

정상 음성의 목소리 특성의 정성적 분류와 음성 특징과의 상관관계 도출

Qualitative Classification of Voice Quality of Normal Speech and Derivation of its Correlation with Speech Features

김 정 인¹⁾ · 권 철 홍²⁾

Kim, Jungin · Kwon, Chulhong

ABSTRACT

In this paper voice quality of normal speech is qualitatively classified by five components of breathy, creaky, rough, nasal, and thin/thick voice. To determine whether a correlation exists between a subjective measure of voice and an objective measure of voice, each voice is perceptually evaluated using the 1/2/3 scale by speech processing specialists and acoustically analyzed using speech analysis tools such as the Praat, MDVP, and VoiceSauce. The speech parameters include features related to speech source and vocal tract filter. Statistical analysis uses a two-independent-samples non-parametric test. Experimental results show that statistical analysis identified a significant correlation between the speech feature parameters and the components of voice quality.

Keywords: normal speech, voice quality, speech features, qualitative classification, correlation

1. 서론

음성장애 환자들을 대상으로 진단과 치료를 위하여, 음의 높낮이(pitch), 음의 세기(intensity), 음성 품질(voice quality)과 유동성(flexibility) 등을 음성 검사를 통하여 측정하고 분석하는 연구들이 진행되어 왔다. 그런데 음성장애 환자가 아닌 정상적인 목소리를 갖고 있는 사람들을 대상으로 목소리 특성을 연구하는 시도는 많지 않다. 음성은 상기 각각의 요소들이 적절히 조화를 이루어야 좋은 음성으로 평가받을 수 있다.

사람의 음성을 평가할 때 청취 평가를 통한 주관적 평가방법과 음성 분석을 활용한 객관적 평가방법을 사용할 수 있다. 청취 평가 방법은 사람의 목소리 특성을 몇 가지 항목으로 나

누고 청취 평가자가 각 항목에 점수를 할당하는 방식으로 이루어진다. 객관적 평가 방법은 음성 처리 알고리즘을 직접 프로그래밍 하거나 음성 분석 툴을 사용하여 음성 특징 파라미터를 추출하고 연속적인 값으로 수치화 하여 나타낸다.

주관적 청취 평가 방법으로 GRBAS scale, CAPE-V (Consensus Auditory Perceptual Evaluation of Voice), VPA(Vocal Profile Analysis) 등이 있는데, 이 중에서 일본음성언어의학회에서 제안한 음성 검사법인 GRBAS 음성평가가 가장 널리 사용되고 있다[1][2][3]. 이 방법은 목소리 특성을 다섯 개의 항목으로 나누어 평가하는데 Grade(전반적인 애성의 정도), Roughness(거친 소리, 성대의 불규칙한 진동), Breathiness(기식성, 바람 세는 소리), Asthenia(무력성, 연약한 가냘픈 소리), Strain(긴장성, 힘을 들여서 내는 소리) 등으로 구성되어 있다. 각 항목에 대하여 0, 1, 2, 3 등 네 단계로 음성을 평가하는데, 0은 정상 음성이고 3은 가장 좋지 않은 상태를 말한다. 이 방법은 음성장애 환자를 대상으로 평가하는 방법이고, 본 논문에서는 정상인의 목소리 특성에 초점을 맞추므로 목소리 특성을 다르게 분류하는 방법에 대한 연구가 필요하다.

본 논문에서는 정상인의 목소리 특성을 여러 가지로 분류

1) 대전대학교 in10408@naver.com

2) 대전대학교 chkwon@dju.ac.kr, 교신저자

이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2013R1A1A4A01005974).

접수일자: 2014년 2월 5일

수정일자: 2014년 3월 6일

게재결정: 2014년 3월 11일

하고, 이 특성과 관계 깊은 음성 특징 파라미터를 도출해 보고자 한다. 즉, 목소리 특성과 음성 특징 파라미터 간의 상관 관계를 규명하고자 한다.

2. 목소리 특성 분류

사람의 목소리 특성은 기본적으로 빠르기(느리다/빠르다), 높낮이(낮다/높다), 세기(작다/크다) 등으로 분류해 볼 수 있는데, 이들 특성은 비교적 객관적인 특징을 갖는 항목들로서 각각 발화속도, 피치, 음의 크기 등으로 간단하게 측정 가능하다. 이와는 다른 분류로, 탁하다, 웅장하다, 침중하다, 경쾌하다, 강하다, 맑다, 낭랑하다, 명랑하다, 무겁다, 편안하다, 가볍다, 급하다, 우렁차다, 차분하다, 방정하다 등으로 말하는 이의 목소리 특성을 분류해 볼 수 있는데, 이러한 목소리 특성과 객관적으로 관련 있는 음성 특징 파라미터를 연관 짓기는 쉽지 않다.

이와 같이 사람의 목소리 특성을 다양하게 분류할 수 있지만, 본 논문에서는 목소리 특성을 나눌 때 다음과 같은 사항을 고려한다. 서론에서 기술한 GRBAS 음성평가 방법에서 분류한 항목을 고려한다. 분류 항목의 개수를 무리하게 많이 하지 않고 항목을 나누는 기준이 비교적 분명하게 설명될 수 있는 특성을 선정한다. 항목 간에 목소리 특성이 유사하여 평가자에게 혼란을 줄 수 있는 항목을 통합하여 정의가 중복되지 않는 것을 선정한다. 음성을 청취 평가할 때 분류하기가 용이하고 청취 평가자 사이에서 의견 일치율이 높은 목소리 특성을 선정한다. 본 연구의 목적에 맞게, 객관적으로 음성 특징 파라미터와 관련지을 수 있는 목소리 특성을 선정한다. 본 논문에서 이러한 기준으로 선정한 목소리 특성은 숨소리, 짜내는 소리, 거친 소리, 콧소리, 굽기(가늘다/굵다) 등이다. 이러한 목소리 특성의 분류와 GRBAS 분류를 비교해 보면, 숨소리는 Breathiness, 짜내는 소리는 Strain, 거친 소리는 Roughness와 유사하다고 볼 수 있으나, GRBAS에서 Asthenia가 제외되고 콧소리와 굽기는 GRBAS에서는 분류하지 않는 목소리 특성이다. 정상인의 목소리 중에서 평균에 가까운 음색을 중간으로 분류하고 그보다 많은지 적은지 세 단계로 판정한다. 여기에서 많다고 판정한 것이 장애인의 목소리 수준을 의미하는 것은 아니고 정상인의 평균 보다 많음을 의미한다.

3. 음성 특징 파라미터

앞에서 살펴보았듯이 빠르기, 높낮이, 세기와 관련된 음성 특징 파라미터는 쉽게 찾을 수 있으나, 본 논문에서 분석할 목소리 특성과 관련된 음성 특징 파라미터가 무엇인지에 대해서는 연구할 필요성이 있다. 따라서 음성공학 분야뿐만 아니라 인문음성학 분야, 음성의학 분야 등의 논문을 조사하여 이

러한 목소리 특성과 관련이 깊은 음성 특징 파라미터를 찾아 보고자 한다.

본 논문에서는 음성신호의 음원과 관련된 파라미터와 성도 필터와 관계된 파라미터들을 이용하여, 목소리 특성과 관련 있는 음성 특징 파라미터를 도출하고자 한다. <표 1>에 보이는 음성신호의 음원과 관련된 파라미터에는 기본주파수에 관한 파라미터(F0_mean, F0_std), 에너지 크기(Intensity), 기본주파수 변화율(Jitter), 진폭 변화율(Shimmer), 잡음과 하모닉의 에너지 비율(NHR, VTI), 숨소리(breathiness)의 정도(SPI), 주기성의 강도(CPP), 하모닉에 관한 파라미터(H1-H2) 등이 있다. 음성신호의 성도필터와 관계된 파라미터에는 포먼트 주파수 F1~F4, 각 포먼트의 대역폭 BW1~BW4, LTAS_slope 등이 있다[4][5].

표 1. 음성 특징 파라미터

Table 1. Speech feature parameters

음성 특징 파라미터	설명
F0_mean	기본주파수 평균
F0_std	F0 표준편차
Intensity	에너지 크기
Jitter	기본주파수 변화율
Shimmer	진폭 변화율
NHR	잡음 대 하모닉의 에너지 비율
VTI	잡음 대 하모닉의 에너지 비율
SPI	숨소리의 정도
CPP	주기성의 강도
H1-H2	첫 번째와 두 번째 하모닉의 진폭 차이
F1~F4	포먼트 주파수
BW1~BW4	각 포먼트의 대역폭
LTAS_slope	주파수에 따른 스펙트럼 기울기

음성신호의 음원과 관련된 파라미터에서, F0_mean은 기본주파수의 평균값이고, F0_std는 기본주파수의 표준편차이다. Intensity는 음성신호의 에너지 크기(dB)를 나타내는 변수로 목소리 크기를 나타낸다. Jitter는 기본주파수 변화율로서 연속적인 피치주기의 평균변화율을 보여 주고 MDVP에서 정상인과 음성장애인의 경계 값을 1.04%로 제시한다[6]. Shimmer는 진폭 변화율로서 연속적인 진폭변화율의 평균값을 나타내고 MDVP에서 제시하는 정상인의 경계 값은 3.81%이다[6]. 음성신호에 존재하는 잡음의 정도를 보여주는 파라미터에는 NHR(Noise to Harmonic Ratio)과 VTI(Voice Turbulence Index)가 있는데, NHR은 1,500 ~ 4,500 Hz의 잡음 에너지와 70 ~

4,500 Hz의 하모닉 에너지의 비율이고, VTI는 NHR 보다 주파수 대역이 높은 2,800 ~ 5,800 Hz의 잡음 에너지와 70 ~ 4,500 Hz의 하모닉 에너지의 비율이다. SPI(Soft Phonation Index)는 숨소리의 정도를 보여 주는 변수로, 저주파(70 ~ 1,600 Hz) 대역의 하모닉 에너지와 고주파(1,600 ~ 4,500 Hz) 대역의 하모닉 에너지의 비율이다. CPP(Cepstral Peak Prominence)는 주기성의 강도를 보여 주는 변수로, 피치주기에 해당하는 위치에서 cepstral peak의 진폭 크기를 나타낸다. H1과 H2는 첫 번째와 두 번째 하모닉의 진폭을 나타내고, H1과 H2의 차이(H1-H2)는 기식음(breathy voice)의 정도를 구분하는 데 사용할 수 있다[7][8].

음성신호의 성도필터와 관련된 파라미터에서, F1~F4는 각각 첫 번째부터 네 번째 포먼트 주파수를, BW1~BW4는 각 포먼트의 대역폭을 나타낸다. 공명의 특성과 관련된 포먼트 대역폭은 양질의 발성에서는 포먼트 대역폭이 좁아지는 특징을 갖고 있다[6]. LTAS_slope(Long Term Average Spectrum Slope)은 고주파수로 가면서 장구간 스펙트럼의 기울기를 나타낸다.

4. 실험 방법

4.1 피험자 및 음성 데이터

피험자는 후두질환을 앓은 적이 없는 건강한 20대 남학생 50명을 대상으로 하였다. 모든 음성 데이터는 조용한 연구실에서 수집하였으며 마이크(AKG D880s)를 마이크 스탠드에 고정하여 입과 마이크 거리가 5cm 정도가 되도록 특별한 주의를 기울였다. 피험자는 편안히 앉은 상태에서 ‘아’ 모음을 평상시와 같은 크기와 톤으로 3초 이상 발성하고, 단문과 장문 하나를 연속해서 발성하였다. 음성 데이터는 모노 형식, PCM signed 16 bits, 샘플링 주파수 16 kHz로 수집하였다.

4.2 음성 특징 파라미터 추출 방법

음성 분석을 통하여 음성 특징 파라미터를 추출하는 데 ‘아’ 모음과 단문을 사용한다. 수집한 음성 데이터를 끝점 검출을 통해 묵음구간을 제거하고 음성 구간을 검출하여 3절에서 기술한 음성 특징 파라미터를 추출한다.

추출 대상 음성 특징 파라미터 중에서, F0_mean, F0_std, Intensity, Jitter, Shimmer, LTAS_slope, F1~F4, BW1~BW4는 Praat[9]를 이용하여, NHR, VTI, SPI는 MDVP(Multi Dimensional Voice Program)[10]를 이용하여, CPP, H1-H2는 VoiceSauce[11] 툴을 이용하여 추출하였다. Praat와 MDVP는 음성공학, 음성언어학 및 음성의학 분야에서 널리 사용되고 있는 음성 분석 툴이고, VoiceSauce는 UCLA speech processing and auditory perception lab.에서 Matlab으로 구현한 음성 분석 프로그램이다.

4.3 청취 평가 방법

모음은 개인의 목소리 특성을 파악하기에는 한계가 있으므로 문장을 이용하여 목소리 특성을 청취 평가한다.

수집한 문장 음성 데이터를 청취 평가할 때 숨소리, 짜내는 소리, 거친 소리, 콧소리, 굽기(가늘다/굵다) 각 항목에 대하여 1, 2, 3 등 세 단계로 평가를 하는데, 1은 적다, 2는 보통이다, 3은 많다는 의미이다. 즉, 각 항목에 대하여 특색이 강하지 않는 평균에 가까운 음색을 보통으로 판정하고, 이를 기준으로 적으면 1로, 많으면 3으로 판정한다. 1, 2, 3 등 세 단계로 판정을 내리는 이유는, 이 방법이 단순하면서 평가자 간의 의견 일치율을 높이기 위해서이다.

음성 분야 전문가가 목소리 특성을 5가지로 분류하고 각 항목의 특성에 대해 정의를 내리고, 청취 평가 전에 목소리 특성 각 항목의 대표 음성을 이용하여 충분히 청취 평가 훈련을 하였다. 수집한 문장에 대하여 음성공학 전공자 5인이 청취 평가를 수행하는데, 청취 후 목소리 특성 각 항목에 대하여 분류할 때 각자 판단을 하고 의견이 일치하지 않는 경우에는 수차례 반복하여 청취하면서 최종 판정을 합의하는 절차를 거친다[4].

4.4 통계 처리

목소리 특성 각 항목에 대하여 1, 2, 3 등 세 단계로 청취 평가 판정을 내리므로 분산분석(ANOVA)을 수행할 수 있다. 분산분석 방법은 세 개 이상의 모집단 평균 간의 차이를 검증하는 데 이용하는 분석방법이다. 그런데 1과 2, 2와 3은 적다와 보통, 보통과 많다이므로, 집단 사이에 특성의 차이가 크지 않아 세 집단 사이에 유의미한 파라미터를 찾는데 어려움이 있다. 따라서 특성의 차이가 비교적 크다고 할 수 있는 1과 3 두 집단 사이에 대하여 통계 처리를 수행하여, 목소리 특성 각 항목에 대하여 그 성분이 적고 많을 때 상관관계가 있는 음성 특징 파라미터를 도출하고자 한다.

녹음 대상자의 수가 50명이므로 이를 세 집단으로 분류하므로 각 집단에 속하는 데이터의 수가 적게 되어(< 30), 모집단의 분포를 정규분포로 가정할 수 없어 비모수적 검정 방법을 사용한다. 본 논문에서는 독립 2-표본 비모수적 검정 방법(Two-independent-samples non-parametric test)인 Mann-Whitney U test를 시행하는데, 이 방법은 Wilcoxon rank sum test 라고도 한다[12]. 이와 같은 통계 처리를 위하여 SPSS v.21.0[13]을 사용하는데, 집단변수에 목소리 특성 각 항목을, 검정변수에 음성 특징 파라미터를 설정하면 통계적으로 유의미한 음성 파라미터를 도출해 낼 수 있다.

5. 실험 결과

목소리 특성 각 항목에 대하여 두 집단 1과 3 사이에 독립

2-표본 비모수적 검정 방법을 수행하여 통계적으로 유의미한 파라미터를 구한 결과는 <표 2> ~ <표 6>과 같다. 유의수준은 $p < 0.05$ 를 사용하였다.

숨소리 항목에 대한 통계 분석 결과가 <표 2>에 보인다. 통계적으로 유의미한 음성 특징 파라미터는 모음 ‘아’에서 추출한 SPI, H1-H2, CPP 등이고, 단문에서 추출한 LTAS_slope이다. LTAS_slope는 절뚝값이 클수록 기울기가 급하다는 것을 나타낸다.

숨소리의 정도를 보여 주는 SPI는 집단 1에서 16.7이고 집단 3에서 29.0인데, 이는 기식음(숨소리, breathy voice)에서 SPI가 크다는 [16]의 연구결과와 일치한다. H1-H2는 집단 1에서 3.09이고 집단 3에서 5.97인데, 기식음은 비교적 큰 H1-H2를 보여 준다는 [14]의 보고와 부합한다. CPP는 집단 1에서 28.6이고 집단 3에서는 이보다 작은 27.2를 보여 주는데, 이는 기식음에서 작은 CPP 값을 갖는다는 [17]의 연구결과와 일치한다. 스펙트럼 기울기(LTAS_slope)의 절뚝값은 집단 1에서 17.9이고 집단 3에서 18.1인데, 기식음에서 큰 값을 갖는다는 [15]의 보고와 부합한다.

표 2. 숨소리에 대한 통계 분석 결과
Table 2. Statistical analysis results for breathy voice

음성 파라미터		집단 1 평균값 (표준편차)	집단 3 평균값 (표준편차)	p 값
모음 ‘아’	SPI	16.7(9.41)	29.0(2.28)	0.021
	H1-H2	3.09(1.95)	5.97(0.88)	0.007
	CPP	28.6(1.51)	27.2(0.16)	0.024
문장	LTAS_slope	-17.9(3.05)	-18.1(3.10)	0.028

짜내는 소리 항목에 대한 통계 분석 결과가 <표 3>에 보인다. 모음 ‘아’에서 추출한 BW1, Jitter, SPI 등과, 단문에서 추출한 LTAS_slope이 통계적으로 유의미한 음성 파라미터이다.

Jitter는 집단 1에서 0.337이고 집단 3에서는 이보다 큰 0.393을 보여 주는데, 과도한 Jitter는 짜내는 소리(creaky voice)를 만든다는 설명[18]과 부합한다. LTAS_slope의 절뚝값은 집단 1에서 22.2이고 집단 3에서는 이보다 작은 17.9인데, 이것은 짜내는 소리는 LTAS_slope이 작다는 연구결과[15]와 일치한다. 이 외에도 SPI는 집단 1에서 보다 집단 3에서 작은 값을 가지며, BW1은 집단 3 보다 집단 1에서 작은 값을 갖는다는 것은 다른 연구에서 도출하지 못한 의미 있는 실험결과이다.

표 3. 짜내는 소리에 대한 통계 분석 결과
Table 3. Statistical analysis results for creaky voice

음성 파라미터		집단 1 평균값 (표준편차)	집단 3 평균값 (표준편차)	p 값
모음 ‘아’	BW1(Hz)	118(64.1)	258(148.9)	0.017
	Jitter(%)	0.337(0.079)	0.393(0.038)	0.036
	SPI	24.5(11.7)	14.9(6.14)	0.029
문장	LTAS_slope	-22.2(3.28)	-17.9(2.72)	0.012

거친 소리 항목에 대한 통계 분석 결과가 <표 4>에 보인다. 통계적으로 유의미한 음성 특징 파라미터는 모음 ‘아’에서 추출한 BW1, Jitter, Shimmer, NHR, VTI 등이고, 단문에서 추출한 F0_std이다.

다른 목소리 특성 항목과는 달리, 잡음과 관련된 파라미터인 NHR, VTI가 포함되어 있는 것이 특징적이며, 특히 NHR이 거친 소리에 대한 좋은 지표라는 [19]의 연구결과와 일치한다. Jitter, Shimmer, NHR, VTI, F0_std 모두 거친 소리가 강한 집단 3이 집단 1보다 크다는 실험결과는 기존의 연구결과들에 부합한다. BW1이 집단 1 보다 집단 3에서 큰 값을 갖는다는 것은 거친 소리가 강할 때 발음의 명료성이 떨어진다는 것을 의미하는 실험결과이다.

표 4. 거친 소리에 대한 통계 분석 결과
Table 4. Statistical analysis results for rough voice

음성 파라미터		집단 1 평균값 (표준편차)	집단 3 평균값 (표준편차)	p 값
모음 ‘아’	BW1(Hz)	121(44.2)	231(92.5)	0.016
	Jitter(%)	0.316(0.080)	0.504(0.115)	0.028
	Shimmer(%)	2.30(0.876)	3.41(0.861)	0.028
	NHR	0.134(0.013)	0.167(0.022)	0.028
	VTI	0.032(0.006)	0.055(0.010)	0.004
문장	F0_std	21.6(3.89)	27.6(1.08)	0.004

콧소리 항목에 대한 통계 분석 결과가 <표 5>에 보인다. 통계적으로 유의미한 음성 파라미터는 모음 ‘아’에서 추출한 BW3, BW4, H1-H2 등이고, 단문에서는 통계적으로 유의미한 파라미터가 없다.

BW3과 BW4 모두 집단 1에서 보다 집단 3에서 큰 값을 갖는다는 실험결과가 나왔는데, 이것은 비음의 포먼트 대역폭은

상대적으로 넓은 설명[18]과 부합한다. 이 외에도 H1-H2는 집단 1에서 보다 집단 3에서 큰 값을 갖는다는 것은 의미 있는 실험결과이다.

표 5. 콧소리에 대한 통계 분석 결과
Table 5. Statistical analysis results for nasal voice

음성 파라미터		집단 1 평균값 (표준편차)	집단 3 평균값 (표준편차)	p 값
모음 '아'	BW3(Hz)	246(75.7)	371(170.1)	0.002
	BW4(Hz)	172(68.5)	335(209.0)	0.003
	H1-H2	2.26(1.79)	3.88(1.91)	0.024

가늘다/굵다 항목에 대한 통계 분석 결과가 <표 6>에 보인다. 상대적으로 가는 목소리가 집단 1이고 상대적으로 굵은 목소리는 집단 3이다. 모음 '아'에서 추출한 F0_mean, H1-H2 등과, 단문에서 추출한 F0_mean, F0_std 등이 통계적으로 유의미한 음성 특징 파라미터이다.

가는 목소리는 상대적으로 여성스러운 소리를 굵은 목소리는 상대적으로 남성스러운 목소리를 가리킨다고 볼 수 있다. F0_mean이 집단 1 보다 집단 3이 작은 값을 갖고, F0_std도 집단 3이 작다는 것은, 앞 문장에서 제시한 관점과 일치하는 결과이다. H1-H2는 집단 1에서 5.16이고 집단 3에서 2.95인데, 이것은 여성이 상대적으로 큰 H1-H2 값을 갖는다는 연구결과 [8]와 유사하다.

표 6. 가늘다/굵다에 대한 통계 분석 결과
Table 6. Statistical analysis results for thin/thick voice

음성 파라미터		집단 1 평균값 (표준편차)	집단 3 평균값 (표준편차)	p 값
모음 '아'	F0_mean(Hz)	157(9.8)	125(10.2)	0.000
	H1-H2	5.16(1.24)	2.94(1.81)	0.017
문장	F0_mean(Hz)	159(6.8)	125(10.4)	0.000
	F0_std	27.7(1.73)	20.9(4.60)	0.002

상기에서 기술한 실험결과 이외에도 포먼트 주파수와 Intensity 등도 통계적으로 유의미한 결과가 나왔으나, 여기에 음성학적으로 별다른 의미를 부여하기가 어려워 실험결과를 제시하지 않았다.

6. 결론

본 논문에서는 먼저 정상인의 목소리 특성을 몇 가지 항목으로 분류하고, 이 특성과 상관관계가 높은 음성 특징 파라미터를 도출해 보고자 하였다.

사람의 목소리 특성을 다양하게 분류할 수 있지만, 본 논문에서는 몇 가지 기준을 정하여 이 기준에 부합하는 목소리 특성을 선정하였다. 이러한 기준으로 선정한 목소리 특성은 숨소리, 짜내는 소리, 거친 소리, 콧소리, 굵기(가늘다/굵다) 등이다. 수집한 문장을 이용하여 목소리 특성 각 항목에 대하여 적다, 보통, 많다 등 세 단계로 청취 평가하였다.

이렇게 선정한 목소리 특성과 음성 특징 파라미터간의 상관관계를 규명하기 위하여, '아' 모음과 단문에서 추출한 음원 정보와 관련된 파라미터와 성도필터와 관련된 파라미터를 이용하였다.

목소리 특성 각 항목에 대하여 그 성분이 적고 많을 때 상관관계가 높은 음성 특징 파라미터를 도출하기 위하여, 특성의 차이가 비교적 크다고 할 수 있는 적다와 많다 두 집단 사이에 대하여 통계 처리하였는데, 독립 2-표본 비모수적 검정 방법인 Mann-Whitney U test를 시행하였다.

실험 결과 통계적으로 유의미한 음성 특징 파라미터는, 숨소리 항목에서 SPI, H1-H2, CPP, LTAS_slope, 짜내는 소리 항목에서 BW1, Jitter, SPI, LTAS_slope, 거친 소리 항목에서 BW1, Jitter, Shimmer, NHR, VTI, F0_std, 콧소리 항목에서 BW3, BW4, H1-H2, 가늘다/굵다 항목에서 F0_mean, H1-H2, F0_std 등이다.

앞으로, 본 연구에서 구한 음성 특징 파라미터를 이용하여 사람의 목소리를 자동으로 분류하여 제시해 주는 알고리즘을 연구할 계획이다. 적용 알고리즘으로는 음성인식 분야에서 사용하거나 기계학습 방법인 HMM(Hidden Markov Model), 신경망(Neural Net), CART(Classification and Regression Trees), SVM(Support Vector Machine) 등을 고려하고 있다.

참고문헌

[1] Fex, S. (1992). Perceptual evaluation. *Journal of Voice*, 6(2), 155-158.

[2] Sohn, J. H. (2008). GRBAS and voice handicap index. *Journal of the Korean Society of Logopedics and Phoniatrics*, 19(2), 89-95.

(손진호 (2008). GRBAS 음성평가와 음성장애지수, 대한음성언어의학회지, 19(2), 89-95.)

[3] Pyo, H. Y., Choi, S. H., Lim, S. E., Sim, H. S., Choi, H. S., Kim, K. M. (1999). The correlation between GRBAS scales and MDVP parameters on the pathologic voices of the patients

- with vocal polyps. *Journal of the Korean Society of Logopedics and Phoniatrics*, 10(2), 154-163.
- (표화영, 최성희, 임성은, 심현섭, 최홍식, 김광문 (1999). 성대 폴립 환자를 대상으로 한 GRBAS 척도와 MDVP 측정치 간의 상관관계 연구. *대한음성언어의학회지*, 10(2), 154-163.)
- [4] Han, S. M., Kim, S. B., Kim, J. Y., Kwon, C. H. (2011). A preliminary study on correlation between voice characteristics and speech features. *Phonetics and Speech Sciences*, 3(4), 85-91.
- (한성만, 김상범, 김종열, 권철홍 (2011). 목소리 특성의 주관적 평가와 음성 특징과의 상관관계 기초연구. *한국음성학회, 말소리와 음성과학*, 3(4), 85-91.)
- [5] Kwon, C. H., Song, S. K., Kim, J. Y., Kim, K. H., Jang, J. S. (2012). Extraction of speech features for emotion recognition. *Phonetics and Speech Sciences*, 4(2), 73-78.
- (권철홍, 송승규, 김종열, 김근호, 장준수 (2012). 감정 인식을 위한 음성 특징 도출. *한국음성학회, 말소리와 음성과학*, 4(2), 73-78.)
- [6] Hwang, Y. S., Seong, C. J. (2008). A comparative study on the acoustic characteristics of the Korean vowel /a/ before and after LMS. *Malsori*, 67, 33-60.
- (황연신, 성철재 (2008). 후두미세수술 전후 /아/의 음향적 특성 비교. *대한음성학회, 말소리*, 67, 33-60.)
- [7] Hillenbrand, J. & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, 39, 311-321.
- [8] Park, H. S. (2007). An acoustic study of phonation types in vowels following consonant clusters in Korean. *Malsori*, 64, 53-76.
- (박한상 (2007). 한국어 자음군의 후행모음에 나타난 발성유형의 음향음성학적 연구. *대한음성학회, 말소*, 64, 53-76.)
- [9] Praat. (2014). *Phonetic Sciences*, Univ. of Amsterdam, <http://www.fon.hum.uva.nl/praat/>.
- [10] MDVP. (2014). KayPentax, <http://www.kayelemetrics.com>.
- [11] VoiceSauce. (2014). A program for voice analysis, UCLA, <http://www.seas.ucla.edu/spapl/voicesauce/>.
- [12] Seo, E. H. (2010). *Statistical analysis using SPSS 18.0*. Free Academy, 403-415.
- (서의훈 (2010). SPSS 18.0을 이용한 통계분석, 자유아카데미, 403-415.)
- [13] IBM SPSS Statistics. (2014). SPSS Korea, <http://www.spss.co.kr>.
- [14] Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *Speech Communication Group Working Papers I*, 73-83.
- [15] Klatt, D. H. & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male speakers. *Journal of the Acoustical Society of America*, 87(2), 820-857.
- [16] Mathew, M. M. & Bhat, J. S. (2009). Soft phonation index - a sensitive parameter?. *Indian Journal of Otolaryngology Head Neck Surg.*, 61, 127-130.
- [17] Hillenbrand, J., Cleveland, R. A., Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37, 769-778.
- [18] Kerstens, J., Ruys, E., Zwarts, J. (1996-2001). *Lexicon of Linguistics*. Utrecht Institute of Linguistics, Utrecht Univ.
- [19] Dekrom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech Hearing Res.*, 38, 794-811.

• **김정인 (Kim, Jungin)**

대전대학교 정보통신공학과
대전광역시 동구 용운동 96-3
Tel: 042-280-2567 Fax: 042-280-2559
Email: in10408@naver.com
관심분야: 음성기술
2013~현재 정보통신공학과 대학원 석사과정 재학 중

• **권철홍 (Kwon, Chulhong)** 교신저자

대전대학교 정보통신공학과
Tel: 04-280-2555 Fax: 042-280-2559
Email: chkwon@dju.ac.kr
관심분야: 음성기술, 음성과 의학과 융합연구
1997~현재 정보통신공학과 교수