

식도발성화자 음성의 spectral & cepstral 분석 Spectral and Cepstral Analyses of Esophageal Speakers

심 희 정¹⁾ · 장 효 령²⁾ · 신 희 백³⁾ · 고 도 흥⁴⁾

Shim, Hee-Jeong · Jang, Hyo-Ryung · Shin, Hee-Baek · Ko, Do-Heung

ABSTRACT

The purpose of this study was to analyze spectral versus cepstral measurements in esophageal speakers. The comparison between the measurements in thirteen male esophageal speakers was compared with the control group of thirteen normal speakers using the sustained vowel /a/. The main results can be summarized as below: (a) the CPP and L/H ratio of the esophageal group were significantly lower than those of the control group (b) the CPP was significantly correlated with the spectral parameters such as jitter, shimmer, NHR and VTI, and (c) the ROC analysis showed that the threshold of 10.25dB for the CPP achieved a good classification for esophageal speakers, with 100% perfect sensitivity and specificity. Thus, it was known that cepstral-based acoustic measures such as CPP, may be more reliable predictors than other spectral-based acoustic measures such as jitter and shimmer. And it was found that cepstral-based acoustic measures were effective in distinguishing esophageal voice quality from normal voice quality. This research will contribute to establishing a baseline related to speech characteristics in voice rehabilitation with laryngectomees.

Keywords: total laryngectomy, voice quality, esophageal speakers, cepstral analysis, CPP, L/H ratio, mean CPP Fo, ADSV

1. 서론

2012년 한국중앙암등록본부 자료에 의하면 후두암(larynx cancer)은 1,444건으로 전체 암 발생률의 0.6%를 차지한다. 후두암의 주된 요인은 과다한 흡연 및 음주, 공기오염 등으로, 목소리 변화, 이물감, 삼킴 문제, 통증 등의 전조증상을 보일 수 있다[1]. 현재 후두암의 치료방법으로 외과적 수술과 방사선 및 약물치료가 병행되고 있으며, 수술의 경우 후두전적출술(total laryngectomy)과 후두부분적출술(partial laryngectomy)이 주로 시행되고 있다. 위와 같은 수술을 받은 환자들은 수술 후 일부 또는 전체 후두 및 성대 구조의 소실로 인해 정상

적인 발성에 제한을 받기 때문에 음성재활이 요구된다.

후두적출자의 음성재활 유형으로 식도발성(esophageal speech)이 가장 많이 사용되는데, 이는 트림과 같은 원리로 식도를 통해 공기를 벨어 인두-식도 부위(pharyngo-esophageal segment, PES)의 진동을 통해 발생하는 방법이다. 식도발성은 숙련될 경우 정상인과 유사한 음질, 음높이와 목소리 크기 또한 조절할 수 있기 때문에 후두적출술을 받은 환자들에게 많이 선호되지만, 후두 및 성대의 절제로 목소리를 산출할 수 있는 근원을 잃게 되어 이전처럼 정상적인 의사소통을 하기에는 제한점이 많다[2, 5, 8, 9].

식도발성화자는 발성기관의 부재와 발성에 필요한 에너지 원인 공기의 부족으로 특히 저하된 음질 문제를 호소하게 된다. 이들의 음질 특성은 주로 spectral 측정치인 기본주파수(Fo), 주파수 변동률(jitter), 진폭 변동률(shimmer), 소음 대 배음 비율(NHR) 등의 음성 매개변수에 대한 연구 결과를 보고한다[2-9]. 특히 식도발성화자는 심각한 음질 문제로 인해 jitter, shimmer, NHR 수치가 높은 양상을 보이는데[2, 4, 5], 이는 신성대(neoglottis)의 역할을 하는 PES가 근본적인 발성 기관이 아니므로 성대만큼의 진동수 확보에 어려움이 있기 때문이다. 위와 같이 공통적으로 식도발성화자의 심각한 음질

-
- 1) 한림대학교 대학원 언어병리청각학과, amy2020@hallym.ac.kr
2) 한림대학교 대학원 언어병리청각학과, gyufd0601@hanmail.net
3) 한림대학교 대학원 언어병리청각학과, yj2000102@nate.com
4) 한림대학교 언어청각학부, dhko7@hallym.ac.kr, 교신저자
이 논문은 2013년도 한림대학교의 교내 학술 연구비(HRF-201312-019) 지원으로 수행되었습니다.

접수일자: 2014년 4월 28일
수정일자: 2014년 5월 28일
게재결정: 2014년 5월 28일

문제를 보고한 연구들에 비해 Fo에 대한 연구 결과는 서로 다른 양상을 보인다[2, 6, 7, 8, 9]. 정상 성인과 유사한 Fo 값을 보고한 [2], [6], [7] 연구에서는 식도발성화자의 Fo 값이 137.2-176.4Hz로 다소 높은 양상을 보였다. 일반적으로 높은 Fo는 청자에게 더욱 수용적이고 자연스럽게 들리는 것으로 판단될 수 있지만, 식도발성화자의 경우 가변적인 주파수와 진폭 변이의 양과 소음관련 요소들의 영향으로 나쁜 음질로 판단될 수 있다. 반면 [3, 8, 9] 연구에서는 64.6-80Hz까지 낮은 Fo 값을 보고하였으며 이는 식도발성의 불안정한 음정산출로 인한 비주기성으로 측정 또는 분석이 불가능하기 때문으로 볼 수 있다. 위와 같이 소리의 근원을 나타내는 Fo 값에서 서로 다른 양상을 보고하는 것은 식도발성의 불규칙적이고 비주기적인(aperiodic) 음성 특성을 보여주며 이는 주파수 관련 음성 매개변수의 특성을 살펴보는 데 제한점을 제시한다[3]. 이러한 식도발성화자의 비주기적인 음성 특성은 CSpeech 프로그램을 통해서도 살펴볼 수 있는데, 자기상관함수의 피크와 일치하는 주기를 계산하는데 실패한 분석알고리즘의 수를 나타내는 Err count가 0인 정상 성인에 비해, 식도발성화자는 86.9로 높은 수치를 보여 비주기적 음성의 음향분석에 신뢰로운 수치가 모색되어야 함을 강조하였다[10].

일반적으로 jitter, shimmer와 NHR 등의 spectral 측정치들은 Fo 주기의 정확한 측정을 요구하기 때문에 식도발성처럼 심각한 음성장애의 경우 발성의 불규칙성으로 Fo 측정이 어려워 측정된 수치들에 대한 신뢰성이 떨어진다. 최근에는 이러한 점을 보완하고자 전체 신호음 중 조화음(harmonics)의 정도를 나타내는 캡스트럼 분석에 대한 관심이 대두되고 있다[11]. cepstral 측정치는 주파수영역에서 성대의 주기진동 신호와 성도의 공명주파수를 분리하여 정확한 Fo 측정을 요하지 않으므로 심각한 음질 분석의 경우에서도 신뢰성 있는 주기성 결과를 얻을 수 있다[12]. 특히 음질문제가 심각한 경우 비주기적인 특성 때문에 spectral 분석에 비해 cepstral 분석이 높은 민감도와 특이도를 보이며[13, 14], 모음연장발성과 연결발화 과제에서 캡스트럼 방식의 타당성과 신뢰성을 보고한다[11, 13, 14]. 따라서 본 연구에서는 식도발성화자 음성분석에 spectral 분석과 cepstral 분석을 적용하여 모음 /a/ 연장발성에서 다음 세 가지의 음향 파라미터들을 살펴보고자 한다.

첫째, CPP(cepstral peak prominence)는 조화음의 정도를 나타내는 매개변수로 정상 음성의 경우 주기적 신호에서 조화음 구조가 잘 나타나기 때문에 높은 CPP 값을 나타내며[15], 음질문제가 심각한 경우에는 비주기성 때문에 캡스트럴 피크가 완만하고 낮은 CPP 값을 나타내게 된다. 특히 기식 음성을 주요 음성 특징으로 보이는 성대마비나 성대결절 환자의 경우도 낮은 CPP 값을 나타냈으며, 이는 성대의 불충분한 폐쇄로 캡스트럴 피크의 비정상적인 감소를 나타낸다[16, 17]. 둘째, mean CPP Fo는 가장 큰 조화음 에너지를 가지는 CPP의 주파

수로 기본주파수를 나타내는데 이때 캡스트럼 상에서 제일 두드러진 캡스트럴 피크(cepstral peak) 정점은 원음의 기본 주기인 Fo를 반영한다[18]. 마지막으로, L/H ratio는 4kHz를 컷오프(cutoff)하여 주파수영역을 기점으로 4kHz이상과 이하의 비율 에너지(dB)를 비교함으로써 계산된다[14]. 즉 스펙트럼 상에서 고주파수 에너지에 대한 저주파수 에너지의 비율로 정상 음성의 L/H ratio 값이 더 크게 나타난다. 이때 고주파수 영역에 높은 에너지를 나타내는 것은 기식성(breathiness), 조조성(strainedness)과 상관성이 높기 때문에 병리적인 음성에서 낮은 L/H ratio 값을 보고한다[13, 19].

cepstral 측정치는 전체 신호음 중 소음으로부터 구분되는 조화음의 주기성을 나타내기 때문에 주파수 관련 음성 매개변수인 jitter, shimmer, NHR과의 관련성도 살펴볼 수 있다. [15]는 성대마비 환자의 CPP 값과 MDVP의 매개변수인 NHR(noise-to-harmonic ratio), APQ(amplitude perturbation quotient), RAP(relative average perturbation), sPPQ(smoothed pitch perturbation)의 높은 상관성을 보고하였으며, 이러한 매개변수들은 기본 주기의 불규칙성에 관한 변수이기 때문에 병적음성을 감별 진단하는 음향학적 변수라고 하였다[20].

지금까지 식도발성화자의 음질 특성과 캡스트럼 분석에 관한 국내외 연구들을 살펴본 결과[2-20], 식도발성화자의 음질 특성을 반영하는 데 섭동분석(perturbation analysis)에 근거한 spectral 음향학적 매개변수의 제한점을 살펴볼 수 있었다. 또한 심각한 음질문제를 호소하는 식도발성화자의 비주기적인 음성을 새롭게 정의하기 위해 캡스트럼 분석을 적용하여 살펴볼 필요가 있음을 알 수 있었다. 따라서 본 연구에서는 캡스트럼 분석을 적용하여 식도발성화자의 음질 특성의 다방면적인 이해와 그들을 위한 음성재활접근 및 효율적인 임상적 중재 방안에 기초가 되는 자료를 제공하기 위해 다음의 연구문제들을 살펴보고자 한다.

첫째, 집단(식도발성/정상노인)에 따른 cepstral 측정치(CPP, mean CPP Fo, L/H ratio)에 차이가 있는가?

둘째, 집단(식도발성/정상노인)에 따른 spectral 측정치(Fo, jitter, shimmer, NHR, VTI, SPI)에 차이가 있는가?

셋째, 집단(식도발성/정상노인)에 따른 cepstral 측정치와 spectral 측정치의 상관성은 어떠한가?

넷째, cepstral 측정치가 식도발성화자의 음질문제에 대한 감별 진단적 가능성이 있는가?

2. 연구 방법

2.1 연구 대상

본 연구는 서울지역 병원의 음성재활실을 이용하는 후두적출술을 받은 환자 13명(M=13)을 대상으로 하였으며, 환자와

연령을 일치시킨 노인 13명, 총 26명을 군집하였다. 대상자는 후두암으로 인해 전체후두적출을 한 60-80대 식도발성화자로, 식도발성의 숙련도에 대한 통계학적 변인에 차이가 없음을 확인하였다. 또한 시각·청각 및 언어문제가 없는 화자를 대상으로 하였으며, 이들에 대한 배경정보는 <표 1>에 제시하였다.

표 1. 대상자
Table 1. subjects

	식도발성	정상노인
대상자 수	13 (M=13)	13 (M=13)
연령	73.53±6.33	73.69±5.57
숙련기간	11.92±4.97	.
수술유형	후두전적출술	.
수술원인	후두암	.

2.2 검사 도구

KayPENTAX사의 ADSV™(Analysis of Dysphonia in Speech and Voice Model 5109)와 MDVP(Multi-Dimensional Voice Program)가 탑재된 데스크탑에 지향성 마이크(SHURE SM48)를 연결하여 샘플링 속도(sampling rate)를 22,000Hz, 44,000Hz로 녹음 및 디지털화하였다.

2.3 검사 절차

본 연구에서 환자군 음성 녹음은 소음이 40dB 이하인 병원 내 조용한 방에서 실시하였으며, 의자와 책상이 배치된 공간에서 환자와 마주 보는 상태로 음성녹음을 실시하였다. 음성 녹음 시, 마이크는 화자의 입과 10cm 떨어진 위치에서 90도 각도로 고정된 마이크로 녹음하였다. 또한 화자에게 최대한 편안하고 자연스러운 상태에서 모음 /a/를 3회 2초간 연장발성 하였다. 대조군 또한 같은 검사 절차를 실시하였다.

2.4 자료 분석

cepstral과 spectral 측정치의 자료 분석은 각각 ADSV™와 MDVP를 사용하였으며 대상자가 모음 /a/를 1회 연장 발성한 샘플마다 모음의 시작부에서 약 1/3 지점이면서 안정적으로 펄스 신호(pulse bar)가 생기고, 포먼트, 강도 그리고 피치선이 일정하게 수평이 되는 부분을 분석하였다. 모음 자료 처리는 3회 반복한 값을 평균 처리하여 통계에 적용시켰다. <그림 1>은 ADSV로 분석한 식도발성화자와 정상노인의 캡스트럼 그래프이다. 앞서 언급했던 것처럼 정상 음성의 경우 주기적 신호에서 조화음 구조가 잘 나타나 높은 CPP 값과 캡스트럴 피크를 나타내며, 음질문제가 심각한 경우에는 비주기성 때문에 낮은 CPP 값과 완만한 캡스트럴 피크를 나타내게 된다.

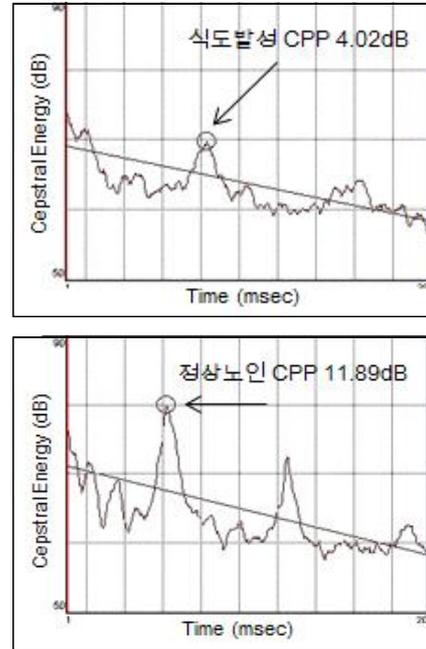


그림 1. 집단 간 CPP 비교
Figure 1. CPP between groups

2.5 통계 분석

집단 간 음성 매개변수의 차이를 보기 위해 SPSS(version 19.0)을 사용하여 집단을 변인으로 하는 Mann-Whitney 검정을 실시하였으며, cepstral과 spectral 측정치의 상관성을 살펴보고자 Spearman 서열상관계분석을 실시하였다. 모든 측정값은 M±SD로 나타내었으며 유의수준이 .05 이하인 경우를 통계적으로 유의한 차이가 있다고 평가하였다. 또한 cepstral 측정치가 식도발성화자의 음질문제에 대한 감별 진단적 가능성이 있는지 살펴보고자 ROC 곡선 분석을 실시하였다.

2.6 신뢰도

자료 분석에 대한 신뢰도 검증을 위해 전체 음성자료의 20%를 임의로 선정하여 평가자내 신뢰도를 측정된 결과, cepstral 측정치의 평가자내 신뢰도는 CPP 99.47%, mean CPP Fo 99.34%, L/H ratio 99.21%, spectral 측정치는 Fo 92.84%, jitter 84.42%, shimmer 88.62%, NHR 87.02%, VTI 89.32%, SPI 96.73%로 다소 낮은 신뢰도를 보였다. 이는 앞서 언급했던 것처럼 음질이 나쁜 음성의 경우 섹션분석에 기초한 측정치들이 낮은 신뢰도를 보일 수 있음을 제시한다고 볼 수 있다.

3. 연구 결과

3.1 집단 간 cepstral 측정치 분석

집단 간 cepstral 측정치에 차이가 있는지 살펴보기 위해 Mann-Whitney 검정을 실시한 결과, 집단 간 CPP, L/H ratio 값

에 유의한 차이가 있는 것으로 나타났으며 그 결과는 <표 2>와 같다. CPP는 식도발성화자 3.85dB, 정상노인 12.87dB, L/H ratio는 식도발성화자 21.72dB, 정상노인 33.55dB로 정상노인, 식도발성 순으로 높게 나타났다. 반면 mean CPP Fo는 통계적으로 유의한 차이는 없었지만 식도발성화자 158.83Hz, 정상노인 129.91Hz로 식도발성, 정상노인 순으로 높게 나타났다. 집단 간 cepstral 측정치들의 값을 <그림 2>에 도식화 하였다.

표 2. 집단 간 cepstral 측정치
Table 2. cepstral parameters between groups

매개 변수	집단	N	M±SD	평균 순위	U
CPP (dB)	식도발성	13	3.85±2.62	7.00	.00***
	정상노인	13	12.87±1.26	20.00	
L/H ratio (dB)	식도발성	13	21.72±8.87	8.77	23.00**
	정상노인	13	33.55±5.73	18.23	
mean CPP Fo (Hz)	식도발성	13	158.83±61.22	14.77	68.00
	정상노인	13	129.91±21.76	12.23	

*p<.05 **p<.01 ***p<.001

3.2 집단 간 spectral 측정치 분석

집단 간 spectral 측정치에 차이가 있는지 살펴보기 위해 Mann-Whitney 검정을 실시한 결과, 집단 간 jitter, shimmer, NHR, VTI 값에 유의한 차이가 있는 것으로 나타났으며 그 결과는 <표 3>과 같다. jitter는 식도발성화자 7.79%, 정상노인 .57%, shimmer는 식도발성화자 20.64%, 정상노인 13.66%, NHR은 식도발성화자 .56, 정상노인 .13, VTI는 식도발성화자 .18, 정상노인 .03로 식도발성, 정상노인 순으로 통계적으로 유의하게 높게 나타났다. 반면 Fo는 식도발성화자 120.95Hz, 정상노인 127.74Hz, SPI는 식도발성화자 11.43, 정상노인 14.42였으며 정상노인, 식도발성 순으로 통계적으로 유의한 차이는 없었지만 높게 나타났다. 집단 간 spectral 측정치들의 값을 <그림 3>에 도식화 하였다.

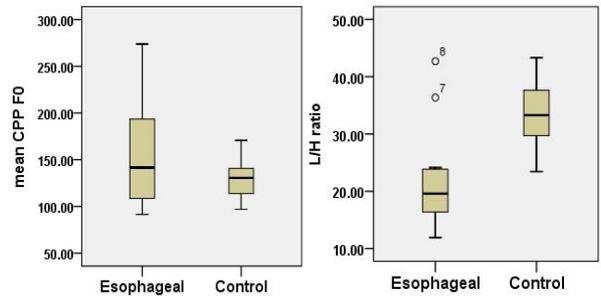
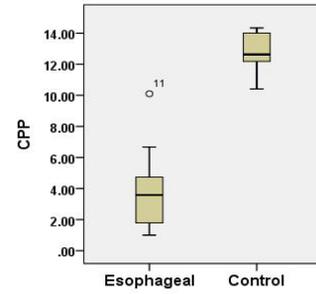


그림 2. 집단 간 cepstral 측정치
Figure 2. cepstral parameters between groups

표 3. 집단 간 spectral 측정치
Table 3. spectral parameters between groups

매개 변수	집단	N	M±SD	평균 순위	U
Fo (Hz)	식도발성	13	120.95±35.02	11.54	59.00
	정상노인	13	127.74±20.68	15.46	
jitter (%)	식도발성	13	7.79±3.82	20.00	.00***
	정상노인	13	.57±.27	7.00	
shimmer (%)	식도발성	13	20.64±6.48	20.00	.00***
	정상노인	13	3.66±.94	7.00	
NHR	식도발성	13	.56±.19	20.00	.00***
	정상노인	13	.13±.01	7.00	
VTI	식도발성	13	.18±.13	19.73	3.50***
	정상노인	13	.03±.01	7.27	
SPI	식도발성	13	11.43±7.67	11.77	62.00
	정상노인	13	14.42±7.44	15.23	

*p<.05 **p<.01 ***p<.001

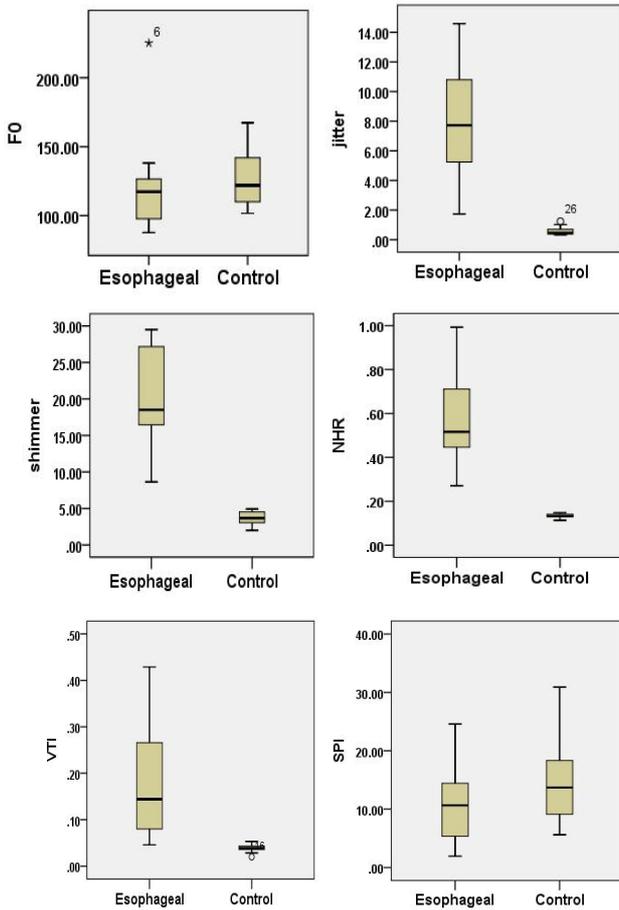


그림 3. 집단 간 spectral 측정치
Figure 3. spectral parameters between groups

3.3 집단의 cepstral 측정치와 spectral 측정치의 상관 분석

Spearman 상관분석을 통해 cepstral과 spectral 측정치의 상관성을 살펴본 결과, 측정치들 간에 통계적으로 유의한 상관

성이 있는 것으로 나타났으며 그 결과는 <표 4>와 같다.

측정치들 간의 상관성을 자세히 살펴보면 CPP와 jitter, shimmer, NHR, VTI 간에 통계적으로 유의하게 매우 높은 부적 상관관계가 있었으며(CPP-jitter, $\rho = -.826, p < .01$; CPP-shimmer, $\rho = -.899, p < .01$; CPP-NHR; $\rho = -.833, p < .01$; CPP-VTI, $\rho = -.677, p < .01$), 특히 CPP와 shimmer는 매우 높은 부적 상관관계를 보였다. 또한 L/H ratio는 shimmer, VTI와 통계적으로 유의하게 높거나 매우 높은 부적 상관관계를 보였다(L/H ratio-shimmer, $\rho = -.482, p < .05$; L/H ratio-NHR, $\rho = -.596, p < .01$). 마지막으로 mean CPP Fo는 Fo, shimmer, NHR과 통계적으로 유의하게 높거나 매우 높은 정적 상관관계를 보였다(mean CPP Fo-Fo, $\rho = .587, p < .01$; mean CPP Fo-shimmer, $\rho = .440, p < .05$; mean CPP Fo-NHR, $\rho = .421, p < .05$).

3.4 cepstral 측정치의 ROC 곡선 분석

cepstral 측정치가 식도발성화자의 음질문제에 대한 감별 진단 가능성이 있는지 살펴보고자 ROC 곡선 분석을 실시한 결과는 <표 5>, <그림 4>와 같다.

측정치들의 ROC 곡선 아래의 면적(area under curve, AUC) 값을 살펴보면, CPP는 1.000, L/H ratio는 .864, mean CPP Fo는 .402으로 CPP가 식도발성화자의 음질문제의 감별 진단에 가능성이 매우 높은 유용한 변수임을 살펴볼 수 있었다. 이때 각 측정치의 Youden Index를 구하여 기준점(cut off point)을 정한 뒤 민감도와 특이도를 살펴본 결과, Youden Index는 CPP 1.000, L/H ratio .769, mean CPP Fo는 .154로 CPP의 값이 가장 높았다. 또한 CPP의 민감도와 특이도는 각 1.000로 식도발성화자의 음질을 감별할 수 있는 최적 여치 값으로 ≤ 10.25 를 사용할 수 있음을 확인하였다. 또한 L/H ratio의 민감도는 .923, 특이도는 .846이었으며, mean CPP Fo의 민감도는 .846,

표 4. 집단의 cepstral 측정치와 spectral 측정치의 상관관계
Table 4. Correlation of cepstral and spectral parameters in esophageal and control groups

	L/H ratio	mean CPP Fo	Fo	jitter	shimmer	NHR	VTI	SPI
CPP	.601**	-.533**	-.163	-.826**	-.899**	-.833**	-.677**	.275
L/H ratio		-.055	.269	-.362	-.482*	-.381	-.596**	.312
mean CPP Fo			.587**	.302	.440*	.421*	.157	-.130
Fo				.064	.230	.373	.093	-.228
jitter					.913**	.866**	.530**	-.253
shimmer						.950**	.653**	-.277
NHR							.570**	-.286
VTI								-.507**

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

특이도는 .307이었다.

표 5. cepstral 측정치의 ROC 곡선 분석
Table 5. ROC curve analysis of cepstral parameters

	AUC	Std. Error	Sig.	95% CI	
				Lower	Upper
CPP	1.000	.000	.000	1.000	1.000
L/H ratio	.864	.084	.002	.699	1.000
mean CPP Fo	.402	.119	.397	.168	.636

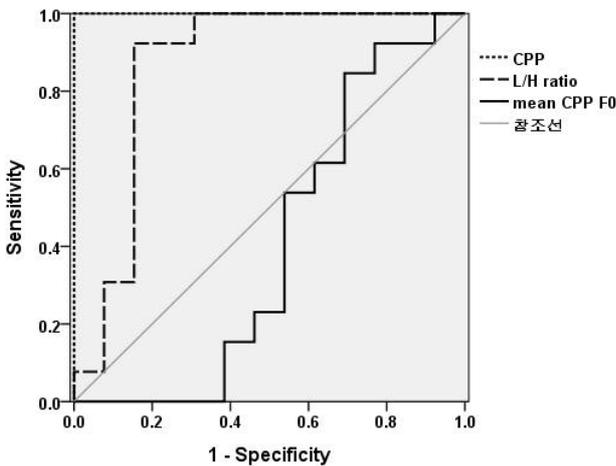


그림 4. cepstral 측정치의 ROC 곡선
Figure 4. ROC curve of cepstral parameters

4. 논의 및 결론

본 연구는 캡스트럼 분석을 통해 식도발성화자의 음성 특성에 대해 다음과 같은 결론들을 도출할 수 있었다.

첫째, CPP 값은 식도발성화자 3.85±2.62dB, 정상노인 12.87±1.26dB로 정상노인이 높았다. CPP는 조화음의 정도를 나타내는 매개변수이므로 정상 음성은 주기적 신호에서 조화음 구조가 잘 나타나기 때문에 높은 값을 나타낸 것으로 볼 수 있다[15]. 반면 식도발성화자가 발성에 사용하는 인두-식도 부위(PES)는 근본적인 발성기관이 아니므로 성대만큼의 진동수 확보에 어려움이 있기 때문에 비주기적인 음성 특성으로 조화음 구조가 잘 정의되지 않아[3] 낮은 CPP 값을 나타낸 것으로 생각할 수 있겠다. 이러한 식도발성화자의 비주기적인 음성 특성은 높은 SD 값을 통해서도 살펴볼 수 있으며 본 연구의 CPP SD 값은 식도발성화자 2.62dB, 정상노인 1.26dB로 식도발성화자가 더 높았다. 높은 CPP SD 값은 음성문제의 중증도와 관련성이 높기 때문에[17] 음질 문제가 심각한 식도발

성화자의 CPP SD 값이 더 크게 나타난 것으로 사료된다.

둘째, L/H ratio 값은 식도발성화자 21.72±8.87dB, 정상노인 33.55±5.73dB로 정상노인이 더 높았다. 이는 4kHz 이하의 에너지가 식도발성화자보다 정상노인이 더 높음을 의미하는데 본 연구과제인 /a/모음은 F1과 F2가 1-2kHz 사이에 위치하기 때문에[21] 더 큰 L/H ratio 값의 차이를 나타냈다고 볼 수 있다. 또한 스펙트럼 에너지 분포 측면에서 청지각적으로 좋은 음성이라고 판단되는 경우 3-4kHz의 주파수 대역에서 강한 에너지를 보이고[22], 특히 기식성(breathiness)을 갖는 음성은 고주파영역 에너지가 낮고, 조조성(roughness) 음성은 고주파영역 에너지가 높기 때문에[23], 이러한 음질 특성을 모두 보이는 식도발성화자의 경우 주파수 에너지가 동시에 영향을 주므로 정상노인의 L/H ratio와 유의한 차이를 보인 것으로 사료된다. 이는 L/H ratio가 정상인과 후두 및 신경 질환환자의 기식성 및 조조성과 관련성이 높으며, 전반적인 음질 문제의 중증도, 치료 후의 음성 변화 측정을 확인할 수 있는 측정치임을 알 수 있다[24, 25, 26].

셋째, 집단 간 mean CPP Fo 값의 차이를 비교한 결과 식도발성화자 158.83±61.22Hz, 정상 노인 129.91±21.76Hz로 식도발성화자가 더 높은 mean CPP Fo 값을 보였다[2, 7]. 일반적으로 mean CPP Fo는 가장 큰 조화음 에너지를 가지는 CPP의 주파수로 기본주파수(Fo)를 나타내기 때문에[18] 정상 노인보다 더 높은 Fo를 보인 점은 식도발성화자의 음질 특성을 통해 논의할 수 있겠다. 조조성이 높은 음성과 정상 음성을 비교한 [14]는 조조성이 높은 화자의 CPP Fo가 227.61±23.58Hz로, 정상 음성 178.40±24.21Hz에 비해 매우 높게 측정됨을 보고하였으며, 이는 CPP Fo가 꺾어짜는 음성에서 증가된 패턴을 보인다고 하였다. 이러한 양상은 꺾어짜는 음질을 보이는 근긴장 발성장애의 경우에서도 높은 Fo 값을 보이므로 인두-식도 부위의 내전을 통해 조조성 음질 특성이 두드러지는 식도발성에서 또한 높은 CPP Fo를 보인 것으로 생각할 수 있겠다.

넷째, jitter, shimmer, NHR, VTI는 식도발성화자가 높은 수치를 보였으며 Fo와 SPI는 정상성인이 높은 수치를 보였다. 이는 식도발성화자의 신성대의 역할을 하는 PES가 근본적인 발성기관이 아니므로 성대만큼의 진동수 확보에 어려움이 있기 때문에 식도발성화자의 심각한 음질 문제로 인해 jitter, shimmer, NHR 수치가 높은 양상을 보인다는 [2, 4, 5]와 동일한 결과를 나타냈다. 이러한 결과는 식도발성화자의 비주기적이고 가변적인 음성 특성 때문에 음도 및 강도 조절에 어려움을 보이며[3, 10], 안정된 음성을 산출하기 위해 필요한 많은 양의 공기 확보에 어려움을 보이기 때문에 정확한 Fo 값을 반영하는 잡음 변수들의 수치가 높게 나타난 것으로 생각할 수 있겠다[7]. 또한 임상적 관찰 측면에서 대상자들의 spectral 측정치들의 SD 값이 높고, 평가자내 신뢰도가 낮은 양상을 보였으며 이는 식도발성화자처럼 음질이 나쁜 음성의 경우 섭동분

석에 기초한 spectral 측정치들이 낮은 신뢰도를 보일 수 있을 것을 제시한다. 또한 잡음 관련 변수인 SPI(soft phonation index)는 통계적으로 유의하지는 않았지만 식도발성화자에 비해 정상노인이 높은 수치를 보인 것은 노화로 인해 성대의 폐쇄가 불완전하고 약한 음성과 관련성을 보였기 때문으로 사료된다.

넷째, cepstral 측정치와 spectral 측정치들 간의 상관관계를 살펴본 결과, CPP와 jitter, shimmer, NHR, VTI 간에 매우 높은 부적 상관관계를 보였다. 이는 CPP가 잡음 변수 NHR, 강도 변수 APQ, 음도 변수 RAP, sPPQ와 높은 상관성을 보였으며[15], CPP가 음도, 강도, 잡음 관련 변수 모두와 상관성을 보인 본 연구 결과와 동일한 양상을 보였다. CPP는 조화음의 정도를 나타내는 매개변수로 주기적인 신호일 경우 조화음 구조가 잘 나타나게 된다. 이때 jitter, shimmer, NHR, VTI는 기본 주기에 기초한 변수들이기 때문에[20] 모두 CPP와 높은 상관성을 보였다고 생각할 수 있겠다. 또한 L/H ratio는 shimmer, VTI와 통계적으로 유의하게 높거나 매우 높은 부적 상관관계를 보였다. L/H ratio는 4kHz 이하와 이상의 에너지 비율을 측정할 값이며, VTI(voice turbulence index)는 70-4500Hz 대역의 배음과 2.8-5.8kHz 대역의 잡음과의 비율을 나타낸 측정치이므로 컷오프 주파수 대역이 유사하며, 또한 VTI가 잡음 관련 변수이자 거친 음성과 관련성이 높기 때문에 두 측정치간의 상관성이 높게 나타난 것으로 사료된다. 마지막으로 mean CPP Fo와 Fo의 높은 상관관계는 기본 주파수 분석에 있어서 조화음의 정도를 나타내는 cepstral 측정치가 spectral 측정치 Fo만큼 유용하다고 볼 수 있음을 시사한다[11, 17, 20].

다섯째, cepstral 측정치의 ROC 곡선 분석 결과, CPP의 민감도와 특이도 1.000, AUC 1.000, L/H ratio의 민감도 .923, 특이도 .846로 식도발성화자의 음질문제의 감별 진단에 가능성이 높은 유용한 변수임을 알 수 있었다. AUC가 1.000이라는 것은 그 검사의 모든 가능한 컷오프에 대하여 100%의 정확성을 갖는다는 것을 의미하기 때문에 CPP가 식도발성화자 음성의 객관적인 평가측정치로도 유용함을 보여준다. [11, 13, 15, 17, 20]과 같은 선행연구들에서도 CPP가 정상음성과 후두관련 음성장애의 변별을 위한 측정치로 유용할 뿐만 아니라, 음성장애의 중증도, 기식성과 조조성과 같은 청지각적인 특성에도 높은 상관관계를 보여 장애음성의 분석에 cepstral 분석의 효용성을 살펴볼 수 있었다. 또한 L/H ratio는 기식성과 높은 상관관계가 있고 민감도와 특이도 또한 높게 보고되어[13] 음성문제의 변별을 위한 측정치로서 유용하다.

결론적으로 캡스트럼 분석을 통한 식도발성화자의 음질 특징을 종합해보면, 식도발성화자가 정상노인에 비해 CPP, L/H ratio 값이 낮고, mean CPP Fo 값은 높은 양상을 보였으며, 이것은 식도발성화자의 음성이 조화음의 정도가 낮고, 기식성과 조조성으로 인한 고주파수 에너지 비율이 높으며, 비주기적인 특성을 보인 결과라고 볼 수 있겠다. 또한 CPP와 기존의 음성

측정치들과의 상관성을 알아보고, ROC 곡선을 통해 감별 진단 가능성을 살펴봄으로써 CPP가 식도발성화자 음성의 객관적인 평가측정치로도 유용함을 알 수 있었다. 지금까지 국내외적으로 식도발성화자 음성에 캡스트럼 분석을 적용한 연구는 전무했으며 스펙트럼 분석 측정치와의 상관성과 유용성을 함께 살펴본 경우는 본 연구가 유일하다. 따라서 본 연구를 통해 가변적인 식도발성화자의 음성분석에 캡스트럼 측정치의 활용 가능성과 유용성을 살펴볼 수 있었으며, 이는 후두적출자 음성의 다방면적인 이해와 음성재활접근 및 효율적인 임상적 중재 방안에 기초가 되는 자료 확립에 기여할 것이다.

참고문헌

[1] Website, <http://www.cancer.go.kr/>, September 6, 2013.

[2] Jang, H. R., Shim, H. J., & Ko, D. H. (2013). Acoustic Characteristics of Patients with Total Laryngectomees via Voice Rehabilitation Techniques. *Phonetics and Speech Sciences*, 5(4), 25-32.
(장효령, 심희정, 고도홍 (2013). 후두적출술 환자의 발성법에 따른 음향학적 특성. *말소리와 음성과학*, 5(4), 25-32.)

[3] Debruyne, F., Delaere, P., Wouters, J., & Uwents, P. (1994). Acoustic analysis of tracheo-oesophageal versus oesophageal speech. *Journal of Laryngology and Otology*, 108, 325-328.

[4] Most, T., Tobin, Y., & Mimran, R. C. (2000). Acoustical and perceptual characteristics of esophageal tracheoesophageal speech production. *Journal of Communication Disorder*, 33(2), 165-181.

[5] Choi, S. H., Choi, H. S., Kim, H. S., Lee, S. E., & Pyo, H. Y. (2003). Analysis of acoustic characteristics of vowel and consonants production study on speech proficiency in esophageal speech. *Speech Sciences*, 10(3), 7-27.
(최성희, 최홍식, 김한수, 임성은, 이성은, 표화영 (2003). 식도발성의 숙련 정도에 따른 모음의 음향학적 특징과 자음 산출에 대한 연구. *음성과학*, 10(3), 7-27.)

[6] Kang, S. K. Aerodynamic analysis of esophageal speech of the laryngectomees. Ph.D. dissertation Taegu University.
(강수균 (2002). 후두적출자의 식도발성시 기류역학적 연구, 대구대학교 박사학위 논문.)

[7] Pyo, H., Choi, H., Lim, S., & Choi, S. (1999). The comparison of the acoustic and aerodynamic characteristics of provox voice and esophageal voice produced by the same laryngectomee. *Speech Sciences*, 5(1), 121-139.
(표화영, 최홍식, 임성은, 최성희. (1999). 동일 후적자가 산출하는 기관식도 발성과 식도 발성에 대한 음향학적 및 공기역학적 특성 비교. *음성과학*, 5(1), 121-139.)

- [8] Blood, G. W. (1984). Fundamental frequency and intensity measurements in laryngeal and alaryngeal speakers. *Journal of Communication Disorders*, 17, 319-324.
- [9] Štajner-Katušić, S., Horga, D., Mušura, M., & Globlek, D. (2006). Voice and speech after laryngectomy. *Clinical Linguistics & Phonetics*, 20(2/3), 195-203.
- [10] MacCallum, J., Cai, L., Zhou, L., Zhang, Y., & Jiang, J. (2007). Acoustic analysis of aperiodic voice: perturbation and nonlinear dynamic properties in esophageal phonation. *Journal of Voice*, 23(3), 283-290.
- [11] Moers, C., Mobius, B., Rosanowski, F., Noth, E., Eysholdt, U., & Haderlein, T. (2012). Vowel- and text-based cepstral analysis of chronic hoarseness. *Journal of Voice*, 26(4), 416-424.
- [12] Noll, A. M. (1967). Cepstrum pitch determination. *Journal of the Acoustic Society of America*, 41, 293-309.
- [13] Watts, C. R., & Awan, S. N. (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.
- [14] 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 and Laryngology*, 121(8), 539-548.
- [15] Heman-Ackah, T., Michael, D., & Goding, G. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, 16(1), 20-27.
- [16] Kumar, B., Bhat, J., & Prasad, N. (2009). Cepstral analysis of voice in persons with vocal nodules. *Journal of Voice*, 24(6), 651-653.
- [17] Awan, S., Roy, N., & Dromey, C. (2009). Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. *Clinical linguistics & phonetics*, 23(11), 825-841.
- [18] Hillenbrand, J., Cleveland, R., & Erickson, R. (1994). Acoustic Correlates of Breathy Vocal Quality. *Journal of Speech, Language, and Hearing Research*, 37, 769-778.
- [19] Seo, I., & Seong, C. (2013). Voice quality of dysarthric speakers in connected speech. *Phonetics and Speech Sciences*, 5(4), 33-41.
(서인효, 성철재 (2013). 연결발화에서 마비말화자의 음질 특성. *말소리와 음성과학*, 5(4), 33-41.)
- [20] Wolfe, V, Fitch J, & Cornell R. (1995). Acoustic prediction of severity in commonly occurring voice problems. *Journal of Speech, Language & Hearing Research*, 38, 273 - 279.
- [21] Shim, H., Jang, H., & Ko, D. (2013). Speech Intelligibility and Vowel Space Characteristics of Alaryngeal Speech. *Phonetics and Speech Sciences*, 5(4), 17-24.
(심희정, 장효령, 고도홍 (2013). 무후두음성의 말 명료도와 모음 공간 특성. *말소리와 음성과학*, 5(4), 17-24.)
- [22] Timo, L. (2009). Long term average spectrum in screening of voice quality in speech: untrained male university students. *Journal of Voice*, Vol. 23, No. 6, 671-676.
- [23] Kent, R. D., & Read, C. (2002). *Acoustic Analysis of speech*(2nd ed). Singular Thomson Learning.
- [24] Hillenbrand, J., & Houde, R. (1996). Acoustic correlates of breathy vocal quality dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39, 311-321.
- [25] Awan, S., & Roy, N. (2009). Outcomes measurement in voice disorders: application of an acoustic index of dysphonia severity. *Journal of Speech, Language and Hearing Research*, 52(2), 482-499.
- [26] Lee, M. (2011). A Study of Acoustic Measurement in Connected Speech with Dysphonia. *Phonetics and Speech Sciences*, 3(4), 109-115.
(이명순 (2011). 음성장애에 연속구어의 음향학적 분석. *말소리와 음성과학*, 3(4), 109-115.)
- **심희정 (Shim, Hee-Jeong)**
한림대학교 대학원 언어청각학과
강원도 춘천시 한림대학길 1
Tel: 033-248-2227
Email: amy2020@hallym.ac.kr
관심 분야: 음성학, 음성장애
 - **장효령 (Jang, Hyo-Ryung)**
한림대학교 대학원 언어청각학과
강원도 춘천시 한림대학길 1
Tel: 033-248-2227
Email: gyufd0601@hanmail.net
관심 분야: 음성학, 음성장애
 - **신희백 (Shin, Hee-Back)**
한림대학교 대학원 언어청각학과
강원도 춘천시 한림대학길 1
Tel: 033-248-2227
Email: yj2000102@nate.com
관심 분야: 음성학, 음성장애
 - **고도홍 (Ko, Do-Heung)**
한림대학교 언어청각학부 교수
강원도 춘천시 한림대학길 1
Tel: 033-248-2212
Email: dhko7@hallym.ac.kr
관심 분야: 음성과학, 음성장애