

발성장애 평가 시 /a/ 모음연장발성 및 문장검사의 캡스트럼 분석 비교

성균관대학교 의과대학 강북삼성병원 이비인후과학교실
김태환 · 최정임 · 이상혁 · 진성민

= Abstract =

Comparison of Vowel and Text-Based Cepstral Analysis in Dysphonia Evaluation

Tae Hwan Kim, MD, Jeong Im Choi, SLP, Sang Hyuk Lee, MD, PhD and Sung Min Jin, MD, PhD

Department of Otorhinolaryngology-Head and Neck Surgery, Kangbuk Samsung Hospital,
Sungkyunkwan University School of Medicine, Seoul, Korea

Background : Cepstral analysis which is obtained from Fourier transformation of spectrum has been known to be effective indicator to analyze the voice disorder. To evaluate the voice disorder, phonation of sustained vowel /a/ sound or continuous speech have been used but the former was limited to capture hoarseness properly. This study is aimed to compare the effectiveness in analysis of cepstrum between the sustained vowel /a/ sound and continuous speech. **Methods** : From March 2012 to December 2014, total 72 patients was enrolled in this study, including 24 unilateral vocal cord palsy, vocal nodule and vocal polyp patients, respectively. The entire patient evaluated their voice quality by VHI (Voice Handicap Index) before and after treatment. Phonation of sustained vowel /a/ sample and continuous speech using the first sentence of autumn paragraph was subjected by cepstral analysis and compare the pre-treatment group and post-treatment group. **Results** : The measured values of pre and post treatment in CPP-a (cepstral peak prominence in /a/ vowel sound) was 13.80, 13.91 in vocal cord palsy, 16.62, 17.99 in vocal cord nodule, 14.19, 18.50 in vocal cord polyp respectively. Values of CPP-s (cepstral peak prominence in text-based speech) in pre and post treatment was 11.11, 12.09 in vocal cord palsy, 12.11, 14.09 in vocal cord nodule, 12.63, 14.17 in vocal cord polyp. All 72 patients showed subjective improvement in VHI after treatment. CPP-a showed statistical improvement only in vocal polyp group, but CPP-s showed statistical improvement in all three groups ($p < 0.05$). **Conclusion** : In analysis of cepstrum, text based analysis is more representative in voice disorder than vowel sound speech. So when the acoustic analysis of voice by cepstrum, both phonation of sustained vowel /a/ sound and text based speech should be performed to obtain more accurate result.

KEY WORDS : Cepstrum · Dysphonia · Acoustic analysis · Voice.

서 론

음성장애(voice disorder)는 평균적으로 일생 동안 50%의 인구에서 경험할 정도의 높은 유병률을 가지고 있으며¹⁾ 음성 장애가 만성화 되었을 경우에는 개인에게 심리적, 사회적으로 부정적인 영향을 줄 수 있다.²⁾ 이러한 음성장애는 만성 후두

염, 성대의 양성질환 및 악성종양 등의 다양한 원인에 의해서 생길 수 있으며 이를 정확하게 진단하고 정량화 하기 위하여 여러 가지 음향 지표들이 사용된다.

흔히 사용되는 음성 분석법으로는 주파수와의 연관성을 나타내는 jitter, 진폭과의 연관성을 나타내는 shimmer, 잡음 대 배음비(noise-to-harmonic ratio) 등과 같은 음향지표가 사용되며 경한 발성장애인 경우 유용하게 사용될 수 있지만 중등도 이상의 발성장애의 경우 상기 음향지표들로는 기본주파수(fundamental frequency)를 추적하기 힘들기 때문에 검사의 신뢰도가 떨어진다는 단점이 있다.³⁾ 이러한 단점을 보완하기 위해 최근에 발성장애(dysphonia)를 평가하는데 있어서 캡스트럼(cepstrum)이 소개되어 왔으며 여러 연구에서 그 유

논문접수일 : 2015년 11월 15일
심사완료일 : 2015년 11월 21일
책임저자 : 이상혁, 03181 서울 중로구 새문안로 29
성균관대학교 의과대학 강북삼성병원 이비인후과학교실
전화 : (02) 2001-2269 · 전송 : (02) 2001-2273
E-mail : entlsh@hanmail.net

용성이 증명되었다.³⁻⁵⁾ 캡스트럼은 기본주파수의 영향을 많이 받는 이전 검사들과는 달리 음성신호의 푸리에 변형(Fourier transformation) 과정을 시행하여 얻어진 기저선(baseline)을 중심으로 값을 측정하기 때문에 중등도 이상의 발성장애에서도 신뢰할만한 결과 값을 얻을 수 있다.⁶⁾ 캡스트럼을 이용한 음성 분석에는 Hillenbrand가 개발한 Cepstral Peak Prominence(CPP)와 Computerized Speech Laboratory(CSL) 프로그램에서 분석하는 Cepstral Mean Value(CMV) 방법이 있지만 CPP를 이용한 음성분석이 보다 신뢰도가 높다.⁷⁾

일반적으로 발성장애를 평가하기 위한 음성샘플 녹음은 /a/ 모음연장발성 및 문장검사, 문단검사 등으로 평가를 하게 된다. 임상에서는 모음연장발성을 가장 많이 사용하는데 이는 개인의 발성습관 및 말의 속도 등에 대해서 비교적 빠르고 객관적인 정보를 얻을 수 있다는 장점이 있기 때문이다.⁸⁾ 하지만 /a/ 모음연장발성만을 이용한 평가는 실제 대화할 때의 동적인 측면을 반영하는데 있어서 충분하지 않으며 기식화된 음성을 분석하는 것에 있어 문장검사보다는 민감도가 떨어진다는 보고들이 있어왔다.^{8,9)} 반면 문장검사는 일상생활에서의 목소리 평가를 하는데 있어서 더 현실적이며 음성의 시작 및 종결, 기본 주파수, 음의 강도 등의 변이에 대해서 더 효과적으로 평가할 수 있다.^{9,10)}

모음연장발성과 문장검사는 음성평가를 하는데 있어서 각각의 장점을 가지고 있으나 현재까지 이 두 검사간의 유용성에 대한 직접적인 연구는 많지 않으며 특히 캡스트럼 검사에서 두 가지 방법에 따른 검사의 효용성은 아직 많은 연구가 이루어 지지 못하였다. 따라서 저자들은 본 연구를 통하여 발성장애를 보이는 일측 성대마비, 성대결절, 성대 폴립 환자에서 캡스트럼 분석 시 /a/ 모음연장 발성 및 문장검사가 기식적 음성을 얼마나 객관적이고 정량적으로 평가하는지에 대해 비교 분석하고자 하였다.

대상 및 방법

1. 환자군

2012년 3월부터 2014년 12월까지 강북삼성병원 이비인후과에서 편측 성대마비, 성대결절, 성대 폴립으로 진단받은

환자군 각각 24씩 총 72명의 환자가 본 연구에 포함되었다. 성대마비와 성대결절 및 폴립의 진단은 후두 내시경과 스트로보스코피 검사를 통하여 이비인후과 전문의가 진단하였다. 또한 이전에 현수후두하 미세수술 및 성대 내 주입술 시행 받은 과거력 있거나 약물을 복용중인 환자는 대상에서 제외하였다. 평균 연령은 성대마비가 47.9세(16~81세), 성대결절은 42.1세(17~68세), 성대폴립은 45.6세(23~75세)였으며 평균적으로 증상을 호소하였던 기간은 성대마비가 8.7개월, 성대결절이 10.1개월, 성대폴립이 8.8개월이었다. 남녀 비는 성대마비가 남녀 12명씩, 성대결절과 성대폴립 모두 남자 11명 여자 13명으로 세 군에서 유의한 차이는 없었다(Table 1). 환자군의 주관적인 음성의 질 평가를 위해 VHI(voice handicap index)를 수술 전 모든 환자에게 시행하였으며 치료 후 한달 뒤 특이 합병증 없는 환자에 한해서 VHI 재평가를 실시하여 목소리 호전에 대한 주관적인 평가를 실시하였다. 본 연구는 강북삼성병원 임상시험심사위원회의 검토와 승인을 받았고 모든 환자의 동의 하에 진행되었다.

2. 캡스트럼 분석

음향학적인 분석을 시행할 때 음성 샘플의 녹음은 아무도 없는 조용한 방에서 약 10 cm의 거리를 둔 마이크로폰을 사용하였고, 평상시의 편안한 크기와 속도로 /a/모음 연장 발성을 약 5초, '가을'문단 중에서 '우리나라의 가을하늘은 매우 아름답다'의 문장을 약 4초간 사용하여 녹음하였다. 대상자가 일상 생활에서의 발성과 다르다고 생각할 때는 음성 녹음을 재시도 하였고, 2회 반복 녹음하여 그 중 평상시 음성에 가까운 것을 골랐으며 스펙트럼상에서 가장 안정적인 2초의 연속된 구간을 선택하여 분석하였다. 문장검사를 할 때 환자가 웃거나 기침을 하는 부분은 모두 제거되었다.

캡스트럼은 Hillenbrand가 개발한 CPP 방법을 이용하였으며 CPP는 캡스트럼에서 가장 높은 지점으로 정의하였고, 캡스트럼에서 평균값의 높이로 나타나는 회귀선과 그래프의 가장 높은 지점을 평가하여 그 차이를 계산하였다. Hillenbrand 방법은 홈페이지(<http://homepages.wmich.edu/~hillenbr/>) 접속을 통하여 CPP값을 구하였다. 또한 전반적인 음성평가를 위해 음성 분석 프로그램인 CSL 중 Multiple Dimensional Voice Program(MDVP)를 이용하여 jitter, shimmer, 잡음 대

Table 1. Demographic characteristics of patient data

	VC Palsy (24)	VC Nodule (24)	VC polyp (24)
Sex (M/F)	12/12	11/13	11/13
Age (Mean, year)	47.9	42.1	45.6
Age (Min/Max)	16/81	17/68	23/75
Sx duration (Month)	8.7	10.1	8.8

VC : vocal cord

배음비를 측정하였다.

통계학적 검정은 세 그룹간의 치료 전과 후에 대해서 paired samples t-test를 사용하였으며 각 그룹간에는 one-way ANOVA를 사용하였다. 통계프로그램은 SPSS version 18.0 (IBM Co. New York, NY, USA)를 이용하였다. p-value 값이 0.05 이하인 경우는 통계적으로 유의하다고 판정하였다.

결 과

치료 전 후에 시행한 VHI(voice handicap index)검사상 환자 군들은 주관적인 음성 호전을 보였다. 또한 성대마비, 성대결절, 성대폴립 환자들 각각 24명을 대상으로 켈스트럼을 포함한 음성분석 실시하였으며 켈스트럼 분석을 통한 CPP 값은 치료전의 경우 /a/ 모음연장발성에서 일측 성대 마비 군에서 13.80(±4.69, M=13.04, F=14.80), 성대결절 군에서 16.62(±3.20, M=17.35, F=16.17), 성대폴립 군에서 14.19(±2.46, M=14.48, F=13.86)으로 측정되었으며 치료 후에는 일측 성대마비 군에서 13.91(±4.41, M=13.99, F=13.81), 성대결절 군에서 17.99(±3.28, M=20.03, F=16.73), 성대폴립 군에서 18.50(±2.34, M=19.66, F=17.22)으로 측정되어 성대폴립 군에서만 통계적으로 유의한 호전을 보였다(Table 2).

반면 치료 전의 문장검사에서는 일측 성대 마비 군에서 11.11(±1.64, M=10.80, F=11.54), 성대결절 군에서 12.11(±1.50, M=11.39, F=12.55), 성대폴립 군에서 12.63(±1.23,

M=12.42, F=12.86)으로 측정되었으며 치료 후 일측 성대 마비 군에서 12.09(±2.27, M=11.95, F=12.26), 성대결절 군에서 14.09(±2.01, M=14.57, F=13.79), 성대폴립 군에서 14.17(±1.15, M=14.31, F=14.01)으로 측정되었으며 통계적으로 세 군에서 모두 유의한 치료효과를 보였다(Table 2).

요약하자면 환자 군들은 음성의 주관적인 호전을 보이고 있으나 CPP를 이용한 켈스트럼 검사법에서 /a/ 모음연장발성은 성대폴립에서만 유의한 호전을 보이는 반면 문장검사는 세 그룹에서 유의한 호전을 보이고 있어 문장검사가 기식화된 음성을 더 잘 대변한다고 볼 수 있었다. 또한 일반적으로 많이 사용되는 음성분석지표와의 신뢰도를 비교하기 위하여 치료 전후의 jitter, shimmer, 잡음 대 배음역비(noise to hearing ratio) 값도 분석하였으며 성대결절과 성대폴립에서만 통계적으로 유의한 치료결과를 보였다(Table 3).

고 찰

기식성 음성을 유발하는 후두질환들의 음성을 보다 정확하고 객관적으로 분석하기 위해 보편적으로 사용되는 방법은 jitter, shimmer, 잡음 대 배음비(noise-to-harmonic ratio)가 있다. 하지만 이러한 방법들은 기본주파수 변화에 많은 영향을 받기 때문에 음성이 매우 기식적이거나(breathy) 기본주파수를 추적하기 힘든 경우에는 주파수의 작은 변화에도 변수들의 측정치가 크게 변할 수 있으며³⁾ Carding 등의 보

Table 2. Pre and Post-treatment cepstral analysis with sustained vowel /a/ sample and text-based continuous speech

	VC palsy		VC nodule		VC polyp	
Pre CPP-a	13.80 (4.69)	M : 13.04 (4.79) F : 14.80 (4.63)	16.62 (3.20)	M : 17.35 (2.25) F : 16.17 (3.68)	14.19 (2.46)	M : 14.48 (1.57) F : 13.86 (3.24)
Post CPP-a	13.91 (4.41)	M : 13.99 (4.59) F : 13.81 (4.43)	17.99 (3.28)	M : 20.03 (2.15) F : 16.73 (3.28)	18.50 (2.34)*	M : 19.66 (1.98) F : 17.22 (2.08)
Pre CPP-s	11.11 (1.64)	M : 10.80 (1.78) F : 11.54 (1.44)	12.11 (1.50)	M : 11.39 (1.29) F : 12.55 (1.50)	12.63 (1.23)	M : 12.42 (0.99) F : 12.86 (1.48)
Post CPP-s	12.09 (2.27)*	M : 11.95 (2.45) F : 12.26 (2.14)	14.09 (2.01)*	M : 14.57 (2.06) F : 13.79 (1.99)	14.17 (1.15)*	M : 14.31 (1.13) F : 14.01 (1.20)

* : p<0.05. VC : vocal cord, CPP-a : Cepstral Peak Prominence in /a/ vowel sound, CPP-s : cepstral peak prominence in text-based speech

Table 3. Pre and Post treatment voice analysis (sustained vowel /a/ sample)

	VC palsy		VC nodule*		VC polyp*	
	Pre	Post	Pre	Post	Pre	Post
Jitter	5.36 (7.00)	4.00 (5.06)	2.57 (1.73)	1.59 (1.31)	2.72 (1.17)	0.96 (0.67)
Shimmer	9.58 (8.21)	8.04 (8.27)	6.55 (4.49)	3.54 (1.46)	6.03 (1.99)	2.86 (1.35)
NHR	0.25 (0.30)	0.26 (0.31)	0.15 (0.03)	0.12 (0.02)	0.18 (0.08)	0.12 (0.03)

* : p<0.05

고에 따르면 심한 발성장애 환자에서 20%정도가 상기의 검사들에 오류가 있었다.¹¹⁾ 그렇기 때문에 기본주파수에 영향 받지 않는 객관적인 음성검사방법의 필요성이 대두되어 왔다.

켈스트럼은 음성신호의 푸리에 변형 과정을 두 번 시행하여 얻어지는 검사 값으로 그래프에서 컴퓨터 프로그램을 통해 잡음의 기저선(baseline)의 평균값이 측정된다. 켈스트럼은 이처럼 기저선을 중심으로 값을 구하기 때문에 음파의 기본주파수나 진폭을 측정하여 평가하는 jitter, shimmer와는 측정 방법이 다르며¹²⁾ 때문에 심한 발성장애환자에서 켈스트럼 분석을 통해 배음구조(harmonic structure)의 정도를 보는 것이 발성평가에 있어 기존의 음성분석법에 비해 객관적인 검사 결과를 얻을 수 있다.¹³⁾

켈스트럼을 이용한 음성분석에는 Hillenbrand에 의해 만들어진 컴퓨터 프로그램으로 측정하는 Cepstral peak prominence(CPP)를 이용하는 방법과 CSL 프로그램을 사용하여 Cepstral mean value(CMV) 값을 계산하여 이용하는 방법이 있다. 이 두 가지 방법은 임상적으로 켈스트럼 분석을 하는데 자주 사용되는 방법이지만 CPP값이 CMV값에 비해 통계적으로 유의하다는 보고가 있다.³⁾ CPP값은 켈스트럼 그래프에서 얻어진 cepstral peak와 켈스트럼의 소리 평균을 나타내는 회귀선(regression line) 사이의 거리를 의미하는 것으로 (Fig. 1) 기식성 음성이 심할수록 낮은 값의 CPP가 측정된다.¹³⁾

음향분석을 시행할 때 음성 샘플의 녹음은 일반적으로 /a/ 발음과 같은 모음연장발성을 하는 방법과 문장 및 문단을 읽으면서 녹음하는 방법이 있는데 음성장애를 평가할 때 어떤 방법이 더 신뢰도가 있는지에 대한 연구들이 진행되어 왔다.⁸⁾ 일반적으로 모음연장발성이 검사상의 편의성 때문에 음성검사에서 많이 사용되고 있지만 음성 평가를 하는데 있어서 일상적인 대화를 할 때 음성의 질을 평가하는데 한계가 있다.

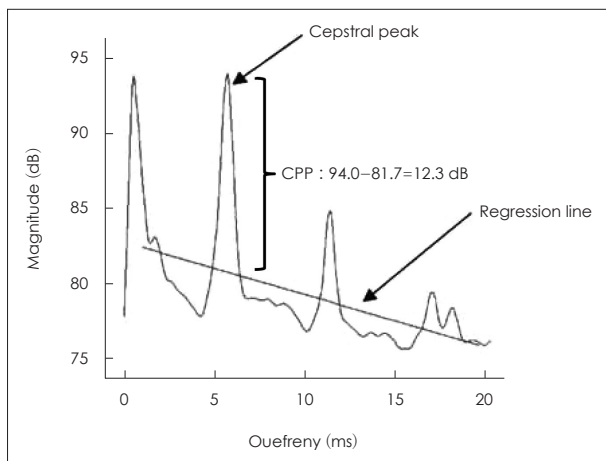


Fig. 1. Cepstral representations of cepstral peak prominence and regression line.

Klingholtz 등은 모음연장발성은 일상대화보다는 노래 할 때의 발성을 더 잘 반영하기 때문에 임상에 적용하기에는 한계가 있다고 보고하였다.¹⁴⁾ 또한 쉰 목소리(hoarseness)는 모음발성시간이 길수록 더 두드러진다는 특성이 있기 때문에 /a/ 모음연장발성 시 심한 발성장애를 가진 환자의 경우 정확한 진단에 오류가 생길 수 있다(diagnostic gap).⁸⁾ 반면 문장검사는 모음연장발성에 비하여 기본주파수의 변화 및 발성의 시작과 중지(onset and pause)에 대한 평가에서 이점이 있다는 보고가 있다.⁹⁾

Hillenbrand 등에 따르면 모음연장발성과 문장검사는 cepstrum을 이용한 음성분석을 하는데 모두 통계적으로 유의한 결과를 보였지만²⁾ Moers 등에 따르면 모음연장발성검사에서의 p값은 -0.45, 문장검사에서의 p값은 -0.73으로 문장검사만이 통계적으로 유의한 결과를 보였으며⁵⁾ Klingholtz 등도 문장검사가 모음연장발성검사에 비해서 SNR(signal-to-noise ratio)의 평가에 있어서 더 효과적이라고 보고하였다.¹⁴⁾ 이처럼 두 방법의 유용성에 대한 일관적인 평가가 이루어 지지 않고 있어 어느 방법이 cepstrum음성분석 결과를 더 잘 반영하는 지에 대한 연구의 필요성이 있어왔다.

본 연구에서는 치료 후 주관적인 음성효전이 있는 환자에서 모음연장발성검사를 시행했을 때는 성대폴립에서만 유의한 치료효과를 보이는 반면에 문장검사를 시행했을 경우 성대마비, 성대결절, 성대폴립 모두에서 통계적으로 유의한 치료효과를 보여 문장검사가 모음연장발성검사에 비해서 기식성 음성을 유발하는 성대질환을 평가하는데 적합하다는 평가를 내릴 수 있었다. 이는 모음분석을 할 때 안정화된 발음이 필요하며(stable phonation) 검사들마다 모음연속발음 하는 시간이 다르기 때문에 심한 발성장애가 있는 환자들의 경우 검사의 신뢰도가 떨어질 수 있기 때문이라고 사료된다.⁸⁾ 또한 MDVP를 이용한 jitter, shimmer, 잡음 대 배음비 값을 측정하였을 때도 성대결절, 성대폴립에서는 치료효과를 반영하였지만 성대마비는 치료효과를 반영하지 못하여 성대마비와 같은 기식성 음성이 큰 경우에는 cepstrum을 이용한 검사가 가장 유용하다고 판단할 수 있었다.

결 론

본 연구를 통해서 /a/ 모음연장발성은 켈스트럼을 이용한 발성장애 평가에서 제한점을 보인 반면, 문장검사는 기식화된 음성을 보다 더 잘 변별하였다. 하지만 임상적으로 여전히 /a/ 모음연장발성을 많이 사용하고 있으며 각각의 장점이 존재하기 때문에 기식화된 음성분석을 위해 켈스트럼 분석을 사용할 경우 모음연장발성검사만 하는 것이 아니라 문장검

사를 같이 이용하는 것이 발성장애 평가에 유용한 정보를 제공해 줄 수 있을 것으로 사료된다.

중심 단어 : 켈스트럼 · 언어장애 · 음성분석 · 음성.

REFERENCES

- 1) Roy N, Stemple J, Merrill RM, Thomas L. *Epidemiology of voice disorders in the elderly: preliminary findings. Laryngoscope* 2007; 117(4):628-33.
- 2) Ruben RJ. *Redefining the survival of the fittest: communication disorders in the 21st century. Laryngoscope* 2000;110(2 Pt 1):241-5.
- 3) Heman-Ackah YD, Heuer RJ, Michael DD, Ostrowski R, Horman M, Baroody MM, et al. *Cepstral peak prominence: a more reliable measure of dysphonia. Ann Otol Rhinol Laryngol* 2003;112(4):324-33.
- 4) Heman-Ackah YD. *Reliability of calculating the cepstral peak without linear regression analysis. J Voice* 2004;18(2):203-8.
- 5) Zieger K, Schneider C, Gerull G, Mrowinski D. [Cepstrum analysis in voice disorders]. *Folia Phoniatr Logop* 1995;47(4):210-7.
- 6) Radish Kumar B, Bhat JS, Prasad N. *Cepstral analysis of voice in persons with vocal nodules. J Voice* 2010;24(6):651-3.
- 7) Park MC, Mun MK, Lee SH, Jin SM. *Clinical Usefulness of Cepstral Analysis in Dysphonia Evaluation. Korean J Otorhinolaryngol-Head Neck Surg DE* - 2013-09-27 2013;56(9):574-8.
- 8) Moers C, Mobius B, Rosanowski F, Noth E, Eysholdt U, Haderlein T. *Vowel- and text-based cepstral analysis of chronic hoarseness. J Voice* 2012;26(4):416-24.
- 9) Parsa V, Jamieson DG. *Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. J Speech Lang Hear Res* 2001;44(2):327-39.
- 10) Askenfelt AG, Hammarberg B. *Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. J Speech Hear Res* 1986;29(1):50-64.
- 11) Carding PN, Steen IN, Webb A, MacKenzie K, Deary IJ, Wilson JA. *The reliability and sensitivity to change of acoustic measures of voice quality. Clin Otolaryngol Allied Sci* 2004;29(5):538-44.
- 12) Hillenbrand J, Houde RA. *Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. J Speech Hear Res* 1996;39(2):311-21.
- 13) Hillenbrand J. *A methodological study of perturbation and additive noise in synthetically generated voice signals. J Speech Hear Res* 1987;30(4):448-61.
- 14) Klingholtz F. *Acoustic recognition of voice disorders: a comparative study of running speech versus sustained vowels. J Acoust Soc Am* 1990;87(5):2218-24.