음성과 인상의 관계규명을 위한 실험적 연구*

문 승재(아주대)

<차 례>

<Abstract>

Voice and Image: A Pilot Study

Seung-Jae Moon

When we hear someone's voice, even without having met the person before, we usually make up a certain mental image of the person. This study aims at investigating the relationship between the voice and the image information carried within the voice. Does the mental picture created by the voice closely reflect the real image and if not, is it related with the real image at all?

To answer the first question, a perception experiment was carried out. Speech samples reading a short sentence from 8 males and 8 females were recorded and pictures of subjects were also taken. Ajou University students were asked to participate in the experiment to match the voice with the corresponding picture.

Participants in the experiment correctly match 1 female voice and 4 male voices with their corresponding pictures. However, it is interesting to note that even in cases of mismatch, the results show that there is a very strong tendency. In other words, even though participants falsely match a certain voice with a certain picture, majority of them chose the same picture for the voice. It is the case for all mismatches. It seems that voice does give the listener a certain impression about physical characteristics even if it might not be always correct.

By showing that there is a clear relationship between voice and image, this study provides a starting point for further research on voice characteristics: what characteristics of the voice carry the relevant information? This kind of study will contribute toward the understanding of the affective domain of human voice and toward the speech technology.

* 이 논문은 1998년도 대우장학재단 연구비 지원에 의해 연구되었음.

1. 문제의 제기: 목소리는 인상과 관계가 있을까?

사람들은 저마다 특색이 있는 목소리를 가지고 있다. 그래서 한 번 만나서 이야기를 나눠 본 사람은 나중에 목소리만 들어도 누구인지를 금방 알 수 있고, 전화로 이야기를 하는 중에도 거의 자동적으로 머리 속에 그 사람의 모습을 그리게 된다.

그러나 얼굴을 모르는 사람에 대해서는 어떨까? 대부분의 사람들은, 전혀 모르는 사람과 전화로 대화를 하거나 혹은 만나 본 적이 없는 사람의 목소리를 뒤로부터 들을 때, 그 사람의 외모에 대해서 무의식적으로 어떤 막연한 모습을 떠올린 경우가 있었을 것이다. 그래서 때로는 그 목소리의 주인공을 직접 만나게 된 경우, 예상하던 모습과 실제의 모습이 너무나 달라서 의아해 한 경우가 있었을 것이다.

여기에서 몇 가지 의문점이 떠오른다. 우선 첫째는, 목소리를 듣고 머리 속에 그리던 모습과 실제의 모습과 과연 관계가 있긴 있는 것인가? 달리 말하면 과연 그 두 모습이 비슷한가 하는 것이다. 직접 물어보았을 때 많은 사람들은 목소리만을 듣고 예상했던 모습과 실제의 모습은 전혀 다른 경우가 더 많았던 것 같다고들 기억했다. 그러나 이것은 혹시 실제로 비슷할 때보다 다를 때의 경우가 더 기억에 많이 남아있을 수 있기 때문은 아닐까? 일상 생활에서 보면 평범한 경우보다는 특별한 경우가 더 기억에 남는 것과 같은 이치는 아닐까?

만일 어떤 우연 이상의 관계가 있다면, 두 번째 의문은 어떻게 우리는 목소리로부터 그 주인공의 모습을 추측해 내는 것일까? 하는 것이다.

평상시에 우리가 소리와 체격적인 조건을 연계시킨다는 가능성은 이미 시사된 바 있다. 현대 음성학에서 정설로 받아들여지고 있는 source-filter theory(Fant, 1970)에 의하면, 성대에서 만들어진 소리는 성도를 지나며 변형되어 특정한 소리를 내게 된다. 따라서 최종적인 소리는 소리를 내는 사람의 성대 및 성도의 특성에 따라서 달라지게 된다. 그렇기 때문에 체격이 큰 사람과 작은 사람의 소리가 다르며 남성과 여성의 소리가 다르게 된다(Hillenbrand, Getty, Clark & Wheeler, 1995; Peterson & Barney, 1952). 그런데 여기서 한가지 재미있는 현상은, 이러한 다양함 속에서도 우리는 일관된 인지를 한다는 것이다. 즉, 남성과 여성이 같은 [이] 소리를 냈을 때, 두 사람의 모음의 formant 값은 물리적으로 전혀 다른데도 불구하고, 우리는 이를 아무 문제없이 모두 '같은' 모음으로 인지한다는 것이다. 이것이 어떻게 가능할까? 이러한 normalization 현상에서 가장 중요한 것 요소 중의 하나가 바로 화자의 성도의 크기라는 것은 일찍부터 확인된 바 있다(Fant, 1967). 즉, 듣는 이는 무의식적으로 화자의

체격을 고려하여 자기가 들은 formant의 절대값을 수정한다고 가정할 수도 있을 것이다.

그렇다면 이처럼 체격적인 정보를 바탕으로 소리를 듣기 때문에 역으로 그 목소리로부터 체격적인 정보를 추출해 내는 것인가? 즉, 단순히 개별적인 소리 뿐 아니라 개개인의 음성도 혹시 체격과 연관지어서 일종의 normalization을 거침으로써 반대로 소리를 듣고 체격의 조건을 추측해 내는 것은 아닐까 하는 추측을 가능케 한다.

그러나 이것은 위에서 제기한 첫 번째 질문에 대한 긍정적인 대답이 나올 때에야 비로소 제기할 수 있는 의문이다. 즉, 소리를 듣고 머리 속에 떠올리는 그 인상이 과연 실제 목소리의 주인공과 일치한다고 할 때에야 비로소 우리는 '어떻게 그럴 수 있을까'하는 질문에 대한 탐구를 할 수 있을 것이다.

본 연구의 목적은 바로 이 첫 번째 질문, 즉'소리를 듣고 떠오르는 인상과 실제의 모습이 과연 일치하는가?'하는 질문에 계량적인 답을 찾아보는 것이다. 단순히, '일치하지 않는 경우가 많았다,' 혹은 '일치하는 경우가 많았다' 하는 식의 추측이 아닌, 실험을 통한 통계적인 추세를 먼저 확인해보고자 하는 것이다.

지금까지 음성학의 많은 연구들은 성대 위에 있는 성도에 대한 연구(자모음의 음가 등)가 주를 이루어왔다. 이에 대한 문헌은 너무 많아서 이루 열거할 수가 없을 정도이다. 그리고 최근 들어 음성 합성이나 인식 등 음성과학분야에서는, 단순히 정확한 음소의 전달에만 신경을 쓰는데서 한 발 더 나아가서 자연스러운 합성 소리를 위하여 이러한 정서적인 면을 연구하는 추세가 매우 활발한 실정이다(Iida & Yasumura, 1998; Jo, 1998; Leinonen & Hiltunen, 1997; Protopapas & Lieberman 1997).

그러나 음성의 특성(timber)을 결정짓는 성대 자체에 대한 연구는 매우 드물고 (예: Sundberg, 1987) 특히 우리말에 관해서는 더욱 그러한 실정이다. 그러나 이러한 음성에 대한 가장 근본적인 차원의 연구는 그 중요성이 날로 인정받아가고 있으며, 특히 자연스러운 음성합성에 반드시 필요한 요소라고 하겠다.

본 연구는 목소리가 전달하는 언어적인 정보 (어떤 모음, 자음인가 하는) 이외에도 목소리의 주인공에 대한 다른 정보를 어떻게, 어느 정도로 전달하는가 하는 것을 밝히기 위한 처음 단계의 시도라고 할 수 있겠다. 그러한 연구는 음성합성에 있어서 상황에 적절하고 듣기 좋으며 자연스러운 소리를 만들고, 화자 인식 등 음성인식 분야에도 기여를 할 수 있을 것으로 기대된다.

2. 인지실험

사람의 목소리가 주는 인상과 그 목소리 주인공의 실제 모습이 얼마나 일치하는지를 알아보기 위하여 인지실험을 실시하였다. 이 인지실험은 여러 명의 화자를 무작위 선정하여 그 사람들의 말을 녹음하고 사진을 찍은 후, 화자를 모르는 사람들로 하여금 녹음을 듣고 사진과 짝을 짓게 하는 식으로 진행하였다.

2.1 실험자료

일반적으로 어떤 사람과 오래 대화를 나누다 보면, (전화로도) 그 목소리 이외에도 말투나 억양 등에서 그 사람에 대한 여러 가지 정보를 얻을 수 있게 마련이다. 본 연구에서는, 그러한 소리 외적인 정보를 배제한, 오로지 목소리의 색(timber)만을 듣고 추측한 모습의 신빙성을 알아보고자 하는 것이 목적이므로, 긴 문장을 쓰지 않고 다음과 같이 아주 짧은 문장을 선택하였다.

> 안녕하세요, 제가 지금은 전화를 받을 수 없으니 메시지를 남겨주세요. 감사합니다.

이 짧은 문장은 여성화자들이 읽도록 하였다. 전화 응답기에서 흔히 들을 수 있는 문장을 선택한 것은 가능한 한 자연스러운 발화로 들릴 수 있도록 하기 위한 고려였다. 대부분의 사람들이 마이크를 들이대면 자연스러운 말투를 잘 구사하지 못하는 점을 생각할 때, 원래 녹음되는 줄 알고 하는 전화 응답기 인사말이라면 듣는 입장에서도 그러려니 하는 마음이 있을 것이라고 생각한 것이다.

시기적으로 후에 실시한 남자화자의 경우는 문장 길이에 따라 인지도에 변화가 있는지를 확인하기 위하여 전화 응답기 문장보다는 조금 긴 (평균 19초 정도), 다음과 같은 동화의 한 부분을 읽도록 하였다.

> 옛날에는 '과거 시험'이란 것이 있었어요. 글을 읽는 선비들은 저마다 '과거 시험'을 보기 위해 지금의 서울인 '한양'으로 몰려 왔답니다. 한양이 아닌 다른 지방의 선비들은 시험 날에 맞춰 먼길을 떠나야 했어요. 어느 시골에 글 솜씨 뿐 만이 아니라 활 쏘는 솜씨 또한 뛰어난 젊은 선비가 있었습니다.

전화 응답기의 문장보다는 길지만, 여전히 이 동화를 통하여 목소리 이외의 화자의 어떠한 다른 정보를 얻기는 쉽지 않을 것으로 판단되었다. 이 경우 역시, 아이들에게 이야기해주듯이 자연스럽게 읽어달라고 부탁했지만, 아주 자연스러운 대화체라기보다는 낭독체에 가까운 분위기였다.

2.2 화자 선정 및 녹음

인지실험 자료 녹음에 참여한 화자는 아주대학교 교육대학원생 남·여 각각 8명이었다..나이가 목소리에 큰 영향을 미칠 것이라고 생각할 수 있었기 때문에, 가능한한 비슷한 나이의 사람을 고르도록 노력하였다. 이들의 나이는 <표1>과 같다.

<표 1> 화자의 나이

남자		여자	
화자	나이	화자	나이
1	34	1	40
2	32	2	43
3	36	3	40
4	38	4	47
5	31	5	39
6	35	6	28
7	39	7	33
8	32	8	36

사전에 미리 녹음을 하겠다는 말을 하지 않은 상태에서 어느 날 갑자기 수업 시간중에 칠판에 문장을 적은 후 녹음을 실시하였으며 실험의 목적에 대한 설명도 없었다. 녹음은 TASCAM PA-1 DAT 기종과 Electro-Voice 635 N/D dynamic omni-directional 마이크를 사용하여 이루어졌으며, 녹음 당시 강의실은 (야간 수업이므로) 양질의 녹음이 가능할 정도로 충분히 조용한 상태였다.

모든 사람의 녹음이 끝난 후에는 같은 강의실에서 사진을 찍었다. 사진을 찍을 때에는 출입문을 배경으로 하여 같은 위치에 선 상태에서 사진기는 삼발이를 이용하여 고정하고 줌 역시 같은 값에 고정한 채 전신을 촬영하였다. 이는 나중에 서로의 체격적인 특징이 비교될 수 있도록 하기 위함이었다. 촬영에는 35-105mm zoom 렌즈를 장착한 Minolta Maxxum 7000 사진기와 ROKUNAR AF 920 TZ flash를 45° 각도

로 이용하였으며, shutter speed나 노출은 자동으로 하고 Kodak ASA 400 필름을 이용하였다. 후에 이 필름은 현상소에 맡겨서 3.5″ × 5″ 크기로 인화하였다.

2.3 자료 처리

녹음된 자료는 Kay Elemetrics의 CSL 5000B를 이용하여 8kHz low-pass filter를 통하여 20kHz sampling rate로 digitize하여 개인별로 소리 파일을 저장하였다. 사진은 보통 크기로 현상하여 8명을 한꺼번에 볼 수 있도록 clear file 2면에 걸쳐 배치하였다.

2.4 인지실험 참여자

실제 인지실험에는 연인원 95명의 아주대학교 학부 학생들이 참여하였다. 실험상의 문제로 우선 여성 화자들의 자료를 45명의 학생이 참여하여 실시하였고, 남성 화자들의 자료는 그 이후에 별도로 50명의 학생이 참여하여 실시하였다. 여성 화자에 대한 45명의 참여자 중 26명이 여성이었으며 19명이 남성이었고, 남성 화자에 대한 50명의 참여자는 35명이 여성, 15명이 남성이었다. 이러한 성적인 불균형은, 이번 인지실험에 참여한 학생들의 대다수가 인문학부 학생이었기 때문인 것으로 풀이된다. 또한 실험을 여성 화자와 남성 화자에 대하여 시기적으로 달리 실시하였기 때문에 각각의 참여자 45명과 50명중에는 약간의 중복이 있을 수 있었다. 그러나 그러한 중복은 본 실험의 의도에 비추어 볼 때 아무런 문제도 되지 않았을 것이라 생각된다.

2.5 인지실험 실시 방법

본 인지실험은 기본적으로 참여자들에게 남 · 여 각 8명 화자의 소리를 다 들려 준 후, 8명의 사진과 짝을 짓도록 하는 것이다. 여기에서 가장 큰 문제는 사진을 보여주는 방법이었다. 소리는 여러 개의 헤드폰이나 스피커를 통해서 양질의 소리를 여러 명의 참여자들에게 동시에 들려주는 것이 가능한 반면, 사진은 그처럼 여러 명에게 함께 제시한다는 것이 여러 가지로 무리가 많았다. 사진을 확대한다 해도 여러 명이 한꺼번에 볼 수 있도록 확대하는 데에는 한계가 있었고, 색상을 지닌 채 복사하여 overhead projector로 투사한다고 해도 8명의 사진을 한꺼번에 투사하는데 어려움이 많았던 것이다. 그래서 결국은 참여자 한 명씩 개별적으로 실험을 실시하기로 하였다.

우선 개별적으로 저장된 화자의 소리 파일을 CSL 5000B를 이용하여 불러서 14인치 컴퓨터 모니터에 A부터 H까지의 8개의 창으로 무작위로 배치하였다. 각 창에

나타난 개별 파일은 각각 M1…M8, F1…F8로 이름을 붙였다. (CSL 5000B가 최고 8개까지의 창을 열 수 있게 되어있기 때문에 화자의 수를 각 성별 8명으로 제한했던 것이다.) 이 모니터 옆에 8장의 사진을 clear file에 1-8까지 배열하였다. 참여자에게 간단한 CSL 조작법 (playback 기능과 창 사이를 움직이는 법)에 대하여 설명을 하면, 참여자는 모니터 앞에 앉아서 헤드폰 (Beyerdynamic DT 211)을 쓰고 사진을 보면서 각각의 목소리 A-H가 사진 1-8 중 누구의 목소리인지를 답하였다. 목소리를 듣는 횟수는 제한하지 않았고 전체 실험에 소요되는 시간도 제한하지 않았기 때문에 참여자는 자기가 원하는 만큼 소리를 들어볼 수 있었다.

이처럼 목소리와 사진을 짝 짓는 외에도, 목소리 중 가장 좋은 목소리를 하나 고르고, 사진 중 가장 좋은 인상을 하나씩 고르도록 하였다. 이 때에 "가장 좋은" 목소리나 인상에 대하여는 전혀 부가적인 설명을 하지 않고 다만 "가장 마음에 드는 좋은" 목소리와 인상을 고르도록 하였다. 이것은 나중에 혹시 가장 마음에 드는 목소리와 가장 마음에 드는 모습을 짝 짓는 경향이 있는지를 보기 위함이었다.

이러한 것을 다음 <표 2>와 같은 응답지에 답하도록 하였다.

<표 2> 인지실험의 응답지

Subject Name					Date			
소리	A	B	C	D	E	F	G	H
사진								
가장 좋은 소리				가장 좋은 인상				

3. 결과 및 논의

여성화자와 남성화자에 대한 인지실험 결과는 각각 다음의 <표 3>, <표 4>와 같다.

표의 세로 행은 화자이며 가로 열은 각 화자의 목소리를 누구의 사진과 짝지었는지를 나타낸 것이며, 실제 참여자 수 외에도 백분율을 괄호 안에 표시했다. 각 화자에 대해 가장 많은 응답수가 있는 부분을 검은 부분으로 진하게 표시했고, 옳은 응답에 해당되는 부분은 밑줄을 쳐서 표시해 놓았다. 예를 들면, <표 4>에서 M1의 목소리를 실제 M1의 사진과 짝지은 사람은 10명으로서 전체의 20%였으며, 13명 (26%)의 참여자가 M1의 목소리를 M7의 사진과 짝지었음을 알 수 있다.

<표 3> 여성화자 인지실험 결과

화자 \ 인지	F1	F2	F3	F4	F5	F6	F7	F8	계
F1	6(13)	10(22)	2(4)	**13(29)**	7(16)	2(4)	2(4)	3(7)	45(100)
F2	**11(24)**	4(9)	4(9)	7(16)	6(13)	7(16)	3(7)	3(7)	45(100)
F3	1(2)	5(11)	8(18)	**13(29)**	8(18)	4(9)	2(4)	4(9)	45(100)
F4	**8(18)**	6(13)	8(18)	8(18)	5(11)	2(4)	1(2)	7(16)	45(100)
F5	4(9)	8(18)	**11(24)**	1(2)	10(22)	3(7)	1(2)	7(16)	45(100)
F6	8(18)	6(13)	2(4)	2(4)	3(7)	6(13)	**15(33)**	3(7)	45(100)
F7	5(11)	3(7)	5(11)	1(2)	4(9)	**15(33)**	8(18)	4(9)	45(100)
F8	2(4)	3(7)	5(11)	0(0)	2(4)	6(13)	13(29)	**14(31)**	45(100)
계	45(100)	45(100)	45(100)	45(100)	45(100)	45(100)	45(100)	45(100)	

여성화자의 경우, 확실하게 목소리와 사진을 올바로 짝지은 경우는 F8 한 사람의 경우뿐이었으나 남성의 경우는 8명 중 4명 (M3, M5, M6, M8)의 목소리가 사진과 올바로 짝지어진 모습을 보이고 있다. 이처럼 옳은 짝짓기의 큰 차이에 대하여는 두 가지 가능한 설명이 있을 수 있겠다. 우선 한 가지는 화자의 성별에 따른 차이가 원인일 수 있을 것이다. 위의 2.4에서 밝힌 바와 같이 여성화자에 대한 인지실험을 먼저 실시하고 남성화자에 대한 실험은 나중에 실시하였는데, 여성화자에 대한 실험 이후, 몇몇 참여자가 "남자 목소리라면 훨씬 더 자신 있게 맞출 수 있었을 텐데"라는 아쉬움을 표현했었다. 이에 대하여 연구자는 별 의미 없는 아쉬움일 것이라고 생각했었는데 실제로 결과가 이처럼 남자에 있어서 더 정확하게 나온 것이 그러한 언급과 관계가 있을지 모르겠다는 생각이 들었다. 또 상당수의 참여자들이 여성화자들의 목소리가 서로 비슷비슷하여 매우 구별하기가 어려웠다고 대답했던 것과도 관계가 있을 법하다.

다른 한 가지 가능성은 여성화자의 경우 말이 짧았던 반면, 남성화자의 경우는 말이 훨씬 길었다는 것이다. 이 두 가지 가능성 중에 어느 것이 옳은 답인지 본 연구의 결과만을 가지고 논의하기에는 불가능하다고 판단된다. 만일 여성화자가 같은 동화를 읽었을 경우의 인지도에 대한 자료가 있다면 결정적인 근거가 될 수 있을 것이지만, 이것은 후속 연구에서 밝혀야 할 과제가 될 것이다.

제대로 짝이 지어진 화자의 경우, 어떤 특별한 점을 곧바로 찾기는 매우 어려웠다. 이를테면 여성화자의 경우 F8이 특별히 나이가 많거나 아주 젊은 사람도 아니었으며, 남성화자들의 경우는 서로 나이가 매우 유사해서 더더욱 나이에서 그 원인을 찾을 수 없었다.

<표 4> 남성화자 인지실험 결과

화자 \ 인지	M1	M2	M3	M4	M5	M6	M7	M8	계
M1	10(20)	4(8)	3(6)	9(18)	2(4)	4(8)	**13(26)**	5(10)	50(100)
M2	5(10)	7(14)	11(22)	6(12)	0(0)	**20(40)**	0(0)	1(2)	50(100)
M3	6(12)	5(10)	**16(32)**	7(14)	1(2)	7(14)	3(6)	5(10)	50(100)
M4	7(14)	5(10)	8(16)	8(16)	4(8)	0(0)	**13(26)**	5(10)	50(100)
M5	2(4)	2(4)	0(0)	6(12)	**31(62)**	2(4)	3(6)	4(8)	50(100)
M6	8(16)	5(10)	6(12)	7(14)	3(6)	**14(28)**	3(6)	4(8)	50(100)
M7	10(20)	**14(28)**	1(2)	5(10)	5(10)	2(4)	8(16)	5(10)	50(100)
M8	2(4)	8(16)	5(10)	2(4)	4(8)	1(2)	7(14)	**21(42)**	50(100)
계	50(100)	50(100)	50(100)	50(100)	50(100)	50(100)	50(100)	50(100)	

그러나 올바른 짝이 지어지지 않은 화자의 목소리 경우도 개별적으로 들여다보면 특이할 만한 점이 있음을 알 수 있다. 각 화자의 목소리에 대하여, 비록 단순한 짝 지움은 틀렸다 할지라도, 다수의 참여자는 특별히 어떤 특정한 모습과 짝을 지었다는 것이다. 여성화자의 경우, 적게는 24%에서부터 (F1-F4, F2-F1, F3-F4, F5-F3) 많게는 30%가 넘는 (F6-F7, F7-F6) 참여자가 특정한 목소리에 대하여, 비록 틀린 짝이지만, 공통된 모습을 목소리의 주인공으로 추측했으며, 남성의 경우도 약 25% 이상의 참여자가 M1, M2, M4, M7의 목소리를 듣고 그 주인공으로 M7, M6, M7, M2를 각각 꼽은 것이다.

이처럼 어떤 목소리의 주인공을 무작위로 꼽지 않고 일관되게 짝짓는 현상은, 역으로 여성과 남성 모두 2명의 사진은 전혀 다수로 선택되지 않은 것에서도 알 수 있다. 즉 여성 중 F2와 F5, 남성 중 M1과 M4의 사진은 다수에 의하여 어떤 목소리의 주인공으로도 꼽히지 않았음을 볼 수 있다.

위의 결과들에서 내릴 수 있는 잠정적인 결론은 참여자들이 단순히 무작위로 목소리와 사진을 짝지은 것이 아니라는 것이다. 참여자들이 목소리를 듣고 사진을 찾을 때에 무엇인지는 모르지만 무언가 일관된 점이 있었다는 것을 부인하기 어려운 결과라고 하겠다.

그렇다면 인지실험 참여자들은 (비록 틀렸다 할지라도) 과연 목소리의 어떤 특성으로부터 그 목소리의 주인공의 모습을 추측한 것일까? 하는 의문이 당연히 제기된다. 여기에 대한 구체적인 대답은 본 논문의 범위를 벗어나지만, 한 가지 확인할 수 있는 것이 있다. 혹시 이들은 무조건 "좋은" 소리와 "좋은" 모습을 연관시키는 것은 아닐

까? 앞서 밝혔듯이 인지실험의 응답 중 하나는 "좋은" 목소리를 꼽는 것이고 다른 하나는 "좋은" 인상을 꼽는 것이었다. 만일 어떤 사람이 "좋은" 목소리를 A라고 했고 "좋은" 인상을 3이라고 했는데, 응답지 윗 부분에서 A 목소리의 주인공을 3이라고 했다면, 이런 경우, 이 응답자는 "좋은" 목소리와 인상을 연관지었다고 볼 수 있을 것이다. 그러나 위 응답자가 만일 A 목소리의 주인공으로 3이 아닌 다른 사람을 썼다면, 이런 경우는 "좋은" 목소리와 "좋은" 인상을 구태여 연관시키지 않았다고 볼 수 있을 것이다. 이러한 식으로 각 응답자가 "좋은" 목소리라고 꼽은 목소리의 주인공과 "좋은" 인상으로 꼽은 사람이 일치하는 경우와 일치하지 않는 경우를 조사해 보니 그 결과는 다음의 표와 같았다.

<표 5>에서 볼 수 있듯이 단순히 "좋은" 목소리를 "좋은" 인상과 결부시키는 것은 아님을 알 수 있다. 이 결과는 본 연구를 시작하기 전 여러 사람들에게 물어 보았던 결과와 아주 상반되는 것이어서 매우 흥미롭다. 많은 사람들에게 목소리를 듣고 생각했던 인상과 실제의 인상이 같은 경우와 다른 경우 중 어느 것이 많은 것 같았는지를 물었을 때, 대부분의 사람들이 다른 경우가 더 많았다고 응답했다. 그리고 이어 좋은 목소리와 좋은 인상에 대해서 물었을 때 대부분의 사람들이 좋은 목소리를 좋은 인상과 결부시켜 생각하는 경향이 있는 것 같다고 대답했었다. 그러나 이번 인지실험 결과는 전혀 그렇지 않다는 것을 보여주고 있는 듯 하다. 이것은 앞서 '1. 문제의 제기'에서 언급한 것처럼, 일반적으로 우리가 알고 있다고 생각하는 것이 사실은 그렇지 않을 수 있다는 것을 보여주는 또 한 가지의 예라고 할 수 있을 것이다.

<표 5> "좋은" 목소리와 "좋은" 인상과의 일치여부

남자 화자		여자 화자	
일치	불일치	일치	불일치
13(26%)	37(74%)	16(36%)	29(64%)

4. 결론 및 제언

본 인지실험의 결과를 요약하면, 목소리를 듣고 머리에 떠오르는 인상은 그것이 설령 실제 목소리의 주인공과는 다를 수 있을지라도, 무엇인가 일관된 면이 있음을 알 수 있었다. 그리고 또 한 가지 확인할 수 있었던 것은 무조건 "좋은" 목소리를 "좋은" 인상과 연관시키지는 않는다는 것이었다.

이처럼 일단 목소리와 그로 인하여 만들어지는 인상에 관계가 있음이 밝혀진 이상, 다음 단계는 '1. 문제의 제기'에서 제시했던 두 번째 질문에 대한 답을 찾기 위하여 노력하는 것일 것이다. 즉, '음성의 어떤 부분이 그러한 시각적인 정보를 주는 것인가' 하는 것이다. 단순히 "좋다"는 주관적인 판단조차도 목소리와 인상간의 관계를 이해하는 실마리를 제공해 주지 않는다면 과연 목소리의 어떤 속성이 이처럼 많은 사람들에게 비슷한 선택을 하도록 하는 것일까?

그러한 답을 본격적으로 찾기 위해서는 본 연구를 바탕으로 몇 가지 고려할 사항이 있다. 본 연구에서는 응답자의 성별 등을 전혀 고려하지 않았다. 앞서 밝힌 바와 같이 응답자 중 여자의 수가 남자보다 훨씬 많았던 것이 이번 실험에 어떤 영향을 미쳤는지는 알 수가 없는 상황이다. 이러한 점을 밝히기 위해서는 남자와 여자가 같은 수로 분포하는 더 큰 집단의 참여가 필수적일 것이다.

그리고 남·여 화자 모두 같은 문장을 읽도록 해야 할 것이다. 이는 '3. 결과 및 논의'에서 언급한 바와 같이, 화자의 성별과 문장의 길이를 두 개의 별개의 변수로 놓을 수 있도록 하고자 함이다.

또한 만일 체격적인 정보, 즉 키나 몸무게 등에 대한 정보를 음성으로부터 추출하는 것이라면, 이를 확인하기 위하여 두 가지의 다른 사진을 이용하면 좋을 것이다. 즉, 한 집단은 체격적인 정보가 풍부한 전신 사진을 보고 고르도록 하고, 다른 한 집단은 그러한 체격적인 정보가 최소화될 수 있도록 얼굴만을 찍은 사진을 보고 고르도록 하여, 그 두 집단간의 결과를 비교한다면, 과연 그러한 체격적인 정보가 중요한 역할을 하는지에 대한 결론을 얻을 수 있을 것이다. 물론 이럴 경우, 두 집단간의 참여자는 반드시 다른 사람이어야 할 것이다.

이처럼, 응답자 성별에 따른 인지도, 화자 성별에 따른 인지도, 사진의 체격적 정보 정도에 따른 인지도 등을 비교한다면, 음성의 어떤 특성이 그 주인공에 대한 시각적인 모습의 특성을 전달하는지를 찾는 데 한 발 더 가까이 나아갈 수 있을 것이다. 그리고 그러한 연구는 좀 더 자연스럽고 듣기 좋은 소리를 합성하는데 기여할 수 있을 것이며 궁극적으로 인간의 음성의 본질을 이해하는데 큰 도움을 줄 수 있을 것이다.

<참고문헌>

Fant, G. (1967). Descriptive Analysis of the Acoustic Aspects of Speech, in Lehiste, I. (ed) *Reading in Acoustic Phonetics,* 93-107, Cambridge: The MIT Press.

Fant, Gunar (1970). *Acoustic Theory of Speech Production,* (2nd edition), The Hague: Mouton.

Hillenbrand, J., Getty, L., Clark, M. & Wheeler, K. (1995). Acoustic characteristics of American English Vowels, *Journal of the Acoustical Society of America, 97 (5).* 3099-3111.

Iida, A. & Yasumura, M. (1998). Designing and testing a corpus of emotional speech, *Proceedings of the First International Workshop on East-Asian Language Resources and Evaluation,* 32-37.

Jo, Cheol-Woo (1998). 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.

Leinonen, L. & Hiltunen, T. (1997). Expression of emotional motivational connotations with a one-word utterancce, *Journal of the Acoustical Society of America, Vol. 102 (3),* 1853-1863.

Peterson, G. E. & Barney, H.L. (1952). Control Methods Used in a Study of the Vowels, *Journal of the Acoustical Society of America, Vol. 24,* 175-184.

Protopapas, A. & Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress, *Journal of the Acoustical Society of America, Vol. 101 (4),* 2267-2277.

Sundberg, Johan (1987). *The Science of the Singing Voice,* Dekalb: Northern Illinois University Press.

접수일자: 1998년 12월 1일
게재결정: 1998년 12월 23일

▶ 문승재(Seung-Jae Moon)
주소: 경기도 수원시 팔달구 원천동 산 5
소속: 아주대학교 인문학부 영어영문학전공
전화: 0331) 219-2816
e-mail: sjmoon@madang.ajou.ac.kr