

음성학적으로 본 사상체질*

문승재(아주대), 탁지현(아주대), 황혜정(아주대)

<차 례>

- | | |
|----------------------|----------------|
| 1. 서론 | 3. 결과 |
| 1.1. 연구의 목적 | 3.1. 음성학적 변수 |
| 1.2. 선행연구 및 최근의 연구동향 | 3.1.1. 아노바 테스트 |
| 2. 실험 | 3.1.2. 배음의 형태 |
| 2.1. 화자 및 말자료 | 3.2. GMM |
| 2.2. 분석변수 및 분석방법 | 4. 논의 |
| 2.3. GMM | |

<Abstract>

A Phonetic Study of "Sasang Constitution"

Seung-Jae Moon, Ji-Hyun Tak, Hyejeong Hwang

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 on 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, B2 (the bandwidth of the second formant) 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.

* Keywords: Sasang Constitution, Voice, Voice quality, Emotional speech.

* 이 논문은 2003년도 한국학술진흥재단의 지원에 의하여 연구되었음. (KRF-2003-041-A20183)

1. 서론

1.1. 연구의 목적

사람의 목소리에 언어적인 정보와 부언어적인 정보가 모두 포함되어 있다는 것은 익히 알려진 사실이다. 그러한 부언어적인 정보가 우리 일상생활에 작용하는 예로서 우리는 목소리만 듣고도 그 말한 이의 감정상태를 추측할 수 있고, 한번 들은 목소리로 아주 쉽고 확실하게 사람을 식별하기도 한다.

사상체질의학에서는 이러한 음성의 부언어적인 정보를 의학적인 목적으로 사용하고 있다. 사상체질의학에서는 사람의 체질을 “태양인, 소양인, 태음인, 소음인의 네 가지 체질을 설정하고 각 체질에 대한 생리, 병리, 진단, 변증, 치료와 약물에 이르기까지 서로 연계를 갖고서 임상에 응용할 수 있는 새로운 방향을 제시”[1]하고자 한다. 이 사상체질의학에서 체질을 분류하는 기준 중에는, 체형, 인상 등 사람의 외적인 면이 포함되어 있으며[2] 그 중 하나의 기준이 바로 그 사람의 음성이다. 일례로 사상체질의학 관련 문헌에서는 “태양인은 호흡기가 크므로 소리가 높다. 태음인은 성량이 풍부하여 소리가 무겁다. 소양인은 호흡기가 작으므로 소리가 가볍고 낮다. 소음인은 성량이 넓으므로 소리가 활발하다...”[3] 등, 음성의 질적인 특성에 주목하고 있다. 그런데 최근의 한 연구[4]는, 사상체질의학에 15년 이상 임상에 종사한 전문의 5명으로 하여금 여러 환자의 목소리를 정성평가하도록 한 후, 5명의 평가자내 일치도와 평가자간 일치도가 “상당히 유의”하여, 음성평가를 통한 사상체질의 진단이 가능하고 경험적인 근거가 있음을 보여줌으로써 사상체질의학의 목소리에 근거한 판단이 무작위적인 것이 아님을 보여주었다.

본 연구는 과연 어떤 음성학적 특성이 사상체질의학의 주장처럼 사람의 체질을 구별할 수 있게 하는지를 찾아봄으로써, 음성학적인 연구를 사상체질의학과 접목하고자 하는 시도이다. 목소리의 여러 음성학적 변수 중, 사상체질에서 구분한 체질과 가장 관련이 깊은 변수를 찾아냄으로써, 음성과 체질간의 상관관계를 확인하고자 하는 것이다.)

1.2. 선행연구 및 최근의 연구 동향

지금까지 음성학을 포함한 언어학 분야에서 수행된 연구들은 대부분 개별자·모음이 전달하는 언어적 정보에 매우 편중되어있어 부언어적 정보에 대한 연

1) 본 연구는 사상체질의학 자체에 대한 어떠한 결정적인 판단은 하지 않는다. 다만, 사상체질의학에서 진단한 체질을 그대로 받아들인 상태에서, 과연 사람의 음성의 특색만으로 그와 같이 체질을 분류할 수 있는지를 알아보려고 하는 것이다.

구는 비교적 취약한 실정이다.

이처럼 필터(filter)를 중심으로 한 음성학의 연구경향은 인간의 의사소통을 매우 풍부하게 하는 매우 중요한 부분, 즉 음원(source)을 이루는 성대에 대한 연구를 상대적으로 소홀하게 하였다. 인간의 감정 등 감성적인 요소를 전달하는 대부분의 요소가 성대에서 결정된다는 사실을 생각할 때에 음원에 대한 연구가 더 활발해야한다는 것은 당연한 이치일 것이다.

이러한 추세를 반영하듯 최근에는 음원에 대한 연구가 활성화되기 시작하였다. 예를 들면 서양음악의 성악 연구에서는, 인간의 목소리 자체에 실려있는 풍부한 언어외적인 정보에 대한 연구가 많이 이루어져 있다. 뿐만 아니라 음성 공학 분야에서도, 이제는 정확한 음소의 전달을 뛰어넘어서 좀 더 자연스러운 소리를 만들기 위한 노력이 점차 늘어나고 있다.[5,6,7,8]

국내 음성학계에서도 최근 들어 이와 맥락을 같이 하여 음성 자체에 대한 연구가 고개를 들고 있다. 예를 들면, 어떤 특정한 목소리를 듣고 사람들이 떠올리는 인상은 우연수준이상으로 상당히 공통점이 있다는 일관된 일련의 연구결과가 있다.[9,10] 이 결과는 말소리를 듣는 사람들이 그 음성이 갖는 부언어적 정보로부터 구체적으로 일관된 어떤 인상을 받는다는 것을 의미한다.

한편 사상체질의학계에서는 비록 음성을 체질판단의 한 기준으로 활용하고 있으나, 앞서 언급한 바와 같이 그 방법은 매우 주관적이다. 임상을 하는 의사가 환자를 대하면서 여러 가지 기준들과 함께 목소리를 질적으로 판단하는 것이 현재의 상황이다.

그러나 이 연구에서도 목소리를 “淸/濁, 高/低, 急/緩, 有力/無力”으로 구분하도록 해서, 비록 전문가들 사이에서는 일치된 판단이 내려지긴 했으나, 음성의 어떤 음향적 특징이 이 기준에 해당하는지에 대한 연구는 아직 본격적으로 이루어지지 않았다. 몇몇의 연구에서 음성과 체질의 관계를 규명하고자 하는 과학적인 노력이 시도되었지만 아직까지 이렇다할 결론은 얻지 못하고 있는 실정이다.[11,12,13,14] 이 연구들은 모두 음성분석을 통하여 체질에 따른 객관적인 변수를 찾고자 하는 노력을 기울였으나, 결과는 부정적이었다. 그렇지만 이 연구들은, 정도의 차이는 있으나, 그 방법 면에 심각한 문제가 있어서 (참여인 수의 부족, 음성학적인 변수의 통제 부족 등) 그 부정적인 결과가 곧바로 음성변수와 체질 간에 관계가 없음을 의미한다고 할 수는 없다.

[10]에서 나타난 것과 같이 음성만으로 체질을 분류할 수 있다면, 음성의 어떤 특색이 그러한 분류를 가능하게 하는 것일까?

구체적인 음성학적 분석을 바탕으로 사상체질의학분야에 접근한 시도로는 [15]가 있다고 하겠다. 이 연구에서는 사상체질의학 전문가들로부터 다른 여러 가지 근거에 의하여 체질을 판명받은 23명의 녹음자료를 수집하여, 체질별로 평균 f0, 최고f0, 최저f0, 음역(pitch range), 발성유형(phonation type), jitter, shimmer, 발화속도

등의 변수에서 공통분모를 찾아보았다. 이 연구는 여러 가지 변수 중 음원과 직접적으로 관련이 되는 변수들을 주로 고려하였으나 그 변수들과 체질간의 구체적인 관계를 찾지 못하였다. 다만 발성유형(phonation type)이 체질별로 약간 일관된 패턴을 보였을 뿐이었으나, 이 역시 통계적으로 입증되지 않았다.

본 연구는 음성학적으로 사상체질의학의 검증이 가능한지를 밝히려고 하는 연구로서, [15]의 연장이라 하겠다. [15]에서는 주로 음원에 대한 변수만 다룬 반면, 본 연구에서는 지금까지 음성학계와 심리학계에서 음성의 특성을 연구하는데 사용된 변수²⁾들을 모두 포함하여 연구함으로써, 사상체질의학의 네 체질을 나타내는 음성학적 변수가 있는지를 찾아보고자 하는 것이다.

이를 위하여 본 연구에서는 사상체질의학 임상³⁾에 의하여 체질판명을 받은 사람들의 목소리를 음성학적으로 분석하고 이를 통계 처리하여 체질과 음성학적 변수간의 상관관계를 찾아보고자 한다.

2. 실험

2.1. 화자 및 말자료

본 연구의 분석대상 화자는, 앞선 연구 [4]에서, 15년 이상 경력의 사상의학전공 임상 한의사 5명이 통계학적으로 유의미하게 분류한 화자 73명 중, 녹음자료가 그대로 보존되어 있는 46명이었다.³⁾ 이들은 모두 부산에 소재한 한 대학의 간호학과에 재학하고 있는 여학생들로서, 이 46명은 소음인 12명, 소양인 15명, 태음인 19명으로 구성되어 있었다.⁴⁾

녹음된 말자료는 본 연구자의 통제에 의하여 이루어진 것이 아니었으며, 녹음 내용은 우리나라 전래 동화인 ‘사또와 산딸기’의 한 부분으로서 약 1분 30초가량의 녹음이었다. 이 중, 모든 화자가 큰 실수(예를 들면, 기침을 한다든가, 읽는 단어를 잘못 발음하여 새로 고쳐서 발음하는 등의 불규칙적인 실수)를 하지 않고 녹

2) 음성의 특성을 연구하는데 사용된 변수에 대한 논의는 [16]을 참조.

3) 분석대상 화자의 선정은 본 연구의 가장 큰 제한점이었다. 원래 사상체질의학에서 한 사람의 체질을 제대로 판명하는 데에는 병리, 생리, 체질 등 여러 가지의 진단 및 관찰이 요구되며, 이는 일반적으로 1년 이상의 긴 시간을 요한다. 본 연구자는 사상체질의학에 종사하지 않기 때문에 사상체질판단이 이미 내려진 사람으로 연구의 대상을 국한할 수밖에 없는 상황이다. 따라서, 분석대상 화자를 선정해서 그 자료를 분석하여 결론에 이르는 시간을 고려할 때, 분석대상 화자의 선정은 전적으로 사상체질의학계의 도움에 의존할 수밖에 없었다.

4) 사상체질의학에 의하면 태양인은 우리나라 인구 중 0.01%에도 못 미치는 희귀한 체질로서 매우 찾기 힘들다고 하며, [4]에서 조사한 73명 중 태양인은 한 명도 없었다.

음상태가 고른 구간을 선택하여 실제 분석에 사용하였으며 그 측정부분은 다음과 같다.

“여봐라, 이방, 산딸기를 따오너라.” 이방은 사또의 명령을 듣고 어리둥절하였습니다. 왜냐하면, 겨울에는 산딸기가 없기 때문입니다. 그러나 사또는 막 무가내로 산딸기를 따오라고 하였습니다. 이방은 내년 여름에 산딸기를 듬뿍 따다 드리겠다고 대답하였습니다. “무엇이라고? 지금 당장 산딸기를 따오지 않으면 큰 벌을 내리겠다.” 사또는 다짜고짜로 이방에게 호령을 하였습니다. 사또의 호령을 들은 이방은 어찌할 줄 몰랐습니다. (평균 32.8초)

2.2. 분석변수 및 분석방법

지금까지 음성의 특색(voice quality)을 나타내는 변수로는, 기본주파수의 변화폭, 성문에서 생기는 파형의 스펙트럼형태, 포먼트, 비음의 스펙트럼, long-term spectra, 스펙트럼의 기울기 등이 연구되었다[16]. 본 연구에서는, source 위주의 변수만 분석했던 선행연구[15]의 제한점을 극복하면서, 음성의 특색 연구에 거론된 변수들을 가능하면 많이 포함시키고자 하여, 다음과 같은 변수를 측정하였다. Source 측면의 변수로는 ①평균f0, ②음역, ③평균f0분산률, ④전체발화 중 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 대역폭(bandwidth), ⑩long-term average spectra의 기울기를 측정하였으며, 이 외에도 ⑪발화속도를 측정하였다.

①평균f0, ②음역, ③평균f0분산률의 세 변수는 전체 측정부분에 대하여 각 화자의 기본주파수 변화를 보기 위한 것이다. 특히 ③평균f0분산률은, 의문문, 명령문 등이 포함된 동화를 읽으면서 얼마나 심하게 기본주파수를 변화했는지를 보기 위한 것으로, 체질이 사람의 성격과 차이가 있다면, 그 성격의 차이가 이러한 동화를 읽을 때, 이야기를 구현하는 과정에서 달리 표시되는지를 보기 위함이었다.

본 연구에 포함된 변수들 중 일부는 (예를 들면 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 가정 하에 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 비주기적인 부분, 즉 잡음부분에 있는 에너지에 대한 비율로서, 사람의 발음 상의 전반적인 특성을 나타낼 수 있을 것이라고 판단되어 포함하였다.

⑧의 포먼트 구조는 아마도 가장 전통적인 필터 특성 변수로서 지금까지 수많은

은 연구들이 이를 집중적으로 조명해왔다. 물론 포먼트 구조는 같은 모음이라도 성별, 연령별로 차이가 있음이 입증된 바 있으나 [17, 18], 비슷한 연령대의 여성들 사이에서도 체질에 따라 포먼트 구조가 다를지 확인해보고자 하였다.

⑨각 포먼트의 대역폭은 반드시 필터만의 변수라고 볼 수는 없겠으나, 각 포먼트 별 대역폭이 체질별로 다른지를 확인해보고자 하였다.

⑩[아] 모음 구간의 long-term average spectra의 기울기 변수는 사람의 발음특성과 밀접한 관계를 맺고 있어서, 탁한 소리(breathy voice), 쥐어짜는 듯한 소리(creaky voice)의 차이를 이와 연관지어 생각할 수 있다.[19] 따라서 체질별로 목소리가 달라서, 어떤 체질은 탁한 소리인 반면 어떤 체질은 맑은 소리가 특징이라면, 그 차이는 이 변수와 밀접한 관계가 있을 것으로 보인다. 일반적으로 spectrum은 그 측정위치에 따라 상당한 변화를 보일 수 있기 때문에, 본 연구에서는 [아] 발성 부분 전체를 대상으로 long-term average spectra를 구하였다. 다만, 지금까지 이 변수에 대하여 어떻게 측정하는지에 대하여 확립된 바가 없었으나, 본 연구에서는 Praat (version 4.3.02)의 'Get slope' 기능을 이용하여 측정하였다.

⑦[아] 모음 구간의 h1, h2, h3의 형태는 앞선 [15]에서, 비록 통계학적으로 증명하지는 못했으나, 유일하게 다소 의미가 있는 결과가 나타난 바 있었기에, 본 연구에서는 이 변수를 포함시켰다.

①부터 ⑥까지의 변수는 측정부분 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 32초 가량 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 계산하였다.

위의 변수들에 대한 분석에는 Praat 4.3.02를 사용하였다.

2.3. GMM (Gaussian Mixture Model)⁵⁾

본 연구에서는 음성학적 분석 이외에도 컴퓨터 모델 중의 하나인 GMM을 이용하여 체질과 음성간의 연관성을 찾아보고자 시도하였다.

GMM은 현재 목소리로 화자를 분별하는 화자인식에서 가장 많이 사용되는 방법이다. GMM은 평균과 공분산을 가지고 다차원 가우시안 확률 분포 함수에 의하여 특징 벡터를 모델링한다. D차원 확률분포함수 pdf 는 다음과 같이 계산된다.

5) GMM을 이용한 실험을 위해 수고해주신 서울립대학교 컴퓨터과학부의 유하진 교수님께 감사드립니다.

$$b_i(x) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \times \exp \left\{ -\frac{1}{2} (x - \mu_i)^T (\Sigma_i)^{-1} (x - \mu_i) \right\} \quad (1)$$

여기서 μ_i 는 상태의 평균 벡터이고, Σ_i 는 상태의 공분산행렬이다. 공분산행렬은 상태에서의 특징벡터의 상관관계와 변이성을 표현한다.

M개의 혼합물(mixture)을 가진 화자 모델에서 가우시안 혼합 모델(Gaussian Mixture Model)은 다음과 같이 표현된다.

$$p(x|\lambda) = \sum_{i=1}^M p_i b_i(x) \quad (2)$$

여기서 λ 는 화자 모델의 파라미터를 나타낸다.

$$\lambda = (p_i, \mu_i, \Sigma_i), \text{ for } i=1, \dots, M \quad (3)$$

따라서, 파라미터 λ 의 화자모델에서 특징벡터 x 를 관측할 확률은 x 가 각각의 상태에서 출력될 확률을 각 확률분포에 가중하여 합한 것이다. 모델의 학습에는 EM (Expectation- Maximization) 알고리즘을 이용한다. 등록화자로부터 발생된 음성에서 추출된 특징벡터가 주어지면 EM 알고리즘은 반복적으로 모델 파라미터를 다듬어서 학습데이터와 모델파라미터가 잘 정합 되도록 한다.

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 뿐이므로 이를 다시 학습데이터와 테스트 데이터로 나누면 그 수가 너무 작아서 신뢰성 있는 결과를 얻기 어려우며, 또한 학습 데이터와 테스트 데이터를 어떻게 나누는가에 따라서 결과가 달라 질 수 있다. 따라서, 본 실험에서는 leave-one-out 방법을 사용하였다. 실험 과정은 다음과 같다. 데이터가 수집된 모든 사람 x 에 대하여 다음의 과정을 수행한다. 이 사람 x 가 속한 체질 클래스 C에서 x 를 제외한 모든 사람의 데이터를 이용하여 클래스 C의 모델을 학습시킨다. 테스트는 x 와, 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 x 의 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 대하여 혼합물(mixture)의 수를 1, 2, 4, 8, 16, 32, 64, 128, 256 과 같이 9가지로 하여 학습해 보고 테스트 하였다.

특징 파라미터는 15차의 멜켵스트럼 계수 (melcepstral coefficient)를 사용하였고, 분석창 넓이는 250 ms, 간격은 100 ms로 하였다.

3. 결과

3.1. 음성학적 변수

3.1.1. 아노바 테스트

⑦h1, h2, h3의 변화형태를 제외한 음성학적 변수는 아노바 테스트를 하여 통계적 의미를 검증하였다. 모든 음성학적 변수 중, F2의 대역폭만이 유일하게 5% 수준에서 집단 간 유의미한 차이를 보였을 뿐, 나머지 변수들은 source, 필터에 관계없이 전혀 유의미한 차이를 보이지 않았다. <표 1>에 ⑦h1, h2, h3의 변화형태를 제외한 모든 변수에 대한 아노바 결과를 나타내었다. ($p < 0.05$ 인 경우만 진한 기울기체로 표시하였다.)

<표 1> 아노바 결과

종속변수		Sum of Squares	df	Mean Square	F	Sig.
①평균 기본주파수	Between Groups	337.319	2	168.659	0.471	0.627
	Within Groups	15,381.920	43	357.719		
	Total	15,719.239	45			
②음역	Between Groups	3,838.245	2	1,919.122	1.949	0.155
	Within Groups	42,350.973	43	984.906		
	Total	46,189.217	45			
③평균기본 주파수분산률	Between Groups	19.085	2	9.543	0.132	0.877
	Within Groups	3,108.771	43	72.297		
	Total	3,127.856	45			
④무성발화 비율	Between Groups	17.156	2	8.578	0.304	0.739
	Within Groups	1,212.530	43	28.198		
	Total	1,229.686	45			
⑤Voice Breaks	Between Groups	16.809	2	8.405	0.283	0.755
	Within Groups	1,276.110	43	29.677		
	Total	1,292.919	45			
⑥Harmonicity	Between Groups	0.825	2	0.413	0.180	0.836
	Within Groups	98.647	43	2.294		
	Total	99.472	45			
⑧F1	Between Groups	127.849	2	63.924	0.006	0.994
	Within Groups	432,847.593	43	10,066.223		
	Total	432,975.442	45			
⑧F2	Between Groups	60,687.419	2	30,343.709	2.240	0.119
	Within Groups	582,445.995	43	13,545.256		
	Total	643,133.413	45			

⑧F3	Between Groups	77,768.816	2	38,884.408	0.862	0.429
	Within Groups	1,938,655.089	43	45,085.002		
	Total	2,016,423.905	45			
⑨B1(bandwidth of F1)	Between Groups	34,808.186	2	17,404.093	1.474	0.240
	Within Groups	507,799.063	43	11,809.281		
	Total	542,607.249	45			
⑨B2	Between Groups	129,705.888	2	64,852.944	3.662	0.034
	Within Groups	761,419.388	43	17,707.428		
	Total	891,125.275	45			
⑨B3	Between Groups	11,073.103	2	5,536.552	0.055	0.947
	Within Groups	4,349,327.970	43	101,147.162		
	Total	4,360,401.073	45			
⑩LTA slope	Between Groups	15.793	2	7.897	0.829	0.443
	Within Groups	409.547	43	9.524		
	Total	425.340	45			
⑪발화속도	Between Groups	2,737.309	2	1,368.654	0.682	0.511
	Within Groups	86,313.705	43	2,007.295		
	Total	89,051.014	45			

이러한 결과는, 집단간 분석에서도 똑같은 결과를 나타냈으며, 두 종류의 사후 분석(Scheffe, Bonferroni)을 실시한 결과도 마찬가지였다.6) <표 2>에 그 결과를 나타내었다. ($p < 0.05$ 인 경우만 진한 기울기체로 표시하였다.)

<표 2> 사후분석결과

종속변수	사후분석	Sig.		
		소음-소양	소음-태음	소양-태음
①평균 기본주파수	Scheffe	.716	.670	.999
	Bonferroni	1.000	1.000	1.000
②음역	Scheffe	.167	.355	.830
	Bonferroni	.180	.458	1.000
③평균기본 주파수분산률	Scheffe	.880	.974	.948
	Bonferroni	1.000	1.000	1.000
④무성발화 비율	Scheffe	.975	.893	.751
	Bonferroni	1.000	1.000	1.000
⑤VoiceBreaks	Scheffe	.991	.794	.848
	Bonferroni	1.000	1.000	1.000
⑥Harmonicity	Scheffe	.844	.902	.986

6) 물론 아노바 결과 통계적으로 유의한 결과가 나오지 않았기 때문에 사후분석 자체가 무의미할 수 있으나, 이것은 혹시나 놓쳤을지도 모를 작은 차이라도 검증해보고자 하는, 일종의 절박한 노력의 하나였다.

	Bonferroni	1.000	1.000	1.000
⑧F1	Scheffe	.994	.996	1.000
	Bonferroni	1.000	1.000	1.000
⑧F2	Scheffe	.620	.124	.532
	Bonferroni	.992	.127	.791
⑧F3	Scheffe	.985	.517	.585
	Bonferroni	1.000	.760	.911
⑨B1(bandwidth of F1)	Scheffe	.532	.928	.254
	Bonferroni	.792	1.000	.299
⑨B2	Scheffe	.034	.254	.473
	Bonferroni	.030	.300	.672
⑨B3	Scheffe	.948	.975	.992
	Bonferroni	1.000	1.000	1.000
⑩LTA slope	Scheffe	.458	.871	.698
	Bonferroni	.642	1.000	1.000
⑪발화속도	Scheffe	.961	.752	.537
	Bonferroni	1.000	1.000	.802

아노바 결과 유일하게 유의미한 차이를 보인 변수는 F2의 대역폭(B2)이었다. 그리고 이 변수도 사후검증 결과, 소음인과 소양인 사이에만 유의미한 차이를 보였을 뿐, 소음인과 태음인, 소양인과 태음인 사이에서는 유의미한 차이를 보이지 않았다.

3.1.2. 배음의 형태

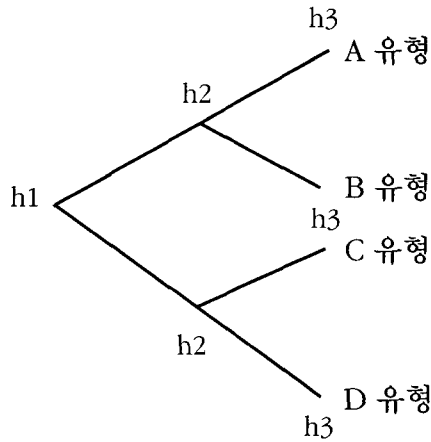
변수 중 ⑦harmonic structure는 정량적인 통계분석을 실시하는 것이 불가능하여 그 양상에 따른 정성적인 분석을 시도하였다.

배음간의 전반적인 양상은 ⑩의 long-term average spectra와 어느 정도 관계가 있을 것이지만, 발생유형을 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 <그림 1>과 같이 네 가지 형태로 나누어서 그 유형별 빈도수를 보면 다음의 <표 3>과 같다.

<그림 1>의 네 가지 형태는 각 유형별로 h1-h2, h2-h3간의 상대적 크기를 도식화한 것으로서, 다른 유형간의 배음의 크기를 비교할 수는 없다. 예를 들면 A유형



<그림 1> 배음구조형태에 따른 분류

의 경우는, h2가 h1보다 크고, h3가 h2보다 큰 형태이며, C유형의 경우는, h2가 h1보다 작지만, h3은 h2보다 큰 것을 의미할 뿐, h1과 h3간의 대소관계는 알 수 없으며, 마찬가지로 A유형의 h3과 C유형의 h3간의 관계 역시 알 수 없다.

<표 3> 배음구조의 형태별 빈도수

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

<표 3>에서 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 예를 들어 “A유형이 나오면 그것은 곧 소음인”이라는 식으로 정량적인 결론을 이끌어 낼 수 있을지에 대해서는 더 많은 자료를 바탕으로 한 분석이 필요할 것이다.

3.2. GMM

<표 4>는 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

<표 4> 평균 순위 백분율

mixtures	1	2	4	8	16	32	64	128	256
태음	45.55	39.87	42.20	45.08	45.94	44.79	44.27	43.21	42.31
소음	44.00	43.67	46.95	51.02	49.33	48.05	50.33	48.57	48.38
소양	45.08	47.96	49.88	46.67	47.06	47.83	49.06	50.27	51.27

포함되어 있으므로 이상적인 경우에는 항상 1위가 나와야 하지만, 실험 결과 혼합 수 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 할 수 있다.

4. 논의

본 연구에서 살펴본 11가지의 변수는 대부분은 사상체질에 의한 세 체질간의 차이를 통계적으로 유의미하게 구별할 수 없었다. 물론 본 연구에서 측정하고 분석한 변수들 중 일부는 워낙 정상에서 벗어나는 범위의 병리학적인 진단을 위한 것이라는 것을 염두에 둘 필요는 있을 것이다. 즉, 사상체질학상으로 어떤 체질이든, 병리학적으로 정상의 범위를 벗어나지 않는다면, 반드시 체질 간 차이는 보이지 않을 수도 있을 것이다.

분산분석에서 유의미한 차이를 보인 F2 대역폭의 경우도, 그 결과로부터 곧바로 F2 대역폭이 사상체질을 나눌 수 있는 변수라고 결론을 내리기에는 매우 조심스러운 면이 있다. 우선 이 변수조차도 소음인과 소양인사이에서만 유의한 차이가 있었을 뿐이었다. 그리고, 원론적으로 볼 때에도, 한 사람의 대역폭이라면 F1, F2, F3의 대역폭간에 어느 정도 관계가 있는 것이 자연스러울 것인데, 다른 포만트는 전혀 대역폭에서 차이를 보이지 않는 반면 유독 F2의 대역폭만 보인다면, 이것을 과연 필연적인 결과로 보아야 할 것인가를 신중하게 검토해야 할 것이다. 즉, F2의 대역폭만이 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 (혹은, 본 연구의 경우처럼 세 가지) 체질은 음성학적으로는 구별할 수 없는 것이라는 해석이고, 다른 하나는 구별은 가능하되, 본 연구를 포함하여 지금까지 살펴본 변수로는 부족하므로 무언가 다른 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 다뤄온 중요 변수였다는[16] 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] 사상체질의학회A. <http://www.esasang.com/dataroom.htm>, 2003.
- [2] 사상체질의학회B. <http://www.esasang.com/dataroom-2.htm>, 2003.
- [3] 김달래, 동의수세보원초고, 서울, 정담, p.134, 1999.
- [4] 유준상. 청각적 성음분석을 통한 사상체질진단에 관한 연구, *상지대학교 한의학과 사상체질의학 박사학위 논문*, 2002.
- [5] A. Iida, and M. Yasumura, "Designing and testing a corpus of emotional speech," *Proceedings of the First International Workshop on East-Asian Language Resources and Evaluation*, 32-37, 1998.
- [6] C-W. Jo, "Aspects on the collection and applications of multimedia emotional speech database," *Proceedings of the First International Workshop on East-Asian Language Resources and Evaluation*, 27-31, 1998.
- [7] L. Leinonen, and T. Hiltunen, "Expression of emotional motivational connotations with a one-word utterance," *Journal of the Acoustical Society of America*, Vol. 102 (3), 1853-1863, 1977.
- [8] A. Protopapas, and P. Lieberman, "Fundamental frequency of phonation and perceived emotional stress," *Journal of the Acoustical Society of America*, Vol. 101 (4), 2267-2277, 1977.
- [9] 문승재, "음성과 인상: 인지실험," *한국음향학회*, 18권 8호, 66-74, 1999.
- [10] S.-J. Moon, "What you hear is what you see?" *Journal of the Acoustical Society of Korea* 21.1E, 31-41, 2002.
- [11] 김달래, 박성식, 권기록, "성문분석법에 의한 사상체질 진단의 객관적 연구(1)" *사상의학회지*, 10(1), 65-80, 1998.

- [12] 김선형, 신미란, 김달래, 권기록, "Laryngograph와 EEG를 이용한 음향 특성과 사상체질간의 상관성 연구," *사상의학회지*, 12(1), 144-156, 2000.
- [13] 박성진, "Harmonics(배음)와 Formant Bandwidth(포먼트 폭)를 이용한 음향특성과 四象體質間의 相關性 研究," 상지대학교 대학원, 박사학위논문, 2002.
- [14] 양상목, 김선형, 유준상, 김형석, 이영훈, 김달래, "Pitch Range 와 Bandwidth를 이용한 음향특성과 사상체질간의 상관성 연구." *사상의학회지*. 13(3), 31-39, 2001.
- [15] 문승재, 박종주, 황혜정, "음성과 사상체질: 음원을 중심으로" *말소리*, 48.1 19-33 2003.
- [16] R. Remez, J. Fellowes and P. Rubin, "Talker identification based on phonetic information," *Journal of Experimental Psychology: Human Perception and Performance* Vol.23 No.3, 651-666, 1997.
- [17] J. Hillenbrand, L. Getty, M. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," *Journal of the Acoustical Society of America*, 97(5). 3099-3111. 1995.
- [18] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America*, 24, 175-184. 1952.
- [19] P. Ladefoged, *A Course in Phonetics*, Orlando; Harcourt College Publishers, 2002.

접수일자: 2005년 7월 7일

게재결정: 2005년 9월 5일

▶ 문승재(Moon, Seung-Jae)

주소: 443-749 경기도 수원시 영통구 원천동 산5번지

소속: 아주대학교 인문대학 영어영문학과

전화: (031)219-2827

E-mail: moon@ajou.ac.kr

▶ 탁지현(Tak, Ji-Hyun)

주소: 443-749 경기도 수원시 영통구 원천동 산5번지

소속: 아주대학교 대학원 영어영문학과

전화: (031)219-2196

E-mail: yongsamo@hanmail.net

▶ 황혜정(Hwang, Hyejeong)

주소: 443-749 경기도 수원시 영통구 원천동 산5번지

소속: 아주대학교 대학원 영어영문학과

전화: (031)219-2196

E-mail: rudder79@hanmail.net