaz-Fresno, E., Álvarez-Fernández, A., Cordero, G. M., & Pendás, J. L. L. (2017). Application of the acoustic voice quality index for objective measurement of dysphonia severity. *Acta Otorrinolaringologica (English Edition)*, 68, 204-211.
- Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., ... & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology*, 27, 887-

- 905.
- Reynolds, V., Buckland, A., Bailey, J., Lipscombe, J., Nathan, E., Vijayasekaran, S., ... & French, N. (2012). Objective assessment of pediatric voice disorders with the acoustic voice quality index. *Journal of Voice*, 26, 672-e1.
- Uloza, V., Barsties, B., Ulozaite-Staniene, N., Petrauskas, T., & Maryn, Y. (2018). A comparison of Dysphonia Severity Index and Acoustic Voice Quality Index measures in differentiating normal and dysphonic voices. *European Archives of Oto-Rhino-Laryngology*, 275, 949-958.
- Uloza, V., Petrauskas, T., Padervinskis, E., Ulozaitė, N., Barsties, B., & Maryn, Y. (2017). Validation of the acoustic voice quality index in the Lithuanian language. *Journal of Voice*, 31, 257-e1.
- Wuyts, F. L., De Bodt, M. S., Molenberghs, G., Remacle, M., Heylen, L., Millet, B., ... & Van de Heyning, P. H. (2000). The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach. *Journal of Speech, Language, and Hearing Research*, 43, 796-809.

## 국문초록

### 한국인 화자에서 새로운 버전의 Acoustic Voice Quality Index 적용

김근호<sup>1</sup>(언어재활사, 제1저자) · 이연우<sup>1</sup>(언어재활사) · 배인호<sup>2</sup>(언어재활사) · 박희준<sup>3</sup>(교수) · 권순복<sup>4</sup>(교수, 교신저자)

<sup>1</sup>부산대학교병원 이비인후과, <sup>2</sup>양산부산대학교병원 이비인후과, <sup>3</sup>춘해보건대학교 언어치료과, <sup>4</sup>부산대학교 언어정보학과

**배경 및 목적:** 본 연구에서는 한국인 화자에서 정상음성과 병리적 음성을 정량화하는 Acoustic Voice Quality Index (AVQI)의 새로운 회귀식을 제시하여 기존의 회귀식(AVQIv2, AVQIv3)과 비교하고자 하였다. 또한 AVQI 측정치와 청지각적 평가 간의 상관관계도 알아보고, 정상과 병리적 음성을 감별할 수 있는 최적의 절사값 및 진단 예측력도 알아보았다. **방법:** 한국인 화자 2,369명으로부터 모음연장 발성 /아/ 5초와 표준문단 '산책' 읽기를 통해서 녹음하였다. 모음의 안정구간 2초와 읽기샘플 25음절을 연결하여 Praat 스크립트를 적용하여 유성음 분절을 추출하고 AVQI를 측정하였다. 또한 음성장애의 중증도를 평가하기 위해서, 5명의 언어재활사가 두 가지 청지각적 평가(GRBAS의 Grade척도, CAPE-V의 overall severity 척도)를 사용하였다. **결과:** 전체 AVQI 측정치는 AVQIv2, AVQIv4-K, AVQIv3 순으로 높았으며, AVQIv3는 나머지 AVQI 측정치들에 비해서 유의미하게 낮았다. 각 버전 모두 정상과 병리적 음성 집단 간에 유의미한 차이를 보였다. AVQI와 청지각적 평가 간에는 .82 이상의 높은 상관관계를 보였고, AVQIv4-K와 상관관계가 가장 높았다. ROC 분석을 통해서 모든 AVQI는 모두 .94 이상의 AUC, 9.5 이상의 LR(+), .2 이하의 LR(-)를 나타내었다. **논의 및 결론:** AVQIv4-K는 기존 AVQI 버전에 상응하는 높은 수준의 진단적 성능을 가지는 것으로 여겨진다.

**핵심어:** 음향학적 음성 질 지수, 음향학적 분석, 음성장애, 청지각적 평가, 프랏

## 참고문헌

- 김근호, 이연우, 이병주, 권순복(2018). 성대용종 환자의 후두미세수술 후 음성회복의 예측인자로서의 음향학적 및 청지각적 평가. *대한이비인후과학회지 두경부외과학*, 61, 361-369.
- 이승진, 임성은, 최홍식(2017). 문단 내 위치에 따른 캡스트럼 및 스펙트럼 측정치 비교. *Communication Sciences & Disorders*, 22, 818-826.
- 최성희(2013). 음성언어재활사의 음성평가와 음성치료의 임상실제: 표준 임상지침과 증거기반증재를 위한 설문조사. *Communication Sciences & Disorders*, 18, 473-485.
- 황민아(2012). 문장따라말하기 검사의 학령 전 단순언어장애 진단 정확도. *언어청각장애연구*, 17, 1-14.