

목소리 특성의 주관적 평가와 음성 특징과의 상관관계 기초연구

A Preliminary Study on Correlation between Voice Characteristics and Speech Features

한 성 만¹⁾ · 김 상 범²⁾ · 김 종 열³⁾ · 권 철 홍⁴⁾

Han, Sungman · Kim, Sangbeom · Kim, Jongyeol · Kwon, Chulhong

ABSTRACT

Sasang constitution medicine utilizes voice characteristics to diagnose a person's constitution. To classify Sasang constitutional groups using speech information technology, this study aims at establishing the relationship between Sasang constitutional groups and their corresponding voice characteristics by investigating various speech feature variables. The speech variables include features related to speech source and vocal tract filter. Experimental results show that statistically significant correlation between voice characteristics and some speech feature variables is observed.

Keywords: Sasang constitution, voice characteristics, speech features, correlation

1. 서론

사상체질의학은 사람에 대하여 태양인, 태음인, 소양인, 소음인의 네 가지 체질로 분류하고, 각 체질에 대해 진단, 치료와 처방을 달리하는 한의학 이론이다. 이 사상의학에서 얼굴 모양, 인상, 체형 등 체질을 분류하는 기준에 사람의 걸모습이 포함되어 있는데, 음성도 체질 분류하는 척도 중의 하나이다. 이와 같은 사실은 음성공학자의 관점에서 볼 때 흥미로운 일로서, 음성공학 기술을 적용하여 체질을 분류하는 연구 주제에 대해 관심을 끌게 한다.

사람의 음성이 의사소통에 필요한 정보 전달 도구로서 뿐만 아니라 음성에 다른 정보도 포함하고 있다는 것은 누구나 알 수 있다. 예를 들면 우리는 목소리를 듣고 그 화자가 누구

인지를 알 수 있고 또한 말한 사람의 감정 상태를 파악할 수 있다. 사상체질의학 논문[1]에서도 목소리 특성에 주목 하여, “태양인은 소리가 높고 맑고 원만하고, 태음인은 소리가 무겁고 탁하고, 소양인은 가볍고 낮고 급하며, 소음인은 느리고 평이하다”라고 기술하고 있다.

사상체질 의학계에서 음성과 체질의 관계를 규명하고자 하는 연구가 진행되어 음성을 통한 체질분류의 가능성을 타진하였다[2-7]. [2]에서는 문장 발음 속도, 기본 주파수의 평균, 빈도 및 정규분포 등을 조사하여 태음인의 저음 사용이 소음인보다 많다는 결과를 확인하여 음성을 통한 체질분류에 대한 가능성을 보여 주었다. [3]에서는 문장 발음 속도, 포먼트와 대역폭, 피치의 평균, 최소값, 최대값, 표준편차, 그리고 에너지 평균과 표준편차 등을 조사하였다. [4]에서는 주로 피치와 관련 있는 변수들을 조사했는데, 남자 소양인의 피치 범위가 넓으며, 여자인 경우 소음인보다 태음인의 피치 범위가 넓고, 소양인보다 태음인의 피치 최대값이 높고, 태음인의 피치 표준편차가 큰 것으로 나타났다. [5]에서는 단위시간당 발화 음절 수, 포먼트와 대역폭, 피치 범위 등을 조사하였다. [6]에서는 피치와 Shimmer, 하모닉스, 포먼트와 대역폭을 조사하여, 첫 번째 포먼트와 대역폭에서 소양인과 소음인 사이에 유의한 차이가 있다는 결과를 얻었다. 상기 연구들에서 조사한 음성 변수는 주로 피치와 포먼트 관련 변수들을 알 수 있다. [7]에서는 사상체질의학전문가 5인이 73명의 목소리를 청취 평가한 결과, 5인의 평가자내 일치도와 평가자간 일치도가 통계

1) (주)베라시스 mchanman@nate.com

2) 대전대학교 sbland@nate.com

3) 한국한의학연구원 ssmmed@kiom.re.kr

4) 대전대학교 chkwon@dju.ac.kr, 교신저자

이 논문은 한국한의학연구원 기관고유 사업인 체질건강수준 표준개발 과제(K10070)와, 2011년 정부(교육과학기술부)의 재원으로 한국연구재단(No. 20110027738)의 지원을 받아 수행된 연구입니다.

접수일자: 2011년 7월 25일

수정일자: 2011년 9월 22일

게재결정: 2011년 10월 19일

적으로 유의미하게 일치하여, 향후 객관적 음성평가에 대한 전망을 밝게 보고 있다.

음성학자가 음성학적 분석을 바탕으로 사상체질 의학 분야에 접근한 시도로는 [8,9]가 있다. 이 연구에서는 체질별로 평균 F0, 최고 F0, 최저 F0, 피치 범위, 발성유형, Jitter, Shimmer 등 음성 특징 중 음원과 직접적으로 관련이 되는 변수들과, 포먼트와 대역폭, 발화속도 등에서 체질분류 변수를 찾고자 하였다. 이 연구에서 소음인과 소양인 사이에서 두 번째 대역폭이 통계적으로 유의미한 차이를 보였고, 발성유형이 체질별로 약간 일관된 패턴을 보인다는 결과를 도출하였다.

본 연구에서는 목소리의 여러 음성학적 변수 중 사상체질에서 구분한 체질과 가장 관련이 깊은 변수를 찾아냄으로써 음성과 체질간의 상관관계를 규명하는 것이 궁극적인 목표이다. 그런데 음성기술을 이용하여 사상체질을 분류하기 위해서는 먼저 목소리 특성과 관계있는 음성 특징 변수에 대해 선행 조사가 이루어져야 한다. 따라서 본 논문은 사람의 목소리 특성과 음성 특징 변수간의 상관관계를 규명하고자 한다.

음성공학적인 관점에서 빠르기, 높낮이, 세기, 청음/탁음, 목소리 톤 항목에서 체질별 목소리 특성을 살펴보면 다음과 같다. 말 빠르기에서 소양인은 빠르고 태음인과 소음인은 느리다. 높낮이에서 태양인과 소양인은 고음이고 태음인과 소음인은 저음이다. 목소리 크기에서 태양인과 태음인은 크고 소음인은 작다. 청음/탁음 항목에서 태양인과 소양인은 청음이고 태음인은 탁음이다. 목소리 톤은 태양인은 카랑카랑하며, 태음인은 중후하며 굵고, 소양인은 가벼운 특징이 있고, 소음인은 음성의 선이 가늘다. 이와 같이 체질별로 목소리 특성에 차이가 있으며 따라서 음성으로 체질을 분류하는 것이 가능한지 여부를 연구할 필요성이 있다고 생각한다.

빠르기, 높낮이, 세기와 관련된 음성 특징 변수는 각각 문장을 읽는 속도, 피치, 음성신호의 크기와 관련이 있다. 그러나 청음/탁음, 목소리 톤과 관련된 음성변수가 무엇인지에 대한 드러난 연구결과는 없다. 따라서 음성공학 분야뿐만 아니라 인문음성학 분야, 의학 분야 등의 논문을 조사하여 이러한 목소리 특성과 관련이 깊은 음성 변수를 찾아보고자 한다.

2. 음성 특징 변수 추출

타 연구자의 기존 연구에서 일부 성과가 있으나 체질분류에 사용하기에는 판별력이 부족하므로, 본 연구에서는 음성 변수에 대한 기존 연구 범위를 확대하여, <표 1>에 보이는 음성신호의 음원 정보와 관련된 변수와 성도필터와 관계된 변수들을 광범위하게 조사한다.

표 1. 음성 특징 변수
Table 1. Speech feature variables

음원 정보 변수	설명
F0_mean	기본주파수 평균
F0_max	F0 최대값
F0_min	F0 최소값
F0_std	F0 표준편차
Intensity	에너지 크기
Jitter	피치주기 변화율
Shimmer	진폭 변화율
HNR (HNR05, HNR15, HNR25, HNR35)	하모닉과 잡음의 에너지 비율
CPP	주기성의 강도
H1-H2	첫 번째와 두 번째 하모닉의 진폭 차이

성도 필터 변수	설명
F1~F4	포먼트 주파수
B1~B4	각 포먼트의 대역폭
A1~A4	각 포먼트의 진폭
H1-A1	첫 번째 대역폭 크기
H1-A3	스펙트럼 기울기
LTAS	주파수에 따른 하모닉 에너지 감소율

2.1 음원 정보와 관련된 변수

<표 1>에 보이는 음성신호의 음원 정보와 관련된 변수에는 기본주파수에 관한 변수(F0_mean, F0_max, F0_min, F0_std), 세기에 관한 변수(Intensity), 피치 변화율에 관한 변수(Jitter), 진폭 변화율에 관한 변수(Shimmer), 하모닉에 관한 변수(HNR, CPP, H1-H2) 등이 있다.

기본주파수에 관한 변수에서, 평균 기본주파수 F0_mean은 평균 피치주기를 주파수에 대한 값으로 변환한 값, 최고 기본주파수 F0_max는 각 피험자의 전체 기본주파수 중 가장 큰 값, 최저 기본주파수 F0_min은 전체 기본주파수 중 가장 작은 값, 표준편차 F0_std는 기본주파수의 표준편차이다. 기본주파수는 고음과 저음 목소리를 구분하는데 중요한 변수이고, F0_std는 문장을 읽으면서 얼마나 심하게 기본주파수를 변화시켰는지를 보기 위한 것으로, 체질이 사람의 성격과 차이가 있다면 그 성격의 차이가 문장을 읽을 때 얼마나 달리 표현되는지를 보기 위한 것이다.

세기에 관한 변수 Intensity는 음성신호의 에너지 크기(dB)를 나타내는 변수로 목소리의 크기를 구분해 준다.

Jitter는 피치 변화율에 관한 변수로 연속적인 피치주기의

평균변화율을 나타낸다. 이 변수는 목소리가 청음인지 탁음인지를 구분해 주는 변수로 생각할 수 있으며, 청음일수록 Jitter 값이 작으리라고 예상된다.

Shimmer는 진폭의 변화율을 나타내는데 사용하는 변수로 연속적인 진폭변화율의 평균값을 나타낸다. 이 변수는 에너지 레벨의 분포를 보여 주는데, 거친 목소리(harshness)는 진폭 변화율이 크고, 맑고 고운 음성일수록 진폭 변화율이 낮다.

하모닉에 관한 변수는 HNR(Harmonics to Noise Ratio), CPP(Cepstral Peak Prominence), H1-H2 등이 있다. HNR은 하모닉과 잡음의 에너지 비를 나타내며 주기성의 정도를 나타낸다. HNR05는 0~500Hz 사이, HNR15는 0~1,500Hz 사이, HNR25는 0~2,500Hz 사이, HNR35는 0~3,500Hz 사이의 HNR을 나타내는데, 고주파로 갈수록 잡음의 에너지가 증가하므로 HNR 값은 작아진다고 생각할 수 있다. 이 변수는 목 쉼 정도(hoarseness)를 나타내는데, 목소리의 쉼 정도는 모음의 고주파 영역에서 하모닉 성분이 줄고 잡음 성분이 증가하는 특징을 가지므로 쉼 정도가 크다면 작은 값을 갖는다[10][11].

CPP는 피치주기에 해당하는 cepstral peak의 진폭 크기로서 주기성의 강도를 나타낸다. 이 변수는 기식음(breathiness)의 정도를 나타내며, 기식의 정도가 크다면 이 변수는 작은 값을 갖는다[12].

H1과 H2는 첫 번째와 두 번째 하모닉의 진폭을 나타내는데, 기식음은 H1이 크고 H2가 작은 값을 갖는다. 따라서 H1과 H2의 차이(H1-H2, open quotient, 개방 지수)가 크면 기식음을 나타낸다. 이 변수는 사람의 발음 특성과 관계를 맺고 있어서 기식음과 귀어짜는 소리(creaky voice)의 차이를 이 변수와 연결 지어 생각할 수 있는데, 기식음은 H1-H2 값이 크고 귀어짜는 소리는 작은 값을 갖는다[13][14]. 따라서 체질별로 목소리가 달라서 어떤 체질은 기식음이 섞인 목소리인 반면 어떤 체질은 맑은 소리가 특징이라면 그 차이는 이 변수와 밀접한 관계가 있을 것으로 보인다.

2.2 성도 필터와 관련된 변수

<표1>에 보이는 음성신호의 성도 필터와 관계된 변수로는 포먼트 주파수 F1~F4, 각 포먼트의 대역폭 B1~B4, 각 포먼트의 진폭 A1~A4, H1-A1, H1-A3, 그리고 LTAS(Long-Term Average Spectrum) 등이 있다.

포먼트는 대표적인 성도필터 변수로서 음성분야에서 지금까지 가장 널리 사용해 왔다. 포먼트 구조는 같은 모음이라도 성별과 연령별로 차이가 있음이 알려져 있으므로, 체질에 따라 포먼트 구조가 다른지 조사한다. 그리고 명료한 음성일수록 대역폭이 작으므로, 각 포먼트의 대역폭은 음성의 명료도를 보여 주는 변수이다. 포먼트는 Burg가 제안한 방식을 이용하여 LPC 스펙트럼에서 추출했다[15].

첫 번째 하모닉과 첫 번째 포먼트의 진폭 차이 H1-A1은 첫

번째 포먼트 대역폭 B1의 크기를 보여 주고 기식음에서 큰 값을 갖는다[16]. 첫 번째 하모닉과 세 번째 포먼트의 진폭 차이가 H1-A3는 스펙트럼 기울기를 보여 주고, 일반적으로 남자보다 여자가 크다[16].

LTAS는 스펙트럼에서 주파수 증가에 따른 하모닉 에너지의 감소율이다. 일반적으로 스펙트럼은 그 측정위치에 따라 상당한 변화를 보일 수 있기 때문에 본 연구에서는 특정 모음 구간 전체를 대상으로 LTAS를 구한다. 기식음은 큰 LTAS 값을, 귀어짜는 소리는 완만한 기울기를 갖는다(보통 목소리(modal)는 연음이다). Aspiration noise의 존재는 고주파 영역에서 잡음 에너지를 증가시키지만 하모닉 에너지는 줄어들므로 LTAS는 큰 값을 갖는다[17][18].

3. 실험 방법 및 결과

3.1 실험 방법

3.1.1 피험자 및 음성 데이터

피험자는 신체적으로 건강한 20대와 60대 남성과 여성을 대상으로 하였다. 20대 남성 231명, 20대 여성 228명, 60대 남성 200명, 60대 여성 218명으로 총 877명이 대상이다. 모든 연령에 대해 음성 데이터를 수집하지 않고 20대와 60대로 국한한 이유는, 청년층과 노년층을 구분하여 목소리 특성을 판별하는 것이 실험적으로 용이하기 때문이다.

모든 음성 데이터는 동일하게 배경 잡음의 SNR이 30dB 이하인 조용한 공간에서 수집하였으며 사용된 마이크는 Sennheiser e-835s를 마이크 스탠드에 고정하여 입과 마이크 거리가 4-5cm가 되도록 특별한 주의를 기울였다. 피험자는 편안히 앉은 상태에서 독립된 5개의 모음(아, 에, 이, 오, 우)을 각각 2초 이상 자연스럽게 발성하고 각 모음 사이는 약 1-2초 정도 공백을 유지하도록 하였고, 연속해서 하나의 문장("우리는 높은 산에 올라가 맑은 공기를 마시고 왔습니다.")을 두 번 반복하도록 하였다. 음성 데이터는 PCM signed 16bits, mono 형식으로 샘플링 주파수 16,000Hz로 수집하였다[19].

3.1.2 음성 특징 변수 추출

총 877명으로부터 5개의 모음(아, 에, 이, 오, 우)에 대한 음성 데이터를 수집하여 끝점 검출을 통해 모음 구간을 검출하여, 모음 당 32개씩 총 160개의 음성 변수를 추출하였다. 추출한 음성 변수는 기본주파수 F0 계열(F0_mean, F0_min, F0_max, F0_std), 세기(Intensity), Jitter, Shimmer, CPP, HNR(HNR05, HNR15, HNR25, HNR35), 하모닉의 진폭(H1, H2, H4), 두 진폭의 차이(H1-H2, H2-H4, H1-A1, H1-A2, H1-A3), 포먼트와 대역폭(F1~F4, B1~B4), 포먼트의 진폭(A1~A3), LTAS 등이다.

3.1.3 청취 평가

877명이 녹음한 문장에 대하여 음성공학을 전공하는 5인이 청취 평가를 수행하여, 앞에서 기술한 체질별 목소리 특징에 근거하여 각 피험자에 대해 높낮이, 세기, 청음/탁음, 목소리 톤을 판정하였다. 청취평가를 하기 전에 목소리 특성 각 항목에 대해 정의를 하고 문장을 들으면서 평가 훈련을 하였고, 평가시에는 각자 판단을 하고 의견이 일치하지 않는 경우에는 문장을 수차례 들으면서 최종 판정을 합의하는 절차를 거쳤다.

목소리 특성 중에서 높낮이, 세기, 청음/탁음 항목에 대해서는 고/중/저 등 세 개의 그룹으로, 목소리 톤 항목에는 카랑카랑하다/굽다/거칠다/가늘다 등 네 개 그룹으로 분류하였다. 청음은 음정이 높고 명료하며 맑은 목소리를, 탁음은 둔탁하고 거친 목소리를 의미한다. 사상체질 의학자가 체질을 구분하는데 음성을 보조 지표로 사용하는데, 체질에 따라 목소리를 ‘카랑카랑하다/굽다/거칠다/가늘다’라고 구분하기도 한다. 사상체질 의학자의 이런 기준이 음성 특징과 어떤 관련이 있는가를 보기 위하여 목소리 톤을 상기와 같이 나누어 실험을 수행해 보았다.

3.1.4 통계 처리

본 연구에서는 앞에서 기술한 160개의 음성 변수를 이용하여 청취 평가 각 항목에 대하여 분산분석(ANOVA)과 다중 비교를 수행하여, 각 항목의 그룹을 분류하는데 어느 변수가 통계적으로 유의미한 차이(p 값은 0.05)가 있는지를 다중 비교하였다. ANOVA는 세 개 이상의 모집단 평균 간의 차이를 검증하는데 이용하는 분석방법이고[20], 통계처리는 SPSS 버전 18.0[21]을 사용하여 분석하였다. 통계처리는 SPSS v.18.0을 사용했는데, 분산분석에서 종속변수에 목소리 특성 항목을, 독립변수에 음성변수를 설정하면 통계적으로 유의미한 변수를 출력한다.

3.2 실험 결과

성별, 연령별로 ANOVA 분석을 수행하여 통계적으로 유의미한 변수를 구한 결과는 <표2>와 같다. 이 결과에서 목소리 톤과 통계적으로 유의미한 음성 특징 변수는 찾지 못했다.

20대 남성의 경우 높낮이 항목에서는 F0_mean, F0_min, Intensity, F1, Jitter, LTAS, A2, A3, CPP, H1-A2, H1-A3, HNR05이고, 세기 항목에서는 Intensity, A1, A2, A3, CPP, H1-A1, H1-A2, H1-A3, H1, H4, HNR15, HNR25, HNR35이고, 청음/탁음 항목에서는 Jitter, LTAS, A2, CPP, H1-A2, HNR05이다.

20대 여성의 경우 높낮이 항목에서는 F0_mean, F1, LTAS 이고, 세기 항목에서는 Intensity, A1, A2, A3, H1, H2, H4, HNR05, HNR15, HNR25, HNR35이고, 청음/탁음 항목에서는

F0_mean, Jitter, Shimmer, LTAS, CPP, HNR05이다.

60대 남성의 경우 높낮이 항목에서는 F0_mean, Intensity, Jitter, A1, A2, A3, CPP, H1-A1, H1-A3이고, 세기 항목에서는 Intensity, A1, A2, A3, H1, H2, H4, HNR15, HNR25, HNR35이고, 청음/탁음 항목에서는 Intensity, F4, Jitter, Shimmer, A2, A3, CPP, H1-A1, H1-A2, H1-A3, H1-H2이다.

60대 여성의 경우 높낮이 항목에서는 F0_mean, F0_min, F1, Jitter, LTAS, A3, H1-A3이고, 세기 항목에서는 Intensity, A1, A2, A3, H1, H2, H4이고, 청음/탁음 항목에서는 F0_mean, F0_min, Jitter, Shimmer, LTAS, A3, CPP, H1-A2, H1-A3, HNR05, HNR15, HNR25, HNR35이다.

표 2. 성별, 연령별 통계적으로 유의미한 변수
Table 2. Statistically significant variables for sex and age

	20대 남성	20대 여성
높낮이	F0_mean, F0_min, Intensity, F1, Jitter, LTAS, A2, A3, CPP, H1-A2, H1-A3, HNR05	F0_mean, F1, LTAS
세기	Intensity, A1, A2, A3, CPP, H1-A1, H1-A2, H1-A3, H1, H4, HNR15, HNR25, HNR35	Intensity, A1, A2, A3, H1, H2, H4, HNR05, HNR15, HNR25, HNR35
청/탁음	Jitter, LTAS, A2, CPP, H1-A2, HNR05	F0_mean, Jitter, Shimmer, LTAS, CPP, HNR05
	60대 남성	60대 여성
높낮이	F0_mean, Intensity, Jitter, A1, A2, A3, CPP, H1-A1, H1-A3	F0_mean, F0_min, F1, Jitter, LTAS, A3, H1-A3
세기	Intensity, A1, A2, A3, H1, H2, H4, HNR15, HNR25, HNR35	Intensity, A1, A2, A3, H1, H2, H4
청/탁음	Intensity, F4, Jitter, Shimmer, A2, A3, CPP, H1-A1, H1-A2, H1-A3, H1-H2	F0_mean, F0_min, Jitter, Shimmer, LTAS, A3, CPP, H1-A2, H1-A3, HNR05, HNR15, HNR25, HNR35

연령을 불문하고 성별로 통계적으로 유의미한 변수를 정리한 결과는 <표3>과 같다.

남성의 경우 높낮이 항목에서는 F0_mean, Intensity, Jitter, A2, A3, CPP, H1-A3이고, 세기 항목에서는 Intensity, A1, A2, A3, H1, H4, HNR15, HNR25, HNR35이고, 청음/탁음 항목에서는 Jitter, A2, CPP, H1-A2이다.

여성의 경우 높낮이 항목에서는 F0_mean, F1, LTAS이고, 세기 항목에서는 Intensity, A1, A2, A3, H1, H2, H4이고, 청음/탁음 항목에서는 F0_mean, Jitter, Shimmer, LTAS, CPP, HNR05이다.

표 3. 성별로 통계적으로 유의미한 변수
Table 3. Statistically significant variables for sex

	남성(20대+60대)	여성(20대+60대)
높낮이	F0_mean, Intensity, Jitter, A2, A3, CPP, H1-A3,	F0_mean, F1, LTAS
세기	Intensity, A1, A2, A3, H1, H4, HNR15, HNR25, HNR35	Intensity, A1, A2, A3, H1, H2, H4,
청/탁음	Jitter, A2, CPP, H1-A2,	F0_mean, Jitter, Shimmer, LTAS, CPP, HNR05

성별을 불문하고 연령별로 유의미한 변수를 정리한 결과는 <표4>와 같다.

20대의 경우 높낮이 항목에서는 F0_mean, F1, LTAS이고, 세기 항목에서는 Intensity, A1, A2, A3, H1, H4, HNR15, HNR25, HNR35이고, 청음/탁음 항목에서는 Jitter, LTAS, CPP, HNR05이다.

60대의 경우 높낮이 항목에서는 F0_mean, Jitter, LTAS, A3, H1-A3이고, 세기 항목에서는 Intensity, A1, A2, A3, H1, H2, H4이고, 청음/탁음 항목에서는 Jitter, Shimmer, A3, CPP, H1-A2, H1-A3이다.

표 4. 연령별로 통계적으로 유의미한 변수
Table 4. Statistically significant variables for age

	20대(남성+여성)	60대(남성+여성)
높낮이	F0_mean, F1, LTAS	F0_mean, Jitter, A3 H1-A3
세기	Intensity, A1, A2, A3, H1, H4, HNR15, HNR25, HNR35	Intensity, A1, A2, A3, H1, H2, H4,
청/탁음	Jitter, LTAS, CPP, HNR05	Jitter, Shimmer, A3, CPP, H1-A2, H1-A3,

목소리의 특성과 음성 변수와의 상관관계가 <표5>에 종합적으로 정리되어 있다. 높낮이와 상관관계가 높은 음성 변수는, F0_mean, Intensity, F1, Jitter, LTAS, A2, A3, CPP, H1-A3 등이다. 세기와 상관관계가 높은 음성 변수는, Intensity, A1, A2, A3, H1, H2, H4, HNR15, HNR25, HNR35 등이다. 청음/탁음과 상관관계가 높은 음성 변수는, F0_mean, Jitter, Shimmer, LTAS, A2, A3, CPP, HNR05, H1-A2, H1-A3 등이다.

표 5. 목소리의 특성과 음성 변수와의 상관관계
Table 5. Correlation between voice characteristics and speech feature variables

		f0mean	f0min	f0max	f0std	int	F1	F2	F3
높 낮 이	20M	고 주2)	고			고	고		
	20F	고					고		
	60M	고				고			
	60F	고	고				고		
세 기	20M					고			
	20F					고			
	60M					고			
	60F					고			
청 / 탁 음	20M								
	20F	고							
	60M					고			
	60F	고	고		저 주3)				

주1) ‘고’ 또는 ‘저’는 상관관계가 높음을, 빈 칸은 낮음을 나타낸다.

주2) 높은 높낮이에서 f0mean이 높음을 의미한다.

주3) 높은 청음에서 f0std가 작음을 의미한다.

		F4	B1	B2	B3	B4	jit	sh m	LT AS
높 낮 이	20M						저		저
	20F								저
	60M						저		
	60F						저		저
세 기	20M								
	20F								
	60M								
	60F								
청 / 탁 음	20M						저		저
	20F						저	저	저
	60M	저					저	저	
	60F						저	저	저

		A1	A2	A3	CP P	H1- A1	H1- A2	H1- A3	H1
높 낮 이	20M		고	고	고		저	저	
	20F								
	60M	고	고	고	고	저		저	
	60F			고				저	
세 기	20M	고	고	고	고	저	저	저	고
	20F	고	고	고					고
	60M	고	고	고					고
	60F	고	고	고					고
청 음/ 탁 음	20M		고		고		저		
	20F				고				
	60M		고	고	고	저	저	저	
	60F			고	고		저	저	

		H1-H2	H2	H2-H4	H4	HN R05	HN R15	HN R25	HN R35
높낮이	20M					고			
	20F								
	60M								
	60F								
세기	20M				고		저	저	저
	20F		고		고	저	저	저	저
	60M		고		고		저	저	저
	60F		고		고				
청음/탁음	20M					고			
	20F					고			
	60M	저							
	60F					고	고	고	고

높은 높낮이에서 F0_mean이 큰 값을 가진다는 결과는 당연하지만, Jitter를 보면 높은 높낮이에서 피치주기의 변화율이 작아짐을, LTAS와 H1-A3를 보면 스펙트럼 기울기는 작아진다는 것은 의미가 있는 결과이다. 남자의 경우 Intensity, A2, A3를 보면 높은 높낮이에서 크게 말하는 경향이 있음을, CPP를 보면 주기성의 강도가 커짐을 알 수 있다.

큰 세기에서 Intensity, A1, A2, A3가 큰 값을 갖는 것을 알 수 있다. 큰 세기에서 H1, H2, H4가 큰 값을 갖고 HNR15, HNR25, HNR35가 작은 값을 갖는 것은, 크게 말할 때 저주파 영역에서 하모닉 성분이 강해지고 중간과 고주파 영역에서 하모닉 성분이 줄어든다는 것을 의미한다.

청음인 경우 피치주기의 변화율(Jitter), 진폭의 변화율(Shimmer), 스펙트럼 기울기(LTAS)가 작아짐을 알 수 있다. 또한 주기성의 강도(CPP)가 강하고, 저주파 영역에서 잡음 대비 하모닉 에너지(HNR05)가 커짐을 알 수 있다. 여자의 경우 청음에서 F0가 큰 값을 갖는 경향이 있다.

4. 결론

본 논문은 음성기술을 이용하여 사상체질을 분류하기 위한 선행 연구로 사람의 목소리 특성과 상관관계가 높은 음성 특징 변수를 찾고자 하였다. 연구에 사용한 음성 변수는 음원정보와 관련된 변수와 성도필터와 관련된 변수를 광범위하게 포함시켰다. 피험자의 음성을 청취하여 높낮이, 세기, 청음/탁음, 목소리 톤 등을 세 개 또는 네 개의 그룹으로 분류하였다. 청취 평가 각 항목과 음성 특징 변수와의 관계를 ANOVA 통계 분석하여 그 둘의 상관관계를 도출하였다.

실험 결과 목소리 톤을 제외하고 목소리 특성의 나머지 항목에서 음성 변수와 상관관계가 높은 변수를 찾아내었다. 그 중에서 특히, 청음/탁음 항목에서 상관관계가 높은 음성 변수

로 Jitter, Shimmer, LTAS, CPP, HNR05를 찾은 것은 의미가 있다고 말할 수 있다.

목소리 톤에서 유의미한 음성 변수를 찾지 못한 것은, 평가 기준을 ‘카랑카랑하다/굵다/거칠다/가늘다’라고 구분한 것이 청취 평가하는 데 어려움이 있었다고 사료된다. 향후 연구에서는 사상체질 전문의의 자문을 구해 목소리 톤을 구분하는 기준을 달리할 계획이다.

앞으로, 본 연구에서 구한 음성 특징 변수를 이용하여 체질을 분류하는 알고리즘에 대한 연구를 진행할 계획이다. 적용 알고리즘으로는 기계학습 방법이나 음성인식에서 사용하고 있는 CART(Classification and regression trees), 신경망, SVM(Support vector machine), HMM(Hidden markov model) 등을 계획하고 있다.

참고문헌

[1] Kim, D. R., (2003). A study about five-sounds of Sasang constitutional sound analysis, *J. of Sasang Constitutional Medicine*, Vol. 15, No. 1, 50-59.
(김달래 (2003), 오음의 사상의학적 음성분석과 고찰, 사상체질의학회지, 15권 1호, 50-59.)

[2] Yang, S. H., Kim, D. R. (1996). A study on the correlation between voice print and Sasang constitution, *J. of Sasang Constitutional Medicine*, Vol. 8, No. 2, 191-202.
(양승현, 김달래, (1996). 성문과 사상체질과의 상관성에 관한 연구, 사상체질의학회지 8권 2호, 191-202.)

[3] Shin, M. R., Kim, D. R. (1999). A study on the correlation between sound characteristics and Sasang constitution by CSL, *J. of Sasang Constitutional Medicine*, Vol. 11, No. 1, 137-157.
(신미란, 김달래 (1999). CSL을 통한 음향특성과 사상체질간의 상관성 연구, 사상체질의학회지 11권 1호, 137-157.)

[4] Kim, S. H., Shin, M. R., Kim, D. R., Kwon, K. R. (2000). A study on the correlation between sound characteristic and Sasang constitution by Laryngograph and EGG, *J. of Sasang Constitutional Medicine*, Vol. 12, No. 1, 144-156.
(김선형, 신미란, 김달래, 권기록 (2000) Laryngograph와 EEG를 이용한 음향특성과 사상체질간의 상관성 연구, 사상체질의학회지 12권 1호, 144-156.)

[5] Yang, S. M., Kim, S. H., Yoo, J. S., Kim, Kim, H. S., Lee, Y. H., Kim, D. R. (2001). A study on the correlation between sound characteristic and Sasang constitution by pitch range and bandwidth, *J. of Sasang Constitutional Medicine*, Vol. 13, No. 3, 31-39.
(양상목, 김선형, 유준상, 김형석, 이영훈, 김달래 (2001). Pitch Range와 Bandwidth를 이용한 음성특성과 사상체질간의 상관

- 성 연구, 사상체질의학회지 13권 3호, 31-39.)
- [6] Park, S. J., Kim, D. R. (2004). A study on the correlation between Sasang constitution and sound characteristics used harmonics and formant bandwidth, *J. of Sasang Constitutional Medicine*, Vol. 16, No. 1, 61-73.
(박성진, 김달래 (2004). Harmonics와 Formant Bandwidth를 이용한 음성특성과 사상체질간의 상관성 연구, 사상체질의학회지, 16권 1호, 61-73.)
- [7] Yoo, J. S., Kim, D. R. (2004). A study on the Sasang constitutional diagnosis by perceptual voice analysis, *J. of Sasang Constitutional Medicine*, Vol. 16, No. 3, 46-58.
(유준상, 김달래 (2004). 청각적 성음분석을 통한 사상체질진단에 관한 연구, 사상체질의학회지 16권 3호, 46-58.)
- [8] Moon, S. J., Park, J. J., Hwang, H. J. (2003). Voice and Sasang constitution: in terms of source functions, *Malsori*, Vol. 48, 19-33.
(문승재, 박종주, 황혜정 (2003). 음성과 사상체질: 음원을 중심으로, 대한음성학회, 말소리 48호, 19-33.)
- [9] Moon, S. J., Tak, J. H., Hwang, H. J. (2005). A phonetic study of Sasang constitution, *Malsori*, Vol. 55, 1-14.
(문승재, 탁지현, 황혜정 (2005). 음성학적으로 본 사상체질, 말소리 55호, 1-14.)
- [10] C. T. Ferrand (2002). Harmonics-to-Noise Ratio: an index of vocal aging, *Journal of Voice*, Vol. 16, No. 4, 480-487.
- [11] Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, *Proceedings of Institute of Phonetic Sciences*, Vol. 17, 97-110.
- [12] Hillenbrand, J., Houde, R. A. (1996) Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech, *Journal of Speech and Hearing Research*, Vol. 39, 311-321.
- [13] Park, H. S. (2007). An acoustic study of phonation types in vowels following consonant clusters in Korean, *Malsori*, Vol. 64, 53-76.
(박한상 (2007). 한국어 자음군의 후행모음에 나타난 발성유형의 음향음성학적 연구, 대한음성학회, 말소리 64호, 53-76.)
- [14] Wayland, R., Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: the case of Khmer, *Journal of Phonetics*, Vol. 31, 181-201.
- [15] Rabiner, L.R., Schafer, R.W. (1978). *Digital processing of speech signals*, Prentice-Hall.
- [16] Iseli, M., Shue, Y. L., Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source, *Journal of Acoustical Society of America*, Vol. 121, No. 4, 2283-2295.
- [17] linville, S. E. (2002). Source characteristics of aged voice assessed from long-term average spectra, *Journal of Voice*, Vol. 16, No. 4, 472-479.
- [18] Mendoza, E., Valencia, N., Munoz, J., Trujillo, H. (1996). Differences in voice quality between men and women: use of the long-term average spectrum, *Journal of Voice*, Vol. 10, No. 1, 59-66.
- [19] Sung, T. J. (2007). *Understanding and application of modern basic statistics*, Kyoyookbook.
(성태제 (2007). 현대 기초통계학의 이해와 적용, 교육과학사.)
- [20] Kang, J. H., Do, J. H., Kim, J. Y. (2010). Voice classification algorithm for Sasang constitution using support vector machine, *J. of Sasang Constitutional Medicine*, Vol. 22, No. 1, 17-25.
(강재환, 도준형, 김종열 (2010). SVM을 이용한 음성 사상체질 분류 알고리즘, 사상체질의학회지, 22권 1호, 17-25.)

• **한성만 (Han, Sungman)**

(주)베라시스
경기도 안양시 호계1동
Tel: 042-280-2567 Fax: 042-280-2559
Email: mchanman@nate.com
관심분야: 음성기술
현재 (주)베라시스

• **김상범 (Kim, Sangbeom)**

대전대학교 정보통신공학과
대전광역시 동구 용운동 96-3
Tel: 042-280-2567 Fax: 042-280-2559
Email: sbland@nate.com
관심분야: 음성기술
현재 정보통신공학과 대학원 석사과정 재학중

• **김종열 (Kim, Jongyeol)**

한국한의학연구원 체질의학연구본부
대전광역시 유성구 전민동 461-24
Tel: 042-868-9489 Fax: 042-868-9480
Email: ssmed@kiom.re.kr
관심분야: 사상체질의학
현재 체질의학연구본부 본부장

• **권철홍 (Kwon, Chulhong)** 교신저자

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