

Effect of Voiced Segment Extraction on Voice of Vocal Cord Paralysis: Focused on Cepstral Analysis

Geun-Hyo Kim¹, Yeon-Yoo Lee¹, Wonjae Cha¹, Byung-Joo Lee¹, and Soon-Bok Kwon² 

¹Department of Otorhinolaryngology-Head and Neck Surgery and Biomedical Research Institute, Pusan National University Hospital, Busan; and

²Department of Humanities, Language and Information, Pusan National University, Busan, Korea

성대마비 음성에 대한 유성음 분절 추출의 영향: 캡스트럼 분석 위주

김근효¹ · 이연우¹ · 차원재¹ · 이병주¹ · 권순복²

부산대학교병원 이비인후과,¹ 부산대학교 언어정보학과²

Received April 30, 2019

Revised June 13, 2019

Accepted June 24, 2019

Address for correspondence

Soon-Bok Kwon, PhD

Department of Humanities,

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

Background and Objectives The purpose of this study was to investigate the effect of unvoiced segments on the cepstral analysis in patients with vocal cord paralysis (VCP).

Subjects and Method A total 302 subjects (173 subjects with VCP and 129 normal voice subjects) participated in this study. The sustained vowel /a/ 2 seconds and one sentence of ‘Sanchaek’ were edited, and analyzed by Praat script. The cepstral analyses were performed using sustained vowel (SV), continuous speech (CS), and extracted continuous speech (EXT) samples. The auditory-perceptual (AP) rating was also completed by three raters.

Results First, there were significant differences in all variables except low-to high spectral ratio (L/H ratio)_EXT between two groups. Second, cepstral peak prominence (CPP), smoothed cepstral peak prominence (CPPS), and L/H ratio showed significant differences in SV, CS, and EXT samples. Third, cepstral measurements were highly correlated with the AP ratings. Finally, the level of discrimination of dysphonia estimated from CPP and CPPS gotten from SV and CS values was more than area under the curve (AUC) of 0.941. AUC of 0.880 or more was also found in EXT.

Conclusion In this study, we confirmed that both CS and EXT are highly predictive of pathologic speeches. Further study will also need to be validated for a more diverse group of voice disorders.

Korean J Otorhinolaryngol-Head Neck Surg 2020;63(3):114-22

Key Words Acoustics · Dysphonia · Vocal cord paralysis.

서론

임상현장에서 음성장애환자의 음질을 평가하기 위해서 음향학적, 공기역학적, 청지각적 평가, 자가 설문지 등을 실시하게 된다.¹⁾ 다양한 평가들을 종합하여 음성장애의 중증도를 파악한다. 이 중 널리 사용되는 방법의 하나인 음향학적 평가는 음성장애 환자의 음질을 간단하게 수치화할 수 있는 방법이다.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

예전부터 Multi-Dimensional Voice Program(MDVP, Model 5105, KayPENTAX, Montvale, NY, USA), Praat(Institute of Phonetic Sciences, University of Amsterdam, Amsterdam, Netherlands) 등이 음질을 평가하는 데 많이 사용되었으며 시간 기반 분석(time-based analysis)을 하게 된다.²⁾ 하지만 매우 심한 음성장애(very severe dysphonia)의 경우에는 분석 신뢰도가 떨어져 이를 보완하기 위해 주파수 기반 분석(frequency-based analysis)인 캡스트럼 분석이 소개되었다.³⁾ 캡스트럼 분석은 Analysis of Dysphonia in Speech and Voice(ADSV, Model 5109, KayPENTAX), SpeechTool(<https://homepages>.

wmich.edu/~hillenbr/), Praat⁴⁾ 등을 통해서 이루어진다. 시간 기반 분석에서는 기본주파수(fundamental frequency, F0), 주파수변동율(jitter), 강도변동율(shimmer), 소음 대 배음비(noise-to-harmonic ratio) 등을, 주파수 기반 분석에서는 캡스트럼 피크 현저성(cepstral peak prominence, CPP), 평활화된 캡스트럼 피크 현저성(smoothed cepstral peak prominence, CPPS), L/H 비율(low-to high spectral ratio, L/H ratio), CPP 기반 주파수(cepstral peak prominence F0, CPP F0) 등을 측정할 수 있다. 이러한 분석들은 모음연장과 연속발화 샘플을 통해서 이루어진다.

최근 모음연장과 연속발화 샘플을 연결하여 분석하는 방법인 acoustic voice quality index가 소개되었는데 이 분석의 주요한 특징은 무성음, 쉼 구간을 제거하여 분석한다는 것이다.⁵⁾ 캡스트럼 분석은 무성음, 쉼 구간에 영향을 많이 받기 때문에 모음연장 샘플보다는 연속발화 샘플에서 CPP의 측정값이 낮게 관찰된다. 기존 연구들에서는 연구자가 임의로 연속발화 내의 무성음, 쉼 구간을 제거하는 것은 주관적이기 때문에 제거하지 말고 포함시켜 분석하도록 권고하였다.⁶⁾ 연속발화 녹음 시 환자들의 상태에 따라 다양한 쉼 길이를 가질 수 있다. 이러한 무성음, 쉼 길이는 상대적으로 원래의 음질보다 더 중증의 수치로 측정된다.

다양한 원인에 의해 성대에 병리적인 문제가 있다면 더 불규칙적, 비주기적, 비선형적 신호를 산출하기 때문에 비선형 동적 분석 혹은 캡스트럼 분석과 같은 방식이 신뢰성 높은 측정을 할 수 있다고 하였다.^{7,8)} 성대마비, 후두암 등과 같은 중증도가 심한 기식음성의 연구에서 캡스트럼 분석에 대한 유용성은 많이 보고되었다.⁹⁻¹²⁾

성대마비 집단은 무성음, 쉼 등을 많이 포함하고 있다. 성대마비 환자의 음성엔 기식음 등이 많이 포함되어 MDVP에서는 정확한 분석이 어렵기 때문에, 캡스트럼 분석을 통해 음성장애의 중증도를 정량화한다.

본 연구에서 캡스트럼 분석을 위해 Praat을 활용하였다. 무료로 배포된 프로그램이기 때문에 임상현장에서 누구나 쉽게 사용해 볼 수 있다. Praat은 다양한 스크립트를 적용하여 많은 수의 음성 샘플을 자동으로 분석할 수 있게 해준다. 또한 결과를 텍스트파일로 출력할 수 있어 결과 분석 및 연구 진행에도 많은 도움을 준다.

따라서 본 연구의 목적은 연속발화 내의 무성음 구간을 제거, 분석한 결과가 음성장애의 감별에 더욱 유용한지 알아보고 이러한 음향학적 측정값과 청지각적 평가와의 상관성을 알아보고자 하였다.

대상 및 방법

대상

본 연구의 대상은 부산대학교병원에서 성대마비로 진단받은 대상자 173명과 대조군으로 129명의 정상 집단을 포함하였다. 연구대상자들에 대한 정보는 Table 1에 요약하여 제시하였다. 본 연구의 음성 샘플은 중재(음성치료, 후두미세수술, 성대 내 주입술 등) 전의 첫 내원 시의 음성 샘플만을 선정하였다. 성대마비 진단은 차트리뷰, 후두내시경, 후두스트로보스코피, 후두근전도를 통해서 최종 확인하였다. 본 연구의 정상 집단은 1차적으로 후두질환 관련 수술병력이 없고 음성장애를 호소하지 않는 연구대상자의 인터뷰를 확인한 후 2차적으로 본 연구의 1저자가 GRBAS 평가를 통해서 grade(G) 척도 0(정상음성)으로 평가된 음성 샘플들로 선정하였다.

본 연구는 획득된 음성 샘플을 후향적으로 분석한 연구로서 부산대학교병원 연구윤리위원회(Institutional Review Board)의 심의를 면제받았다(IRB No. H-1904-011-078).

음성 샘플

음성녹음은 음성검사실 내의 방음부스에서 Computerized Speech Lab(CSL, Model 4500, KayPENTAX)과 단일방향성 마이크인 Shure SM48(Shure Inc., Niles, IL, USA)을 이용하였다. 방음부스 내의 소음은 40 dB 이하로 유지하여 배경소음의 영향을 최대한 배제하였다. 음성 샘플의 저장형태는 표본추출률(sampling rate) 44.1 kHz, 양자화(quantization) 16 bit로 하였다.

발성 과업에서 모음연장은 편안한 음도와 강도로 /아/ 모음을 연장발성, 연속발화는 산책 문단의 25음절을 편안한 속

Table 1. Demographics and patient characteristics

Variables	Normal (n=129)	VCP (n=173)
Age	57.5 ± 10.8	59.3 ± 14.0
Sex		
Male	46 (35.7)	90 (51.9)
Female	83 (64.3)	83 (48.1)
Rule out		
Idiopathic		64 (37.0)
Thoracic surgery		73 (42.2)
Thyroid surgery		24 (13.9)
Neurosurgery		8 (4.6)
Orthopedic surgery		4 (2.3)
Paralyzed side		
Left side		120 (69.4)
Right side		53 (30.6)

Variables are presented as mean ± standard deviation or number (percentage). VCP: vocal cord paralysis

도와 강도로 읽었다. 모음 연장발성은 3회 반복하여 연구자가 판단하기에 안정적인 발성을 포함한 샘플을 선택하여 안정 구간 2초를 편집하였고, 2초 이하인 샘플은 프랏 스크립트를 이용하여 전체 음성파형의 1/4 구간을 시작지점, 3/4 구간을 종료지점으로 설정되도록 구간 편집을 하였다. 연속발화는 선행연구에서 제시한 것과 같이 산책 문단 중 25음절의 문장(예시: 넓게 펼쳐있는 바다를 바라보면 내 마음 역시 넓어지는 것 같다.)을 선택하여 읽도록 지시하였다.^{13,14)} 신뢰성 높은 음성평가를 위한 임상평가지침을 근거로 음향학적 녹음 및 분석, 청지각적 평가를 수행하였다.^{15,16)}

캡스트럼 분석

CSL에서 녹음한 모음연장 샘플은 사운드에디터 창에서 select를 이용하여 정확하게 2초 구간을 편집하고, 연속발화 샘플은 발화의 시작과 끝 부분을 선택하여 불필요한 부분을 편집하였다. 모음연장 샘플은 무성음 구간 제거 스크립트는 적용하지 않고 연속발화 샘플에만 적용하였다. 캡스트럼 분석은 모음연장 샘플(sustained vowel, SV), 연속발화 샘플(continuous speech, CS), 무성음 구간이 제거된 연속발화 샘플(extracted continuous speech, EXT)을 대상으로 실시하였다. 무성음 구간 제거와 캡스트럼 분석의 스크립트는 선행연구에서 제시한 것을 적용하였다.^{17,18)} 유성음 추출 스크립트는 선행연구에서 제시한 것을 본 연구의 목적에 맞게 수정하였다.^{17,18)} 유성음 추출의 기준은 다음과 같다. 1) 사운드 에너지가 전체 신호 에너지의 30%를 초과, 2) 제로 크로싱 속도가 1500 Hz 미만, 3) 정규화된 자기 상관 피크가 0.3보다 큰 경우, 30 ms의 프레임 유성음으로 지정했다. L/H ratio는 4000 Hz 이하의 에너지와 4000~8000 Hz 사이의 에너지 비율을 말하며 너무 큰 값으로 측정되어 자연로그(natural logarithm)를 취하여 측정값을 제시하였다.¹⁹⁾

청지각적 평가

무성음 구간 제거 전의 음성 샘플을 통해서 청지각적 평가를 실시하였다. 음성 평가 및 음성 치료 경력이 7년 이상인 언어재활사 1급 3명이 음성 샘플에 대해서 청지각적 평가를 수행하였다. 평가자들은 소음이 40 dB 이하로 통제된 방에서 대상자에 대한 사전 정보 없이 무작위로 제시된 음성 샘플을 반복 청취하여 평가하였다. 청지각적 평가는 선행연구에서 제시한 방법대로 수행하였다.¹³⁾ 모음연장과 연속발화에 대해서 각각 평가를 수행하였고, 전반적인 음성 중증도 확인을 위해서 GRBAS의 G, Consensus Auditory Perceptual Evaluation-Voice(CAPE-V)의 overall severity(OS)를 평가하였다. 청지각적 평가의 반복 평가로 인한 집중도 저하를 최소화하

기 위해 30개의 음성 샘플 평가마다 짧은 휴식을 취하도록 지시하였다.²⁰⁾ 평가자 내 신뢰도 측정을 위해서 전체 음성 샘플 중 약 10%에 해당하는 60명의 음성 샘플을 초기 평가 후 2주 뒤에 재평가하였다.

자료 처리

캡스트럼 분석에 대한 자료 처리는 R, version 3.5.2(The R Foundation for Statistical Computing, Vienna, Austria)와 RStudio 1.1.463(RStudio Inc., Boston, MA, USA)을 이용하였다. 청지각적 평가에 대한 평가자 간, 내의 신뢰도 측정을 위해서 급내상관계수(intraclass correlation coefficient; 단일측도)를 수행하였다. 그 결과 평가자 간 신뢰도[G: 0.732~0.928(mean: 0.874), OS: 0.721~0.897(mean: 0.838)], 평가자 내 신뢰도[G: 0.746~0.907(mean: 0.865), OS: 0.747~0.853(mean: 0.812)]로 높은 수준이었다.

정상 집단과 성대마비 집단 간의 캡스트럼 측정값 및 청지각적 평가의 차이가 있는지 알아보기 위해 독립표본 t-검정을 실시하였다. 또한 과업별 CPP, CPPS, L/H ratio 측정값의 비교를 위해서 one way analysis of variance를 실시하였다. 캡스트럼 측정값과 청지각적 평가 간의 상관관계를 알아보기 위하여 피어슨 상관분석을 시행하였다. 정상과 성대마비 음성 집단을 선별하기 위한 receiver operating characteristic curve(ROC) 분석을 수행하였다. 본 연구에서는 캡스트럼 분석의 민감도(sensitivity)와 특이도(specificity), 곡선하 영역(area under the curve, AUC), 양성/음성 우도비(likelihood ratio for a positive result; LR+, likelihood ratio for a negative result; LR-)를 산출하였다. 또한 ROC 곡선 쌍대 비교(pairwise comparison)를 사용하여 캡스트럼 변수들의 AUC를 비교하였다.

결 과

정상과 성대마비 집단의 캡스트럼 및 청지각적 평가 결과 비교

정상음성과 성대마비 음성 집단의 섭동, 캡스트럼, 청지각적 평가 결과를 비교하여 Figs. 1, 2와 Table 2에 제시하였다. L/H ratio_EXT를 제외한 모든 변수들에서 두 집단 간에 유의미한 차이가 확인되었다.

과업별 CPP, CPPS, L/H ratio 측정값의 비교

과업에 따른 캡스트럼 측정값을 비교한 결과는 Table 3에 제시하였다. CPP, CPPS, L/H ratio 모두 SV, CS, EXT 과업별에 따른 측정값에 유의미한 차이를 보였다($p < 0.05$).

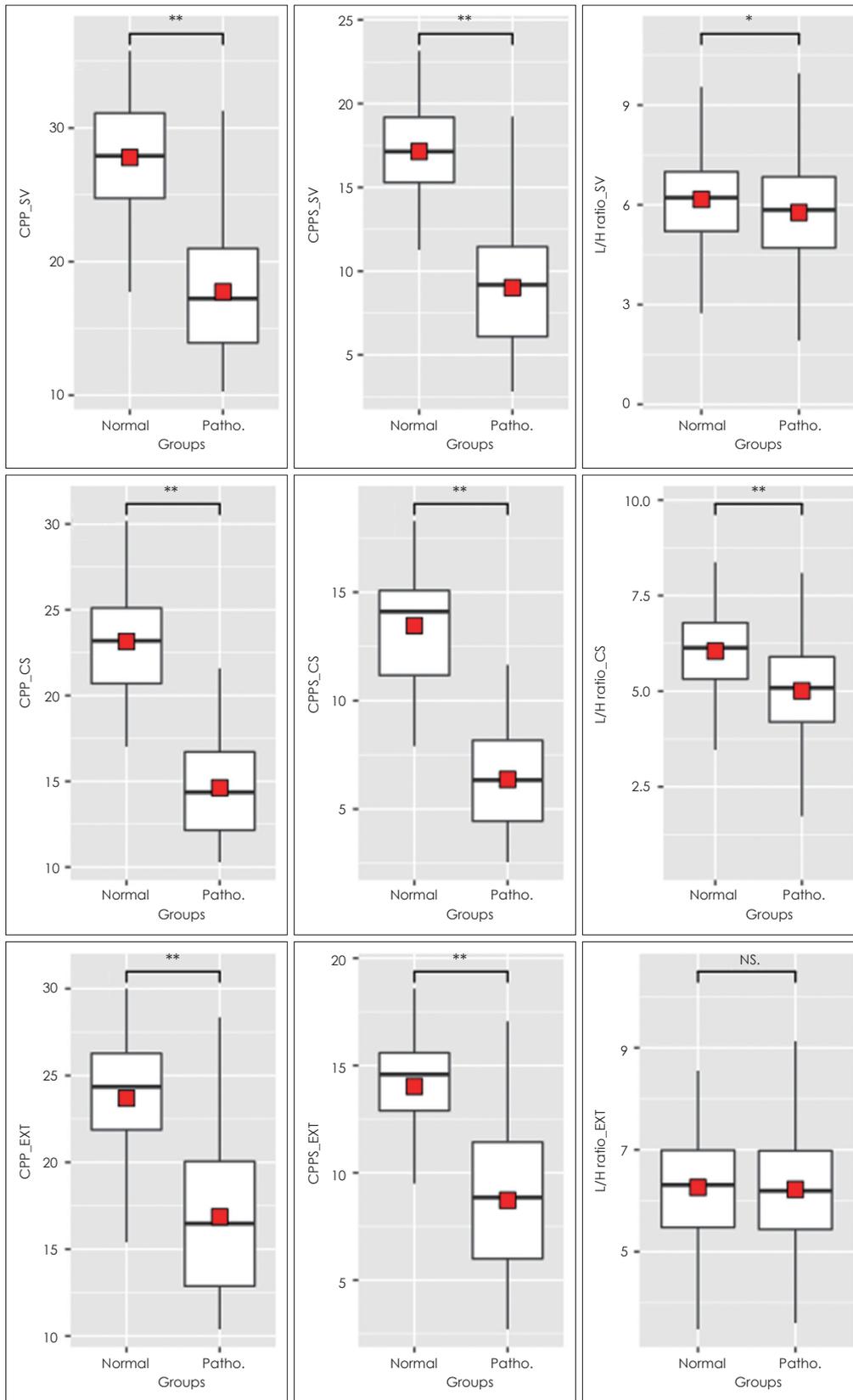


Fig. 1. Comparison of measured variables between normal and vocal cord paralysis group. * $p < 0.05$, ** $p < 0.01$. CPP: cepstral peak prominence, CPPS: smoothed cepstral peak prominence, L/H ratio: low-to high spectral ratio, EXT: extracted continuous speech, SV: sustained vowel, CS: continuous speech, Patho.: pathological voice group, NS: no significance.

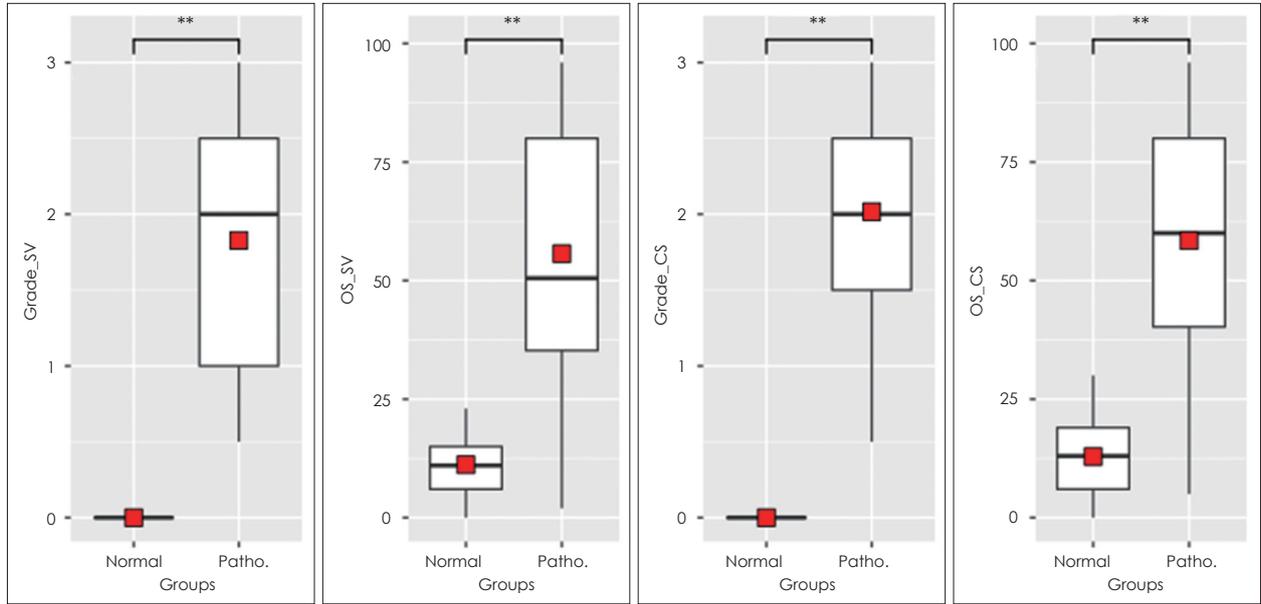


Fig. 2. Comparison of auditory-perceptual ratings between normal and vocal cord paralysis group. ** $p < 0.01$. Grade: degree of grade, OS: overall severity, SV: sustained vowel, CS: continuous speech, Patho.: pathological voice group.

Table 2. Comparison of measured variables between normal and VCP group

Groups	Normal (n=129)	VCP (n=173)	p-value
Age	57.5±10.8	59.3±14.0	0.065
Jitter	1.1±0.4	4.4±1.7	0.01**
Shimmer	4.1±1.6	9.0±5.8	0.01**
NHR	0.1±0.0	0.2±0.1	0.01**
CPP_SV	27.8±4.1	17.8±4.6	0.01**
CPPS_SV	17.1±2.7	9.1±3.7	0.01**
L/H ratio_SV	6.2±1.3	5.9±1.5	0.05*
CPP_CS	23.1±2.9	14.6±2.6	0.01**
CPPS_CS	13.5±2.3	6.4±2.2	0.01**
L/H ratio_CS	6.0±1.1	5.1±1.3	0.01**
CPP_EXT	23.7±3.6	16.9±4.2	0.01**
CPPS_EXT	14.0±2.5	8.7±3.4	0.01**
L/H ratio_EXT	6.3±1.2	6.1±1.1	0.736
Grade_SV	0.0±0.0	1.9±0.7	0.01**
OS_SV	11.2±5.8	55.7±23.6	0.01**
Grade_CS	0.0±0.0	1.9±0.8	0.01**
OS_CS	12.9±6.8	58.3±21.6	0.01**

Variables are presented as mean±standard deviation. * $p < 0.05$, ** $p < 0.01$. VCP: vocal cord paralysis, NHR: noise to harmonics, CPP: cepstral peak prominence, SV: sustained vowel, CPPS: smoothed cepstral peak prominence, L/H ratio: low-to high spectral ratio, CS: continuous speech, EXT: extracted continuous speech, Grade: degree of grade, OS: overall severity

CPP, CPPS는 SV에서 가장 큰 측정값으로 나타났으며 CS는 유성음 구간을 추출한 후(EXT)에 측정값이 증가하였다(SV>EXT>CS). L/H ratio는 SV가 CS보다 큰 값으로 나

타났지만 EXT의 측정값이 SV보다도 크게 나타났다(EXT>SV>CS).

캡스트럼 측정값과 청지각적 평가(G, OS) 간의 상관관계

캡스트럼 측정값과 청지각적 평가 간의 상관관계를 분석한 결과, 캡스트럼 측정값은 연장모음과 연속발화 모두 G, OS와 유의미한 상관관계($p < 0.01$)를 보였으며 그 결과를 Table 4에 제시하였다. CPP, CPPS는 -0.88 이상의 상관관계를 보였으며 CS에서 조금 더 높은 상관성을 나타내었다. L/H ratio는 청지각적 평가와 -0.54 이하의 낮은 상관관계를 보였다.

정상과 성대마비 음성 집단을 감별하기 위한 ROC 분석

음성장애를 선별하기 위한 ROC 분석을 실시한 결과는 Fig. 3, Table 5에 제시하였다. SV, CS에서 CPP, CPPS 측정값을 통해서 음성장애를 감별할 수 있는 수준은 AUC 0.940 이상으로 나타났다. 유성음 구간만을 추출한 EXT에서도 AUC 0.879 이상의 수준을 나타내었다. 반면에 L/H ratio는 모든 발성 과업에서 상대적으로 낮은 AUC가 확인되었다. SV, CS 과업에서, CPP, CPPS의 LR+는 12 이상, LR-는 0.2 이하로 높은 감별 예측도를 보였고, EXT에서는 상대적으로 낮은 예측도를 나타냈다(LR+: 4 이상, LR-: 0.22 이하). 발성 과업에서는 CS가 SV보다 우세하고 음향학적 변수에서는 CPPS가 CPP보다 뛰어난 감별 능력을 가진 것으로 나타났다. AUC를 쌍대 비교(pairwise comparison of ROC curves)를 한 결과, SV에서 CPP, CPPS, L/H ratio 간에는 모두 유의미

Table 3. Comparison of cepstral values according to task type (Bonferroni post hoc test)

	(I) Task	(J) Task	Mean difference (I-J)	Std. error	Sig.	95% Confidence interval	
						Lower bound	Upper bound
CPP	SV	CS	3.533802*	0.308443	0.000	2.79468	4.27292
		EXT	1.667649*	0.308443	0.000	0.92853	2.40677
	CS	SV	-3.533802*	0.308443	0.000	-4.27292	-2.79468
		EXT	-1.866153*	0.307778	0.000	-2.60368	-1.12863
	EXT	SV	-1.667649*	0.308443	0.000	-2.40677	-0.92853
		CS	1.866153*	0.307778	0.000	1.12863	2.60368
CPPS	SV	CS	2.945379*	0.246536	0.000	2.35461	3.53615
		EXT	0.987023*	0.246536	0.000	0.39625	1.57780
	CS	SV	-2.945379*	0.246536	0.000	-3.53615	-2.35461
		EXT	-1.958356*	0.246005	0.000	-2.54786	-1.36886
	EXT	SV	-0.987023*	0.246536	0.000	-1.57780	-0.39625
		CS	1.958356*	0.246005	0.000	1.36886	2.54786
L/H ratio	SV	CS	0.645215*	0.078568	0.000	0.45694	0.83349
		EXT	-0.348041*	0.078568	0.000	-0.53631	-0.15977
	CS	SV	-0.645215*	0.078568	0.000	-0.83349	-0.45694
		EXT	-0.993256*	0.078399	0.000	-1.18112	-0.80539
	EXT	SV	0.348041*	0.078568	0.000	0.15977	0.53631
		CS	0.993256*	0.078399	0.000	0.80539	1.18112

* $p < 0.05$. SV: sustained vowel, CS: continuous speech, EXT: extracted continuous speech, CPP: cepstral peak prominence, CPPS: smoothed cepstral peak prominence, L/H ratio: low-to high spectral ratio, Std. error: standard deviation error, Sig.: significance

Table 4. Correlation analyses between acoustic variables and auditory-perceptual ratings

	Grade_SV	OS_SV	Grade_CS	OS_CS
CPP_SV	-0.90*	-0.88*		
CPPS_SV	-0.92*	-0.90*		
L/H ratio_SV	-0.40*	-0.41*		
CPP_CS			-0.93*	-0.91*
CPPS_CS			-0.93*	-0.92*
L/H ratio_CS			-0.53*	-0.54*

* $p < 0.01$. Grade: degree of grade, OS: overall severity, CS: continuous speech, CPP: cepstral peak prominence, SV: sustained vowel, CPPS: smoothed cepstral peak prominence, L/H ratio: low-to high spectral ratio

한 차이를 보였고, CS, EXT에서 CPP와 CPPS 간에는 유의미한 차이가 없었고 L/H ratio와 유의미한 차이를 보였다.

고찰

본 연구에서는 연속발화의 유성음 구간을 추출, 켈스트럼 분석을 실시하여 병리적 음성의 감별에 어떤 의미가 있는지 확인하고자 하였다. 정상과 성대마비 음성을 정량화하여 비교하고 청지각적 평가와의 상관관계도 보았다. 그리고 두 집단을 구분할 수 있는 최적의 절사값과 진단 예측력을 확인하였다. 본 연구의 결과에 대한 논의는 다음과 같다.

첫째, 정상과 성대마비 음성의 발화 과업 및 유성음 분절 추출에 따른 켈스트럼 측정값에 차이가 있는가? 집단, 발화 과업 및 무성음 분절 제거에 따른 켈스트럼 측정값을 제시하였다. 먼저 정상과 성대마비 집단의 켈스트럼 측정치는 L/H ratio_EXT를 제외한 모든 음향학적, 청지각적 변수들에서 통계적으로 유의미한 차이가 있었다. 정상 집단은 성대마비 집단에 비하여 높은 CPP, CPPS, L/H ratio 측정치를 나타내었다. CPP, CPPS는 음성 신호 내의 배음(harmonics)의 정도를 측정하고 L/H ratio는 4000 Hz 이하의 주파수 스펙트럼 에너지와 4000 Hz 이상의 주파수 스펙트럼 에너지의 비율을 의미하고 병리적 음성의 경우 정상 음성에 비해 낮은 비율을 나타내는 것이다. 기식성이 많고 거친 음성의 특징을 가진 성대마비 음성에 대한 켈스트럼 분석을 통해서 확인할 수 있었다. 기존 연구들에서 병리적 음성은 음성 신호 내의 감소된 배음으로 인하여 낮은 켈스트럼 측정치가 보고되어 왔다. 본 연구에서 수행된 성대마비 음성에 대한 켈스트럼 분석도 기존 연구 보고들을 뒷받침하는 결과를 도출하였다. 이러한 결과는 선행연구에서도 많이 보고되었다.²¹⁻²³⁾

세 가지 발성 과업에 따른 비교에서는 CPP, CPPS는 SV 과업에서 가장 크게 측정되었고 L/H ratio는 EXT에서 가장 크게 측정되었다. 또한 모든 변수들은 유성음 추출 후 측정값(EXT)이 추출 전(CS)보다 증가하였다. 이러한 결과는 선행연

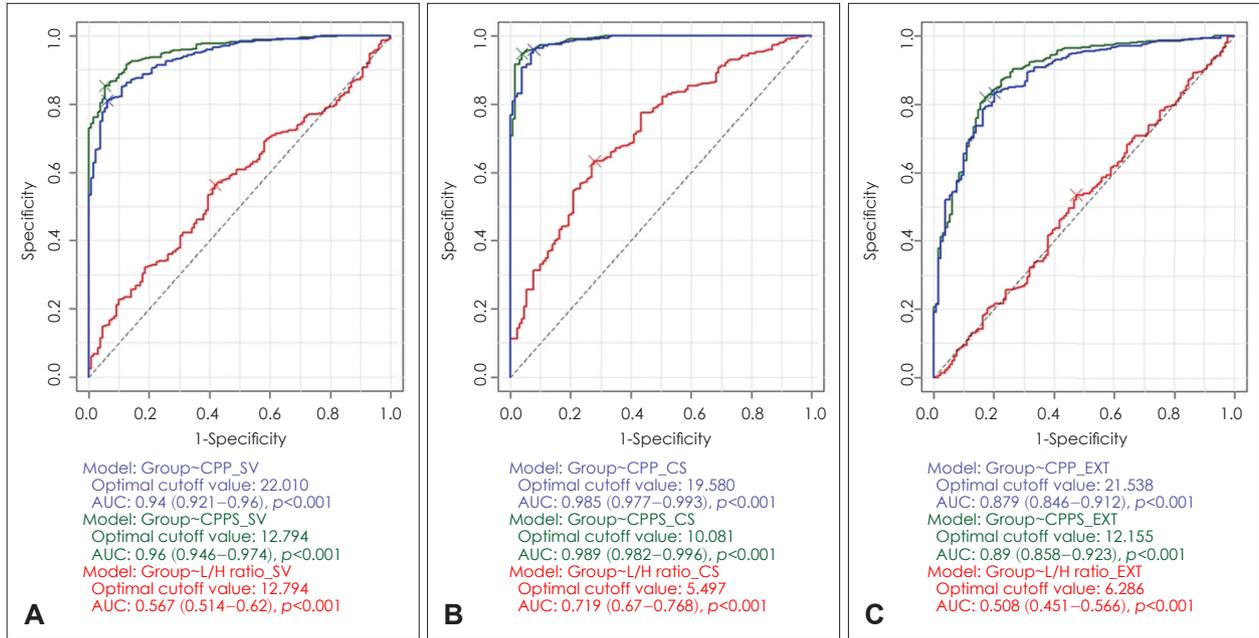


Fig. 3. Receiver operating characteristic curve analysis to discriminate normal and vocal cord paralysis group. SV (A), CS (B), EXT (C). CPP: cepstral peak prominence, CPPS: smoothed cepstral peak prominence, L/H ratio: low-to high spectral ratio, SV: sustained vowel, CS: continuous speech, EXT: extracted continuous speech, AUC: area under the curve.

Table 5. Descriptive statistics of receiver operating characteristic curve according to acoustic parameters and task

Variables	Sensitivity	Specificity	Cutoff	AUC	95% CI	LR+	LR-
CPP_SV	0.811	0.938	21.98	0.940	0.918–0.958	13.07	0.20
CPPS_SV	0.853	0.946	12.73	0.960	0.941–0.974	15.72	0.16
L/H ratio_SV	0.563	0.581	6.06	0.567	0.526–0.608	1.35	0.75
CPP_CS	0.960	0.922	19.55	0.985	0.971–0.993	12.38	0.04
CPPS_CS	0.949	0.961	10.07	0.989	0.976–0.996	24.48	0.05
L/H ratio_CS	0.633	0.721	5.48	0.719	0.680–0.755	2.27	0.51
CPP_EXT	0.833	0.798	21.50	0.879	0.849–0.904	4.13	0.21
CPPS_EXT	0.820	0.829	12.14	0.890	0.862–0.914	4.81	0.22
L/H ratio_EXT	0.535	0.527	6.28	0.510	0.469–0.552	1.13	0.88

AUC: area under the curve, CI: confidence interval, LR+: likelihood ratio for a positive result, LR-: likelihood ratio for a negative result, CPP: cepstral peak prominence, CPPS: smoothed cepstral peak prominence, L/H ratio: low-to high spectral ratio, SV: sustained vowel, CS: continuous speech, EXT: extracted continuous speech

구들의 결과와 일치하였다. 기존의 연구들에서는 CS 샘플로부터 무성음 분절을 제거(혹은 유성음 분절만 추출)하기 위해서 수작업 기법을 사용하였다. 또한 구간 추출에 대한 객관성 확보를 위해서 ADSV 소프트웨어를 통해서 자동화된 유성음 탐지 알고리즘을 적용하였다.^{24,25)} 선행연구에서 수작업 혹은 자동화된 방식에 상관없이 무성음 분절을 제거한 후 분석하였을 때 더 높은 CPP값으로 계산된다고 보고하였다. 이러한 무성음 분절 제거가 CPP, CPPS의 절대값에는 영향을 미치지 않지만, 정상과 병리적 음성 집단 사이의 차이는 비슷한 경향을 보인다고 보고하였다.²⁵⁾

쉽, 무성음 구간을 포함한 CS 샘플은 유성음 분절만을 포

함한 EXT 샘플보다 낮은 캡스트럼 측정치를 산출할 것으로 예측할 수 있었고 본 연구의 결과를 통해서 확인하였다. 무성음 분절이 상대적으로 낮은 캡스트럼 측정치에 영향을 미쳤다고 볼 수 있다.^{8,26,27)} 본 연구에서는 산책 문장에 한정하여 분석하였지만 최근에는 서로 다른 음소로 구성된 표준 문장들도 보고되었다.²⁸⁾ 이러한 문장들은 미국에서 개발된 CAPE-V에 대응하는 한국판 문장이라고 할 수 있다. 추후 연구에서는 음소 구성에 따른 캡스트럼 측정치도 확인하여 다양한 측면에서 음성을 분석하는 것도 중요할 것으로 생각된다.

둘째, 캡스트럼 측정값과 청지각적 평가(G, OS)와의 상관관계는 어떠한가? 음향학적 변수와 청지각적 평가 변수 간의

상관관계는 SV와 CS로 나누어 확인하였다. SV에서 CPP, CPPS는 G, OS와 -0.88 이상, CS에서도 G, OS와 -0.91 이상의 높은 상관관계를 보였다. 하지만 L/H ratio에서는 SV, CS 모두 -0.53 이하의 낮은 상관관계를 보였다. CPP, CPPS와 청지각적 평가(G, OS)와의 높은 상관관계는 선행연구들의 결과들을 지지하였다. L/H ratio는 다른 캡스트럼 측정 변수에 비해 낮은 상관을 보였고 기존 ADSV의 알고리즘과 차이가 있는 것으로 여겨진다. 또한 본 연구에서 측정된 L/H ratio의 값이 너무 큰 범위의 값으로 측정되어 로그함수를 취한 영향도 있는 것으로 생각된다. 후속 연구에서는 ADSV의 분석 알고리즘과 비교하여 유사한 값을 도출할 수 있는 스크립트를 개발해야 할 것이다.

셋째, 유성음 추출 전후의 음성 샘플들을 분석하여 정상과 병리적 음성 집단을 구분할 수 있는 최적 절사값 및 진단 예측력은 어떠한 변화가 있는가? 유성음 추출 전후의 음성 샘플을 이용하여 병리적 음성의 감별에 대해 확인하였다. 유성음 추출 전의 음성 샘플에서는 최적 절사값은 CPP 19.58, CPPS 10.08, L/H ratio 5.50이었고, 추출 후에는 최적 절사값이 모두 증가하였고 선행연구의 결과와 비슷하였다. 본 연구에서 CS를 통해서 병리적 음성을 감별할 수 있는 진단 예측력(AUC)은 0.985, 10 이상의 LR+, 0.04 이하의 LR-로 매우 높은 수준이었다. 또한 EXT에서의 진단 예측력(AUC)은 0.880, 4 이상의 LR+, 0.21 이하의 LR-로 높은 수준을 나타내었다. AUC, LR+, LR-의 결과를 확인하여 CS의 분석이 더 유용한 것으로 확인되었지만 EXT도 여전히 높은 진단 예측력을 유지한다는 것을 확인하였다.

보통 새로운 검사 방법을 평가하는 척도로 민감도, 특이도, 우도비, 예측도(predictivity) 등을 사용한다. 대상 집단에 대해 민감도와 특이도를 통해 선별검사(screening test)와 심층검사(deep test)를 수행하게 된다.²⁹⁾ LR+는 민감도값을 1에서 특이도를 뺀 값으로 나누며, LR-는 1에서 민감도를 뺀 값을 특이도값으로 나누는 것이다.³⁰⁾ LR+는 1 이상, LR-는 0~1 사이의 범위로 측정되며, LR+의 값이 증가할수록, LR-의 값이 감소할수록 검사의 예측도가 증가하는 것으로 보고되었다.^{31,32)}

본 연구에서 측정된 CS, EXT의 캡스트럼 측정치들은 성대마비 환자들의 병리적인 특성에 대한 강력한 예측인자로 확인되었다. 성대마비 음성의 중증도는 매우 다양하게 산출이 되기 때문에 유성음 추출 후 진단 예측력이 감소한 것으로 생각된다. 이러한 결과가 음성 신호 내에 기성음이 많이 포함되어 있는 성대마비 음성만의 특성인지 후속 연구를 통해서 확인해야 할 것이다. 또한 유성음 추출 후 감별진단에 대한 성능을 향상시키기 위해서는 선행연구에서 제시한 유성음 추출 스크립트를 수정해야 할 것으로 생각된다. 무성음

을 필터링하는 에너지의 초과값, 제로 크로싱 속도, 프레임의 길이(ms)를 조정하여 더 높은 진단 예측력을 획득할 수 있는 옵션값을 찾아야 할 것이다.

종합해보면, 본 연구의 결과들은 Praat 스크립트를 통해서 연속발화 내의 무성음 구간을 분리하여 유성음 분절만을 캡스트럼 분석하여 음성장애의 중증도를 정량화하고 이러한 무성음 추출 방식이 병리적 음성의 감별에 도움이 되는지 확인하였다는 점에서 의의가 있다. 유성음 추출 전후의 음성 샘플, CS, EXT 모두 병리적 음성을 감별하는 데 높은 예측력이 확인되었다. 후속연구에서는 임상 현장에서 접할 수 있는 다양한 음성장애군들을 포함하여 이러한 유성음 추출 방식이 유효한지를 검증해야 할 것이다. 이러한 부분이 검증된다면 대상자들의 발화 속도, 억양, 쉼, 호흡특성에 상관없이 음성장애의 중증도를 정량화하고 비교할 수 있는 객관적인 정보를 제공할 수 있을 것으로 생각된다.

Acknowledgments

This work was supported by a 2-Year Research Grant of Pusan National University.

Author Contribution

Conceptualization: Geun-Hyo Kim. Data curation: Yeon-Woo Lee. Formal analysis: Geun-Hyo Kim, Yeon-Woo Lee. Methodology: Wonjae Cha, Byung-Joo Lee. Project administration: Soon-Bok Kwon. Resources: Soon-Bok Kwon. Supervision: Soon-Bok Kwon. Validation: Byung-Joo Lee. Visualization: Geun-Hyo Kim. Writing—original draft: Geun-Hyo Kim. Writing—review & editing: Wonjae Cha, Byung-Joo Lee.

ORCID

Soon-Bok Kwon <https://orcid.org/0000-0002-9424-0077>

REFERENCES

- 1) Ravi SK, Shabnam S, George KS, Saraswathi T. Acoustic and aerodynamic characteristics of choral singers. *J Voice* 2019;33(5): 803.e1-803.e5.
- 2) Dejonckere PH, Remacle M, Fresnel-Elbaz E, Woisard V, Crevier-Buchman L, Millet B. Differentiated perceptual evaluation of pathological voice quality: Reliability and correlations with acoustic measurements. *Rev Laryngol Otol Rhinol (Bord)* 1996;117(3):219-24.
- 3) Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J Speech Hear Res* 1996;39(2):311-21.
- 4) Boersma P. Praat, a system for doing phonetics by computer. *Glott international* 2002;5.
- 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):161-74.
- 6) Awan SN, Roy N, Dromey C. Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. *Clin Linguist Phon* 2009;23(11):825-41.
- 7) Choi SH, Choi CH. The Utility of Perturbation, non-linear dynamic, and cepstrum measures of dysphonia according to signal typing. *Phonetics and Speech Sciences* 2014;6(3):63-72.

- 8) Heman-Ackah YD, Michael DD, Baroody MM, Ostrowski R, Hillenbrand J, Heuer RJ, et al. Cepstral peak prominence: A more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol* 2003; 112(4):324-33.
- 9) Awan SN, Roy N. Outcomes measurement in voice disorders: Application of an acoustic index of dysphonia severity. *J Speech Lang Hear Res* 2009;52(2):482-99.
- 10) Awan SN, Roy N. Toward the development of an objective index of dysphonia severity: A four-factor acoustic model. *Clin Linguist Phon* 2006;20(1):35-49.
- 11) Kim GH, Lee YW, Bae IH, Park HJ, Lee JS, Wang SG, et al. A cepstral analysis of voices with glottic cancer and laryngeal leukoplakia: Sustained vowels and continuous speech. *J Speech-Lang, Hear, Disord* 2016;25(3):135-45.
- 12) Kang Y-A, Seong C-J. A cepstral analysis of breathy voice with vocal fold paralysis. *Phonetics and Speech Sciences* 2012;4(2):89-94.
- 13) Kim GH, Lee YW, Bae IH, Park HJ, Wang SG, Kwon SB. Validation of the Acoustic Voice quality Index in the Korean language. *J Voice* 2019;33(6):948.e1-9.
- 14) Maryn Y, Kim HT, Kim J. Auditory-perceptual and acoustic methods in measuring dysphonia severity of Korean speech. *J Voice* 2016;30(5):587-94.
- 15) Choi SH. Speech-language pathologists' voice assessment and voice therapy practices: A survey for standard clinical guideline and evidence-based practice. *Commun Sci Disord* 2013;18(4):473-85.
- 16) Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyiski D, Eadie T, et al. Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *Am J Speech Lang Pathol* 2018;27(3):887-905.
- 17) 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):540-55.
- 18) Reynolds V, Buckland A, Bailey J, Lipscombe J, Nathan E, Vijayasekaran S, et al. Objective assessment of pediatric voice disorders with the acoustic voice quality index. *J Voice* 2012;26(5):672.e1-7.
- 19) Park J, Seong C. The implementation of children's automated formant setting by Praat scripting. *Phonetics and Speech Sciences* 2018;10(4): 1-10.
- 20) Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *J Speech Hear Res* 1993;36(1):21-40.
- 21) Hartl DM, Hans S, Vaissière J, Riquet M, Brasnu DF. Objective voice quality analysis before and after onset of unilateral vocal fold paralysis. *J Voice* 2001;15(3):351-61.
- 22) Balasubramaniam RK, Bhat JS, Fahim III S, Raju III R. Cepstral analysis of voice in unilateral adductor vocal fold palsy. *J Voice* 2011;25(3):326-9.
- 23) Watts CR, Awan SN. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *J Speech Lang Hear Res* 2011;54(6):1525-37.
- 24) Lowell SY, Colton RH, Kelley RT, Hahn YC. Spectral- and cepstral-based measures during continuous speech: Capacity to distinguish dysphonia and consistency within a speaker. *J Voice* 2011;25(5): e223-32.
- 25) Lowell SY, Colton RH, Kelley RT, Mizia SA. Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech. *J Voice*. 2013;27(4):393-400.
- 26) Awan SN, Giovinco A, Owens J. Effects of vocal intensity and vowel type on cepstral analysis of voice. *J Voice* 2012;26(5):670.e15-20.
- 27) Awan SN, Helou LB, Stojadinovic A, Solomon NP. Tracking voice change after thyroidectomy: Application of spectral/cepstral analyses. *Clin Linguist Phon* 2011;25(4):302-20.
- 28) Choi SH. Development of Korean standardized sentences on voice quality evaluation for dysphonia. *Audiol Speech Res* 2018;14(2):128-42.
- 29) Linn S. New patient-oriented diagnostic test characteristics analogous to the likelihood ratios conveyed information on trustworthiness. *J Clin Epidemiol* 2005;58(5):450-7.
- 30) Ransohoff DF. Challenges and opportunities in evaluating diagnostic tests. *J Clin Epidemiol* 2002;55(12):1178-82.
- 31) Deeks JJ, Altman DG. Diagnostic tests 4: Likelihood ratios. *BMJ* 2004;329(7458):168-9.
- 32) Dollaghan CA. Evidence-based practice in communication disorders: What do we know, and when do we know it? *J Commun Disord* 2004;37(5):391-400.