

음성인식프로그램을 이용한 무후두 음성의 말 명료도와 병적 음성의 수술 전후 개선도 측정

이화여자대학교 의과대학 이비인후과학교실,¹
연세대학교 의과대학 이비인후과학교실, 연세대학교 음성언어의학연구소,² KT 서비스개발 연구소 음성언어연구팀³
김한수¹ · 최성희² · 김재인³ · 임재열² · 최홍식²

= Abstract =

Speech Intelligibility of Alaryngeal Voices and Pre/Post Operative Evaluation of Voice Quality using the Speech Recognition Program(HUVOIS)

Han Su Kim, MD¹, Seong Hee Choi, MD², Jae-In Kim³,
Jae-Yol Lim, MD² and Hong-Shik Choi, MD²

¹Department of Otorhinolaryngology, Ewha Womans University College of Medicine, Seoul; and

²Department of Otorhinolaryngology, The Institute of Logopedics & Phoniatics, Yonsei University College of Medicine, Seoul; and

³Service Development Laboratory, Spoken Language Research Team, KT, Seoul, Korea

Background and Objectives : The purpose of this study was to examine objectively pre and post operative voice quality evaluation and intelligibility of alaryngeal voice using speech recognition program, HUVOIS.

Materials and Methods : 2 laryngologists and 1 speech pathologist were evaluated 'G', 'R', 'B' in the GRBAS scale and speech intelligibility using NTID rating scale from standard paragraph. And also acoustic estimates such as jitter, shimmer, HNR were obtained from Lx Speech Studio.

Results : Speech recognition rate was not significantly different between pre and post operation for pathological voice samples though voice quality (G, B) and acoustic values (Jitter, HNR) were significantly improved after post operation. In Alaryngeal voices, reed type electrolarynx 'Moksori' was the highest both speech intelligibility and speech recognition rate, whereas esophageal speech was the lowest. Coefficient correlation of speech intelligibility and speech recognition rate was found in alaryngeal voices, but not in pathological voices.

Conclusion : Current study was not proved speech recognition program, HUVOIS during telephone program was not objective and efficient method for assisting subjective GRBAS scale.

KEY WORDS : Speech intelligibility · Speech recognition · Voice quality.

서 론

사람의 음성은 개인마다 매우 다른 특징을 가지고 있다. 따라서 아름다운 목소리와 듣기 거북한 목소리에 대한 경계는 객관적이기 보다는 개인의 나이, 성별, 직업, 문화 등에 따라서 다소 주관적인 것이 사실이다. 하지만 일반적으로 사

회에 통념화 되어 있는 정상음성과 병적음성이 존재하는 것은 명확한 사실이고 의학적으로 이에 대한 분석과 치료에 대한 많은 연구가 이루어져 왔다.

병적음성의 치료에 있어서 치료의 기본 조건은 음성의 질을 평가하는 것으로 전체적인 음성을 평가하는 방법으로는 일본 음성언어의학회에서 제창한 듣기 평가(Psycho-acoustic evaluation ; GRBAS grading system)가 널리 사용되고 있다(Table 1). 이는 검사자가 피검사자의 음성을 듣고 심한 정도를 0(정상), 1(경도), 2(중등도), 3(고도)의 4등급으로 표시는 방법이다(예 : G₂R₁B₂A₀S₀) 이 GRBAS grading system은 음성을 전반적으로 평가한다는 장점이 있으나, 검사자에게 많은 임상 경험을 요하며 검사자마다 평

논문접수일 : 2004년 10월 4일
심사완료일 : 2004년 10월 25일
책임저자 : 최홍식, 135-720 서울 강남구 도곡동 146-92
연세대학교 의과대학 이비인후과학교실, 음성언어의학연구소
전화 : (02) 3497-3461 · 전송 : (02) 3463-4750
E-mail : hschoi@yumc.yonsei.ac.kr

Table 1. Psycho-acoustic evaluation ; GRBAS grading system

G(grade)	Hoarseness(애성)의 정도 (hoarseness=roughness+breathiness)
R(rough)	음성의 거친소리(roughness) 정도 (roughness=jitter+shimmer)
B(breathy)	음성의 바람새는 소리(breathiness) 정도 (breathiness=SNR or HNR or NNE)
A(asthenic)	음성의 허약한(asthenic) 정도
S(strained)	음성의 힘들어가는(strained) 세기 정도

가의 정도가 다를 수 있으므로 객관적으로 비교 연구를 위한 수치로 사용하는 데는 장애가 있다.

한편, 음성 인식에 대한 연구는 40여 년 전부터 시작된 분야로 국내에서도 1980년도에 들면서부터 본격적인 음성 인식에 관한 연구가 이루어져 왔다. 이미 자동통역시스템, 증권정보 안내시스템, 음성구동 퍼스널 컴퓨터, 음성구동 셀룰러폰, 음성메모 장치 등이 상용화 중에 있다. 이 중 KT(舊 한국통신)에서 개발한 HUIVOIS-VAD(Voice Activated Dialing)는 전화번호 자동안내를 위해 개발된 시스템으로 현재 상용화되고 있다. 이에 본 연구자들은 이 시스템을 trained listener로 활용하여 음성의 질을 평가하는 객관적인 방법을 모색하고자 한다.

대상 및 방법

1. 대 상

1) 제 1실험군

2002년 7월부터 2003년 5월까지 영동세브란스 병원 이비인후과에 음성장애를 주소로 내원한 환자 중 후두 내시경상 병적 소견(성대폴립, 성대결절, 라인케스부종, 성대마비)으로 후두 현미경하미세수술 또는 제 I형 갑상연골성형술 등 음성수술을 시행 받고 술 전과 술 후 2개월에 음성정밀검사와 음성인식을 검사를 시행한 환자를 대상으로 하였다.

2) 제 2실험군

후두암으로 후두 전절제술을 시행 받은 환자로 현재 전기인공후두, 기관식도천자관, 피리형 인공후두 등 보조 장치 또는 식도 발성을 사용하고 있는 후적자들을 대상으로 하였다.

2. 방 법

1) 청지각적 평가

약 1분 20여 초의 길이에 해당하는 '가을' 문장 읽기를 Lx Speech studio(Laryngograph Ltd, London, UK) 프로그램에서 녹음한 후 이를 숙련된 음성외과의 2명과 음

Table 2. Phonetically balanced words

청와대	그야말로	됐습니다	자유와	과연이
컴퓨터	예컨대	아니냐는	주위의	되풀이
그에게	분야에서	야당의	최악이	넋물을
위대한	어두운	낙화암	의욕을	의외로
당뇨병	소프트웨어	요컨대	뭐라고	거예요

성 언어 치료사 1명이 평가하였다.

(1) 음질 평가

평가항목은 GRBAS scale의 'G', 'R', 'B' 세 항목으로 하였으며 평가척도는 0~3점까지 4점 척도(정상=0점, 경도=1점, 중등도=2점, 고도=3점)를 기본으로 0.5단위 평정도 허용하였으며 두 평가치의 평균을 분석에 이용하였다. 평가자 간 1.5점 이상 차이가 나는 평정의 경우 재평가를 하였으며 조정될 경우 조정된 점수를 사용하였다. 재평가 후에도 같은 평정 결과를 보이면 그대로 사용하였다.¹⁾

(2) 말 명료도 평가

말 명료도를 평가하기 위하여 NTID(National Technical Institute for the Deaf, USA) Intelligibility Rating²⁾을 사용하였다. 평가척도는 1~5점까지 5점 척도인데, 1점은 말을 전혀 이해할 수 없다. 2점은 말을 이해하기가 매우 어려우나, 특정단어나 절에서만 이해할 수 있다. 3점은 말을 이해 하기가 어렵지만, 내용의 요지는 이해할 수 있다. 4점은 몇몇 단어나 구를 제외하고 말이 명료하다. 5점은 말이 매우 명료하다는 기준으로 평가하였다.

2) 음향학적 검사

양측 갑상연골 부위에 전기성분과형검사를 위한 전극 밴드를 착용한 후 '가을' 문장 읽기를 Lx Speech studio에 녹음하여 SPEAD 프로그램을 이용하여 분석하였다. 분석으로 추출한 변수는 jitter, shimmer, 신호대잡음비(HNR)이었다.

3) 음성인식률검사

HUIVOIS-VAD는 전화사용자를 대상으로 개발된 시스템이므로 자료 입력은 전화선을 통하여 KT 본사의 시스템에 직접 접속하는 방식을 이용하였다. 어음 자료는 주어진 25개의 PB word(PBW-Phonetically Balanced Word)(Table 2)를 전화 수화기에 대고 읽은 후 몇 개의 단어를 인식하는지를 백분율로 구하였다. 후적자의 경우 전화 수화기를 통한 자료 입력이 힘들어 디지털녹음기의 일종인MD(Sony, Japan)로 녹음한 후 HUIVOIS-ASR(Automatic Speech Recognizer) 시스템에 곧바로 입력하였다.

4) 신뢰도

청각적 평가의 경우, 평가자 간, 평가자 내 신뢰도를 평가하기 위해 Cohen의 Kappa공식을 사용하였으며, 평가자 내 신뢰도를 측정하기 위해 총 측정치 중 20%를 1주일 후 재평가하였다. 평가자 간, 평가자 내 신뢰도는 음질평가(G-RABS)에서는 각각 .75, .84였고, 말명료도에서는 각각 .92, .97이었다.

5) 통계

병적 음성에서 술 전후 음성인식률과 말 명료도, 음향학적 측정치 간에 차이가 있는지 확인하기 위해 paired t-test를 실시하였으며, 술 전·후 음성인식률 변화와 청지각적인 측정치와 음향학적 측정치의 개선도와의 관련성을 확인하기 위하여 상관분석을 실시하였다.

무후두 음성에서는 음성인식률과 말 명료도 간의 관계를 살펴보기 위하여 상관 분석을 실시하였고, 음성인식률과 음향학적 측정치(jitter, shimmer, HNR)의 관계, 말 명료도와 음향학적 측정치(jitter, shimmer, HNR) 간의 관계를 살펴보기 위하여, 회귀 분석을 실시하였다.

결 과

1. 대 상

1) 제 1실험군

총 21명으로 남자 13명 여자 8명이었다. 성대마비가 10명으로 가장 많았으며 성대폴립이 7명, 후두 과각화증이 1명이었고 성대결절, 성대구증, 라인케스부종이 각 1예씩 있었다. 성대마비 환자 10명은 모두 제 I형 갑상연골성형술을 시행 받았으며 성대구증은 성대 내 주입술을, 이외 나머지 환자들은 후두 현미경하미세수술을 시행받았다.

2) 제 2실험군

식도발성이 5명으로 가장 많았으며 기관식도천자관 발성이 3명, 경부형 전기인공후두가 2명, 피리형 인공후두가 1명으로 총 11명이었다. 전기인공후두는 사용기간이 평균 9.3년, 기관식도 발성은 5.6년, 피리형 인공후두는 6년, 식도발성은 6.7년으로 숙련된 무후두 음성 사용자였다.

2. HUVOS-VAD 음성인식프로그램을 이용한 병적 음성의 술 전·후 인식률

술 전 음성인식률은 89.81%였고, 술 후 음성인식률은 90.76%로 술 후 음성인식률이 증가하였으나, 통계적으로 유의한 차이는 없었다($p>.05$).

Table 3. Speech recognition rate and speech intelligibility of pre and post operation

	Pre-op	Post-op
Speech recognition rate (%)	89.81 ± 5.06	90.76 ± 4.83
Speech intelligibility (%)	4.05 ± .74	4.38 ± .67

Table 4. GRBAS scale of pre and post operation

	Pre-op	Post-op
G	1.83 ± .67	1.05 ± .65*
R	1.17 ± .83	0.81 ± .62
B	1.52 ± 1.08	0.67 ± 0.14*

* $p<.05$

Table 5. Acoustic values of pre and post operation

	Pre-op	Post-op
Jitter (%)	14.04 ± 22.0	1.49 ± 2.39*
Shimmer (%)	20.10 ± 21.44	11.44 ± 11.37
HNR (dB)	14.84 ± 8.46	21.52 ± 4.63*

3. 병적 음성의 술 전·후 청지각적 평가

1) 말 명료도

술 전 청지각적 말 명료도는 4.05, 술 후 말 명료도는 4.38로 증가하였으나, 통계적으로 유의한 차이는 없었다($p>.05$) (Table 3).

2) 음 질

전체적인 애성 정도를 나타내는 'G'는 술 전은 1.833, 술 후는 1.048로 통계적으로 유의하게 음질이 개선되었다($p<.05$). 음성의 거친 정도를 나타내는 'R'은 술 전은 1.17, 술 후는 0.81로 음질이 개선되었으나, 통계적으로 유의한 차이는 없었다($p>.05$). 음성의 기식성 정도를 나타내는 'B'는 술 전은 1.52, 술 후는 0.67로 통계적으로 유의하게 음질이 개선되었다($p<.05$) (Table 4).

4. 병적 음성의 술 전·후 음향학적 측정치

Jitter는 술 전 14.01%, 술 후 1.49%로 통계적으로 유의하게 개선 되었으며($p<.05$), shimmer는 술 전 20.11%, 술 후 11.44%로 개선되었으나, 통계적으로 유의한 차이는 없었다($p>.05$). HNR은 술 전 14.84%, 술 후 21.52%로 통계적으로 유의하게 개선되었다($p<.05$) (Table 5).

5. 음성인식률과 청지각적 측정치, 음향학적 측정치 간의 상관관계

술 전·후 음성인식률 변화와 청지각적 말 명료도 개선도와의 상관 계수는 .303으로 통계적으로 유의한 상관관계가 없었으며, 음성인식률 변화와 G, R, B간의 상관계수는 각각 -.16, .04, -.25로 이들 음질 측정치와 말 인식률 간에 통계적으로 유의한 차이가 없었다($p>.05$).

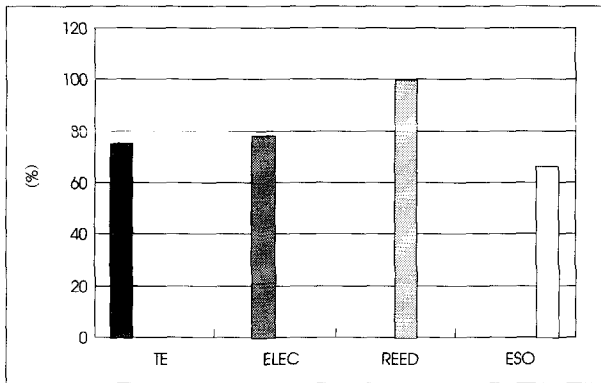


Fig. 1. Speech recognition rate of alaryngeal speech using HUVOIS. The recognition rate of the reed type voice is 100%. TE : Tracheoesophageal shunt voice, ELEC : Electric larynx (neck type), REED : reed type voice, ESO : Esophageal voice.

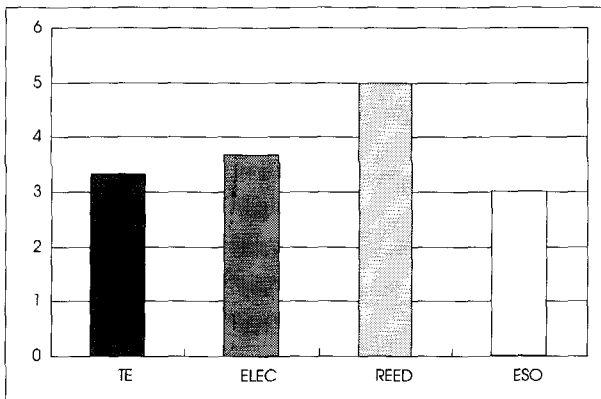


Fig. 2. Speech intelligibility of alaryngeal speech (NTID rating scale). TE : Tracheoesophageal shunt voice, ELEC : Electric larynx (neck type), REED : Reed type voice, ESO : Esophageal voice.

술 전·후 음질 G, R, B의 개선도와 말 명료도 개선 정도와의 상관 계수는 각각 -.21, -.39, -.04로 통계적으로 유의한 차이가 없었다($p>.05$).

6. HUVOIS-ASR 음성인식프로그램을 이용한 무후두 음성의 인식률

무후두 음성의 인식률은 피리형 인공후두가 100%로 가장 인식률이 높았고, 전기인공후두 78.75%, 기관식도천자관 발성 75.17%, 식도발성 66.34% 순이었다(Fig. 1).

7. 무후두 음성의 말 명료도

무후두 음성의 말 명료도는 피리형 인공후두가 5점으로 말이 매우 명료하였으며, 전기인공후두 3.5점, 기관식도천자관 발성 3.33점, 식도발성 3점 순이었다(Fig. 2). 무후두 음성의 HUVOIS-ASR을 이용한 음성인식률과 청지각적 말명료도 간의 상관계수는 .96로 유의한 상관관계를 보였다.

8. 무후두 음성의 음향학적 측정치와 말 명료도

무후두 음성의 음향학적 측정치는 jitter(%)의 경우 3.68%

Table 6. Acoustic values of alaryngeal speech

	Jitter(%)	Shimmer(%)	HNR(%)
ELEC	.55 ± .48	.05 ± .87	12.69 ± 1.18
ESO	3.68 ± 1.86	12.25 ± 7.30	11.41 ± 6.88
REED	.27 ± .32	1.02 ± .42	23.27 ± 3.26

TE : Tracheoesophageal shunt voice, ELEC : Electric larynx (neck type), REED : Reed type voice, ESO : Esophageal voice

로 식도 발성이 가장 높았고, shimmer(%)도 12.25%로 식도 발성이 가장 높았다. HNR은 식도발성이 제일 낮고, 피리형 인공후두 '목소리'가 23.27%로 가장 높았다(Table 6). 무후두 음성의 말 명료도와 음향학적 측정치(jitter, shimmer, HNR)과의 관련성을 살펴보기 위하여 회귀 분석을 실시한 결과, 세 음향학적 측정치의 변수가 말 명료도를 설명하는 정도는 21.8%($r^2=.218$ 였고), 음성인식률을 설명하는 정도는 26.4%($r^2=.264$)로 낮아 말 명료도, 음성인식률에 영향을 미치는 더 중요한 음향학적 변수가 있을 것으로 보인다.

고 찰

사람의 목소리를 구성하는 요소는 매우 다양하고 복잡하다. 따라서 음성의 질을 평가하고 음성장애를 이해하기 위해서는 먼저 음성에 대한 정량적인 검사가 이루어져야 한다. 현재 이용되고 있는 음성 기능 평가 방법으로는 청지각적검사, 음향학적검사, 공기역학적검사 등이 있다.³⁾

현재 상업적으로 생산되어 사용되고 있는 각종 음성 분석 기기는 음성의 구성요소를 미분화하여 음향학적 또는 공기역학적 변수를 정의하고 이를 분석하는 데는 많은 발전을 가져왔으나 음성의 질을 종합적으로 청취자적 입장(Listener's viewpoint)에서 객관적으로 평가하는 방법으로는 아직 미흡한 실정이다. 현재 널리 사용되는 있는 대표적인 청지각적 검사인 GRBAS 스케일은 주관적인 검사방법으로 검사자에게 많은 임상 경험을 요하며 검사자마다 평가의 정도가 다를 수 있으므로 객관적으로 비교 연구를 위한 수치로 사용하는 데는 장애가 있다.

본 연구에 사용한 HUVOIS는 KT에서 개발한 음성인식 시스템으로 인식기내에 데이터베이스화되어 있는 기존의 단어 정보와 입력된 음성 정보를 비교하여 음성을 인식하는 방법이다. 이런 시스템의 경우 비슷한 단어가 입력되거나 입력 신호에 포먼트를 교란시키는 잡음 구성비가 높을 때 인식률이 낮아지는 한계가 있다. 본 연구자들은 이런 한계에 착안하여 성대점막 병변 또는 성대마비 환자에서 수술치료로 음성의 질이 향상되었을 때 음성인식률이 향상될 것이라는 가정하에 음성인식기를 객관적 청취자(Objective listener)

로 활용하는 방법을 모색하고자 하였다.

수술 전후에 음성 장애의 정도를 나타내는 변수로 음향학적 검사 변수인 jitter, shimmer, HNR을 측정하였는데 이는 기존의 연구에서 이 변수의 변화가 의미가 있는 것으로 보고되었기 때문이다.⁴⁾⁵⁾ 공기역학적 검사도 시행하였으나 이를 분석에 이용하지는 않았는데 이는 성대점막 병변 환자와 성대마비 환자의 측정치가 상당히 상이하기 때문이다.

음성인식이란 전화, 또는 마이크를 통하여 입력된 사람의 음성을 컴퓨터가 분석하여 특징을 추출하고 미리 입력된 단어 또는 문장에서 가장 근접한 결과를 출력하는 최첨단 소프트웨어 기술이다. 음성인식 기술을 분류해 보면 적용 화자에 따라 화자종속과 화자독립으로 나뉘며, 인식대상의 어휘 수에 따라서 소어휘 인식기술과 대어휘 인식기술로, 인식 수준에 따라 단어 단위로 인식하는 단어인식기술, 음소 단위로 인식하는 가변어인식기술, 문장단위로 인식하는 연속음성인식기술, 문장에서 핵심단어만 인식하는 핵심어인식기술 분류할 수 있다.⁶⁾ 본 연구에 이용된 HUVOIS 시스템은 전화로 걸려 오는 불특정 다수의 대용량 단어를 음소 단위로 분석하므로, 화자독립, 대어휘, 가변어 인식기술이라 할 수 있다.

결과를 보면 제 1실험군의 경우 술 전에 비해 술 후에 인식률의 향상이 통계적으로 유의하지가 않다. GRABS scale에서는 음성의 질이 호전되었으나 인식률에는 차이가 없는 이유는 무엇일까? 이는 아마도 음성인식 기술의 원리 때문일 것이다. 본 연구에 이용된 HUVOIS 시스템 뿐만 아니라 현재 사용되고 있는 대부분의 음성인식기는 패턴 매칭 알고리즘(pattern matching algorithm)을 이용한다. 이 알고리즘은 음성 패턴의 특징이 발생자 및 발생시간에 따라 변하는 것이 아니라 음성의 의미에 따라서만 변한다는 가정을 전제로 한다.⁷⁾ 즉 사람이 발생하는 특정 단어의 신호를 패턴화 하여 컴퓨터 메모리 안에 기억시킨 다음 새로운 음성이 입력되면 입력된 음성이 저장되어 있는 패턴들 중 어느 것과 가장 유사한 지를 판단하여 인식하게 되는 것이다.⁸⁾ 따라서 각 음성신호가 가지고 있는 특성을 계수화하여 특정 벡터를 추출 함으로써 패턴을 만드는 과정이 중요한데 현재까지는 보통 네 가지의 방법이 이용되고 있다.⁹⁾ 이 중 본 연구에서 사용한 시스템은 LPC(Linear Predictive Coding)에 기반을 둔 cepstral 계수를 이용한 방법을 쓰고 있다. 이 방법에서 기준이 되는 패턴정보는 음소내 '시간적인 포먼트정보'와 '음소길이정보'이다. 이 과정은 보통 20msec 단위로 음성신호를 블록화하여 각 frame별로 특징벡터를 계산해 내는데 그 각각의 포먼트가 시간적으로 어떻게 변동되는지를 수식적으로 표현한 것이 시간적인 포먼트정보이

다. 분석 시 보통 10msec씩 중첩을 시켜 분석하기 때문에 예를 들어 1초가 넘는 단어의 경우 100 frame이 넘는 포먼트정보가 추출된다. 이후 삼음소(triphone) 별로 묶은 후 중심 음소를 기준으로 좌우측에 어떤 음소가 존재하는지에 따라 음소정보를 새로이 구성하여 인식대상 단어들에 대한 음성 모델들을 형성한다. 이 후 임의의 음성정보가 들어오면 이것과 음성 모델들을 비교하여 가장 비슷한 것을 찾아 인식하게 되는 것이다.

그러므로 성대 점막 병변 또는 성대 마비를 가진 환자가 의사소통이 가능한 발성을 하였다면 비록 이것이 귀에 듣기에 나쁜 음질을 가진 음성정보 일지라도 자체 내에는 포먼트 정보와 음소정보를 가지고 있으므로 음성인식을 하는 데는 별문제가 없는 것이다. 이런 이유로 제 1실험군의 술 전 인식률이 높게 나오게 된 것이다. 그런데 제 2실험군의 경우에는 인식률이 제 1실험군에 비해 좋지 않을 것을 볼 수 있다. 본 연구에서 술 전·후 말 명료도도 통계적으로 의미 있는 차이를 보이지 않았는데, 음질의 측정 도구였던 GRBAS경우, 애성의 정도를 나타내는 'G'와 'B'는 유의하게 음질이 개선되었으나, 'R'은 유의한 차이를 보이지 않았으며, 이는 음성장애 환자의 음질 특히, 기식성(breathiness)의 정도는 말 명료도에 큰 영향을 주지 않음을 시사해준다. 후적자의 재할 음성의 말 명료도를 살펴본 결과, 피리형 인공후두기인 'Moksori'가 명료도가 매우 높았고, 전기인공후두, 기관식도천자관 발성, 식도발성 순으로 나타났으며, 본 연구에 사용된 음성인식프로그램에서도 피리형 인공후두기 'Moksori'는 인식률이 100%였으며, 전기인공후두, 기관식도천자관발성, 식도발성 순으로 동일한 순위 분포를 나타내었다. 후적자의 재할 음성은 잘 알고 있는 것처럼 일반인에 비해 유창성이 떨어지고 실제로 발화시간도 길다. 무후두 음성의 말 명료도에 영향을 미치는 요인은 기본주파수, 말속도, 강도, 거친 음질, 소음, 발화기간, 최대 발생시간과 같이 여러 요인이 있지만,¹⁰⁾ 그 중에서 말속도와 같이 시간과 관련된 요인과 signal-to-noise ratio가 무엇보다도 중요하다고 하였다.¹¹⁾ 위에서도 설명했지만 패턴 매칭 알고리즘은 음성 패턴의 특징이 발생자 및 발생시간에 따라 변하는 것이 아니라 음성의 의미에 따라서만 변한다는 가정을 전제로 하는데 후적자의 음성정보는 이러한 가정 자체를 교란 시키기 때문에 음성 인식률이 저하되는 것으로 보인다. 또한, 무후두 음성의 말 명료도, 음성인식률과 음향학적 측정치(jitter, shimmer, HNR)와의 관련성은 낮은 것으로 나타났는데, 말 명료도와 음성인식률에 영향을 미치는 다른 중요한 음향학적 변수들이 더 있을 것으로 보이며, 이러한 변수를 찾는 일은 무후두 음성의 음성합성 및 음성

인식, 말 명료도 개선에 도움을 줄 것으로 보인다.

결 론

이번 연구에서는 음성인식 시스템이 GRABS 스케일의 단점인 주관적 측정 방법을 보완하는 객관적이고 사용이 편한 효율적인 방법으로 확인되지는 않았다. 하지만 현재 상용화가 연구되고 있는 사람의 청각 특성을 주파수 영역에서 고려한 MFCC(Mel-frequency cepstral coefficients) 방법이 현실화 된다면 좋은 결과가 있을 것으로 생각한다. 또한 본 연구에서 보듯이 후적자의 음성은 아직 현 음성인식 시스템에서는 미완의 영역으로, 이에 대한 음성인식률 향상 연구는 현재 음성 인식 분야의 큰 걸림돌인 병적 음성 및 소음 환경음성의 인식 기술 연구에 도움이 될 것으로 기대된다.

중심 단어 : 말명료도 · 음성인식 · 음질.

본 연구는 2002년도 연세대학교 학술연구비의 지원으로 이루어 졌음.

REFERENCES

- 1) Pyo HY, Choi SH, Lim SE, Sim HS, Choi HS, Kim KM. *The correlation between GRABS scale and MDVP parameters on the pathologic voices of the patients with vocal polyps. J Korean Soc Logo Phon* 1999;10:154-2.
- 2) Kent RD. *Intelligibility in speech disorders: theory, measurement and management. 1992 John benjamiins publishing com. Philadelphia.*
- 3) Hong KW. *Aerodynamics of speech using Aerophone II. J Korean Logo Phon* 1995;6:165-72.
- 4) Wolfe V, Fitch J, Cornell R. *Acoustic prediction of severity in commonly occurring voice problems. J Speech Hear Res* 1995;38:273-4.
- 5) Kim YM, Cho JI, Kim CH, Kim YJ, Ha HR. *Vocal dynamics studies before and after laryngeal microsurgery. Korean J Otolaryngol* 1999; 42:1174-8.
- 6) 김형순. 음성인식. 제 1회 음성학 학술대회;1994. p.156-65.
- 7) 이기희 · 박두석. 음성인식기술과 현황. 한국OA 학회지 1997; 4:1-11.
- 8) 김희린. 음성인식 기술. 한국멀티미디어학회지 2003;7(2):16-22.
- 9) 이기희 · 박두석. 음성인식기술과 현황. 한국OA 학회지 1997; 4:1-11.
- 10) Finizia C, Dotevall H, Lundstrom E, Lindstrom J. *Acoustic and perceptual evaluation of voice and speech quality: a study of patients with laryngeal cancer treated with laryngectomy vs irradiation. Arch Otolaryngol Head Neck Surg* 1999;125(2):157-63.
- 11) Prosek RA, Vreeland LL. *The intelligibility of time-domain-edited esophageal speech. J Speech Hear Res* 2001;44:525-34.