

# 음성학적으로 본 사상체질<sup>1)</sup>

문승재, 탁지현, 황혜정  
아주대학교 영어영문학과

## A Phonetic Study of "Sasang Constitution"

Seung-Jae Moon, Ji-Hyun Tak, Hyejeong Hwang  
Department of English, Ajou University

### Abstract

Sasang Constitution, one branch of oriental medicine, claims that people can be classified into four different "constitutions:" Taeyang, Taeum, Soyang, and Soeum. This study investigates whether the classification of the "constitutions" could be accurately made solely based on people's voice by analyzing the data from 46 different voices whose constitutions were already determined. Seven source-related parameters and four filter-related parameters were phonetically analyzed and the GMM(gaussian mixture model) was tried with the data. Both the results from phonetic analyses and GMM showed that all the parameters (except one) failed to distinguish the constitutions of the people successfully. And even the single exception, the bandwidth of F2, did not provide us with sufficient reasons to be the source of distinction. This result seems to suggest one of the two conclusions: either the Sasang Constitutions cannot be substantiated with phonetic characteristics of peoples' voices with reliable accuracy, or we need to find yet some other parameters which haven't been conventionally proposed.

### I. 서론

이제마가 창시한 사상체질의학에서는 사람의 체질을 그 사람의 생리에 따라 태양, 태음, 소양, 소음의 네가

<sup>1)</sup> 이 논문은 2003년도 한국학술진흥재단의 지원에 의하여 연구되었음 (KRF-2003-041-A20183)

지로 분류한다.[1] 본 연구의 목적은, 임상한의사들이 진단한 사상체질을 그대로 인정한 상황에서, 사람의 음성의 특색만으로 그와 같이 사상체질을 분류할 수 있는지를 알아보려 하는 것이다.

유준상[2]에 의하면 음성평가를 통한 사상체질진단의 전망은 매우 밝으나, 그 과정을 더욱 객관화시키는 노력이 필요하다고 말하고 있다.

만일 사람을 그 음성의 객관적 특색만으로 한의사들이 분류한 것과 같은 체질대로 분류할 수 있다면, 그것은 음성과 그 음성이 만들어내는 이미지 사이에 우연을 넘는 관계가 있다는 연구[3]와 더불어, 음성이 그 음성의 주인공에 대하여 구체적이고 풍부한 정보를 포함하고 있음을 입증할 것이다.

이미 앞서 이와 같은 연구가 실시된 바 있었다 [4], 이 연구에서는 23명의 자료를 음성의 변수 중 음원(source)과 관계된 변수를 중심으로 분석하였으나, 체질간에 뚜렷한 차이를 찾지 못하였다. 그래서 본 연구에서는, 음원이외에도 filter에 관계된 변수를 포함하고, 더 많은 인원을 대상으로 연구를 진행하였다.

### II. 연구 방법

#### 1. 화자

본 연구의 분석대상 화자는, 사상체질의학 전공의 박사과정 논문[2]에서 분류하고 사용한 화자 73명 중, 녹음자료가 그대로 보존되어 있는 46명이었다.<sup>2)</sup> 이들은

<sup>2)</sup> 분석대상 화자의 선정은 본 연구의 가장 큰 제한점이었다. 원래 사상체질의학에서 한 사람의 체질을 체질대로 판명하는 데에는 병리, 생리, 체질 등 여러 가지의 진단 및 관찰이 요구되며, 이는 일반적으로 1년 이상의 긴 시간을 요한다. 그리고 본 연구자는 사상체질의학에 종사하지 않기 때문에 사상체질판단이 이미 내

모두 부산에 소재한 한 대학의 간호학화에 재학하고 있는 여학생들로서, 이 46명은 소음인 12명, 소양인 15명, 태음인 19명으로 태양인은 한 명도 없었다.<sup>3)</sup> 이들 화자의 사상의학적 체질은 사상의학전공 임상 한의사 5명이 일치하게 판단한 결과이다.

## 2. 녹음자료

분석에 사용한 녹음자료는 유준상[2]의 자료를 사용하였다.

## 3. 분석변수

지금까지 음성의 특성을 나타내는 변수로는 흔히 기본주파수의 변화폭이나 성문에서 생기는 파형의 스펙트럼형태, 포먼트, 비음의 스펙트럼, long-term spectra, 스펙트럼의 기울기 등이 사용되었다 (음성의 특성을 연구하는데 사용된 변수에 대한 논의는 Remez, Fellowes, & Rubins[5]의 논의를 참조.)

본 연구에서는 이들 중, 다른 특별한 장비없이 측정이 가능한 변수를 모두 포함시키고자 다음과 같은 변수를 측정하였다. Source 측면의 변수로는 ①평균기본주파수, ②음역, ③평균기본주파수분산율, ④전체발화중 무성발화의 비율, ⑤voice break, ⑥harmonicity, ⑦ [아] 모음 구간의 h1, h2, h3의 형태를 측정하였고, filter 측면의 변수로서 [아] 모음 구간의 ⑧F1, F2, F3, ⑨각 포먼트의 bandwidth, ⑩long-term average spectra의 기울기를 측정하였으며, 이 외에도 ⑪발화속도를 측정하였다.

이 변수들 중 일부는 (일테면 ③, ④, ⑤, ⑥) 원래 [아] 모음을 지속적으로 발화한 자료를 분석하여 언어병리학적인 이상을 찾을 때 쓰는 변수들이지만, 그 성격상, 똑같은 문장을 발화한 자료를 분석한 결과라면, 그러한 특성들이 사람간의 특성을 구분지을 수 있을 것이라는 가정하에 사용하였다. 예를 들면, 사람에 따라 같은 문장이라 하더라도 voice break를 더 많이 보일 수도 있고, 더 적게 보일 수도 있을 것이다. 또한 harmonicity는, 주기적인 부분에 있는 음향학적인 에너지의 부주기적, 즉 잡음부분에 있는 에너지에 대한 비

려진 사람에 연구의 대상을 국한할 수 밖에 없는 상황이다. 따라서, 분석대상 화자를 선정해서 그 자료를 분석하여 결론에 이르는 시간을 고려할 때, 분석대상 화자의 선정은 전적으로 사상체질의학계의 도움에 의존할 수 밖에 없었다.

<sup>3)</sup> 사상체질의학에 의하면 태양인은 우리나라 인구 중 0.01%에도 못 미치는 희귀한 체질로서 매우 찾기 힘들다고 한다.

율로서, 사람의 발음 상의 전반적인 특성을 나타낼 수 있을 것이라고 판단되어 포함하였다.

위 변수들의 측정을 위하여, 약 1분 30초 정도의 전체 발화 중, 어느 화자도 실수하거나 녹음 상태가 고르지 못한 부분이 없는 부분(평균 32.8초)을 선택하였다.

①부터 ⑥까지의 변수는 이 부분 전체를 구간으로 하여 측정하였다. ⑧, ⑨의 변수는 선택 부분 중, 가장 먼저 나오는 [아] 모음 세 군대를 정하여 화자 당 3번의 측정치를 얻었다. 포먼트는 스펙트로그램과 DFT를 참조하여 개별적으로 측정하였다. ⑦, ⑩의 변수는, 위의 35초 가량의 부분 중, 최소한 200ms 이상 지속되는 [아] 모음 전체를 구간으로 측정하였다. ⑪은 선택부분의 음절수(233음절)를 걸린 시간으로 나눈 후 이를 바탕으로 1분에 읽는 음절수를 측정하였다.

위의 변수들에 대한 분석에는 Praat 4.3.02를 사용하였다.

## 4. GMM (Gaussian mixture model)<sup>4)</sup>

본 연구에서는 음성학적 분석 이외에 GMM을 이용하여 체질과 음성간의 연관성을 찾아보았다.

GMM을 학습하기 위해서는 각 클래스의 특징을 잘 표현할 수 있도록 충분한 양의 데이터가 있어야 한다 그리고 모델을 테스트하기 위한 데이터 또한 충분히 있어야 한다. 그런데, 본 실험에서 사용한 데이터는 한 클래스 당 최소 12명 뿐 이므로 이를 다시 학습데이터와 테스트 데이터로 나누면 그 수가 너무 작아서 신뢰성 있는 결과를 얻기 어려우며, 또한 학습 데이터와 테스트 데이터를 어떻게 나누는가에 따라서 결과가 달라질 수 있다. 따라서, 본 실험에서는 leave-one-out 방법을 사용하였다.

실험 과정은 다음과 같다. 데이터가 수집된 모든 사람 x에 대하여 다음의 과정을 수행한다. 이 사람 x가 속한 체질 클래스 C에서 x를 제외한 모든 사람의 데이터를 이용하여 클래스 C의 모델을 학습시킨다. 테스트는 x와, 클래스 C에 속하지 않은 모든 사람의 데이터를 이용한다. 즉, 클래스 C를 학습하는 데 사용된 데이터가 클래스 C의 독특한 성질을 잘 표현하고 있다면 다른 클래스의 사람들의 데이터보다 x의 데이터가 클래스 C에 속할 우도(likelihood) P(x|C)가 가장 높게 나와야 할 것이다. 본 실험에서는 46명의 모든 사람에 대하여 이와 같은 실험을 반복하여 실제로 체질별 클래스가 잘 모델링 되는지 알아보았다. 또한, 각 화자에

<sup>4)</sup> GMM을 이용한 실험을 위해 수고해주신 서울시립대학교 컴퓨터과학부의 유하진 교수님께 감사드립니다.

대하여 혼합물(mixture)의 수를 1, 2, 4, 8, 16, 32, 64, 128, 256 과 같이 9가지로 하여 학습해 보고 테스트 하였다.

특징파라미터는 15차의 멜켄스트럼 계수(melcepstral coefficient)를 사용하였고, 분석창 넓이는 250msec, 간격은 100msec으로 하였다.

### III. 분석 결과

#### 1. 음성학적 변수

##### 1.1 ANOVA test 결과

⑦H1, h2, h3의 변화형태를 제외한 음성학적 변수는 ANOVA 테스트를 하여 그 의미를 검증하였다. 모든 음성학적 변수 중, F2의 bandwidth만이 유일하게 5% 수준에서 집단 간 유의미한 차이를 보였을 뿐, 나머지 변수들은 source, filter에 관계없이 전혀 유의미한 차이를 보이지 않았다. 이러한 결과는, 집단간 분석에서도 똑같은 결과를 나타냈으며, 혹시 놓쳤을지 모를 두 집단간의 차이를 검증하기 위한 두 종류의 사후분석(Scheffe, Bonferroni)에서도 마찬가지였다.

ANOVA 결과 유일하게 유의미한 차이를 보인 변수는 F2의 bandwidth였다. 그리고 이 변수도 사후검증 결과, 소음인과 태음인, 소양인과 태음인 사이에 유의미한 차이를 보일 뿐, 소음인과 소양인 사이에서는 유의미한 차이를 보이지 않았다.

##### 1.2 harmonic structure

변수 중 ⑦ harmonic structure는 통계적으로 처리하기에는 다소 부족한 면이 있어서 정성적인 분석을 시도하였다.

배음간의 전반적인 양상은 ⑩의 long-term average spectra와 어느 정도 관계가 있을 것이지만, phonation type을 알아보기 위하여 본 연구에서는, ⑦을 h1-h2와 h1, h2, h3의 상대적인 크기의 변화를 보는 두 가지로 다시 나누어 분석을 시도하였다.

h1-h2의 경우에는, 다른 변수들과 마찬가지로 세 집단간에 유의한 차이를 볼 수 없었다. 다른 변수들과는 달리 소음:소양, 소음:태음 사이에 어느 정도 차이가 있어보였지만 (각각  $p < 0.08$ ,  $p < 0.07$ ), 이 역시 통계학적으로는 무의미한 것이기 때문에 중요성을 둘 수는 없다

첫 세 배음의 관계를 정성적으로 평가하기 위하여 본 연구에서는 세 배음의 관계를 다음의 네 가지 형태로 나누어보았다.

A 유형:  $h1 < h2 < h3$

B 유형:  $h1 < h2 > h3$

C 유형:  $h1 > h2 < h3$

D 유형:  $h1 > h2 > h3$

이 유형별 빈도수를 보면 다음 표 1과 같다.

표 1 배음구조의 형태별 빈도수

유형 체질	A	B	C	D
소음	3	2	5	2
소양	0	2	11	1
태음	0	2	15	2

표1에서 보듯이, 소음인의 경우만 유형 A가 보였고 그 외의 체질에서는 전혀 보이지 않았으며, 소양, 태음인의 경우에는 유형 C가 압도적으로 많았음을 볼 수 있었다. 그러나 이것을 보고 일테면 "A유형이 나오면 그것은 곧 소음인"이라는 결론을 이끌어 낼 수 있을 수 있는 것인지에 대해서는 신중하게 생각해야 할 것이다. 이러한 현상을 정량적으로 어떻게 결론지을 수 있을지에 대해서는 더 많은 노력이 필요할 것이다

#### 2. GMM

표1은 각각의 체질 클래스에 대하여 혼합수를 바꾸어서 실험했을 때 평균 순위 백분율이다. 즉, 클래스 C의 학습에 참가하지 않은 x 및 모든 다른 클래스의 데이터의 클래스 C에 대한 우도를 계산하였을 때 x가 몇 번째로 큰 값을 가지는 지를 백분율로 표시한 것이다. 이때 1위일 경우 100%로 표시하였다. 여기서 x만이 클래스 C에 포함되어 있으므로 이상적인 경우에는 항상 1위가 나와야 하지만, 실험 결과 혼합수 256에서 모든 순위백분율의 평균이 46.5%가 되었다. 그리고, 순위백분율이 100%일 확률은 2.2%밖에 되지 않았다. 그러므로, 본 실험에서 알 수 있는 것은 현재 가지고 있는 데이터로는 사상체질과 음성간의 연관성을 찾기 어렵다고 하겠다.

표 2. 평균 순위 백분율

mixture	4	8	16	32	64	128	256
태음	42.20	45.08	45.94	44.79	44.27	43.21	42.31
소음	46.95	51.02	49.33	48.05	50.33	48.57	48.38
소양	49.88	46.67	47.06	47.83	49.06	50.27	51.27

### IV. 논의

본 연구에서 살펴본 11가지의 변수는 F2 bandwidth 한 가지를 제외하고는 어느 것도 사상체질에 의한 새

체질간의 차이를 통계적으로 유의미하게 구별할 수 없었다. 분산분석에서 유의미한 차이를 보인 F2 bandwidth의 경우도, 그 결과로부터 곧바로 F2 bandwidth가 사상체질을 나눌 수 있는 변수라고 결론을 내리기에는 매우 조심스러운 면이 있다. 한 사람의 bandwidth라면 F1, F2, F3의 bandwidth간에 어느 정도 관계가 있는 것이 자연스러운 것인데, 다른 포만트는 전혀 bandwidth에서 차이를 보이지 않는 반면 유독 F2의 bandwidth만 보인다면, 이것을 과연 필연적인 결과로 보아야 할 것인가를 신중하게 검토해야 할 것이다. 즉, F2의 bandwidth만이 다르게 나타날 이유를 납득할 수 없는 상태이기 때문에 성급한 결론을 내릴 수 없다고 본다. 이에 대한 확정적인 결론을 얻으려면 더 많은 자료에 대한 분석이 필요할 것이다.

그렇다면, 이 결과를 어떤 의미로 해석해야 할 것인가? 두 가지의 해석이 가능할 것이다.

첫째는 사상체질의학의 네 가지 체질은 음성학적으로는 구별할 수 없는 것이라는 해석이고, 다른 하나는 구별은 가능하되, 본 연구를 포함하여 지금까지 살펴본 변수로는 안 되므로 새로운 변수를 찾아야 한다는 해석이다.

이 중 어느 것이 타당한 결론인지에 대해서는 더 많은 자료와 분석이 필요할 것이다. 특히 앞서 언급한 것처럼 사상체질의학상의 판단이 전제되는 연구인만큼, 사상의학계와의 밀접한 연관이 전제된다면, 더 많은 자료를 가지고 연구하는 것이 가능할 것이다.

그러나 사람들의 목소리로부터, 그 사람들을 사상체질의학에서 구분하는 네 가지 체질, 그 중에서도 현실적으로 관찰이 가능한 세 가지 체질로 구분할 수 있는 공통분모를 찾는다는 대명제는, 얼핏 보면 쉬운 것 같지만, 오히려 십여가지의 체질로 나누는 것보다 더 어려운 일일 수 있을 것이다. 게다가 본 연구에서 살펴본 변수들은 지금까지 목소리에 대한 연구들에서 주로 다루은 중요 변수였다는[5] 점, 그 분석과정에서 매우 세심한 주의를 기울인 점 등을 고려할 때, 본 연구는 사상체질의학에서 진단한 세 가지 체질을 사람의 목소리만으로 구분한다는 것은 매우 어려운 일이 아닐까 하는 조심스러운 결론을 내려본다.

## 참고문헌

[1] 사상체질의학회 홈페이지  
(<http://www.esasang.com>), 2003.  
[2] 유준상, 청각적 음성분석을 통한 사상체질진단에 관한 연구, 상지대학교 한의학과 박사학위논문,

2002.  
[3] 문승재, “음성과 인상: 인지실험” 한국음향학회지, 18권 8호, 66-74, 1999.  
[4] 문승재, 박종주, 황해정. “음성과 사상체질: 음원을 중심으로,” 말소리 48, pp 15-33. 2003  
[5] Remez, R., Fellowes, J. & Rubin, P. “Talker Identification Based on Phonetic Information,” *Journal of Experimental Psychology: Human Perception and Performance* Vol.23 No.3, 651-666, 1997.