



## Classification of muscle tension dysphonia (MTD) female speech and normal speech using cepstrum variables and random forest algorithm\*

Joowon Yun<sup>1</sup> · Heejeong Shim<sup>2</sup> · Cheoljae Seong<sup>1\*\*</sup>

<sup>1</sup>Department of Speech & Language Pathology, Chungnam National University, Daejeon, Korea

<sup>2</sup>Division of Speech Pathology & Audiology, Hallym University, Gangwon, Korea

### Abstract

This study investigated the acoustic characteristics of sustained vowel /a/ and sentence utterance produced by patients with muscle tension dysphonia (MTD) using cepstrum-based acoustic variables. 36 women diagnosed with MTD and the same number of women with normal voice participated in the study and the data were recorded and measured by ADSV™. The results demonstrated that cepstral peak prominence (CPP) and CPP\_F0 among all of the variables were statistically significantly lower than those of control group. When it comes to the GRBAS scale, overall severity (G) was most prominent, and roughness (R), breathiness (B), and strain (S) indices followed in order in the voice quality of MTD patients. As these characteristics increased, a statistically significant negative correlation was observed in CPP. We tried to classify MTD and control group using CPP and CPP\_F0 variables. As a result of statistic modeling with a Random Forest machine learning algorithm, much higher classification accuracy (100% in training data and 83.3% in test data) was found in the sentence reading task, with CPP being proved to be playing a more crucial role in both vowel and sentence reading tasks.

**Keywords:** muscle tension dysphonia (MTD), cepstral peak prominence (CPP), CPP\_F0, sentence reading task, Random Forest, machine learning

### 1. 서론

음성의 병리적 문제는 객관적 또는 주관적으로 평가될 수 있으며, 언어치료사들은 이를 확인하기 위하여 자가평가, 지각적 평가, 기기적 평가를 시행한다. 특히 기기적 평가는 대상자 음성을 정량화하여 수치로 나타내어주며, 다양한 분석방법에 따

른 음향학적 측정치들에 의해 음성장애 환자들과 정상 음성을 감별하는데 효율적인 방법으로 꼽힌다.

장애음성의 분석에 켈스트럼(cepstrum) 및 스펙트럼 분석의 임상적 유용성이 지속적으로 보고되고 있다. 스펙트럼의 로그 변형인 켈스트럼은 성문음(source)과 성도에서의 공명성분을 큐프런시(quefreny) 축에서 각각 따로 분리하여 출력한다. 큐프

\* This work was supported by 2020 research fund of Chungnam National University.

\*\* cjsong49@gmail.com, Corresponding author

Received 1 August 2020; Revised 25 September 2020; Accepted 25 September 2020

© Copyright 2020 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://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.

런시 축의 낮은 영역은 성도 공명 성분이 응축되어 있다(Kim & Seong, 2017). Hillenbrand et al.(1994)은 캡스트럼 관련 측정치가 기식음성(breathy voice)의 특성을 잘 반영하며 이와 높은 상관성이 있다고 보고하면서 캡스트럼의 임상적 적용 가능성을 제시하였다.

Cepstral peak prominence (CPP)는 캡스트럼 에너지의 선형회귀선과 캡스트럴 정점값 사이의 편차를 나타내는 값이다. 조화음구조가 잘 형성되는 정상화자들에게서 높은 값을 보이는 반면, 기식성(breathiness)이나 긴장성(strain) 등의 음성 특성을 보이는 환자들에게서는 두드러지게 낮은 값을 보이기 때문에 음성장애의 감별에 유용한 분석 및 측정치라고 보고되고 있다(Awan & Roy, 2006; Heman-Ackah et al., 2002; Heman-Ackah et al., 2003; Kim et al., 2017; Kumar et al., 2010; Seo & Seong, 2013; Shin, et al., 2018; Yu et al., 2018). 특히, 모음뿐 아니라 음도변화를 포함하는 연결발화도 분석이 가능하여 비주기적인 음성 신호도 분석이 가능한 장점이 있다(Choi & Choi, 2016; Heman-Ackah et al., 2003).

캡스트럼 및 스펙트럼 분석으로 도출된 값들은 청지각적 음질평가 결과와도 높은 상관성을 보이며(Kumar et al., 2010), 음성 중증도(severity)가 높은 음성 및 마비말장애의 음성을 타당하게 분석할 수 있다(Seo & Seong, 2013). 또한 파킨슨병환자의 Lee Silverman Voice Treatment(LSVT®LOUD) 중재 후 음성의 변화 정도를 비교하거나(Alharbi et al., 2019), 음성장애 환자의 음성치료 중재 과정을 모니터링하고 중재효과를 비교하였으며(Peterson et al., 2013), 음성장애 진단에서 높은 민감도(sensitivity)와 특이도(specificity)를 보여(Watts & Awan, 2011), 캡스트럼 측정치가 갖는 임상적 가치가 주목되고 있다.

캡스트럼 및 스펙트럼 분석의 유용성은 환자의 음성문제가 두드러지게 나타나는 연결발화(connected speech) 과제에서 두각을 나타낸다(Hillenbrand & Houde, 1996; Lowell et al., 2012; Shim et al., 2016; Watts & Awan, 2011). 임상현장에서 ‘모음발성’은 객관적 음성 평가의 대표 과제로 사용되고 있다. 그러나 모음 발성 시에는 화자가 순간적으로 힘을 주어 일시적으로 성대 내전을 높이는 양상을 보이므로 환자의 음성문제가 과대평가될 수 있다(Pyo & Shim, 2007). 따라서 환자의 평소 음성 상태를 반영하기 위해 문장 과제에 대한 객관적 음성 평가가 필수적으로 요구된다. 후두 근육의 과긴장에 의해 음성 문제가 나타나는 기능적 음성장애는 모음뿐만 아니라 보편적 연결발화인 문장읽기에서의 음성 문제가 두드러지므로 문장에 대한 객관적 음성 평가도 이루어져야 한다.

기능적 음성장애의 대표 사례인 근긴장성 발성장애(muscle tension dysphonia, MTD)는 음성 오용과 남용 원인으로 후두와 후두 주변 근육의 과도한 긴장에 의해 나타나는 음성장애이다(Rubin et al., 2006). MTD 환자들이 보이는 후두 긴장은 성대진동(vibration)에 제한을 주어 애성(hoarseness), 쥐어짜고 긴장된 목소리(strained-strangled), 음성 떨림(tremor) 등의 병리적인 음성 문제로 발현된다. MTD 환자들은 특히 발성 시 음성노력(vocal effort)이 두드러지고, 이는 성대의 압축 정도와 관련성이

높기 때문에 산출되는 음성 특성에 차이를 가져온다.

Lowell et al.(2012)은 긴장된 음성 특성을 보이는 음성장애 환자의 음성을 감별하는데 캡스트럼 분석의 유용성을 보고한바 있다. 지금까지 음성장애 환자의 캡스트럼과 스펙트럼 분석 연구는 주로 기식성이 두드러지는 성대 마비나 성대폴립 및 결절, 노인성 후두 등의 환자를 대상으로 한 연구들이 주를 이루었다. 그러나 캡스트럼은 조조성(roughness), 애성 등의 청지각적 평가 결과와도 높은 상관성을 보이기 때문에(Shim et al., 2016), 후두근육의 과긴장으로 인해 성대 진동에 제한이 있는 MTD 환자의 음성 특성과 음성 중증도를 객관적으로 예측할 수 있게 한다.

따라서 본 연구에서는 캡스트럼 변수를 이용하여 MTD 환자 음성의 음향음성학적 특성을 살펴보고, 이것이 청지각적 평가와 어떠한 상관성이 있으며, 캡스트럼 측정치가 머신러닝 분류 알고리즘인 랜덤포레스트(Random Forest) 모델에서 장애음성을 얼마나 잘 감별 진단할 수 있는지 확인하고자 한다.

## 2. 연구방법

### 2.1. 연구 대상

2015년 8월부터 2015년 11월까지 서울 역삼동에 소재한 ○○이비인후과병원에 내원하여 이비인후과 전문의와 1급 언어재활사에게 MTD로 진단받은 음성장애환자 36명(F=36)을 대상으로 하였으며, 환자와 연령을 일치시킨 정상성인 36명(F=36)을 합쳐 총 72명을 군집하였다.

MTD 실험군은 성대의 긴장 유형에 따라(Koufman & Blalock, 1988) 1형 41.67%(15명), 2형 13.89%(5명), 3형 16.67%(6명), 1+3형 27.77%(10명)로 구성되었다. MTD 음성의 음향분석 결과, 하위유형에 따른 음향학적 특성에 통계학적으로 유의한 차이가 없어서 36명 모두 본 연구 대상으로 포함하였다(Park, 2011).

연구 대상은 호흡, 시각, 청각 및 언어 문제가 없으며, 후두의 신경학적, 구조적 이상이 없는 환자들로 전문음성사용자는 대상에서 제외하였다. 본 연구는 이비인후과의 검토와 승인을 받았으며, 모든 환자의 동의하에 진행되었다. 연구 대상의 세부적인 정보는 표 1에 제시하였다.

표 1. 연구 대상  
Table 1. Subjects information

	MTD (n=36)	Control (n=36)	t-value	Degree of freedom
Age±SD (range)	33.92±6.32 (21-48)	31.39±4.85 (22-39)	1.904	70

MTD, muscle tension dysphonia; SD, standard deviation.

### 2.2. 녹음 및 검사 절차

마이크는 SHURE사의 지향성 마이크 SM48 사용하였으며, 녹음 프로그램은 Computerized Speech Lab(CSL, KayPENTAX, Model 4150B)의 MDVP를 사용하였으며 44,100 Hz 표본추출률, 16 bit 양자화 조건에서 녹음하였다. MTD 환자의 음성 녹음은 40 dB 이하의 소음 상황인 병원 내 음성검사실에서 실시하였으

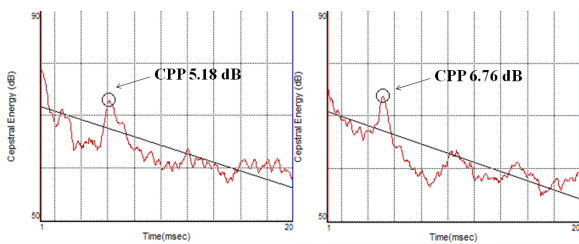
며 환자 입과 마이크가 10 cm 떨어진 거리에서 90도 각도로 고정하여 사용하였다.

환자는 검사 초반에 /아/ 모음 연장발성과제를 수행하였다. 과제 수행 시 평상시 음도와 강도로 최대한 편안하고 자연스러운 상태에서 1회 3초 이상 발성하도록 유도하였다. 검사 후반에는 ‘산책’의 첫 문장(높은 산에 올라가 맑은 공기를 마시며 소리를 지르면 가슴이 활짝 열리는 듯하다)을 정서를 배제한 중립적 상태로 1회 읽도록 하였다. 검사 문장은 자음 42개 56.76%(유성자음 25개 59.52%+무성자음 17개 40.48%), 모음 32개 43.24% 비율로 구성되었다. 청지각 평가는 GRBAS 척도를 이용하여 0점(normal)에서 3점(severe)까지 4점 척도로 실시하였다. 상기 검사 절차는 정상 성인 집단에게도 동일하게 적용하였다.

### 2.3. 음향 분석

본 연구에서 수집한 음성샘플은 ADSV™ (Analysis of Dysphonia in Speech and Voice, Model 5109, Kay Pentax Medical, Montvale, NJ, USA)를 사용하여 분석하였다. 모음은 음도와 강도가 일정하게 유지되는 2초 이상의 안정구간을 편집하였으며, 문장은 검사 문장인 12어절(32음절)이 포함된 구간을 편집한 뒤, 자동으로 음성 분석 구간을 잡아주는 Apply Automatic Data selection 기능을 활용하여 분석하였다. 분석에 사용된 음향 변수는 모음 연장발성에 대한 CPP, mean CPP\_F0, CSID(cepstral spectral index of dysphonia)와 읽기 문장의 CPP, mean CPP\_F0, CSID다.

CPP 변수는 음성장에 환자의 경우 불규칙한 성대 접촉으로 인해 주기성 형성에 제한을 받아 그 값이 낮게 나타난다(Heman-Ackah et al., 2002; Heman-Ackah et al., 2003; Shim et al., 2015, 그림 1). mean CPP\_F0(CPP\_F0)는 전체 발성의 60-300Hz 범위에 해당하는 CPP의 평균 주파수를 나타내는 측정치로, 성대의 긴장성이 증가할수록 CPP\_F0도 커진다고 보고되어 있다(Lowell et al., 2012). 켈스트럼과 스펙트럼 측정치들로 산출된 회귀식에 의해 음성의 비정상 정도를 나타내는 음향학적 비정상 음성지수(CSID)는 음성 중증도가 높을수록 증가하는 양상을 보인다(Awan et al., 2016).



CPP, cepstral peak prominence; MTD, muscle tension dysphonia

그림 1. 문장읽기에서 MTD 환자(왼쪽)와 정상 성인(오른쪽)의 CPP  
Figure 1. The CPP in the MTD(left) and the control groups(right) in sentence utterance

MTD의 청지각적 평가는 음성장애 진단 및 치료경험이 3년 이상인 언어재활사 3명이 수행하였다. 평가자들은 녹음된 전체 음성 샘플 72개(MTD 모음연장발성 과제 36개+문장 읽기 과제

36개)를 무작위순으로 듣고, GRBAS 평가척도로 0(normal)에서 3점(severe)까지 4점 척도로 평가하였으며, 평가된 결과를 평균하여 통계에 적용하였다.

### 2.4. 통계 분석

MTD 환자와 정상성인 집단의 음향학적 특성을 비교하기 위해 SPSS version 19.0(IBM, Armonk, NY, USA)으로 독립표본 t검정(independent t-test)을 실시하였다. 유의수준 알파는 95%로 하였다. MTD 환자의 GRBAS 평가 결과를 살펴본 뒤, 음향변수 측정치와 GRBAS 청지각적 평가 결과 간의 Pearson 상관분석(Pearson Product Moment Correlation, PPMC)을 시행하였다.

데이터를 70% 무선(random) 선택하여 트레이닝 데이터로 삼아 랜덤포레스트 통계 모델을 만든 후 이 모델을 나머지 30% 테스트 데이터에 적용하여 그 분류정확도(classification accuracy)를 살펴보았다.

### 2.5. 신뢰도

청지각적 평가 결과에 대한 평가자 간 신뢰도를 측정된 결과, 급간 내 상관계수(ICC: Intraclass Correlation Coefficient) 중 평균 측정도(AMI: average measure intraclass Coefficient)가 .758로 통계적으로 유의하게 높은 평가자 간 신뢰도를 얻었다( $p < .001$ ). 음향학적 평가 결과에 대한 평가자 내 신뢰도를 측정하기 위해 음성 분석 일주일 후, 모음과 문장 과제의 20%(MTD 환자 음성 6개+정상 성인 6개)에 해당하는 음성샘플을 재분석하였으며,  $r = 1.000$ 으로 통계적으로 유의한 결과를 얻었다( $p < .001$ ).

## 3. 결과

### 3.1. MTD 환자와 정상 성인의 음향학적 특성

MTD 환자와 정상 성인의 음향학적 특성은 표 2와 같다. 전반적으로 CPP와 CPP\_F0는 정상 성인에게서 높은 값을 보였으며, CSID는 MTD 환자들에게 두드러졌다. 과제별로 살펴보면, 모음 과제에서 정상 성인의 CPP, CPP\_F0 값이 통계적으로 유의하게 컸으며, 읽기 과제에서도 정상 성인의 CPP, CPP\_F0 값이 큰 반면, CSID는 MTD 환자에게서 높게 나타났다.

표 2. MTD 환자와 정상 성인의 음향학적 특성

Table 2. Acoustic measurements in the MTD and the control groups

	Variables	Group	N	Mean±SD	t-value	p-value
Sustained vowel /a/	CPP	MTD	36	10.11±2.24	-3.632	.001**
		CG	36	11.56±1.16		
	CPP_F0	MTD	36	199.65±29.82	-3.126	.003**
		CG	36	218.12±19.18		
	CSID	MTD	36	12.14±12.91	1.220	.226
		CG	36	9.15±7.02		
Connected speech	CPP	MTD	36	5.21±1.15	-7.341	.000***
		CG	36	6.86±0.71		
	CPP_F0	MTD	36	201.59±19.93	-5.591	.000***
		CG	36	224.23±13.88		
	CSID	MTD	36	15.91±19.68	5.512	.000***
		CG	36	-3.36±7.22		

\* $p<.05$ , \*\* $p<.01$ , \*\*\* $p<.0001$ .

MTD, muscle tension dysphonia; CPP, cepstral peak prominence; CG, control group; CSID, cepstral spectral index of dysphonia.

### 3.2. MTD 환자의 GRBAS 특성과 음향학적 특성의 상관관계

MTD 환자의 모음연장발성과 문장 읽기 과제 수행 시 GRBAS 청지각 평가 결과는 표 3과 같다. 표에서 살펴볼 수 있듯이 MTD 환자는 전반적인 음성중증도(G) 항목에서 가장 두드러졌으며, 다음으로 긴장되고 거칠며(R), 바람 세는 목소리(B)와 쥐어짜는 목소리(S) 항목 순으로 나타났다. 과제에 따라 음성 특성이 크게 달라지지 않고 유사하게 나타나는 임상적 특징을 보였다.

표 3. MTD 환자의 청지각적 평가 결과

Table 3. Auditory-perceptual rating results for the clinical group

Perceptual measures	Mean±SD	
	Sustained vowel /a/	Connected speech
G	1.74±.59	1.67±.48
R	1.56±.58	1.55±.51
B	1.41±.57	1.33±.55
A	.11±.32	.15±.36
S	.89±.80	.85±.66

MTD 환자의 경우 전반적인 음성중증도(G), 조조성(R), 기식성(B)이 높아질수록 CPP는 감소하는 부적 상관관을 보였고, CSID는 증가하는 정적 상관관이 관찰되었다. G와 B 항목에서 CPP와 CSID는 청지각적 평가 결과와 중간 이상의 상관관계를 보였다(표 4).

표 4. MTD 환자의 음향학적, 청지각적 측정치의 상관

Table 4. Pearson correlation(r) between acoustic and perceptual measures

Task	Acoustic measures	G	R	B	A	S
Sustained vowel /a/	CPP	-.483*	-.450*	-.646***	-.286	-.011
	CPP_F0	.093	.149	.178	.048	.055
	CSID	.691***	.484*	.541**	-.101	.204
Connected speech	CPP	-.462*	-.386*	-.433*	-.350	.290
	CPP_F0	-.056	-.174	.118	-.114	.137
	CSID	.572**	.260	.355	-.094	-.110

\* $p<.05$ , \*\* $p<.01$ , \*\*\* $p<.001$ .

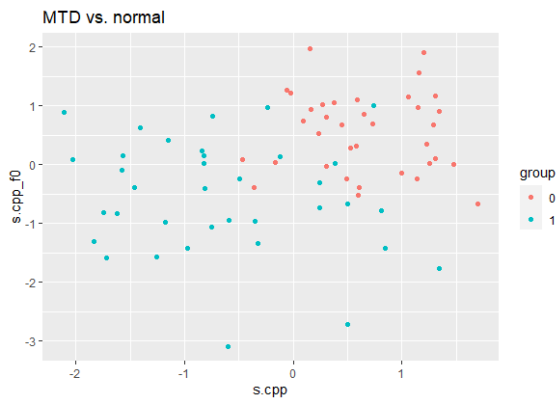
F0, fundamental frequency; G, grade; R, rough; B, breathy; A, asthenic; S, strained.

### 3.3. Random Forest 알고리즘을 이용한 MTD 구분

다수의 머신 러닝 알고리즘 중 랜덤포레스트는 데이터 분류에 주로 사용되며 앙상블 학습 방법으로 알려져 있다. 훈련과정에 참여하는 다수의 결정트리에서 출력되는 평균 예측치에 의해서 데이터를 분류하는 방식을 취한다(Breiman, 2001). 초기 개념은 Breiman(2001)에서 시작되었는데 랜덤 노드 최적화(randomized node optimization)와 배깅(bootstrap aggregating)을 결합한 CART(classification and regression tree) 개념으로 데이터 분류 방법을 제시하였다.

R(ver. 4.0.2) 통계 프로그램의 sample() 함수를 이용하여 트레이닝 데이터 70%와 테스트 데이터 30%를 무선(random) 구분한 다음 Random Forest 패키지를 이용하여 트레이닝 데이터를 모델링하였다. 모음과 문장읽기 모두에서 공통으로 집단 간 유의한 차를 보인 변수 CPP와 CPP\_F0 중(표 2 참조) 우선순위로 작용하는 변수를 알아보기 위해 importance() 함수를 적용하여 지니(Gini) 계수를 살펴본 결과 CPP 변수는 11.80970, CPP\_F0 변수는 11.66738로 CPP 변수의 중요도가 조금 더 높았다. 모음 데이터의 경우, 이 2개의 변수를 이용하여 트레이닝 데이터를 대상으로 두 집단(MTD, 일반) 분류한 결과 분류정확도가 100%로 나왔으며( $p<.0001$ ), 테스트 데이터의 분류 정확도는 75%로 나왔다( $p<.05$ ). 민감도와 특이도도 각각 75%로 동일하였다.

같은 방식으로 살펴본 문장읽기의 경우 CPP 변수는 지니(Gini) 계수가 12.95160, CPP\_F0 변수는 10.56657로 CPP 변수의 중요도가 모음에 비해 조금 더 높았다. 2개의 변수를 이용하여 트레이닝 데이터에서 두 집단(MTD, 일반) 분류한 결과 분류정확도가 모음의 경우와 마찬가지로 100%로 나왔으며( $p<.0001$ ), 테스트 데이터의 분류 정확도는 83.3%로 나왔다( $p<.01$ ). 민감도와 특이도도 각각 83.3%로 동일하였다. 문장읽기 데이터를 대상으로 두 변수를 x(normalized CPP), y(normalized CPP\_F0) 축에 배치하여 그려본 산점도는 그림 2와 같다. x축에 배치된 정규화된 CPP(s.cpp) 변수가 y축에 배치된 정규화된 CPP\_F0(s.cpp\_f0) 보다 데이터를 더 분명하게 구분하는 모습을 보여준다.



F0, fundamental frequency.

그림 2. 정규화된 CPP와 CPP\_F0의 집단(MTD vs. normal)별 분포 산점도(문장읽기), 0(붉은 점): 정상 집단

Figure 2. Scatter plot showing normalized CPP and CPP\_F0 according to MTD and normal group(sentence reading), 0(red dot): normal group

#### 4. 맺음말

본 연구는 MTD 여성 환자 36명과 정상 성인 여성 36명을 대상으로 모음 발성과 문장 읽기 과제에서 캡스트럼 기반 변수의 분석 결과를 비교하였다. 또한 음성장애 환자의 청지각적 특성과 음향학적 특성의 상관성을 살펴보고, MTD 감별과 관련하여 머신러닝 방법론을 적용해보았다.

모음연장발성과 문장읽기 과제를 수행한 MTD환자 음성의 CPP 값은 모음 10.11(±2.24) dB, 문장 5.21(±1.15) dB로 정상 성인의 모음 11.56±1.16, 문장 6.86±0.71에 비해 통계적으로 유의하게 낮은 양상을 보였다. 이는 다양한 음성장애군과 정상 성인의 음성특성을 캡스트럼 분석을 통해 비교한 Yu et al.(2018)의 연구와 일치하는 결과다. 정상화자의 경우 음성 산출 시 주기적인 음향신호로 인해 조화음구조가 잘 형성되는 반면, 병리적 음성의 경우 성대 내전의 불완전 및 불규칙성으로 인해 주기성 확보에 제한을 받아 낮은 캡스트럼 값을 보인다(Heman-Ackah et al., 2002; Heman-Ackah et al., 2003; Hillenbrand & Houde, 1996; Hillenbrand et al., 1994).

모음과제 시 장애군의 CPP\_F0는 199.65(±29.82) Hz로 정상군(218.12(±19.18) Hz)에 비해 통계적으로 유의하게 낮은 값을 보였다. 문장읽기에서도 장애군 201.59(±19.93) Hz, 정상군 224.23(±13.88) Hz로 장애군의 CPP\_F0값이 정상군에 비해 낮게 나타났다. CPP\_F0는 60-300 Hz 범위의 CPP 평균 주파수를 나타내며, 비정상적인 발성노력이 심할수록 높아지는 양상을 보인다고 하였다(Seo, 2014). 반면 후두의 긴장 정도가 낮은 상태에서 산출되는 기식음성의 경우 정상음성에 비해 낮은 CPP\_F0를 보인다(Seo & Lee, 2015). 3.2 장에서 언급한 GRBAS와 음향변수 사이의 상관관계를 살펴보면 CPP\_F0는 어떤 항목과도 유의한 상관관계를 맺지 못했다. 이 논문의 데이터에만 국한해서 판단한다면 CPP\_F0는 GRBAS 척도와는 무관하게 작동하는 변수라는 해석이 가능하다. MTD 환자는 발성 시의 과도한 긴장이 인두 수축근이나 심경부의 다른 근육에도 영향을 미쳐 애성,

노력성 발성 외에도 음이탈, 음성 단절과 고주파수 대역에서의 잡음이 같이 발생하게 된다(Noh et al., 2017). 이러한 음질특성은 MTD 환자 고유의 특징이라 할 수 있는데 이로 인해 MTD 환자의 CPP\_F0가 더 낮은 결과를 보인 것으로 해석되며 이런 복합적 특성이 GRBAS 청지각 평가에도 복합적으로, 혼동스럽게 반영된 것이 아닌가 생각된다. 또 한 가지 덧붙일만한 소견은 본 연구의 대상자는 모두 여성으로 남성에게 비해 기식성(breathiness) 음질이 높은 대상군 특성(Choi & Choi, 2016) 때문이지 않을까 하는 것이다. CPP\_F0의 표준편차 값은 MTD가 정상군에 비해 더 넓게 나타났는데, 장애음성의 비정상성이 심할수록 표준편차 값이 증가하는 양상을 보이므로, MTD에 비해 정상군의 음질이 더 좋을 것을 시사하는 결과다.

CSID는 음성의 비정상 정도를 나타내는 음향변수로 음성 중증도가 높을수록 증가하는 양상을 보인다(Awan et al., 2016). Shim et al.(2016)은 모음 연장발성을 대상으로 한 CSID가 정상군에 비해 MTD 환자에게서 유의하게 높은 값을 보였고, 모음과제 CSID가 MTD의 음성특성을 반영하는데 유용한 측정치라고 설명하였다. 본 연구의 경우 모음 연장발성 데이터는 통계적으로 유의한 차이를 보여주지 않았다. 오히려 MTD 환자의 문장과제 CSID값이 정상군에 비해서 유의하게 높은 양상을 보였는데, 이는 MTD환자의 음성특성이 문장과제에서 더 잘 드러날 수 있음을 의미한다고 할 수 있다.

MTD 환자의 청지각적 음질 특성은 캡스트럼 기반 변수와 높은 상관관계가 있다(Awan & Roy, 2006; Heman-Ackah et al., 2003; Hillenbrand et al., 1994; Lowell et al., 2012). 본 연구결과에서는 모음연장발성 과제에서 전반적인 음성중증도(G), 조조성(R), 기식성(B)이 높아질수록 CPP는 감소하고, CSID는 증가하는 양상이 관찰되었다. 이러한 양상은 문장 읽기 과제에서도 유사하였다.

음성장애 환자들의 음성 치료 전·후를 종단적으로 살펴본 Awan & Roy(2006)는 음성 치료 후 환자들의 CPP가 증가하고 CPP SD값이 유의하게 감소되었을 뿐만 아니라, 시각적 아날로그 척도(VAS)를 통한 청지각적 평가에서도 음질문제가 감소되는 양상을 보여 캡스트럼과 스펙트럼 측정치들이 음성장애의 중증도와도 밀접하게 관련됨을 보고하였다(Awan & Roy, 2009).

상기와 같은 결과는 신경질환 환자에게서도 살펴볼 수 있다. Seo & Seong(2013)은 연결발화에서 마비말장애 환자의 청지각적 측정치 G, B, S와 캡스트럼 측정치인 CPPs(smoothed CPP), CPP\_F0가 높은 상관성을 보여 청지각적 측정치와 캡스트럼 분석의 관련성을 제시하였다. 목 원 음성과 거친 음성은 MTD 환자에게서 두드러지는 음질 특성으로 과도한 성대근육의 긴장으로 인해 불충분한 성대내전 또는 발성노력(vocal effort) 등의 보상적 발성을 보일 수 있다(Altman et al., 2005). 이러한 성대의 불규칙적인 운동에 의해 목소리의 조화음 형성은 제한을 받게 된다. 음성 중증도가 높아짐에 따라 CPP는 감소하고 CSID는 증가하는 양상을 보인다.

CSID는 CPP와 L/H ratio, 성별(sex)에 대한 정보를 다중회귀 방식으로 가중치를 주어 도출된 지수로, 청지각적인 평가 결과

와 높은 상관성이 있다(Jalalinajafabadi et al., 2013). 본 연구에서는 모음과 문장과제 수행 시 MTD 환자의 CSID가 정상 성인에 비해 높게 나타났다. 이는 음성장애의 중증도가 객관적으로도 높게 나타났음을 의미하며, MTD 환자의 음성장애 중증도가 정상화자보다 더 높게 나타났다는 선행연구들과 일치하는 결과다(Awan et al., 2014; Awan et al., 2016; Peterson et al., 2013).

CSID와 GRBAS의 상관성을 살펴본 Jalalinajafabadi et al.(2013)은 음성장애환자의 청지각적 평가 결과와 CSID가 moderate 수준의 상관성이 있다고 보고하였고, Peterson et al.(2013)은 CSID가 중재 전후를 비교한 변화를 청지각적으로 평가한 VAS 결과와도 높은 상관성을 보여준다고 하였다. Awan et al.(2014)은 332명의 음성장애환자가 본인의 음성문제로 인해 느끼는 불편함의 정도를 VHI(voice handicap index) 설문지로 살펴본 것이었는데, CSID와 통계적으로 유의미하게 높은 상관성이 있다고 하였다.

MTD 환자와 정상 음성의 분류를 머신러닝 계열의 랜덤포레스트 알고리즘으로 시행한 결과 모음에 비해 문장읽기의 테스트 데이터에서 조금 더 높은 분류 정확도가 나왔으며(83.3%) 연장발성 모음과 문장읽기 둘 다 CPP 변수의 중요성이 조금 더 높은 결과를 보여주었다. 그림 2의 집단별 산점도에서도 CPP(x 축)가 집단별로 더 분명하게 데이터가 나누어진 모습을 보여준다.

분석 결과에서 알 수 있듯이 CPP는 모음 및 문장읽기 모두에서 MTD 음성 특성을 감별하기에 적절한 음향변수임이 입증되었다. 문장읽기에서 CPP와 CPP\_F0 두 변수의 조합으로 구성된 랜덤포레스트 모델이 보여준 분류정확도는 괄목할 만 하였다(트레이닝 데이터: 100%, 테스트 데이터: 83.3%). 이 2개의 변수는 서론에서 언급한 집단별 구분 감별 음향변수로 모자람이 없어 보이며 그 실행에 랜덤포레스트 알고리즘을 사용할만 하다는 결론을 내릴 수 있다.

본 연구에서는 다루지 않았지만 음성장애관련 연구에서 중요한 역할을 하는 변수인 LH ratio를 랜덤 포레스트 분류 작업에 포함시킬 경우 모음 연장발화에서는 트레이닝 데이터 100%, 테스트 데이터 75%의 분류 정확도를 보여주었으며, 문장읽기의 경우 트레이닝 데이터 100%, 테스트 데이터 83.3%의 분류 정확도를 보여주었다. LH ratio는 스펙트럼 객체를 대상으로 측정하는 변수이므로 켈스트럼 변수만을 다루는 본 연구에 포함하는 것이 적절하지 않으나 전체 데이터수(72개)가 변수 3개를 독립변인으로 활용해도 과다적합에 해당되지 않으므로 포함시켜보았다. 변수 2개(CPP, CPP\_F0)만을 사용할 때와 차이점이 없으므로 분류 정확도를 높여준다는 면에서 LH ratio의 기여도는, 이 논문의 데이터에서만큼은, 평가하기가 어렵다.

본 연구는 성인 여성만을 대상으로 연구하였다는 약점이 있다. 성인 남성집단에 대한 연구를 시행하여 추후 반영해본다면 결과를 일반화할 수 있을 것으로 생각된다.

## References

Alharbi, G. G., Cannito, M. P., Buder, E. H., & Awan, S. N. (2019).

Spectral/cepstral analyses of phonation in Parkinson's disease before and after voice treatment: A preliminary study. *Folia Phoniatrica et Logopaedica*, 71(5-6), 275-285.

Altman, K. W., Atkinson, C., & Lazarus, C. (2005). Current and emerging concepts in muscle tension dysphonia: A 30-month review. *Journal of Voice*, 19(2), 261-267.

Awan, S. N., & Roy, N. (2006). Toward the development of an objective index of dysphonia severity: A four-factor acoustic model. *Clinical Linguistics & Phonetics*, 20(1), 35-49.

Awan, S. N., & Roy, N. (2009). Outcomes measurement in voice disorders: Application of an acoustic index of dysphonia severity. *Journal of Speech, Language & Hearing Research*, 52(2), 482-499.

Awan, S. N., Roy, N., & Cohen, S. M. (2014). Exploring the relationship between spectral and cepstral measures of voice and the Voice Handicap Index (VHI). *Journal of Voice*, 28(4), 430-439.

Awan, S. N., Roy, N., Zhang, D., & Cohen, S. M. (2016). Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: Development of clinical cutoff scores. *Journal of Voice*, 30(2), 130-144.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

Choi, S.H., & Choi, C. H. (2016). The effect of gender and speech task on cepstral- and spectral-measures of Korean normal speakers. *Audiology and Speech Research*, 12(3), 157-163.

Heman-Ackah, Y. D., Michael, D., & Goding, G. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, 16(1), 20-27.

Heman-Ackah, Y., Heuer, R., Michael, D., Ostrowski, R., Horman, M., Baroody, M., Hillenbrand, J., & Sataloff, R. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otolaryngology, Rhinology & Laryngology*, 112(4), 324-333.

Hillenbrand, J., & Houde, R. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39(2), 311-321.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4), 769-778.

Jalalinajafabadi, F., Gadepalli, C., Ascott, F., Homer, J., Luján, M., & Cheetham, B. (2013, November). Perceptual evaluation of voice quality and its correlation with acoustic measurement. *Proceedings of the 2013 European Modelling Symposium* (pp. 283-286). Manchester, UK.

Kim, G. H., Lee, Y. W., Park, H. J., Bae, I. H., & Kwon, S. B. (2017). A study of cepstral peak prominence characteristics in ADSV, speech tool and praat. *Journal of Speech-Language & Hearing Disorders*, 26(3), 99-111.

Kim, N. S., & Seong, C. J. (2017). The acoustic characteristics and classification variables of two Hyponasal groups. *The Linguistic Society of Korea*, 78, 31-61.

- Koufman, J. A., & Blalock, P. D. (1988). Vocal fatigue and dysphonia in the professional voice user: Bogart-bacall syndrome. *The Laryngoscope*, 98(5), 493-498.
- Kumar, B., Bhat, J., & Prasad, N. (2010). Cepstral analysis of voice in persons with vocal nodules. *Journal of Voice*, 24(6), 651-653.
- Lowell, S. Y., Kelley, R. T., Awan, S. N., Colton, R. H., & Chan, N. H. (2012). Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. *Annals of Otology, Rhinology & Laryngology*, 121(8), 539-548.
- Noh, S. H., Kim, S. Y., Cho, J. K., Lee, S. H., & Jin, S. M. (2017). Differentiation of adductor-type spasmodic dysphonia from muscle tension dysphonia using spectrogram. *Journal of Korean Society of Laryngology, Phoniatrics and Logopedics*, 28(2), 100-105.
- Park, J. H. (2011). *A study on aspects of vocal cord vibration and acoustic characteristics according to types of muscle tension dysphonia* (Master's thesis). Daegu University, Daegu, Korea.
- Peterson, E. A., Roy, N., Awan, S. N., Merrill, R. M., Banks, R., & Tanner, K. (2013). Toward validation of the cepstral spectral index of dysphonia (CSID) as an objective treatment outcomes measure. *Journal of Voice*, 27(4), 401-410.
- Pyo, H. Y., & Shim, H. S. (2007). A study for the development of Korean voice assessment model for the patients with voice disorders: A qualitative study. *Phonetics and Speech Sciences*, 14(2), 7-22.
- Rubin, J. S., Sataloff, R. T., & Korovin, G. S. (2006). *Diagnosis and treatment of voice disorders*. San Diego, CA: Plural.
- Seo, I. (2014). *Acoustic measures of voice quality and phonation types across speech conditions in dysarthria* (Doctoral dissertation). Chungnam National University, Chungnam, Korea.
- Seo, I. H., & Lee, O. B. (2015). Cepstral and spectral analysis of whispery voice by healthy adults: Preliminary study. *Journal of Speech-Language & Hearing Disorders*, 24(4), 259-266.
- Seo, I. H., & Seong, C. J. (2013). Voice quality of dysarthric speakers in connected speech. *Journal of the Korean Society of Speech Science*, 5(4), 33-41.
- Shim, H. J., Jang, H. R., Shin, H. B., & Ko, D. H. (2015). Cepstral, spectral and time-based analysis of voices of esophageal speakers. *Folia Phoniatrica et Logopaedica*, 67(2), 90-96.
- Shim, H. J., Jung, H., Lee, S. A., Choi, B. H., Heo, J. H., & Ko, D. H. (2016a). Cepstral and spectral analysis of voices with adductor spasmodic dysphonia. *Phonetics and Speech Sciences*, 8(2), 73-80.
- Shin, H. B., Shim, H. J., Jung, H., Ko, D. H. (2018). Characteristics of voice quality on clear versus casual speech in individuals with Parkinson's disease. *Phonetics and Speech Sciences*, 10(2), 77-84.
- Watts, C., & Awan, S. (2011). Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech, Language, and Hearing Research*, 54(6), 1525-1537.
- Yu, M., Choi, S. H., Choi, C. H., & Choi, B. (2018). Predicting normal and pathological voice using a cepstral based acoustic index in sustained vowels versus connected speech. *Communication Sciences & Disorders*, 23(4), 1055-1064.

• **윤주원 (Joowon Yun)**

충남대학교 대학원 언어병리학과 박사과정  
대전광역시 유성구 대학로 991  
Tel: 042-821-5114  
Email: voiceslp@naver.com  
관심분야: 음성장애, 음성치료, 장애음성분석

• **심희정 (Heejeong Shim)**

한림대학교 일반대학원 언어청각학부 박사  
강원도 춘천시 한림대학길 1  
Tel: 033-248-2227  
Email: amygmlwd@gmail.com  
관심분야: 말장애, 음성장애, 장애음성분석

• **성철재 (Cheoljae Seong)** 교신저자

충남대학교 언어학과 교수  
대전광역시 유성구 대학로 991  
Tel: 042-821-6395  
Email: cjseong49@gmail.com  
관심분야: 분절음 및 운율 분석

# 캡스트럼 변수와 랜덤포레스트 알고리즘을 이용한 MTD(근긴장성 발성장애) 여성화자 음성과 정상음성 분류\*

윤주원<sup>1</sup> · 심희정<sup>2</sup> · 성철재<sup>3</sup>

<sup>1</sup>충남대학교 언어병리학과, <sup>2</sup>한림대학교 언어청각학부, <sup>3</sup>충남대학교 언어학과

## 국문초록

근긴장성 발성장애(cepstral peak prominence, MTD) 환자의 모음 발성과 문장읽기 과제를 캡스트럼 기반 변수를 이용하여 분석하였으며 음성장애 환자의 GRBAS 청지각적 특성과 음향학적 특성의 상관관계를 살펴보고, 랜덤포레스트 머신러닝 분류 알고리즘을 이용한 MTD 감별 진단 가능성을 논의하였다. 내원 시 MTD로 진단받은 여성 36명과 정상음성을 사용하는 여성 36명이 연구에 참여했으며, 수집한 음성샘플은 ADSV™ 를 사용하여 분석하였다. 연구 결과, 음향학적 측정치 중 MTD의 CSID(cepstral spectral index of dysphonia)는 대조군보다 높았으며, CPP(cepstral peak prominence), CPP\_Fo 값이 대조군보다 유의하게 낮았다. 이는 모음 발성과 읽기 과제에서 모두 동일하게 나타났다. MTD 환자의 음질 특성은 전반적인 음성중증도(G)가 가장 두드러졌으며, 조조성(R), 기식성(B), 노력성(S)순으로 음성 특성을 보였다. 이 특성이 높아질수록 CPP가 감소하는 부적 상관을 보이고, CSID는 증가하는 정적 상관이 관찰되었다. 캡스트럼 변수 중 모음과 문장읽기과제 모두에서 집단간 유의한 차이를 보여준 CPP와 CPP\_F0를 이용하여 MTD와 대조군의 음성분류를 시도하였다. 머신러닝 알고리즘인 랜덤포레스트로 모델링한 결과 문장읽기 과제에서 모음연장발성보다 조금 더 높은 분류 정확도(83.3%)가 나왔으며, 모음 발성과 문장 읽기 과제 모두에서 CPP변수가 더 중심적 역할을 수행하였음을 알 수 있었다.

**핵심어:** 근긴장성발성장애, 음성장애, 문장읽기, CPP(cepstral peak prominence), CSID(cepstral spectral index of dysphonia), 머신러닝, Random Forest

## 참고문헌

김근호, 이연우, 박희준, 배인호, 권순복 (2017). ADSV, Speech-Tool, Praat 간의 CPP 특성 연구. *언어치료연구*, 26(3), 99-111.

김난숙, 성철재 (2017). 과소비성(hyponasality) 집단의 음향특성과 분류변인. *언어학*, 78, 31-61.

노승호, 김소연, 조재경, 이상혁, 진성민 (2017). 스펙트로그램을 이용한 내전형 연속성 발성 장애와 근긴장성 발성 장애의 감별. *대한후두음성언어의학회지*, 28(2), 100-105.

박진향 (2011). *근긴장성 발성장애의 유형에 따른 성대 진동 양상 및 음향학적 특성 연구*. 대구대학교 석사학위논문.

서인호 (2014). *마비말장애의 발화조건에 따른 음질 및 발성유형의 음향음성학적 접근*. 충남대학교 박사학위논문.

서인호, 성철재 (2013). 연결발화에서 마비말화자의 음질 특성. *말소리와 음성과학*, 5(4), 33-41.

서인호, 이옥분 (2015). 연속성 발성장애 음성에 대한 캡스트럼과

스펙트럼 분석: 예비연구. *언어치료연구*, 24(4), 259-266.

심희백, 심희정, 정훈, 고도홍(2018). 명료발화와 보통발화에서 파킨슨병환자 음성의 캡스트럼 및 스펙트럼 분석. *말소리와 음성과학*, 10(2), 77-84.

심희정, 정훈, Lee sue ann, 최병훈, 허정화, 고도홍 (2016). 내전형 연속성 발성장애 음성에 대한 캡스트럼과 스펙트럼 분석. *말소리와 음성과학*, 8(2), 73-80.

유미옥, 최성희, 최철희, 최병(2018). 모음과 연결발화에서 캡스트럼 음향 지표의 정상 및 음성장애 예측. *한국언어청각임상학회지*, 23(4), 1055-1064.

최성희, 최철희. (2016). 한국 정상 화자의 캡스트럼과 스펙트럼 측정치의 성별과 발화 과제 효과. *청능재활*, 12(3), 157-163.

표화영, 심현섭 (2002). 음성장애 진단 및 평가에 관한 질적 연구: 진단 및 평가 모형 정립을 위한 기초연구. *음성과학*, 14(2), 7-22.

\* 이 논문은 2020년도 충남대학교 자체연구비의 지원을 받아 수행되었습니다.