

Acoustic Voice Analysis in Patients with Penetration/Aspiration Via Videofluoroscopic Swallowing Study

Young Ae Kang^{1,3}, Sung Ju Jee², and Bon Seok Koo^{1,3}

¹Departments of Otolaryngology-Head and Neck Surgery, ²Rehabilitation Medicine, ³Research Institute for Medical Sciences, Chungnam National University Hospital, Daejeon, Korea

비디오투시조영검사를 통한 침습/흡인에 따른 음성의 음향적 분석

강영애^{1,3} · 지성주² · 구본석^{1,3}

충남대학교병원 이비인후-두경부외과, ¹ 재활의학과, ² 의학연구소³

Received March 18, 2017
Revised April 20, 2017
Accepted June 2, 2017
Address for correspondence
 Bon Seok Koo, MD, PhD
 Department of Otolaryngology-Head and Neck Surgery,
 Chungnam National University Hospital,
 282 Munhwa-ro, Jung-gu,
 Daejeon 35015, Korea
Tel +82-42-280-8692
Fax +82-42-253-4059
E-mail bskoo515@cnuh.co.kr

Background and Objectives The present study aimed to investigate the effects of penetration/aspiration (P/A) on voice acoustic parameters.

Subjects and Method Twenty-seven patients were analyzed with the videofluoroscopic swallowing study (VFSS) and then divided into two groups based on the modified Penetration and Aspiration Scale results. Ten patients (5 males and 5 females) were included in the Non-P/A group, and 17 patients (12 males and 5 females) in the P/A group. Stroke was the major cause of swallowing disorders. Three sustained /a/ vowels recorded in pre- and post-VFSS were analyzed. Mann-Whitney U-test was used to compare acoustic values before and after VFSS, and the receiver operating characteristics (ROC) curve with combination of significant parameters was also conducted.

Results Among acoustic parameters, the length of analyzed sample ($p=0.010$), number of segments computed ($p=0.018$), total number detected pitch periods ($p=0.017$), and second formant ($p=0.013$) in pre- and post-VFSS were significantly different between Non-P/A and P/A groups. In the P/A group after VFSS, the means of these significant parameters decreased. According to ROC combined with four significant parameters, the probability of predicting P/A condition was 84% ($p=0.005$), the sensitivity was 80%, and the specificity was 80%.

Conclusion Voice acoustic analysis can reflect voice changes by penetration/aspiration and the combination of significant parameters can also detect swallowing disorders. Therefore, voice analysis can be a reliable screening tool for patients with swallowing disorders.

Korean J Otorhinolaryngol-Head Neck Surg 2017;60(9):454-62

Key Words Acoustic analysis · Aspiration · Penetration · Swallowing disorders · Videofluoroscopic swallowing study.

서 론

삼킴(연하, swallowing)은 '무엇을 입에 넣어서 목구멍으로 넘긴다'라는 뜻으로,¹⁾ 생명 연장을 위한 가장 중요한 기능적

요소이면서 삶의 질을 결정짓는 핵심적 요소이다.²⁾ 삼킴 관련 해부학적 구조물은 입술부터 식도까지 모든 기관이 관여되지만 음식을 혀뿌리 아래로 내려보내는 구강 이동시간이 1초 내외이고 후두덮개가 내려가고 성대가 닫히며 후두가 전상방으로 올라가는 이후 이동시간도 1초 내외일 정도로, 정상 삼킴은 각 기관의 정확한 협응이 순간적으로 진행된다. 이런 각 기관의 유기적 협응이 깨질 때 삼킴장애가 발생한다. 이후

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.

두 단계의 삼킴장애로는 먹은 음식물이 식도로 가지 않고 기도 쪽으로 들어가 성대 아래까지 내려가는 흡인(aspiration)과 성대를 통과하지 않지만 성대 주위에 머무는 침습(penetration)이 대표적이다.¹⁾ 또한 적절한 삼킴반사로 음식물이 인후두까지 잘 내려갔더라도 하인두의 괄약근(inferior pharyngeal constrictor muscle), 즉 상부식도괄약근(upper esophageal sphincter)이 열리지 않거나 작게 열려 후두전정(laryngeal vestibule) 구간에 모여 있던 음식물이 피열연골을 넘어 다시 성대 안쪽으로 흘러들어가는 2차 흡인(삼킴 후 흡인, aspiration after swallowing)도 있다.

삼킴장애의 심각성은 성대 안쪽으로 내려간 음식물이 흡인성 폐렴을 유발한다는 것이고, 특히 기침반사가 소실된 무증상 흡인(silent aspiration) 환자에서 흡인성 폐렴의 위험성은 매우 크다. 무증상 흡인은 전체 삼킴장애 환자의 40% 정도 차지하며,³⁾ 재활병원 입원 환자 중 무증상 흡인 환자는 적게는 58%⁴⁾ 많게는 82%⁵⁾까지 보고될 정도로 뇌병변 이후 감각저하를 보이는 뇌졸중 환자의 무증상 흡인은 빈번하게 관찰된다.

삼킴장애를 진단하기 위해서는 특수한 장비가 필요하며, 초음파검사, 광섬유내시경 삼킴검사, 근전도검사, 내압검사, 비디오투시조영삼킴검사(videofluoroscopic swallowing study, VFSS) 등 다양한 삼킴검사가 있다. 사용 및 타당도 측면을 고려할 때 VFSS가 기본적 검사(golden standard)로 평가되지만^{6,7)} 이것은 방사능 피폭, 고비용 장비 필요, 병원 환경 밖에서의 제한된 접근, 짧은 검사시간으로 장시간에 걸친 기능적 섭식 능력 미반영, 영상자료의 해석 오류,⁷⁾ 침상검사의 불가능⁶⁾ 등의 단점이 있다.

삼킴장애 발생 시 쉰 목소리(hoarse vocal quality)를 동반한 젖은 목소리(wet voice, gurgle voice)가 관찰되며⁸⁻¹⁰⁾ 흡인 환자의 90%가 젖은 목소리를 보일 정도로 그 빈도는 높다.⁵⁾ 임상에서는 물삼킴검사(water swallow test)를 통해 음성변화를 확인하여 이를 선별적 삼킴진단검사로 활용하고 있다. 이런 음성변화를 연구한 선행연구 중, 뇌졸중 환자 93명의 VFSS 전후 모음 /아/ 발성을 분석한 결과, 침습/흡인 민감도가 91%로 음성분석 변수가 삼킴장애를 예측할 수 있다고 주장한¹¹⁾ 반면 동일한 연구 방법을 적용한 다른 선행연구¹²⁾에서는 오히려 현재의 음성분석 변수로는 삼킴장애를 예측할 수 없다고 상반된 보고가 이루어졌다. 동일한 연구 방법론을 적용하여 상반된 결과를 보고한 두 선행연구의 공통적인 문제는 음성분석 변수의 폭이 매우 한정적이란 점이다. 두 선행연구에서는 주파수 기반 분석인 5개 변수[average fundamental frequency(F0), relative average perturbation(RAP), shimmer, noise-to-harmonic ratio(NHR), voice turbulence(VTI)]만 선택하여 multi-dimensional voice program(MDVP)으로 분석

하였는데, 이 주파수 기반 분석은 성대의 진동주기를 계산하고 이를 바탕으로 다른 변수의 주기성을 측정하는 방식이므로 주기성이 많은 음질의 음성분석에 적당하다. 하지만 병리적 음성(pathological voice)은 Titzer¹³⁾의 분류에서 Type 1보다는 Type 2~4로 분류될 정도로 일반적으로 비주기성이 많다. 삼킴장애 음성은 삼킴장애를 유발시킨 질병으로 인한 음질 변화와 침습/흡인으로 인한 성대진동의 변화를 모두 내포하고 있다. 이에 본 연구에서는 침습/흡인에 따른 음성의 변화를 기존의 주파수 기반 음성분석 변수(MDVP 분석) 이외 시간 기반 음성분석 변수(캡스트럴 분석)와 성도의 여과 과정(filtering)의 음성분석 변수(포먼트, 밴드위스)를 조사하고자 한다.

대상 및 방법

대상자

연구 대상자는 2015년 9~10월까지 VFSS가 예정되어 있고 녹음에 동의한 환자 27명(남 17명, 여 10명)을 대상으로 하였다. 보청기 착용자, 성대 마비자, 기관루(stoma) 존재자, 전반 실어증 및 구어 실행증 환자는 제외시켰다. 병력은 뇌혈관질환(cerebrovascular accident)이 19명으로 대부분을 차지하였고 그 외 파킨슨병 2명, 두경부암 4명, 알츠하이머 치매 1명이었다. 19명의 뇌혈관질환을 세부적 진단으로 살펴보면 뇌경색 12명, 뇌대출혈 5명, 거미막하출혈 2명이었다.

VFSS는 방사선 기기(Flexavision FD RF X-ray system, Shimadzu Corp., Kyoto, Japan)를 사용하며 다양한 검사 식이(요거트, 죽, 물 5 mL, 물 10 mL, 밥)에 조영제(레디스액, 황산바륨, 동인당제약주식회사, 40%w/v)를 일정 비율(음식물 3:조영제 7)로 섞어 환자에게 순서적으로 제공하였다. 검사자의 지시에 따라 환자가 제공받는 음식을 삼키면 검사자는 실시간 촬영 모니터를 보면서 수정된 침습-흡인 척도(Modified Penetration and Aspiration Scale, MPAS)를 매겼다. 각 식이 단계에서 흡인이 확인되면 다음 단계로 넘어가지 않고 환자의 안전을 위해 검사는 중단되었고, 음성녹음을 바로 시작하였다. MPAS(1~5점, 5점 척도)는 8점 척도인 Penetration and Aspiration Scale의 수정판으로, 구체적인 점수 척도는 다음과 같다; 기도로 흘러 들어감 없음(1점), 기도로 흘러 들어가며 성대 위쪽에 남음(2점), 기도로 흘러 들어가 성대에 닿음(침습 상태, 3점), 기도로 흘러 들어가고 성대 아래쪽까지 내려가며 빼내기 위한 노력 반응 보임(기침반사 등)(증상 흡인 상태, 4점), 기도로 흘러 들어가고 성대 아래쪽까지 내려가지만 빼내려는 반응 없음(무증상 흡인 상태, 5점).¹⁴⁾ 본 연구에서는 1~2점을 정상 삼킴군[Non-penetration/aspiration(P/A)

group], 3~5점은 침습/흡인군(P/A group)으로 분류하였다. 두 집단 간 연령 및 성별 차이는 통계적으로 없었으며, P/A군 17명 중 침습은 5명, 나머지 12명은 무증상 흡인(silent aspi-

ration)이었다(Table 1). 본 연구는 임상연구심의위원회의 승인(Institutional Review Board No. CNUH 2016- 05-020-001)하에 이루어졌다.

Table 1. Demographic data of study patients

Variables	Groups (n=27)			
	Non-P/A	P/A	t	p
Number & age				
No. of patients	10	17		
Male:female	5:5	12:5	1.05	0.303
Mean age (range) (yrs.)	71.6 (56-83)	66.8 (49-86)	0.615	0.544
Pathogenesis				
CVA				
CI				
Unspecified	2	3		
MCA	1	1		
BA	1	2		
PCA		2		
ICH	1	4		
SAH		2		
Parkinson's disease	2			
Head-neck cancer	2	3		
Alzheimer's disease	1			

P/A: penetration/aspiration, CVA: cerebro-vascular accident, CI: cerebral infarction, MCA: middle cerebral artery, BA: basilar artery, PCA: posterior cerebral artery, ICH: intracerebral hemorrhage, SAH: subarachnoid hemorrhage

Table 2. List of parameters according to analysis tool

Category	No.	Parameters
MDVP	24	<ul style="list-style-type: none"> - F0-related measurements F0, mF0, T0, Fhi, Flo, SD, Tsam, PFR, SEG, PER - Frequency Perturbation measurements Jita, Jitt, RAP, PPQ, sPPQ, vF0 - Amplitude Perturbation measurements ShdB, Shim, APQ, sAPQ, vAm - Noise measurements NHR, VTI, SPI
ADSV	16	<ul style="list-style-type: none"> - Cepstral peak-related measurements CPP, CPP_SD, CPP_Max, CPP_Min, CPP avg, SD, Mean CPP_F0, Mean CPP_F0 SD - Spectral ratio-related measurements L/H spectral ratio, L/H spectral ratio SD, L/H spectral ratio Max, L/H spectral ratio Min - Slope-related measurements Regression slope, Regression slope SD - Cepstral/spectral calculation-related measurements CSID_female, CSID_male
PRAAT	6	<ul style="list-style-type: none"> - Formant measurements F1, F2, F3 - Bandwidth measurements BW1, BW2, BW3

F0: average fundamental frequency (Hz), mF0: mean fundamental frequency (Hz), T0: average pitch period (ms), Fhi: highest fundamental frequency (Hz), Flo: lowest fundamental frequency (Hz), SD: standard deviation of F0 (Hz), PFR: phonatory F0-range in semi-tones, Tsam: length of analyzed sample (s), SEG: number of segments computed, PER: total number detected pitch periods, Jita: absolute Jitter (μ s), Jitt: jitter percent (%), RAP: relative average perturbation (%), PPQ: pitch perturbation quotient (%), sPPQ: smoothed pitch perturbation quotient (%), vF0: fundamental frequency variation (%), ShdB: shimmer in dB (dB), Shim: shimmer percent (%), APQ: amplitude perturbation quotient (%), sAPQ: smoothed Amplitude perturbation quotient (%), vAm: peak-to-peak amplitude (%), NHR: noise to harmonic ratio, VTI: Voice Turbulence Index, SPI: Soft Phonation Index, F1: first formant (Hz), F2: second formant (Hz), F3: third formant (Hz), BW1: first bandwidth (Hz), BW2: second bandwidth (Hz), BW3: third bandwidth (Hz), MDVP: multi-dimensional voice program, ADSV: analysis of dysphonia in speech and voice, CPP: cepstral peak prominence, L/H: low/high, CSID: cepstral/spectral index of dysphonia

녹음 및 분석 방법

음성은 VFSS 직전과 직후 편안한 모음 /아/ 발성 3회 녹음하였고, 녹음 장소는 VFSS 방사선 기기의 소음이 최대한 통제되도록 방음문이 설치된 공간에서 이루어졌다. 녹음 장비는 보이스 레코드(PCM-50, Sony corporation, Tokyo, Japan)에 헤드셋 마이크(Sennheiser pc151, Sennheiser electronic GmbH & Co. KG, Wedemark, Germany)를 연결시켜 사용하였고, 48 kHz 표본화율과 24 bit 양자화 조건에서 디지털 변환시켰다. 마이크와 입 간의 거리는 입술에서 5 cm 정도로 위치시켰고 절삭(clipping)이 생기지 않는 범위에서 입력 볼륨을 4단계로 고정시켰다.

분석구간은 모음 /아/의 첫 신호(impulse of wave signal)부터 끝 신호까지 선택하였고, 분석변수는 MDVP(model 5105,

KayPentax Inc., Lincoln Park, NJ, USA)의 변수 24개, Analysis of dysphonia in speech and voice(ADSV™, model 5109, KayPentax Inc.)의 변수 16개, 그리고 PRAAT(ver. 6021)의 포먼트 및 밴드위스 6개 변수를 분석하였다(Table 2). 포먼트와 밴드위스 분석방식을 간략히 기술하면, 성별 간 다운샘플링(downsampling)을 달리 적용한 후 linear predictive coding(burg)(window length 0.025, time step 0.005)으로 전환하여 프리-엠퍼시스(pre-emphasis) 처리 후, 음원(source)의 기율기 특성을 제거하고 성도의 공명특성(filtering)만을 추출하였다.

통계 적용

VFSS 검사 전후 모음 /아/ 3회 발성을 평균 처리하여 통계

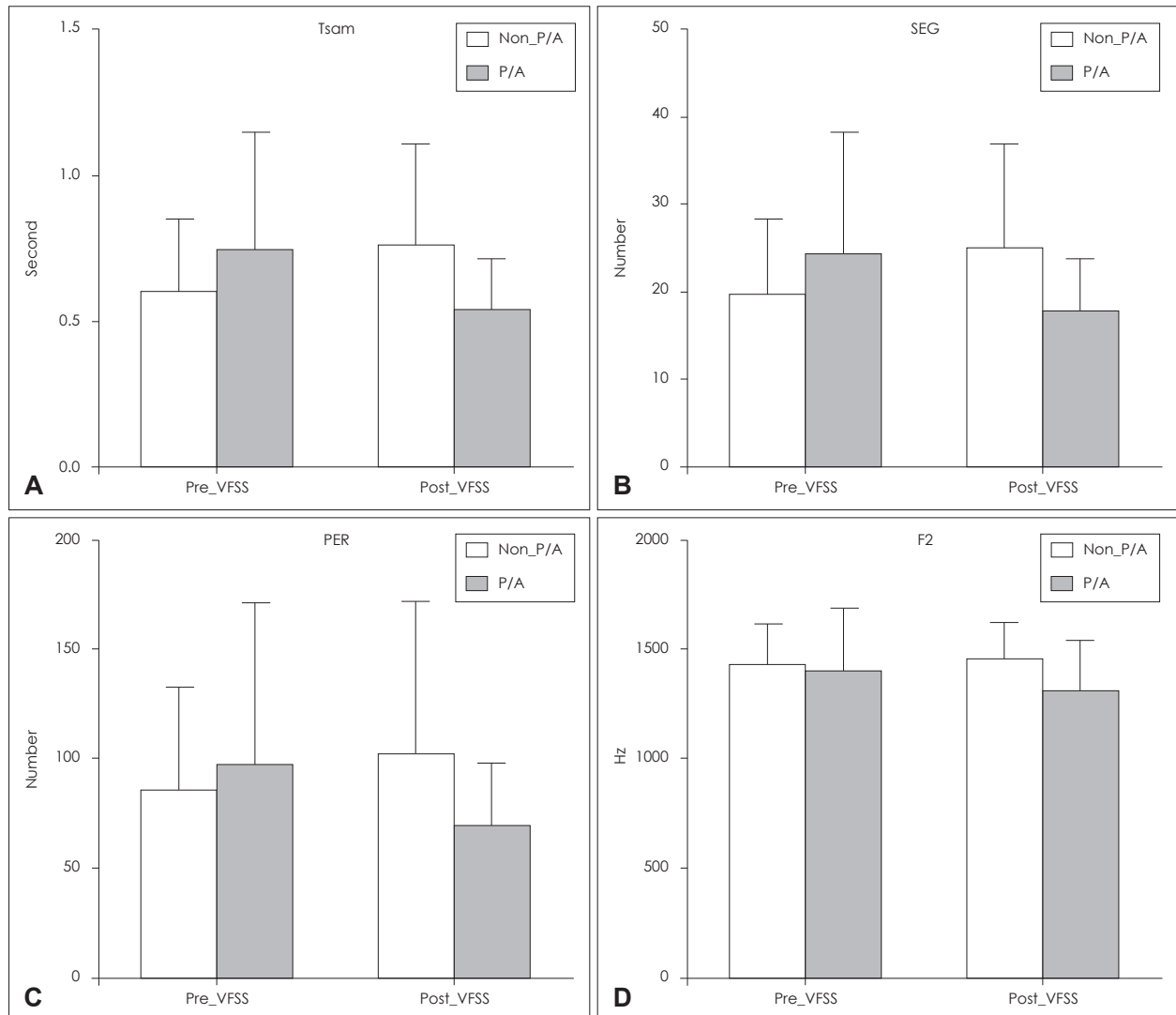


Fig. 1. Mean & SD of 4 significant parameters pre-and post-VFSS (A Tsam, B SEG, C PER, D F2). A gray bar is for P/A group and a white bar is for Non-P/A group. Tsam: length of analyzed sample, SEG: number of segments computed, PER: total number detected pitch periods, F2: second formant (Hz), P/A: penetration / aspiration, VFSS: videofluoroscopic swallowing study.

Table 3. Differences in significant acoustic parameters pre-and post-VFSS

Parameters	Z	p	Mean (\pm SD)*	
			Non-P/A	P/A
Tsam	-2.35	0.010†	0.15 (\pm 0.31)	-0.20 (\pm 0.41)
SEG	-2.38	0.018*	5.21 (\pm 10.41)	-7.03 (\pm 13.89)
PER	-2.16	0.017*	16.96 (\pm 44.01)	-31.31 (\pm 70.64)
F2	-2.56	0.031*	25.59 (\pm 188.25)	-92.40 (\pm 227.91)

* $p < 0.05$, † $p < 0.01$, *post value minus pre value. VFSS: videofluoroscopic swallowing study, Tsam: length of analyzed sample (sec.), SEG: number of segments computed (number), PER: total number detected pitch periods (number), F2: second formant (Hz), SD: standard deviation, P/A: penetration/aspiration

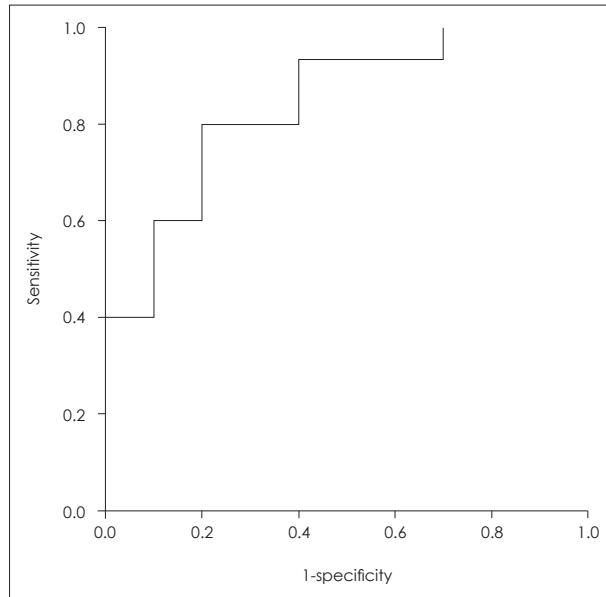


Fig. 2. ROC curve of four acoustic parameters combination (Tsam+SEG+PER+F2). X axis is 1-specificity and Y axis is sensitivity. ROC: receiver operating characteristics, Tsam: length of analyzed sample, SEG: number of segments computed, PER: total number detected pitch periods, F2: second formant (Hz).

에 적용하였으며, 두 군 모두 고령인 관계로 음질 저하가 있었고, 특히 검사 전에 P/A군이 Non-P/A군보다 청지각적으로 음질 상태가 더 나빴기 때문에 두 집단 간 음질 차이를 보정하기 위해, VFSS 후 음성분석 값에서 VFSS 전 음성분석 값을 뺀 차이값에 맨-휘트니 검정(Mann-Whitney U-test)을 적용하였다. 그리고 통계적으로 유의미한 변수를 조합하여 수신 조작 특성(receiver operating characteristics, ROC) 분석을 통해 삼킴장애 분류 정확도를 구하였다.

결 과

MDVP 분석결과

P/A군과 Non-P/A군 간 맨-휘트니 검정 결과, MDVP 24개 변수 중 유의미한 변수는 length of analyzed sample(Tsam) ($p=0.02$), number of segments computed(SEG) ($p=0.017$),

total number detected pitch periods(PER) ($p=0.031$)이었다. P/A군의 Tsam, SEG, PER 평균(\pm 표준편차)이 검사 전 각 0.73 (\pm 0.42), 23.99(\pm 14.28), 96.43(\pm 76.20)초에서 검사 후 각 0.53 (\pm 0.19), 17.38(\pm 6.47), 67.77(\pm 30.31)초로 감소한 반면, Non-P/A군에서는 동일 3개의 변수가 검사 전 0.59(\pm 0.26), 19.38 (\pm 8.80), 83.67(\pm 49.77)초에서 검사 후 각 0.75(\pm 0.36), 24.60(\pm 12.33), 100.63(\pm 71.01)초로 증가하였다. 두 집단 간 24개 변수의 통계결과와 기술통계량은 [Appendix 1]에 제시하였다.

ADSV 분석결과

ADSV 16개 변수에 대한 두 군 간 통계적으로 유의미한 차이는 없었다. ADSV 분석의 대표적인 변수 cepstral peak prominence(CPP)는 VFSS 후 Non-P/A군과 P/A군에서 모두 감소하였다. P/A군 검사 전 CPP 평균(\pm 표준편차) 7.66(\pm 3.92) dB에서 검사 후 7.48(\pm 4.22) dB로 감소하였고, Non-P/A군도 검사 전 9.13(\pm 2.12) dB에서 검사 후 7.59(\pm 3.26) dB로 감소하였다. 두 집단별 16개 변수의 기술통계량과 통계결과는 [Appendix 2]에 제시하였다.

PRAAT 분석결과

포먼트와 밴드위스 관련 총 6개 변수의 통계결과, 제2 포먼트(F2) 변수만이 통계적으로 유의미한 차이를 보였다($p=0.010$). 검사 전 P/A군의 F2 평균(\pm 표준편차) 1359.22(\pm 299.67) Hz에서 검사 후 1302.81(\pm 243.04) Hz로 감소하였고, Non-P/A군은 1420.79(\pm 203.18) Hz에서 검사 후 1446.39(\pm 182.10) Hz로 증가하였다. 두 집단별 6개 변수의 기술통계량과 통계결과는 [Appendix 3]에 제시하였다.

유의미한 4개 변수의 검사 전후의 평균과 표준편차는 Fig. 1에 도식화하였고, 검사 후에서 전을 뺀 차이값의 평균과 표준편차는 Table 3에 제시하였다.

삼킴장애 분류정확도

본 결과에서 유의미한 4개 음향변수(Tsam, SEG, PER, F2)

Table 4. Result of ROC of the acoustic parameter combination

Parameters	p	Cut-off	Sensitivity	Specificity	AUC
Tsam	0.018*	-0.80	0.53	0.70	0.21
SEG	0.017*	-2.50	0.53	0.30	0.21
PER	0.031*	3.25	0.33	0.50	0.24
F2	0.010†	-32.26	0.41	0.20	0.20
Tsam+SEG	0.003†	0.54	0.80	0.60	0.85
Tsam+PER	0.020*	0.51	0.80	0.60	0.78
Tsam+F2	0.017*	0.51	0.75	0.60	0.78
SEG+F2	0.017*	0.51	0.73	0.60	0.78
PER+F2	0.040*	0.53	0.73	0.60	0.74
SEG+PER	0.015*	0.51	0.80	0.60	0.79
Tsam+SEG+PER	0.003†	0.32	0.93	0.60	0.86
SEG+PER+F2	0.013*	0.47	0.86	0.60	0.80
Tsam+PER+F2	0.013*	0.47	0.86	0.60	0.80
Tsam+SEG+PER+F2	0.005†	0.60	0.80	0.80	0.84

* $p < 0.05$, † $p < 0.01$. ROC: receiver operating characteristics, Tsam: length of analyzed sample (sec.), SEG: number of segments computed (number), PER: total number detected pitch periods (number), F2: second formant (Hz), P/A: penetration/aspiration, AUC: area under curve

를 선택하여 이항 로지스틱 회귀분석(입력 방식)을 적용하여 P/A군 선별을 위한 조합 변수의 확률값을 구하고, 이 확률값에 ROC 분석을 적용하였다. 유의미한 4개 음성분석 변수의 조합에서 P/A군을 선별할 곡선하 영역(area under curve, AUC)이 0.84(95% CI 0.682~0.998)로 통계적으로도 유의미하였고($p < 0.01$), 민감도와 특이도는 각 0.80, 0.80으로 높았다(Fig. 2). 독자의 이해를 위해 개별 변수 및 변수 2개 조합, 3개 조합의 AUC를 제시하였다(Table 4). AUC 수치만을 놓고 볼 때, 변수 2개 조합(Tsam+SEG)과 변수 3개 조합(Tsam+SEG+PER)에서 각 0.85, 0.86으로 높은 수치를 보였지만 이 조합들의 민감도와 특이도 수치가 낮았다. 고로 이 모든 점을 고려해 볼 때 변수 4개의 조합(Tsam, SEG, PER, F2)이 안정적인 수치를 보였다.

고 찰

본 연구는 VFSS 전후 모음 /아/를 대상으로 침습/흡인 유무에 따라 변화되는 음향학적 음성변수를 조사한 것이다.

MDVP, ADSV 및 PRAAT 분석에서 두 집단 간 통계적으로 유의미한 변수는 총 4개(Tsam, SEG, PER, F2)였다. 4개의 유의미한 변수 중 MDVP의 주파수 관련 변수인 Tsam, SEG, PER은 VFSS 후 P/A군에서 감소한 반면, Non-P/A군에서는 모두 증가했다. 두 군의 모든 대상자에게 VFSS 전후 편안한 발성을 하도록 동일한 발성 지시를 하였으나, VFSS 이후 P/A군에서만 발성 길이(Tsam) 및 분석구간(SEG)이 짧아졌고 분석구간의 주파수 주기성(PER)이 줄어든 것이다. 침습/흡인으로 인해 성대 및 성도에 음식 잔여물이 남게 되면 인후두에 불

편함이 생기고 이런 불편함을 제거하기 위해 생리적으로 강한 발성(effort phonation)이 유발되어 음식 잔여물을 제거한다.¹⁵⁾ 이런 강한 발성은 공기량 손실을 높여 결과적으로 발성 길이의 감소를 초래하였고, 흡인으로 인해 성대에 묻은 잔여물은 성대진동의 주기성에 악영향을 주어 SEG와 PER 수치를 감소시켰다고 생각한다.

서론에 언급한 두 선행연구^{11,12)}에서는 MDVP의 5개 변수만을 선택하였고 본 연구는 MDVP의 24개 변수를 선택하였기 때문에 본 연구의 분석 변수범위가 더 넓으며 본 연구의 모든 분석변수가 선행연구에서 어떻게 변화를 보였는지는 알 수 없다. 하지만 선행연구¹¹⁾에서 선택한 5개 변수(F0, RAP, Shimmer, NHR, VTI)는 저위험군의 RAP 수치가 검사 후 감소한 것을 제외하고 모든 변수의 수치가 검사 후 증가하였다. 본 연구에서도 검사 후 P/A군의 RAP, NHR 수치가 증가한 점은 선행연구와 일치하며 F0, Shimmer, VTI이 감소한 것은 불일치한다. 본연구와 선행연구¹¹⁾ 결과가 불일치한 이유는 분석구간의 차이로 생각한다. 선행연구¹¹⁾에서는 모음 /아/ 3회의 안정구간만을 선택하여 평균값을 사용한 반면 본 연구에서는 음식물이 성대에 묻은 상황이라면 성대진동의 시작점과 끝점에서 그 양상이 더 민감하게 반영될 것으로 판단하여 모음 /아/ 3회의 발성 전체 구간을 선택하여 평균값을 사용하였다. 일반적인 음성분석 연구에서는 모음의 안정구간만을 선택하여 분석하지만 본 연구처럼, 삼킴장애 음성연구에서는 안정구간 이외 발성 앞부분과 끝부분을 분석에 포함시킬 필요가 있다. 발성 앞부분은 성대의 진동 시작으로 인해 잔여음식물의 흐트러짐이 가장 많이 발생하고 발성 끝부분은 성대 주위에 고이는 잔여물이 다시 성대 쪽으로 흘러 들어가는 삼

김 후 흡인 현상이 포착될 수 있으며 이 두 현상은 모두 성대진동에 영향을 끼치기 때문이다.

삼킴 후 인후두에 남은 잔여 음식물이 포먼트 특성에 변화를 줄 것이라는 가설을 검증하기 위해 분석한 포먼트 결과에서 F2 변수가 통계적으로 유의미한 차이를 보였고, P/A군의 수치가 감소하였다. 60대 이상 정상 노인 20명의 포먼트 연구¹⁶⁾와 비교해 보면, 정상 노인 F2 평균(\pm 표준편차) 1388.4(\pm 169.4) Hz로 본 연구의 VFSS 전 P/A군의 F2 1359.22(\pm 299.67) Hz와 유사하다. 하지만 VFSS 후 P/A군과 Non-P/A군의 F2 변화 양상은 대조적으로 P/A군은 1302.81(\pm 243.04) Hz로 감소하였고 Non-P/A군은 1446.39(\pm 182.10) Hz로 증가하였다. 황산바륨이 인후두에 묻는 코딩 효과와 잔류량으로 인해 성도(vocal tract)의 공명 주파수가 변화되어 이로 인해 두 군 간 상반되는 포먼트의 변화가 나타났다고 본다. 하지만 음식 잔여물의 정도(양)와 포먼트 변화 폭 간의 보강 연구는 추후 더 필요하다.

유의미한 4개의 변수를 조합한 ROC 분석에서 곡선하 영역이 0.84로, 이는 P/A군으로 분류될 확률이 84%임을 의미한다. 한 개의 음향변수보다 유의미한 여러 개의 변수를 조합하면 침습/흡인의 삼킴장애 예측을 높일 수 있기에, 본 연구를 통해 VFSS의 보조적인 도구로써 음성분석이 삼킴장애 진단 도구로 활용 가능성이 있다고 생각한다.

본 연구의 제한점은 VFSS 검사 진행 시 환자의 안전을 위해 침습/흡인이 관찰될 경우 검사를 중단하고 바로 음성녹음을 하였기에 P/A군에서 환자마다 적용된 마지막 식이가 다르다. 이는 식이의 점성도에 통제가 이루어지지 못한 것으로 침습/흡인 시 식이의 점성도에 따른 음질 변화가 다르다는¹⁰⁾ 보고에서 보듯이 본 연구의 음성변수 변화를 침습/흡인에 따른 삼킴장애 영향으로 해석할 수는 있으나 특정 점성도의 식이와 연결 지어 해석하기에는 한계가 있다. 또한 연구 대상자들이 뇌졸중 이외 파킨슨병 및 두경부암 환자들이 혼재되어 있기 때문에 각 질환이 주는 음질 변화를 통제할 수 없었다. 하지만 본 연구는 비디오투시조영삼킴검사 전후 음성을 분석하여 침습/흡인의 삼킴장애에서 발성 길이, 분석구간, 분석구간 내 주파수 주기성 그리고 제2 포먼트의 수치 감소를 확인하였고 이런 유의미한 변수를 조합하여 삼킴장애 예측 확률이 84%로 높았기에 보조적 삼킴 진단검사로써 객관적인 음성분석 변수

의 활용 가능성을 시사한다.

Acknowledgments

I really appreciated our patents and the swallowing team who participated in this VFSS. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5B6914201).

REFERENCES

- 1) Kim HH. Neurologic speech-language disorders. Seoul: Sigmappress; 2012.
- 2) Vainshtein JM, Moon DH, Feng FY, Chepeha DB, Eisbruch A, Stenmark MH. Long-term quality of life after swallowing and salivary-sparing chemo-intensity modulated radiation therapy in survivors of human papillomavirus-related oropharyngeal cancer. *Int J Radiat Oncol Biol Phys* 2015;91(5):925-33.
- 3) Daniels SK, Brailey K, Priestly DH, Herrington LR, Weisberg LA, Foundas AL. Aspiration in patients with acute stroke. *Arch Phys Med Rehabil* 1998;79(1):14-9.
- 4) Splaingard ML, Hutchins B, Sulton LD, Chaudhuri G. Aspiration in rehabilitation patients: videofluoroscopy vs bedside clinical assessment. *Arch Phys Med Rehabil* 1988;69(8):637-40.
- 5) Linden P, Siebens AA. Dysphagia: predicting laryngeal penetration. *Arch Phys Med Rehabil* 1983;64(6):281-4.
- 6) Kwon MS, Kim JS. Evaluation and treatment of swallowing disorders. 2nd ed. Seoul: Hakjisa Publisher; 2007.
- 7) Korean Academy of Dysphagia Rehabilitation. Dysphagia: clinical management in adults and children. Seoul: Koonja Publisher; 2011.
- 8) Murray J, Langmore SE, Ginsberg S, Dostie A. The significance of accumulated oropharyngeal secretions and swallowing frequency in predicting aspiration. *Dysphagia* 1996;11(2):99-103.
- 9) Warms T, Richards J. "Wet Voice" as a predictor of penetration and aspiration in oropharyngeal dysphagia. *Dysphagia* 2000;15(2):84-8.
- 10) Murugappan S, Boyce S, Khosla S, Kelchner L, Gutmark E. Acoustic characteristics of phonation in "wet voice" conditions. *J Acoust Soc Am* 2010;127(4):2578-89.
- 11) Ryu JS, Park SR, Choi KH. Prediction of laryngeal aspiration using voice analysis. *Am J Phys Med Rehabil* 2004;83(10):753-7.
- 12) Chang HY, Torng PC, Wang TG, Chang YC. Acoustic voice analysis does not identify presence of penetration/aspiration as confirmed by videofluoroscopic swallowing study. *Arch Phys Med Rehabil* 2012;93(11):1991-4.
- 13) Lee JY, Choi SH, Jiang JJ, Hahn MS, Choi HS. Perturbation and perceptual analysis of pathological sustained vowels according to signal typing. *Phon Speech Sci* 2010;2(2):109-15.
- 14) Kweon S, Koo BS, Jee S. Change of swallowing in patients with head and neck cancer after concurrent chemoradiotherapy. *Ann Rehabil Med* 2016;40(6):1100-7.
- 15) Santos KW, Scheeren B, Maciel AC, Cassol M. Vocal variability post swallowing in individuals with and without oropharyngeal dysphagia. *Int Arch Otorhinolaryngol* 2015;19(1):61-6.
- 16) Kang YA, Yoon KC, Lee HS, Seong CJ. A comparison of parameters of acoustic vowel space in patients with Parkinson's disease. *Phon Speech Sci* 2010;2(4):185-92.

[Appendix 1]. Mean and SD of MDVP parameters about two groups Pre- and Post-VFSS

Parameters	Non-P/A		P/A		z	p
	Pre-VFSS, mean (SD)	Post-VFSS, mean (SD)	Pre-VFSS, mean (SD)	Post-VFSS, mean (SD)		
F0	151.78 (35.83)	151.21 (41.44)	154.85 (52.63)	146.11 (23.74)	-0.66	0.50
mF0	150.32 (36.16)	148.55 (40.30)	150.11 (40.02)	143.59 (22.95)	-0.49	0.61
T0	7.14 (1.82)	7.17 (1.70)	7.12 (1.25)	7.20 (1.18)	-0.72	0.49
Fhi	176.96 (44.17)	188.13 (68.25)	201.76 (115.30)	182.75 (40.52)	-0.22	0.82
Flo	127.37 (35.70)	124.63 (39.82)	125.40 (22.44)	119.16 (21.94)	-0.27	0.78
SD	11.59 (12.71)	15.22 (20.18)	21.69 (51.30)	14.00 (14.16)	-0.44	0.65
PFR	6.36 (4.53)	8.00 (7.18)	8.11 (6.77)	8.36 (4.87)	-0.41	0.67
Tsam	0.59 (0.26)	0.75 (0.36)	0.73 (0.42)	0.53 (0.19)	-2.33	0.02*
SEG	19.38 (8.80)	24.60 (12.33)	23.99 (14.28)	17.38 (6.47)	-2.38	0.01*
PER	83.67 (49.77)	100.63 (71.01)	96.43 (76.20)	67.77 (30.31)	-2.16	0.03*
Jita	276.72 (358.69)	364.98 (384.07)	314.90 (257.09)	324.19 (311.63)	-0.27	0.78
Jitt	3.33 (3.47)	4.89 (4.70)	4.57 (3.74)	4.59 (4.03)	-0.33	0.73
RAP	1.89 (2.07)	2.86 (2.77)	2.71 (2.24)	2.71 (2.42)	-0.33	0.73
PPQ	1.80 (1.69)	2.92 (2.79)	2.67 (2.41)	2.98 (2.86)	-0.38	0.69
sPPQ	3.09 (3.74)	3.39 (4.39)	1.88 (1.38)	2.45 (3.30)	-0.11	0.91
vF0	7.33 (6.33)	9.38 (11.51)	8.72 (10.93)	9.22 (8.68)	-0.61	0.54
Shdb	0.97 (0.65)	1.14 (0.54)	1.11 (0.50)	1.09 (0.53)	-0.49	0.61
Shim	10.40 (6.71)	12.63 (5.85)	12.67 (6.01)	11.99 (5.93)	-0.77	0.43
APQ	8.97 (6.37)	10.67 (6.92)	8.87 (3.69)	9.53 (4.69)	-0.11	0.91
sAPQ	9.51 (9.82)	10.63 (9.45)	11.14 (6.93)	9.50 (8.56)	-0.95	0.33
vAM	26.22 (11.04)	29.18 (12.10)	24.59 (7.28)	24.71 (10.03)	-0.11	0.91
NHR	0.28 (0.14)	0.38 (0.22)	0.32 (0.20)	0.32 (0.20)	-0.55	0.57
VTI	0.14 (0.12)	0.18 (0.12)	0.14 (0.11)	0.12 (0.08)	-0.30	0.76
SPI	8.39 (7.56)	7.19 (9.52)	6.22 (3.09)	7.02 (2.73)	-1.38	0.16

* $p < 0.05$, F0: average fundamental frequency (Hz), mF0: mean fundamental frequency (Hz), T0: average pitch period (ms), Fhi: highest fundamental frequency (Hz), Flo: lowest fundamental frequency (Hz), SD: standard deviation of F0 (Hz), PFR: phonatory F0-range in semi-tones, Tsam: length of analyzed sample (s), SEG: number of segments computed, PER: total number detected pitch periods, Jita: absolute jitter (μ s), Jitt: jitter percent (%), RAP: relative average perturbation (%), PPQ: pitch perturbation quotient (%), sPPQ: smoothed pitch perturbation quotient (%), vF0: fundamental frequency variation (%), ShdB: shimmer in dB (dB), Shim: shimmer percent (%), APQ: amplitude perturbation quotient (%), sAPQ: smoothed amplitude perturbation quotient (%), vAM: peak-to-peak amplitude (%), NHR: noise to harmonic ratio, VTI: Voice Turbulence Index, SPI: Soft Phonation Index, MDVP: multi-dimensional voice program, VFSS: videofluoroscopic swallowing study, P/A: penetration/aspiration

[Appendix 2]. Mean and SD of ADSV parameters about two groups pre- and post-VFSS

Parameters	Non-P/A		P/A		Z	p
	Pre-VFS mean (SD)	Post-VFSS mean (SD)	Pre-VFSS mean (SD)	Post-VFSS mean (SD)		
CPP	9.13 (2.12)	7.59 (3.26)	7.66 (3.92)	7.48 (4.22)	-1.15	0.24
CPP_SD	1.33 (0.43)	1.37 (0.41)	1.14 (0.48)	1.20 (0.48)	-0.05	0.96
CPP_Max	11.04 (2.06)	10.03 (3.02)	9.70 (4.21)	9.42 (4.32)	-0.60	0.54
CPP_Min	6.23 (2.81)	4.88 (3.16)	5.13 (3.55)	5.04 (4.06)	-0.65	0.51
CPP avg	11.65 (2.92)	10.20 (2.91)	10.08 (3.88)	10.17 (3.74)	-1.30	0.19
CPP avg, SD	1.27 (0.50)	1.27 (0.38)	1.12 (0.46)	1.16 (0.53)	-0.05	0.96
Mean CPP_F0	157.81 (34.64)	165.55 (50.15)	154.13 (28.54)	164.02 (38.87)	-0.55	0.58
Mean CPP_F0 SD	15.76 (18.63)	16.22 (16.60)	13.15 (17.33)	15.07 (15.20)	-0.20	0.84
L/H spectral ratio	27.92 (8.01)	24.35 (6.24)	27.45 (7.10)	24.23 (8.70)	-0.45	0.65
L/H spectral ratio SD	2.28 (0.77)	2.40 (1.14)	2.03 (0.72)	3.01 (1.92)	-1.40	0.16
L/H spectral ratio Max	32.46 (7.62)	29.40 (5.72)	31.29 (7.82)	29.38 (9.08)	-0.60	0.54
L/H spectral ratio Min	22.64 (8.73)	20.02 (6.34)	22.80 (5.97)	17.98 (8.29)	-0.55	0.58
Regression slope	-5.12 (2.42)	-6.74 (3.04)	-7.19 (4.68)	-6.67 (4.04)	-1.55	0.12
Regression slope SD	1.08 (0.57)	1.47 (0.62)	1.25 (0.65)	1.31 (0.76)	-0.57	0.58
CSID_female	35.63 (18.41)	47.35 (21.03)	38.58 (19.69)	50.91 (30.39)	-0.20	0.84
CSID_male	46.31 (18.41)	58.03 (21.03)	49.26 (19.69)	61.59 (30.39)	-0.20	0.84

CPP: cepstral peak prominence, SD: standard deviation (Hz), F0: average fundamental frequency (Hz), L/H: low/high, CSID: cepstral/spectral index of dysphonia, ADSV: analysis of dysphonia in speech and voice, VFSS: videofluoroscopic swallowing study, P/A: penetration/aspiration

[Appendix 3]. Mean and SD of PRAAT parameters about two groups pre- and post-VFSS

Parameters	Non-P/A		P/A		Z	p
	Pre-VFSS, mean (SD)	Post-VFSS, mean (SD)	Pre-VFSS, mean (SD)	Post-VFSS, mean (SD)		
F1	809.04 (203.98)	835.28 (195.32)	825.92 (158.13)	834.44 (112.05)	-0.35	0.72
F2	1420.79 (203.18)	1446.39 (182.10)	1359.22 (299.67)	1302.81 (243.04)	-2.56	0.01*
F3	2939.45 (183.54)	2954.81 (203.22)	2758.20 (158.16)	2703.18 (204.74)	-1.30	0.19
BW1	386.77 (109.57)	432.42 (159.05)	280.45 (214.55)	325.23 (203.45)	-0.05	0.96
BW2	411.25 (210.95)	411.62 (135.78)	370.60 (255.47)	337.48 (223.02)	-0.66	0.51
BW3	478.80 (319.14)	458.23 (347.06)	287.47 (134.52)	227.76 (77.31)	-0.35	0.72

*p<0.01. F1: first formant (Hz), F2: second formant (Hz), F3: third formant (Hz), BW1: first bandwidth (Hz), BW2: second bandwidth (Hz), BW3: third bandwidth (Hz), SD: standard deviation (Hz), VFSS: videofluoroscopic swallowing study, P/A: penetration/aspiration