

음성을 이용한 사상체질 분류 알고리즘

Automated Speech Analysis Applied to Sasang Constitution Classification

강 재 환¹⁾ · 유 중 향²⁾ · 이 혜 정³⁾ · 김 중 열⁴⁾

Kang, Jaehwan · Yoo, Jonghyang · Lee, Haejung · Kim, Jongyeol

ABSTRACT

This paper introduces an automatic voice classification system for the diagnosis of individual constitution based on Sasang Constitutional Medicine (SCM) in Traditional Korean Medicine (TKM). For the developing of this algorithm, we used the voices of 473 speakers and extracted a total of 144 speech features from the speech data consisting of five sustained vowels and one sentence. The classification system, based on a rule-based algorithm that is derived from a non parametric statistical method, presents binary negative decisions. In conclusion, 55.7% of the speech data were diagnosed by this system, of which 72.8% were correct negative decisions.

Keywords: non-parametric, quantitative, Sasang, constitution, SCM, TKM

1. 서 론

사상체질의학은 조선시대 이제마 선생으로부터 창안되어 인간을 사상(태양, 태음, 소양, 소음)체질로 분류해 이에 맞추어 적절한 생리, 병리, 진단, 관리 기준을 적용하는 한의학의 한 학문 분야이다[1]. 예전부터 개인에 대한 사상체질 진단은 전적으로 한의사의 경험에서 나오는 주관적인 판단 기준에 의해서만 결정되기에 꾸준히 사상체질 분류의 표준, 객관화에 대한 요구가 있었으며, 이러한 이유로 사상체질과 과학, 공학기술을 접목하여 체질을 객관적이고 정량적으로 표현하고자 하는 연구는 현재도 계속되고 있다[2],[3]. 이런 연구 추세와 함께 음성은 정량적 사상체질 감별 연구의 하나의 주제이기도 하면서 한의학 병증을 진단하는 가장 기본적인 진단방법인 4진(望聞問切) 중의 하나이기에 음성 연구에 대한 관심 역시 비례적으로 증가하고

있다. 음성의 특징과 이를 객관적인 방법론으로 접목시키고자 하는 다양한 연구는 사상체질의학, 한의학뿐만 아니라 최근에는 중의학 분야에서도 시도되고 있다. Chiu는 중의학 분야에서 음성을 통한 청진법을 정량적으로 표현하는 연구에서 4개의 음성 변수인 평균 영교차점(zero-crossing)개수, 구간 내 꼭지점(peak)과 벨리(valley) 간의 변화량, 첫 번째와 두 번째 포먼트 주파수 차이, 스펙트럼 에너지 분포를 이용하여 40-60명의 피험자 대상으로 정상인과 허증(虛症) 환자를 70% 이상의 정확도로 분류하였다[4]. 또한 FD(Fractal Dimension) 변수와 DTW(Dynamic Time Warping) 기술을 이용하여 건강한 사람과 기허(Qi-vacuity), 양허(Ying-vacuity) 환자를 85% 정확도로 판별할 수 있다는 연구 결과를 발표하였다[5]. 사상체질과 음성과의 관계는 사상임해지남(四象臨海指南)의 사성론(四聲論)에서 “태양인은 호흡기가 크므로 소리가 높다. 성음이 맑고 원만하니 상음(商音)과 화합한다. 태음인은 성량이 풍부하여 소리가 무겁다. 성음이 탁하고 방정하니 궁음(宮音)과 화합한다. 음은 양을 이기므로 음량이 풍성하다. 소양인 호흡기가 작으므로 소리가 가볍고 낮다. 성음이 급하고 멀리 가니 징음(徵音)과 화합한다. 소음인은 성량이 넓으므로 소리가 활발하다. 성음은 느리고 평안하니, 우음(羽音)과 화합한다. 음은 양을 이기므로 음량이 풍성하다”라고 밝히고 있다[6]. 하지만, 이런 사상체질의학 내의 체질과 음성 간의 설명에서도 알 수 있듯이 그 표현이 추상적이고 함축적이라 이를 객관적 연구에 직접적으로 인용하기에는 무리가 따른다. 또한 사상체질 판단에 있어서 음성은 다른 판단

1) 한국한의학연구원 doskian@kiom.re.kr

2) 한국한의학연구원 jhyoo@kiom.re.kr

3) 한국한의학연구원 haejung0614@kiom.re.kr

4) 한국한의학연구원 ssmmed@kiom.re.kr

이 논문은 지식경제부 차세대기술개발사업 중 지능형 한방컨텐츠개발의 지원에 의해 이루어진 것임. (지원번호: 10028438)

접수일자: 2009년 7월 15일

수정일자: 2009년 8월 6일

게재결정: 2009년 9월 7일

요소인 안면, 체형 등과 비교해보면 이들보다 시간과 환경에 따른 가변적인 요인이 많이 작용하기에 음성을 이용하여 특정 체질 정도를 정량적으로 제시한다는 것은 현실적으로 많은 어려움을 수반하고 있다. 이런 이유로 현재까지 음성과 사상체질간의 연구는 체질별 주요 음성 변수에 모수적인 통계 방법을 적용하여 그 가능성 정도만을 제시하고 있으며[7], 그 외 여러 가지 패턴 분류방법을 통한 다양한 시도를 도입하더라도 유의미한 결론을 제시하지는 못하고 있다[8]. 이런 연구 결과는 사상체질 전문의들 역시 음성과 체질간의 유의미한 연관성과 이로 말미암아 정량적인 판별 기기, 연구 필요성을 인정하지만 음성을 통한 체질 판별은 전체가 아닌 일부에 한정되며 안면과 체형, 피부조건보다 우선순위를 낮게 매기는 현실적인 상황에서는 당연한 결과라 할 수 있다. 본 연구에서는 이러한 현실적인 한계 내에서 음성을 이용한 체질 분류라는 기본 목적을 달성하기 위해 기존의 몇몇 음성 변수와 모수적인 통계 방법 적용이라는 틀을 벗어나 100여개 이상의 변수들과 규칙 기반의 비모수 통계 방법을 이용하여 그 결과를 “판단 보류”와 “이진 부정 판단”이라는 형식으로 제공하는 새로운 알고리즘을 구현하였다.

2. 연구 방법

2.1 연구 대상 및 음성 데이터 수집

음성 데이터는 2007년 11월부터 2008년 7월까지 10개 (대학)한방병원과 3개 한의원을 통해 수집되었다. 체질이 명확한 사람을 대상으로 하기 위해 5년 이상 사상체질의학 임상분야 경험이 있는 전문가에 의해 동일 체질 처방을 60첩 이상 투여 받고 최소 4회 이상 방문하여 경과를 관찰한 사람 중 부작용이 없으며 주증(主證)이 일정 정도 이상 호전된 사람, 혹은 동일 체질 처방을 20첩 이상 60첩미만 투여 받고 부작용이 없으며 주증(主證)과 소증이 일정 정도 이상 호전된 사람을 대상으로 선정하였다. 선별된 피험자 나이는 10대부터 80대까지 고루 분포되어 있으며(48.52±15.50yr), 성별과 체질별로는 남자 164명 (소음 32명, 소양 44명, 태음 88명), 여성 309명(소음 82명, 소양 116명, 태음 111명)으로 총 473명이다. 우리나라 전체 인구의 1% 이하에 해당되는 태양인 샘플은 제외하였다. 모든 음성 데이터는 동일하게 배경 잡음이 30dB 이하인 조용한 공간에서 수집되었다. 사용된 마이크는 Sennheiser e-835s이며 마이크 스탠드에 고정하여 입과 마이크 거리가 4-5cm가 되도록 유지하였다. 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 모두 PCM signed 16bits, mono형식으로 표본화율(sampling rate) 44,100Hz로 수집되었다.

2.2 음성 변수 추출

수집된 각 음성 데이터에서 <그림1>과 같이 총 144개의 음성 변수를 추출하였다. 모음에서 추출된 음성 변수는 구분을 위해 변수 이름 앞에 5개의 소문자(a, e, i, o, u)를 이용하여 해당

모음을 표시하였다.

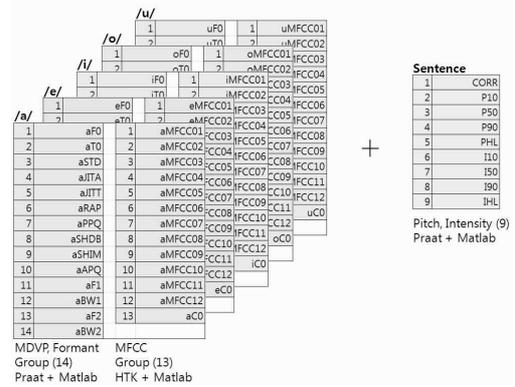


그림 1. 체질 분류에 사용된 144개 음성 변수
Figure 1. 144 Speech Features for Constitution Classification

2.2.1 F0, Intensity, Formant 변수와 MDVP

음성 변수는 기본주파수(F0, fundamental frequency)를 시작으로 강도(intensity), 포먼트(formant), MDVP(Multi-Dimensional Voice Program) 변수까지 추출하였다. 이를 위해 첫 번째 단계로 Praat 스크립트(script) 프로그래밍[9]-[10]을 이용하여 수집된 모든 데이터에 대해서 20msec 간격으로 시간을 동기화하여 F0, 강도, 1차 2차 포먼트(F1, F2), -3dB 대역폭(bandwidth), 최대 진폭(amplitude), 최소 진폭을 구하고 그 결과를 텍스트 파일 형태로 저장하였다. 음성 변수 추출을 위한 Praat 스크립트 프로그램 중 일부 주요 부분만을 아래에 정리하였다. 이 이후 생성된 결과 텍스트 파일을 읽어오는 작업부터 변수 추출, 분석 및 분류 알고리즘까지 관련된 모든 작업은 Matlab 프로그램으로 구현되었다.

```
To Pitch (ac)... 0.010 75 3 yes 0.03 0.45 0.01 0.35 0.14 500
To Intensity... 100 0.005
! 남자의 경우
To Formant (burg)... 0 5 5000 0.025 50
! 여자의 경우
To Formant (burg)... 0 5 5500 0.025 50
deltaT = 0.020
dur = ('offsettime'-'onsettime') / deltaT
iter = 'dur:0'+1
i = 'onsettime'
while i < 'offsettime'
    db = Get value at time... i Cubic
    fl = Get value at time... 1 i Hertz Linear
    w1 = Get bandwidth at time... 1 i Hertz Linear
    f2 = Get value at time... 2 i Hertz Linear
    w2 = Get bandwidth at time... 2 i Hertz Linear
    mimAmp = Get minimum... i i+0.02 Parabolic
    maxAmp = Get maximum... i i+0.02 Parabolic
    fileappend 'wfilename' 'i:3"tab"$f0:1"tab"$db:1"tab"$
    'f1:1"tab"$w1:1"tab"$f2:1"tab"$w2:1"tab"$mimAmp:5
```

```
"tab$"maxAmp:5"newline$
i = 'i'+deltaT
endwhile
```

두 번째 단계로 F₀를 기준으로 하여 독립적인 5모음 자동 분절(segmentation) 작업을 실시하였다. 모음 분절에서는 유효한 피치가 끊임이 없이 500msec 동안 지속되는지 여부와 유효한 전 구간 끝점과 다음 시작점 사이에 750msec 이상의 공백이 존재하는지를 기본 규칙으로 시작점(onset)과 끝점(offset) 후보군(candidates) 구간을 설정하고 이 구간의 20%~80% 지점을 구하여 이때의 시간을 각각 모음의 시작점과 끝점으로 설정하였다. 설정된 5모음의 시작과 끝점은 이후 모든 모음 분석 구간의 기준이 되어 모음 관련 음성 변수 추출은 이 유효한 분절 구간에 대해서만 실시하였다. 세 번째 단계로 마지막 모음인 /u/ 끝점 이후부터는 유효한 F₀를 가지는 부분만을 선택하여 이를 문장으로 판단하고 앞서 Praat으로 생성한 결과 파일에서 이 문장 구간의 F₀와 강도를 따로 분리하였다. 이후 추출한 문장 구간의 F₀와 강도 신호 크기간의 피어슨 상관계수(Pearson correlation coefficient) 구하고, F₀와 강도 신호 크기의 각 10, 50, 90th 백분위수(percentile) 및 이들의 상, 하위 범위 비를 구하였다. 마지막으로 음성의 다양한 측면에 대한 분석을 가장 큰 장점으로 언어병리학 계통의 음향변수 중에서 최근 들어 가장 많이 활용되고 있는 MDVP 변수를 본 연구에도 적용하였다[11]. 이를 위해 지정된 모음구간에 대해서 F₀와 최대/최소 진폭 신호를 이용하여 발성 시 피치의 변화 정도를 나타내는 지터(jitter)계열 변수들(Jita, Jitt, RAP, PPQ)과 진폭의 변화 정도를 나타내는 쉘머(shimmer)계열 변수들(ShdB, Shim, APQ)을 추출하였다.

표 1. MDVP 변수 식

Table 1. Formulas for jitter and shimmer related parameters used in MDVP

STD	$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (F_i - \bar{F})^2}$		
Jita	$\frac{1}{N-1} \sum_{i=1}^{N-1} T_{0i+1} - T_{0i} $	ShdB	$\frac{1}{N-1} \sum_{i=1}^{N-1} 20 \log(A_{i+1} / A_i) $
RAP	$\frac{1}{N-2} \sum_{i=1}^{N-3} \frac{T_{0i+1} + T_{0i+2} + T_{0i+3} - T_{0i}}{3} \times 100$	Shim	$\frac{1}{N-1} \sum_{i=1}^{N-1} A_i - A_{i+1} \times 100$
PPQ	$\frac{1}{N-4} \sum_{i=1}^{N-3} \frac{T_{0i+2} + T_{0i+1} + T_{0i} + T_{0i-1} + T_{0i-2} - T_{0i}}{5} \times 100$		
APQ	$\frac{1}{N-4} \sum_{i=1}^{N-3} \frac{ A_{i+2} + A_{i+1} + A_i + A_{i-1} + A_{i-2} - A_i }{5} \times 100$		

2.2.2 MFCC

음성인식 및 임상분야의 많은 연구에서 멜 스케일(Mel-scale)과 쉘스트럼(cepstrum) 분석 방법을 도입한 MFCC(Mel-Frequency Cepstral Coefficients)를 특징 변수로 사용하여 우수한 인식 성능을 보이는 것으로 알려지고 있다[12]. 본 연구에서도 비록 각

변수들에 대해서 물리적인 의미를 부여하기는 어렵지만 MFCC 변수들을 도입하여 판별 성능 향상을 도모하고자 하였다. 사용된 MFCC 변수는 HTK(Hidden Markov Toolkit)를 이용하였으며, 각 모음별로 유효한 분절 구간에 대해 12차 MFCC 변수와 1개의 에너지 변수로 총 65개의 MFCC 변수를 추출하였다. MFCC 변수 추출에 사용된 HTK 명령어와 설정치를 <표2>에 정리하였다.

표 2. MFCC 변수를 위한 HTK 설정치 (HCopy Command)

Table 2. HTK Coding Parameters for the Extraction of MFCC

OPTION	VALUE
SOURCEKIND	WAVEFORM
SOURCEFORMAT	WAV
SORCERATE	227
TARGETKIND	MFCC_0
TARGETRATE	100000.0
SAVECOMPRESSED	F
SAVETHCRC	F
WINDOWSIZE	400000.0
USEHAMMING	T
PREEMCOEF	0.97
NUMCHANS	30
CEPLIFTER	22
NUMCEPS	12
ENORMALISE	F

2.2.3 음성 변수 정의

최종적으로 <표3>에 본 연구에서 사용되는 144개의 음성 변수에 대한 정의가 정리되어 있다.

표 3. 사용된 음성 변수 정의

Table 3. The Definition of Speech Features

	변수명	변수 정의
모음	xF0	average fundamental frequency
	xT0	period of the average glottal period
	xSTD	standard deviation of F ₀
	xJITA	absolute jitter
	xJITT	jitter percent
	xRAP	relative average perturbation
	xPPQ	pitch period perturbation quotient
	xSHDB	shimmer in dB
	xSHIM	shimmer percent
	xAPQ	amplitude perturbation quotient
	xF1	1st formant
	xBW1	1st 3dB bandwidth
xF2	2nd Formant	
xBW2	2nd 3dB Bandwidth	
xMFCC1~12	1~12th MFCC	
xC0	energy	
문장	CORR	correlation between F ₀ and intensity
	P10	10th percentile of F ₀
	P50	50th percentile of F ₀
	P90	90th percentile of F ₀
	PHL	(P90-P50)/(P50-P10)
	I10	10th percentile of intensity
	I50	50th percentile of intensity
	I90	90th percentile of intensity
	IHL	(I90-I50)/(I50-I10)

※ x : 5모음(a, e, i, o, u)

3. 체질 분류 알고리즘

3.1 분류를 위한 조건 변수 생성

체질 분류 알고리즘은 규칙 기반의 비모수 통계 방법을 응용한 새로운 알고리즘으로 다음과 같은 방법으로 구현하였다. 먼저 수집된 각 음성 데이터에서 144개의 모든 음성 변수를 추출하고 이 중 학습에 필요한 train set에 대해 모든 음성 변수의 체질별 최대값과 최소값을 구한다. 다음으로 체질별 최대값 중의 최소값과 최소값 중의 최대값을 찾아서 전자를 $uMin$, 후자를 $lMax$ 로 설정하였다. 또한 세 가지 체질 중에 $uMin$ 이 발생된 체질의 이름을 ui , $lMax$ 가 발생된 체질의 이름을 li 라는 변수로 정한다. 결과적으로 하나의 음성 변수마다 $uMin$, ui , $lMax$, li 라는 4개의 조건 변수를 새로 생성할 수 있다.

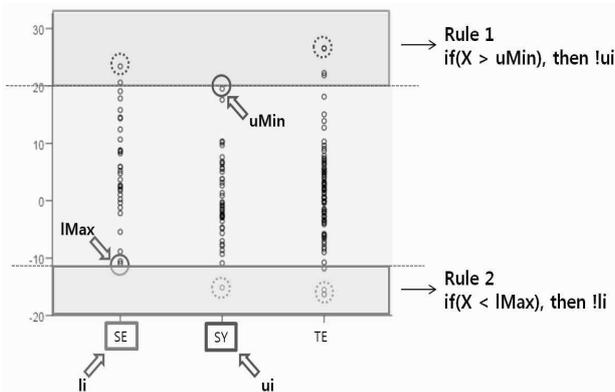


그림 2. 4개의 조건변수와 2개의 논리 규칙
Figure 2. 4 Conditional Variables and 2 Logical Rules

<그림2>에 도식적으로 설명되었듯이 생성된 4개의 조건 변수를 이용하면 다음과 같은 두 개의 논리 규칙을 유추해 볼 수 있다.

- If ($X > uMin$) then NOT ui (1)
- If ($X < lMax$) then NOT li (2)

즉, 먼저 각각의 음성 변수에 대한 4개의 조건 변수 중에서 $uMin$ 을 상위 문턱값(upper threshold), $lMax$ 를 하위 문턱값(lower threshold)으로 설정한다. 이후에 임의의 음성 데이터에서 해당되는 음성 변수 값(X)이 입력되었을 때, 이때의 X 가 $uMin$ 보다 큰 경우에는 “입력된 음성 데이터는 ui 체질은 아닐 확률이 그 반대인 경우보다 더 크다.”라고 할 수 있다. 동일한 조건으로 입력된 변수의 값이 $lMax$ 보다 작다면 “입력된 음성 데이터는 li 체질은 아닐 확률이 그 반대인 경우보다 더 크다.”로 정리할 수 있다. 이러한 규칙은 144개의 음성 변수별로 2개씩, 총 288개의 논리식으로 확장이 가능하며 사용되는 4개의 조건 변수들은 알고리즘 내에서 144×4 행렬(이하 $gRule$ 로 명명)로 효율적으로 저장, 관리할 수 있다. 이번 알고리즘에서는 이러한 4개의 조건 변수 생성 과정에 이상치(outlier) 발생으로 인한 오류를 방지하고 체질간의 변별력을 높이기 위해 체질별 최대값과 최소값과 같은 극한값을 사용하지 않고 체질별 분포에 대한 백분위수를 임계치로 적용하였다.

3.2 음성 변수와 나이, 키, 몸무게, BMI 상관관계

사용된 144개의 음성 변수 중에는 나이, 키, 몸무게와 같은 신체정보와 관련이 있는 변수가 존재하기에 이들의 상관성을 조사하여 다음과 같은 방법으로 알고리즘에 반영하였다. 첫 번째로 남녀별로 모든 음성 변수와 4가지 주요 신체정보인 나이, 키, 몸무게, BMI(Body Mass Index)간의 피어슨 상관계수를 조사하였다. 그 후 남녀별로 상관계수의 절대값이 0.3 이상 되는 변수를 조사하였더니 남자의 경우는 나이에서 9개(eSHDB, eSHIM, uF0, iF2, P10, P50, eMFCC8, iMFCC8, oMFCC8, uMFCC8), 키에서 2개(uF0, P50)가 있었고 몸무게와 BMI에서는 0.3 이상을 갖는 경우가 하나도 없었다. 여성의 경우는 나이에서 21개(aF0, aT0, eF0, eT0, iF0, oF0, oT0, uF0, P10, P50, P90, aMFCC9, aMFCC12, eMFCC6, eMFCC11, eMFCC12, iMFCC6, iMFCC9, oMFCC9, uMFCC7, uMFCC9, uMFCC12), BMI에서는 5개(eF0, P10, P50, P90, aMFCC12)가 존재하였고 키와 몸무게에서는 0.3 이상의 상관계수를 갖는 변수는 없었다. 두 번째로 상관성 조사 결과를 바탕으로 남녀 모두에서 나이와 0.3 이상의 상관계수 절대값을 가지는 변수는 한의학적으로 의미 있는 16세를 기준으로 5개의 나이그룹으로 나누고 키, BMI와 유효한 상관성을 보이는 변수에서는 4개의 그룹으로 나누어 <그림3>와 같이 동일 그룹별로 새로운 하위 $sub-gRule$ 을 생성하였다. 차후 임의의 피험자 음성 신호가 들어왔을 때 피험자 나이와 키, BMI를 조사하여 해당되는 음성 변수에서는 이에 맞는 적절한 하위 $sub-gRule$ 이 상위 $gRule$ 을 대신하도록 하였다. 덧붙여 특정 변수에서 특정 체질의 비이상적인 분포형태를 방지하고자 $gRule$ 의 li 와 ui 가 같은 경우가 발생한 음성 변수에 대해서는 $gRule$ 논리식을 생성하지 않았다.

gRule	#1 (aF0)	#2 (aT0)	...	#63 (p50)	...	#144 (uCO)
uMin	215.79	7.91	...	191.81	...	81.55
ui	2(SY)	3(TE)	...	2(SY)	...	3(TE)
lMax	126.36	4.63	...	143.33	...	0
li	3(TE)	2(SY)	...	1(SE)	...	1(SE)

#63 gRule_Age	Age1	Age2	Age3	Age4	Age5
uMin	241.59	214.17	227.40	233.50	191.81
ui	2(SY)	3(TE)	1(SE)	2(SY)	2(SY)
lMax	254.54	194.85	143.33	182.39	162.26
li	1(SE)	2(SY)	1(SE)	1(SE)	1(SE)

그림 3. 하위 $sub-gRule$ 행렬
Figure 3. $sub-gRule$ Matrix

3.3 체질 점수와 가중치

최종 체질 판단을 위해서 체질 정도를 수치화할 수 있는 체질 점수와 가중치 개념을 도입하였다. 체질 점수는 임의의 음성 데이터에 대해 소음, 소양, 태음 체질별로 부여된 상대적 점수로 판단 알고리즘에서 이를 이용하여 최종 판단을 제공하게 된다. 체질 점수는 다음과 같은 방법으로 생성된다. 입력으로 들어온 음성 신호는 144개의 음성 변수로 변환된 후 소음, 소양, 태음점수가 모두 0으로 초기화된 체질점수 테이블과 함께

2×144개의 *gRule* 논리 규칙을 통과하게 된다. 식(1), (2)와 같은 하나의 논리 규칙을 통과하게 될 때 논리 규칙 조건을 만족하게 되면 그때마다 해당되는 체질 점수 테이블의 *ui* 혹은 *li* 체질에 -1점의 별점을 부여하였다. 결과적으로 입력 음성 데이터는 모든 *gRule* 논리 규칙을 통과함으로써 최종 체질 점수를 얻게 된다. 다음으로 특정 체질에서 비정상적으로 넓은 분포 형태를 보여 오류를 야기할 수 있는 확률을 줄이기 위해 아래와 같이 정규화 벡터(normalized vector) 방법을 응용한 체질별 가중치(W)를 고안하였다. 먼저 *train set*을 대상으로 앞서 서술한 방식대로 *gRule*을 통과시켜 각 샘플들의 체질 점수를 구하고 체질별로 점수를 모두 합하여 하나의 1×3 행렬(S)로 만든다. 체질 점수에 행렬 S를 나누어 체질 점수 비율을 생성하고 식 (3)과 (4)를 이용하여 생성된 배열의 정규화 벡터의 역수를 생성하여 이를 체질별 가중치(W)로 정한다. 결국 *train set*에서 학습을 통해 *sub-gRule*을 포함한 *gRule*과 각 체질별 가중치(W)를 얻을 수 있게 된다.

$$S = [s_{se} \ s_{sy} \ s_{te}] \tag{3}$$

s_{se} : 소음 점수, s_{sy} : 소양 점수, s_{te} : 태음 점수

$$a = \frac{s_{se}}{s_{se} + s_{sy} + s_{te}}, \quad b = \frac{s_{sy}}{s_{se} + s_{sy} + s_{te}}, \quad c = \frac{s_{te}}{s_{se} + s_{sy} + s_{te}}$$

$$W = \sqrt{a^2 + b^2 + c^2} \cdot \left[\frac{1}{a} \ \frac{1}{b} \ \frac{1}{c} \right] \tag{4}$$

3.4 판단 알고리즘

입력된 음성 데이터에 대한 체질 판단은 미리 생성한 *gRule*과 가중치를 이용하여 체질 점수를 얻어내는 부분과 이를 비교하여 최종 판단을 내리는 판단 규칙 알고리즘 부분으로 두 단계에 거쳐 이루어진다.

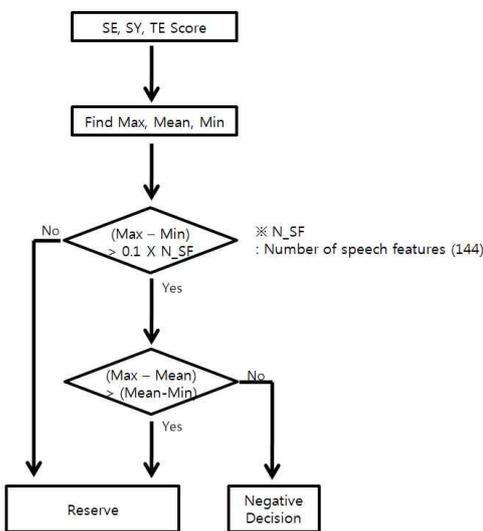


그림 4. 체질 점수와 판단 규칙

Figure 4. Constitutional Score and Decision Rule

첫 번째 단계로, 입력된 음성 신호는 소음, 소양, 태음체질 점수가 모두 0으로 초기화 된 상태에서 *gRule* 논리 규칙을 통과하게 된다. 논리 규칙을 통과하게 될 때 조건식을 만족하게 되면 해당하는 *ui* 혹은 *li* 체질 점수에 가중치(W)만큼의 별점을 부과한다. 두 번째 단계로, 최종적인 체질 판단은 부여받은 체질 점수 관계를 조사하여 “판단 보류”와 “특정 체질은 아니다.”라고 하는 이진 부정 판단 형식의 판단을 내리도록 하였다. 판단 기준은 <그림4>에서 나타나듯이 부여받은 체질 점수의 최대, 중간, 최소값을 찾아서 최대값과 최소값의 차이가 사용된 음성 변수의 개수(N_SF) 즉, 144개의 10%인 14점 이상이 되는 경우만을 대상으로만 판단을 내렸고 그 외의 모든 경우는 모두 음성으로는 판단하기 힘들다는 의미로 “판단 보류” 결과를 제시한다. 판단을 내리지 않는 경우에는 다시 최대값과 중간값의 차이와 중간값과 최소값의 차이를 조사하여 전자가 클 경우에는 판단을 보류하고 후자가 클 경우에는 최소값을 가졌던 체질에 대해서만 “이 체질은 아니다”라는 부정 판단을 내리도록 하였다.

4. 실험 결과

4.1 알고리즘 성능 테스트 결과

음성 데이터를 남녀별로 따로 나누고 성별로 각각 5번의 10-fold 교차 검증(cross validation) 테스트 방법으로 구현한 알고리즘 성능을 확인하였다. 즉, 전체 데이터를 동일한 크기의 10개 블록으로 나누고 이 중 9개 블록에 해당되는 데이터를 *train set*으로 설정, 학습하여 *gRule*과 가중치 행렬을 생성한 후 이를 나머지 한 개의 *test set* 블록에 적용한다. *Test set* 블록의 순서를 중복되지 않게 바꿔가면서 10번 반복하게 되면 모든 데이터가 한 번씩 *test set*에 포함되고, 결과적으로 모든 데이터가 빠짐없이 알고리즘에 의해 판단이 내려지게 되면 비로소 한번의 10-fold 교차 검증이 끝나게 된다. 알고리즘 성능 결과는 본 논문에서 구현된 알고리즘이 얼마나 많은 데이터에 대해서 ‘이진 부정 판단’과 그 반대인 ‘판단 보류’ 결정을 하였는지를 확인하기 위해 전체 데이터 개수를 기준으로 알고리즘으로 특정한 ‘이진 부정 판단’을 받은 수가 얼마나 되는지를 계산하여 이를 판단율로 표기하였다. 그 다음으로 예측된 판단이 사상체질 전문의와 약진으로 확진된 체질을 얼마나 정확하게 맞추었는지를 확인하기 위해 판단이 내려진 샘플 수에서 정확히 맞은 개수를 계산하여 이를 정확률로 나타내었다. 성능 결과 실험은 체질별 변별력을 높이기 위해 *train set*에서 4개의 조건 변수를 추출 시에 10% 백분위수를 적용하여 모든 음성 변수에 대해서 각 체질별 최소값과 최대값 대신 10th와 90th 백분위수를 적용하였다. 남녀별로 각각 5번의 10-fold 교차 검증을 실시하였으며 한 번의 교차 검증마다 무작위로 블록을 새로 지정하였다. 5번의 10-fold 교차 검증 결과, 남성의 경우는 전체 중에서 54.7%에 대해서만 이진 부정 판단을 내렸으며, 이러한 판단의 정확도는 74.0%이었다. 여성의 경우는 전체 중에 56.3%에 대해 판단하고 정확도는 72.1%이었다. 남성과 여성 모두 다른 체질보다 소음/비소음 부분에서 상대적으로 높은 정확률을 보이고 있다.

표 4. 남성을 대상으로 한 결과

Table 4. Algorithm Result of the Male Patients

No	체질	판단율	정확률
1	소음/비소음	0.260	0.844
	소양/비소양	0.191	0.848
	태음/비태음	0.156	0.556
2	소음/비소음	0.214	0.811
	소양/비소양	0.145	0.840
	태음/비태음	0.156	0.630
3	소음/비소음	0.231	0.850
	소양/비소양	0.179	0.839
	태음/비태음	0.150	0.500
4	소음/비소음	0.191	0.879
	소양/비소양	0.197	0.795
	태음/비태음	0.116	0.500
5	소음/비소음	0.225	0.846
	소양/비소양	0.197	0.824
	태음/비태음	0.127	0.545
평균	소음/비소음	0.225	0.846
	소양/비소양	0.182	0.830
	태음/비태음	0.141	0.546

표 5. 여성을 대상으로 한 결과

Table 5. Algorithm Result of the Female Patients

No	체질	판단율	정확률
1	소음/비소음	0.165	0.784
	소양/비소양	0.171	0.642
	태음/비태음	0.248	0.675
2	소음/비소음	0.194	0.750
	소양/비소양	0.142	0.682
	태음/비태음	0.197	0.656
3	소음/비소음	0.187	0.793
	소양/비소양	0.181	0.768
	태음/비태음	0.242	0.667
4	소음/비소음	0.194	0.800
	소양/비소양	0.145	0.689
	태음/비태음	0.216	0.746
5	소음/비소음	0.171	0.755
	소양/비소양	0.165	0.745
	태음/비태음	0.203	0.683
평균	소음/비소음	0.182	0.776
	소양/비소양	0.161	0.705
	태음/비태음	0.221	0.685

4.2 체질 분류에 기여한 음성 변수

알고리즘 성능 테스트에서는 체질 분류 판단 및 정확률뿐만 아니라 144개의 음성 변수 중에서 어떠한 변수가 체질 판별에 가장 크게 작용했는지도 조사하였다. 이는 앞서 서술한 총 5번의 10-fold 교차 검증 과정 중 test set을 대상으로 한 실험에서 gRule 논리 규칙을 만족하여 체질 점수가 발생되었을 때마다 해당되는 음성 변수 발생 횟수를 세어 최종테스트 동안 몇 번이 발생되었는지를 조사한 후 그 결과를 오름차순으로 순위를 매겨 1위부터 20위까지에 해당되는 음성 변수를 <표6>에 나열하였다. 그 결과 남성의 경우에는 F₀계열의 변수, 여성의 경우에는 포먼트 계열의 변수가 상대적으로 높은 순위를 차지하고 있었다. 이는 남성과 여성에서 각각 해당 계열의 변수가 체질 분류에서 다른 변수들에 비해 상대적으로 큰 역할을 담당하고 있다고 결론지을 수 있다.

표 6. 체질 분류에 기여한 음성 변수 순위

Table 6. Rank of Speech Features for Classification

Rank	남성	여성
1	P50[Hz]*	aF1[Hz]**
2	aMFCC12	aMFCC8
3	oF0[Hz]*	oMFCC12
4	oT0[ms]*	uF1[Hz]**
5	P90[Hz]*	oF2[Hz]**
6	aF1[Hz]**	uMFCC11
7	iF0[Hz]*	eMFCC4
8	uT0[ms]*	aMFCC6
9	iT0[ms]*	eMFCC6
10	aT0[ms]*	iMFCC10
11	oMFCC12	oMFCC5
12	IHL	eT0[ms]*
13	iRAP[%]	I50[dB]
14	uF2[Hz]**	eMFCC9
15	eF0[Hz]*	iF0[Hz]*
16	aF0[Hz]*	eF0[Hz]*
17	eT0[ms]*	uCO
18	uBW1[Hz]**	aMFCC9
19	uF0[Hz]*	I10[dB]
20	oBW1[Hz]**	eMFCC10

* : F₀계열 변수, ** : Formant계열 변수

5. 결론 및 고찰

사상체질과 인간의 목소리 간의 관계는 그동안 한의사들의 경험적인 측면에서 상당한 연관성이 짐작되어 왔다. 하지만 체질 진단은 음성뿐만 아니라 안면, 체형, 피부 등 여러 가지 요소들을 살펴서 판단을 내리며, 이러한 진단 과정에서는 한의사 개인의 취향과 경험에 의해 중요하게 여기는 진단 요소가 달라지기도 한다. 대체적으로 음성은 체질 진단에 있어서는 다른 요소들에 비해 추가 아닌 보조적으로 사용되며 서두에서도 언급되었듯이 원전에서 밝히는 체질별 음성 특징이 추상적이고 모호한 표현이 많아 이에 대한 물리적인 해석이 상당히 어렵다. 본 연구는 이러한 현실적인 한계 내에서 음성이 체질분류에 기여할 수 있는 부분을 찾아 객관적이고 정량적인 형태로 결과를 도출하고자 하였다. 이를 위해 판단 결과는 “판단 보류”와 “특정 체질은 아니다.” 라는 이진 부정 판단 형식을 취하였다. 이는 이번 연구와 같이 음성 신호 단독으로 체질을 구분하는 시스템에 직접 도입하기 위한 목적도 있지만, 차후 안면, 피부, 체형 등 복합적인 체질 진단 요소들을 가지고 체질을 진단하는 통합 진단 툴의 일부로 내장되어 특정 체질 분류에 국한되어 사용할 수도 있도록 계획되었다. 알고리즘 결과로는 이진방식으로 약 50% 이상의 판단율과 70% 이상의 정확률을 얻어낼 수 있었다. 체질별로는 다른 체질에서보다 특히 소음/비소음 분류에서 우수한 성능을 보이고 있다. 이러한 결과는 차후 통합 알고리즘에서도 다른 체질 요소들과 함께 하나의 요소가 특정 체질 분류에 강점을 가지는 부분을 맡아서 담당하는 분업화 형태로 알고리즘을 확장시킬 수 있을 것으로 생각된다. 음성 변수 중에서는 체질을 구분하는데 가장 많은 영향력을 준 변수로는 남성에서는 F₀, 여성의 경우는 포먼트 변수로 조사되었다. 이는 이미 여러 연구에서 조사된 피치와 남성호르몬 테스트스테론(testosterone) 수치간의 밀접한 관계[13]-[14]를 고려해 보면 남성의 체질에 의한 목소리 특성도 테스트스테론 수치와 상당한

연관성이 있을 것으로 간접적으로 유추해 볼 수 있다. 또한 여성의 경우는 남성에 비해 테스토스테론의 영향이 적기 때문에 조음기관 및 넓게는 안면의 기하학적인 형태가 반영된 포만트 계열 변수가 체질과 연관성이 있는 것으로 판단된다. 하지만 이런 남성호르몬 및 조음기관의 기하학적인 형태와 체질간의 관계를 밝히기 위해서는 좀 더 깊은 연구가 필요할 것으로 사료된다.

참 고 문 헌

- [1] WHO. (2007). "WHO International Standard Terminologies on Traditional Medicine in The Western Pacific Region", <http://www.who.int>
- [2] Park, S. H., Kim, M. G., Lee, S. J., Kim, J. Y. & Chae, H. (2009). "Temperament and character profiles of sasang typology in an adult clinical sample", *Evid. Based Complement. Altern. Med.*, Advance Access published on April 20; doi:10.1093/ecam/nep034
- [3] Park, S. C. & Kim, D. J. (2004). "Implementation of the automatic pulse-power diagnostic system and the discrimination algorithm of four constitutions", *Journal of IEEK SC*, Vol. 41, No. 2, pp. 53-60
(박승창, 김대진 (2004). "사상 체질 판별 알고리즘과 자동 맥진 시스템의 구현", 전자공학회논문지 제41권 SC편 제2호, pp. 53-60, March)
- [4] Chiu, C. C., Chang, H. H. & Yang, C. H. (2000). "Objective auscultation for traditional Chinese medical diagnosis using novel acoustic parameters", *Comput. Methods Programs Biomed*, Vol. 62, pp. 99-107.
- [5] Chiu, C. C., Yang, M. T. & Lin, C. S. (2002). "Using fractal dimension analysis on objective auscultation of traditional Chinese medical diagnosis", *J. Med.Biol. Eng.*, Vol. 22, pp. 219-225.
- [6] Kim, D. R. (1999). *The Principle of Life Ppreservation in Oriental Medicine*, Chungdam.
(김달래 (1999). 동의수세보원초고, 청담.)
- [7] Cho, D. U., (2006). "Sasang Constitution Classification by Speech Signal Processing", *Journal of KICS*, Vol. 31, No. 5C, pp. 548-555, May.
(조동욱 (2006). "음성 신호 분석에 의한 사상 체질 분류", 한국통신학회논문지 제31권 제5C호, pp. 548-555.)
- [8] Moon, S. J., Tak, J. H. & Hwang, H. J. (2005). "A phonetic study of 'Sasang Constitution' ", *Malsori*, Vol. 55, pp. 1-14.
(문승재, 탁지현, 황혜정 (2005). "음성학적으로 본 사상체질" 대한음성학회지:말소리, 55권, pp. 1-14.)
- [9] Boersma, P. (1993). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound", *Proceeding Institute of Phonetic Sciences* Vol. 17, pp. 97-110.
- [10] Yang, B. G. (2003). *Speech Analysis Using Praat Script*, Mansu (양병곤 (2003). 프라트를 이용한 음성분석의 이론과 실제, 만수출판사.)
- [11] Ko, D. H. & Jeong, O. R. (2001). *The Use of Speech and Language Analyzer*, Hankukmunhwasa
(고도홍, 정옥란 (2001). 음성 및 언어 분석기기 활용법, 한국문화사.)
- [12] Dellar, J. R., Hansen, J. H. L. & Proakis, J. G. (1999). *Discrete- Time Processing of Speech Signals*, Wiley-IEEE Press, pp. 380-385.
- [13] Dabbs, J. M. & Mallinger, A. (1999). "High testosterone levels predict low voice pitch among men", *Personality and Individual Differences*, Vol. 27, pp. 801-804.
- [14] King, A., Ashby, J. & Nelson, C. (2001). "Effects of testosterone replacement on a male professional singer", *Journal of Voice*, Vol. 15, No. 4, pp. 553-557.

• 강재환 (Kang, Jaehwan)

한국한의학연구원 체질생물학 · 의공학연구센터
대전시 유성구 엑스포로 483
Tel: 042-868-9301 Fax: 042-868-9480
Email: doskian@kiom.re.kr
관심분야: 의용생체공학, 음성신호처리

• 유중향 (Yoo, Jonghyang)

한국한의학연구원 체질의학임상연구센터
대전시 유성구 엑스포로 483
Tel: 042-868-9591 Fax: 042-868-9480
Email: jhyoo@kiom.re.kr
관심분야: 임상연구

• 이해정 (Lee, Haejung)

한국한의학연구원 체질생물학 · 의공학연구센터
대전시 유성구 엑스포로 483
Tel: 042-868-9320 Fax: 042-868-9480
Email: haejung0614@kiom.re.kr
관심분야: 다변량통계분석, 신뢰성분석

• 김종열 (Kim, Jongyeol)

한국한의학연구원 체질생물학 · 의공학연구센터
대전시 유성구 엑스포로 483
Tel: 042-868-9489 Fax: 042-868-9480
Email: ssmmed@kiom.re.kr
관심분야: 사상체질의학