

음성과 인상 : 음성학적 인지실험

Voice and Image : A Perception Experiment

문 승 재*
(Moon Seung Jae*)

* 본 연구는 1998년도 학술진흥재단 공모과제 연구비의 지원을 받는 "음성이 전달하는 시각적 정보에 관한 연구"의 일환으로 수행되었음.

요 약

본 연구에서는 남녀 화자 8명의 목소리를 녹음한 후 전신과 얼굴을 촬영하고 100명의 사람들에게 목소리와 사진을 짝짓는 인지실험을 실시하였다. 그 결과, 8명 중 4명 이상의 목소리를 주인공과 올바르게 짝을 지었으며, 전신을 보았을 때 얼굴만 보았을 때보다 정확도가 높았다. 또 좋은 소리는 좋은 인상보다 더 많은 사람들이 동감하는 것을 알 수 있었다.

ABSTRACT

To investigate the relationship between voice and the image created by the voice, a perception experiment was undertaken in which voices of 8 males and 8 females reading a fairy tale were recorded and presented to 100 subjects along with the speakers' pictures. Two kinds of pictures were taken: one, a whole-body picture, and the other, a face-only close-up picture. Subjects were asked to match each voice with its owner. The results showed that out of 8, more than 4 were matched correctly. And the correct rate was higher when the voices were matched against the whole-body pictures than against the face-only pictures. In addition, it was noted that far more people picked a single voice as the most 'favorite' one than they picked one picture as the most 'favorite' image.

I. 배 경

사람의 음성을 들으면 일반적으로 그 음성의 주인공에 대한 인상이 머리 속에 떠오르게 된다. 그 음성의 주인공이 이미 아는 사람일 경우는 이것이 당연하다고 하겠지만, 이 현상이 낯익은 사람에게만 해당되는 것은 아니다. 처음 듣는 목소리에 대해서도 그 목소리의 주인공에 대한 막연한 인상이 그려지게 된다. 왜 그런 것일까? 어쩌서 청각적 자극이 시각적 결과를 초래하는 것일까?

물론 이처럼 음성을 들음으로써 생기는 외모에 대한 인상이 사람마다 제각각 다르다면 이는 더 이상 거론할 여지없이 그저 우연한 현상으로 치부하면 될 것이다. 그러나 문승재[1]에 의하면 음성과 그 음성으로 인하여 생기는 인상의 관계가 무작위적이지 아니라고 할 수 있다. 남·여 각 8명의 목소리를 이들의 전신상과 연결하도록 한 이 연구에 의하면, 목소리를 듣고 머리에 떠오르는 인상은 그것이 실형 실제 목소리의 주인공과는 다를 수 있을지라도, 무엇인가 일관된 면이 있었다. 즉, 어떤 특정한 목소리를 들었을 때 인지실험 참여자들의 대다수는 그

목소리의 주인공으로, 어떤 특정한 사진을 선택했다는 것이다. 응답 중에는 목소리의 주인공으로 선택된 사람이 실제의 목소리의 주인공이 아닌 경우도 많았지만, 특기할 만한 것은 그 틀린 경우에도 인지실험 참여자들은 서로 비슷한 취향을 가지고 어떤 특정한 목소리에 대하여 어떤 특정한 사진을 선택했다는 것이다.

그렇다면 이는 무엇을 의미하는가? 음성과 외모가 서로 무관하지 않아서 한 음성을 듣는 사람들이 서로 비슷한 외모를 머리 속에 그리고 있다면, 이는 음성의 어떠한 면이, 무엇인가 시각적인 정보를 준다고 해석할 수 있을 것이다. 어떻게 이것이 가능할 수 있을까? 음성의 어떤 속성이 외모에 대한 어떤 정보를 전달한다는 것을 물리적으로 파악할 수 있을까? 본 연구는 이러한 문제에 더 체계적으로 도전해보고자 하는 시도의 일환이다. 물론 위에서 제기한 궁극적인 질문에 대한 답을 제시할 수는 없겠지만 그 답을 찾기 위한 방법을 제시하고 구체적인 자료를 확보하고자 하는 것이다.

이처럼 음성과 외모가 관계가 있을 수 있다는 사실은 음성과 관련되는 여러 분야에 영향을 미칠 수 있다. 이론적인 음성학의 관점에서 본다면 이러한 관계에 대한 연구는, 최근 들어 점점 그 중요성이 더 크게 인식되고 있는 음성 source 자체에 대한 이해[2]에 기여를 할 것이다. 뿐만 아

* 아주대학교 영어영문학과
접수일자: 1999년 7월 1일

니라 이러한 연구는 음성학 연관 분야에도 커다란 의미가 있다고 할 수 있다. 최근 들어 음성합성 기술이 많은 부분에서 실용화되고 있다. 아무리 좋은 목소리라 하더라도 우리는 이러한 모든 기계로부터 어떤 한가지 목소리만을 계속 듣는 것을 원치 않을 것이다. 상황에 따라서, 기능에 따라서 적절한 목소리가 따로 있을 것이기 때문이다. (Stanley Kubrick 감독의 2001: Space Odyssey에 등장하는 그 유명한 'HAL 9000' 컴퓨터의 목소리는 매우 부드러운 남성의 목소리로서 그 때까지 생각해왔던, 차갑고 이지적인 컴퓨터의 이미지와는 아주 다른 분위기를 자아내는 목소리였다.) 일테면, 어떤 경우에는 권위적인 목소리가 필요할 수도 있고, 어떤 때는 매력적인 남성이나 여성의 목소리가 좋을 수도 있으며 또 어떤 경우에는 부드럽고 자애로운 '엄마'와 같은 소리가 필요할 경우도 있을 것이다. 우리가 만일 어떤 목소리가 어떤 의도를 생각나게 하는지를 파악할 수 있다면 이것은 특히 rule-based 합성기의 경우에 지대한 영향을 끼쳐서, 합성기의 소리 발생 장치 부분의 변수를 변화시켜 쉽게 여러 가지 상황에 알맞은 여러 목소리를 만드는 것을 가능하게 할 것이다.

II. 인지 실험

2.1 이론적 고찰

일단 앞 선 실험[1]에서 목소리와 인상이 무작위의 관계가 아니라는 결론을 얻어낸 상황에서는 당연히 그 다음 질문은 '어떻게?, 왜?'일 것이다. 어떻게 우리는 특정 목소리로부터 특정한 인상을 생각해내는 것일까? 이 문제와 관련하여 시사하는 바가 있는 것이 소위 source-filter theory[3]이다. 우리는 여러 사람이 내는 각기 다른 소리를 한 소리로 알아듣는데 아무런 어려움을 겪지 않는다. 예를 들면 '가'가 말하는 '아'와 '나'가 말하는 '아'는 서로 다른 formant를 갖는 데에도 불구하고 우리는 그 두 소리를 모두 같은 '아'로 인지한다. Source-filter theory로부터 추론해본다면, 이러한 현상은 듣는 이가 무의식적으로 filter의 특성을 분석함으로써 물리적으로 다른 소리를 하나의 소리로 인식하기 때문에 가능하다고 할 수 있다. 이처럼, 말하는 이의 체격적인 특징으로부터 filter의 특성을 해석해 내고 그 결과, 그 filter로부터 나오는 소리를 재구성할 수 있는 능력이 우리에게 있다면, source가 되는 목소리와 외모를 짝 지을 수 있게 하는 능력도 있지 않을까? 즉, source만을 가지고 filter를 구성하는 외모에 대한 정보를 재구성할 수 있지 않을까? 이러한 이론적인 배경을 가지고 다음과 같은 인지실험을 실시하였다.

2.2 화자

본 인지실험에 사용할 자료의 화자는 앞에 있었던 연구 [1]의 화자와 같은 사람들로써, 아주대학교에서 1998년

하계 영어1급 정교사 자격연수를 받고 있는 현직 중·고등학교 교사들 중, 실험의 목적에 대하여 간단하게 설명한 후 참여를 자원한 사람 중에서 가능한 비슷한 연령층을 선정하였다. 비슷한 연령층을 선정하기 위해 노력한 이유는 나이로 인한 목소리의 자연스러운 변화가 인지에 미칠 수 있는 영향을 배제하고, 오로지 그 외의 다른 요소에 의한 영향만을 관찰하고 싶었기 때문이다. 이들의 나이는 남자의 경우는 32세에서 39세, 여자의 경우는 26세에서 40세까지의 분포를 보였다. 여성의 경우 나이 분포가 넓은 듯 하지만, 40세 한 명을 제외하면 26세에서 35세까지의 분포를 보이고, 나머지 7명 중에도 2명만이 20대로서 화자의 대부분이 30대의 좁은 연령층을 이루고 있었다.

2.3 실험자료 1: 목소리

목소리 자료는 [1]의 실험과 달리 남녀 모두 같은 동화의 한 부분을 읽도록 하였다. 이것은 개인적인 대화를 배제함으로써 개인적인 특성이 드러나는 가능성을 최소화하면서도 너무 짧지 않은 문장을 읽도록 하기 위함이었다. 화자가 읽은 문장은 다음과 같다.

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 떠나야했어요. 어느 시골에 글 습씨 뿐 만이 아니라 팔 쏘는 습씨 또한 뛰어난 젊은 선비가 있었습디다.

자연스럽게 읽으라는 것 이외의 다른 특별한 녹음상의 주의사항은 없었으며, 이렇게 읽은 것은 TASCAM PA-1 DAT 녹음기와 Electro-Voice 635 dynamic omnidirectional 마이크를 사용하여 48kHz의 표본율로 녹음하였다. 대부분 화자들의 경우 이것은 대화체보다는 낭독체에 가까웠다.

2.4 실험자료 2: 사진

녹음을 마친 후 똑같은 칸막이 앞에서 각 화자의 사진을 두 장씩 찍었다. 우선 한 장은 전신사진으로 똑같은 배경을 두고 같은 거리에서 같은 확대율로 찍었으므로 사진을 비교하면 각 화자의 키나 몸집 등 체격에 대한 정보가 분명히 드러나는 사진이었다. 다른 한 장은 각 화자의 얼굴만을 바깥 끌어당겨 찍어서 체격에 대한 정보가 거의 보이지 않는 얼굴 사진이었다. 이것은 체격 정보가 분명하게 드러나는 전신 사진과 그러한 정보가 상대적으로 훨씬 적은 얼굴 사진을 같은 목소리와 짝 짓도록 함으로써 체격 정보의 유무 여부가 정확한 인지에 영향을 주는지를 알아보기 위함이었다. 즉, 우리가 formant 구조로부터 filter의 특성을 추론해낼 수 있었는데, 이번에는 과연 source만을 가지고 그 filter에 관계되는 외모를 찾는 것이 가능한지를 알아보기 위한 것이다.

사진은 35-105mm 줌렌즈를 갖춘 Minolta Maxxum 7000 사진기와 Rokunar AF 920 TZ 플래시(45° 각도)를 이용하여 실내에서 촬영하였다. 후에 이 필름을 3.5" x 5" 크기로 인화하였다.

1) 이 멋진 사상 최대의 컴퓨터 아키텍처 유래에 대하여: HAL이란 이름은 'hell'과는 전혀 무관하며, 당시 최대의 컴퓨터 회사 IBM의 명칭에서 착안하여, 알파벳 상으로 한 자씩 앞자리를 골라 지은 것이라고 한다.

2.5 자료의 처리

개별적으로 한 명씩 실시했던 앞서의 연구[1]와는 달리 본 실험은 여러 사람을 대상으로 한꺼번에 실험을 실시하도록 고안하였다. 이것은 개별적으로 할 때에 걸리는 막대한 시간(한 명당 약 10-15분)을 절약함으로써 더 많은 인지실험 참여자를 확보하고자 하는 의도였다.

이를 위하여 다음과 같이 자료를 처리하였다. 앞에서 녹음된 소리는 남녀별로 세 번씩 반복하여 카세트 테이프에 복사하였다. 첫 번은 위의 녹음 내용 전체를 화자 1부터 화자 8까지 녹음하고, 두 번째와 세 번째에는 위의 동화 낭독 중 첫 두 문장만을 복사하였다. 각 화자의 녹음 앞에는 화자를 밝히기 위하여 실험자의 목소리로 '에이,' '비,' '씨,' 등의 소리를 삽입하였다.

사진은 720dpi의 해상도로 스캔을 한 후, PaintShop Pro를 이용하여 여덟 명의 사진을 한 화면에 나오도록 JPG file로 만들었다. 전신의 경우는 뒤에 있는 칸막이를 기준으로 하여 그 비율을 그대로 살려서 각자의 체격 정보가 그대로 나오도록 하였다. 모두 4개의 JPG file이 생겼다 (남자 전신, 남자 얼굴, 여자 전신, 여자 얼굴). 그 4개의 JPG file은 그림 1부터 4에 나타나 있다. (물론 실제 인지실험에 투사된 그림은 천연색이었다.)



그림 1. 남자 전신
Fig. 1. Male-Whole body.

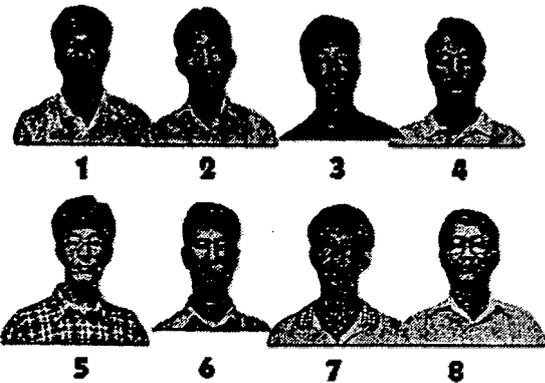


그림 2. 남자 얼굴
Fig. 2. Male-Face only.



그림 3. 여자 전신
Fig. 3. Female-Whole body.

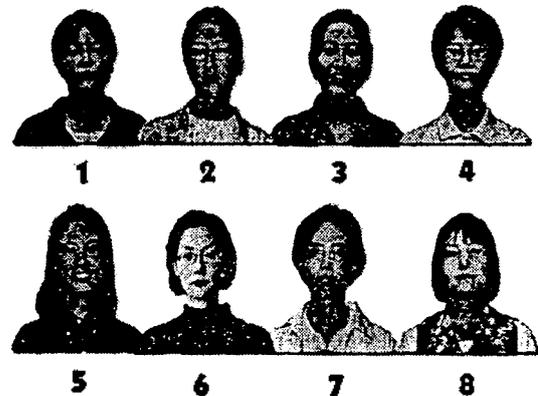


그림 4. 여자 얼굴
Fig. 4. Female-Face only.

2.6 인지실험 방법

인지실험은 아주대학교 학부생들을 대상으로 시청각 수업이 가능한 교실에서 실시하였다. 컴퓨터를 beam projector에 연결하여 그림을 투사하고 천장에 있는 스피커를 통하여 테이프를 톤 후 미리 준비한 답안지를 나눠주고 각 목소리에 걸맞는 사진의 번호를 쓰도록 하였다. 이 인지실험에 사용한 답안지는 표 1과 같다.

표 1. 인지실험 응답지
Table 1. Response Sheet.

실험참여자 이름: _____ 성별 _____ 대상:얼굴 일련번호:

女	소리	A	B	C	D	E	F	G	H
	사진								
	가장 좋은소리				가장 좋은인상				

男	소리	A	B	C	D	E	F	G	H
	사진								
	가장 좋은소리				가장 좋은인상				

목소리는 세 번 반복되므로 잘 듣고 결정하고 반드시 한 목소리에 한 사진을 고르도록 지시하였으며 답은 고쳐서 써도 된다는 것을 주지시켰다. (만일 세 번을 들어서 부족한 경우 요구하면 다시 테이프를 반복할 수 있다고 하였으나 실제로 반복할 것을 부탁한 경우는 없었다.)

한 세트의 인지실험은 같은 종류의 그림으로 구성되었다. 즉, 어느 한 집단에게 남자 전신을 보여주고 남자 목소리와 짝을 짓도록 했으면, 연이어 여자 전신을 보여주고 여자 목소리와 짝을 짓도록 했다. 그리고 다른 집단에게는 남자 얼굴과 목소리, 여자 얼굴과 목소리를 연결하도록 하였다. 물론 이 두 집단 사이에 같은 사람이 없도록 확인하였다. 이 인지실험에 참여한 학생 수는 모두 네 과목의 수강생 200명으로서 각 집단(전신/얼굴)마다 남자 55명 여자 45명이 참여하였으며 각 인지실험에는 약 25분이 소요되었다.

목소리와 얼굴을 짝 짓는 이외에도 '가장 좋은 소리'와 '가장 좋은 인상'을 고르도록 하였다. 그림을 투사한 후 테이프를 틀기에 앞서서 학생들에게 그 그림의 사람 중 '가장 인상이 좋은' 사람, 혹은 '가장 마음에 드는' 사람을 먼저 골라서 '가장 좋은 인상'이라고 마련된 칸에 번호를 쓰도록 하였다. 또한 테이프를 들으면서 '가장 마음에 드는' 목소리를 정하여 그 목소리를 a, b, c 등으로 '가장 좋은 소리'라고 마련된 칸에 적도록 하였다. 이것은 혹시 목소리와 사람을 연관지을 때 단순히 가장 마음에 드는 모습을 가장 마음에 드는 소리와 연결하는 것이 아닌지를 확인하고자 함이었다.

한가지 지적할 수 있는 문제점은 투사된 그림의 선명도가 인화된 사진보다 많이 떨어진다는 것이었다. 그러나 그런 경우에도 모든 사람의 모습을 분명히 알아보고 식별하는데에는 아무런 문제가 없었다.

이렇게 수집한 자료는 Excel의 countif 기능을 이용하여 그 빈도수를 계산하였다.

III. 결과 및 논의

3.1 전신 결과

남성 화자와 여성 화자의 전신을 보여주고 목소리와 짝을 지었을 때의 결과는 다음의 표 2, 3과 같다.

표 2. 남성전신 결과
Table 2. Response from Male-Whole body.

인지 화자	m1	m2	m3	m4	m5	m6	m7	m8	계
a(m1)	13	13	10	4	15	15	5	100	
b(m2)	13	10	7	1	24	8	2	100	
c(m3)	11	13	16	10	17	8	6	100	
d(m4)	9	18	9	13	0	13	100		
e(m5)	2	1	0	10	0	4	30	100	
f(m6)	11	10	1	23	10	10	8	100	
g(m7)	2	20	2	10	2	15	17	100	
h(m8)	17	15	2	15	7	2	16	100	
계	100	100	100	100	100	100	100	100	

표 3. 여성전신 결과

Table 3. Response from Female-Whole body.

인지 화자	f1	f2	f3	f4	f5	f6	f7	f8	계
a(f1)	25	20	4	1	13	19	4	10	100
b(f2)	11	12	20	22	1	6	10	18	100
c(f3)	13	14	1	19	20	2	6	100	
d(f4)	3	1	8	9	0	2	16	11	100
e(f5)	10	5	4	0	29	1	2	100	
f(f6)	14	3	15	6	8	14	100		
g(f7)	5	8	25	5	1	7	11	100	
h(f8)	9	9	11	9	2	11	21	100	
계	100	100	100	100	100	100	100	100	

위 표의 가로 행은 화자이고 세로 열은 각 화자에 대하여 인지실험 참여자가 짝을 지은 사진의 주인공이다. 예를 들면 a 소리는 실제로는 남성화자 m1의 목소리인데 이 소리를 듣고 13명이 m1의 사진과 짝을 지었고, 25명은 m3의 사진과, 4명은 m5의 사진과 짝을 지었다는 것이다. 응답수를 100명씩으로 하였기 때문에 응답수와 백분율은 일치한다. 따라서 각 목소리에 대하여 실제의 주인공과 짝을 지은 경우, 즉 정확히 목소리의 주인공을 알아 맞춘 경우는 표에서 밑줄을 그었으며, 각 목소리에 대하여 가장 많은 응답이 나온 것은 진하게 표시하였다. 따라서 밑줄 친 것과 진하게 표시한 것이 일치한다면 그것은 가장 많은 응답자가 어느 특정한 목소리에 대하여 그 주인공을 바르게 꼽았다는 것을 의미한다.

표 2에서 보면 남성의 경우 모두 8명 중 4명(m3, m5, m6, m8)의 목소리에 대하여 가장 다수와의 인지실험 참여자가 실제의 그 목소리의 주인공을 올바르게 짝지었음을 볼 수 있다. 이들의 올바른 인지율은 m3의 경우가 가장 낮아서 19%이고, 나머지는 1/4이 넘었으며 특히 m5의 경우는 무려 53%로서 절대 다수임을 볼 수 있다.

그러나 표 3에서 보면, 여성의 경우는 상황이 다소 다르다. 여성의 경우는 8명 중 5명(f1, f4, f5, f7, f8)의 목소리가 그 주인공과 제대로 짝지어졌으며, 특히 올바른 인지율도 가장 낮은 경우가 28%(f8)에서 가장 높은 경우는 59%(f4)로, 정확도가 남성보다 훨씬 높음을 알 수 있다.

그리고 남·여 모두 실령 목소리와 사진이 제대로 짝지어지지 않은 경우도 일반적으로 20% 이상의 인지실험 참여자가 어느 특정한 목소리에 대하여는 어떤 특정한 사진을 골랐음을 알 수 있었다.

이러한 결과를 앞 선 실험[1]의 결과와 비교해 보면 매우 흥미있는 현상을 발견할 수 있다. 앞 선 실험[1]에서는 남성의 경우, 화자와 내용이 같았으나 인지실험 참여자와 그 수가 달랐고, 그 방법도 여러 명이 한꺼번에 실시한 것이 아니라 한 명씩 개별적으로 실시했으나 결과는 본 인지실험과 그대로 일치하였다. 즉 8명 중 똑같은 4명에 대하여 올바른 응답이 많이 나왔다. 인지실험에 참여한 사람의 수가 두 배로 늘고, 그 당사자들도 다 달랐음에도 불구하고 이와 같이 유사한 경향을 보인다는 것

은, 최소한 이 8명의 화자에 대해서는 본 연구의 결과가 결코 어느 특정 집단에 의한 우연한 결과가 아니고 일반적인 것임을 확증해주는 것이라 할 수 있을 것이다.

반면 앞선 실험[1]에서, 화자와 내용까지 달라서 동화가 아닌 “안녕하세요, 제가 지금은 전화를 받을 수 없으니 메시지를 남겨주세요, 감사합니다”라는 잘못된 전화응답기 문장을 사용했던 여성의 경우는, 8명 중 불과 1명에 대해서만이 올바른 인지가 이루어져서 이번 결과와 큰 대조를 보이고 있다. 이것은 실험을 고안함에 있어서 발화의 내용이나 분량이 상당한 변수가 될 수 있음을 보여주는 예라고 할 것이다.

3.2 얼굴 결과

그러면 같은 목소리를 들려주고 전체의 체격에 대한 정보가 없이 얼굴 사진과 짝을 짓도록 했을 때의 결과는 어떨까? 만일 앞서 언급한 것처럼 source-filter theory가 암시하는 바가 작용을 한다면 얼굴만 보았을 때에는 전신을 보았을 때에 비하여 정확도가 떨어져야 할 것이다. 다음의 표 4와 표 5는 그러한 예측을 확인해 주고 있다.

표 4. 남성얼굴 결과
Table 4. Response from Male-Face.

인지 화자	m1	m2	m3	m4	m5	m6	m7	m8	계
a(m1)	10	5	23	5	11	10	11	11	100
b(m2)	3	4	22	0	22	3	5	5	100
c(m3)	3	16	18	1	12	16	15	100	
d(m4)	11	4	5	14	3	25	12	100	
e(m5)	37	1	0	1	0	9	5	100	
f(m6)	14	3	7	20	5	8	17	100	
g(m7)	6	4	4	8	22	9	23	100	
h(m8)	16	18	5	5	4	12	100		
계	100	100	100	100	100	100	100	100	

표 5. 여성얼굴 결과
Table 2. Response from Male-Face.

인지 화자	f1	f2	f3	f4	f5	f6	f7	f8	계
a(f1)	14	20	16	1	14	8	2	100	
b(f2)	4	12	19	6	6	14	14	100	
c(f3)	5	15	2	9	32	8	2	100	
d(f4)	2	2	5	0	5	9	25	100	
e(f5)	30	6	2	0	6	0	7	100	
f(f6)	12	11	1	16	15	12	9	100	
g(f7)	7	11	13	12	2	3	23	100	
h(f8)	14	10	13	13	4	8	18	100	
계	100	100	100	100	100	100	100	100	

표 4에서 보면 남성의 경우 목소리를 듣고 그 목소리의 주인공을 얼굴만 보고 옳게 찾은 경우는 2 경우(m5, m6)

로서 전신의 4 경우보다 줄었으며, 표 5의 여성의 경우도 3 경우(f4, f5, f7)로서 전신의 5경우보다 줄었음을 알 수 있다. 이로부터 우리는 목소리를 듣고 그 목소리의 주인공을 추측하는 데에는 체격에 대한 시각적인 정보가 큰 도움이 된다고 말할 수 있을 것이다.

얼굴의 경우도 전신의 경우와 마찬가지로, 틀린 경우에도 적게는 19%에서 많게는 41%까지 어떤 특정한 목소리에 대하여 그 목소리의 주인공으로 일정한 사람을 고르는 현상을 보이고 있어서, 어떠한 추세가 있다는 것은 부인할 수 없을 듯하다.

3.3 “좋은” 소리와 “좋은” 인상

혹시 인지실험 참여자들이 그저 막연히 듣기에 좋은 소리와 보기에 좋은 인상을 연결하는 것이 아닌지 확인하기 위하여 “좋은” 소리와 “좋은” 인상을 하나씩 고르도록 하였다. 이것은 실제로 어떤 목소리나 인상을 선택하는지 보다 “좋은” 목소리라고 꼽은 소리의 주인공이라고 생각한 사람을 “좋은” 인상의 소유자라고 역시 선택하는지를 보기 위함이었다. 일테면, 어떤 사람이 “좋은” 목소리를 A라고 했고, “좋은” 인상을 5라고 했을 때, 응답지에서 A에 대한 답으로 5를 썼다면 이 참여자는 “좋은” 목소리와 “좋은” 인상을 연관시켰다고 보고, 5가 아닌 다른 답을 골랐을 경우는 “좋은” 목소리와 “좋은” 인상을 연관시키지 않았다고 보는 것이다. 이렇게 각 응답자가 “좋은” 목소리라고 고른 목소리의 주인공과 “좋은” 인상으로 고른 사진이 일치하는 경우와 일치하지 않는 경우를 조사해 본 결과가 표 6에 나타나 있다.

표 6. “좋은” 목소리와 “좋은” 인상과의 일치 여부
Table 6. Match-rate of “favorite” voice and “favorite” image.

전신		얼굴	
여성화자	남성화자	여성화자	남성화자
일치	불일치	일치	불일치
47	34	45	33

표 6에서 보듯이 남·여 모두 일치하는 경우보다 불일치하는 경우가 더 많았다. 특히 남성화자에 대한 결과가 여성화자에 대한 결과보다 불일치 비율이 높았다. 따라서 일반적으로 “좋은” 목소리를 “좋은” 인상과 연결시키는 것은 아니라는 것이 밝혀졌다고 하겠다.

그런데 앞에서 이러한 “좋은” 소리나 “좋은” 인상에 대한 객관적인 응답이 중요한 것은 아니라고 했으나 이 결과에 대하여 다른 관점으로 접근을 해 보면 매우 흥미로운 사실을 찾을 수 있다. 즉, “좋은” 와모보다 “좋은” 목소리가 더 많은 사람들이 쉽게 동의할 수 있다는 것이다. 이 점은 표 7과 표 8을 비교해 보면 알 수 있다.

표 7. "좋은" 목소리
Table 7. "Favorite" voice.

	남전신			남얼굴			여전신			여얼굴		
	남	여	계	남	여	계	남	여	계	남	여	계
a	0	0	0	1	0	1	2	9	11	3	12	15
b	1	2	3	1	1	2	4	6	10	3	5	8
c	1	1	2	1	0	1	1	2	3	0	2	2
d	3	1	4	5	2	7						
e		7	27	8	10	18	1	0	1	0	0	0
f	0	0	0	0	0	0	0	1	1	0	0	0
g	14						1	0	1	2	0	2
h	16	10	26	10	7	17	4	1	5	0	3	3
계	55	45	100	55	45	100	55	45	100	55	45	100

표 8. "좋은" 인상
Table 8. "Favorite" picture.

	남전신			남얼굴			여전신			여얼굴		
	남	여	계	남	여	계	남	여	계	남	여	계
1	7	6	13	0	0	0	4	4	8	0	4	4
2	7	5	12				0	0	0	0	0	0
3		7		1	0	1	1	5	6	16	14	30
4	10	11	21	14	9	23	1	2	3	1	1	2
5	2	3	5	3	1	4	5	5	10			
6	0	0	0	7	6	13		14		10	5	15
7	3	1	4	0	1	1	8	0	8	1	0	1
8	10		22	6	2	8	11		26	1	2	3
계	55	45	100	55	45	100	55	45	100	55	45	100

표 7은 "좋은" 목소리로 선택한 목소리 (a-h)에 대한 응답수이고 표 8은 "좋은" 인상으로 선택한 사진 (sp1-sp8)에 대한 응답수로서, 모두 응답자의 성별에 따라 분류하였다. 표 7에서 보듯이 "좋은" 목소리는 남자 화자의 경우, 전신을 보여주었을 때나 얼굴만 보여주었을 때나 모두 e의 목소리가 가장 좋은 소리로 꼽혔다 (전신의 경우 남자 응답자만 예외). 그리고 여성 화자의 경우는 이 추세가 더욱 강하여 전신이나 얼굴 모두, 그리고 남녀 응답자 모두 약 70% 가깝게 d의 목소리를 꼽았음을 볼 수 있다. 반면에 표 8을 보면, "좋은" 인상에 대한 응답은 목소리에 대한 응답보다 훨씬 더 다양하게 분포되어 있음을 볼 수 있다. 그 선택 비율을 봐도 목소리의 경우는 약 70% 가깝게 의견의 일치를 보인 반면, 인상의 경우는 최고가 50%로 대부분 그보다 훨씬 낮은 수치를 나타내고 있다.

이 결과는 매우 재미있는 의미가 있을 듯하다. 우리는 어떤 외모에 대해서 주관적으로 "좋은지"의 여부를 판단하는 것보다 소리에 대해서 주관적으로 "좋은지" 여부를 논할 때 훨씬 의견의 합의를 보기 쉽다는 것이다. 이는 많은 사람이 만족할 목소리를 만드는 것이, 외모를 만드는 것보다 한결 쉬울 것이라는 것을 의미하기도 한다. 할 수 있을 것이다.

IV. 결론 및 제언

본 실험은 우리가 어떤 목소리를 들었을 때 떠오르는 외모에 대한 추측이 전혀 사실과 무관하지는 않으며, 오히려 우리 예상과는 달리, 비슷하게 맞추는 경우가 그렇지 않은 경우보다 많다는 것을 보여 주었다. 또한 본 실험의 결과와 앞 선 실험[1]의 결과와 비교함으로써 틀리는 경우에도 어떠한 일관적인 추세가 분명히 있음을 입증하였다.

그리고 이처럼 청각적인 정보와 시각적인 정보를 연결함에 있어서, 전체의 체계적인 정보가 주어졌을 때에 더욱 정확하게 목소리의 주인공을 알아낼 수 있었음도 발견하였다. 이는 source-filter theory에서 이미 암시했던 바로, 우리는 머리 속에서 우리의 체계적인 정보를 무의식적으로 처리하여 객관적인 정보를 해석하는 데에 능동적으로 이용하고 있음을 짐작케 해 주는 부분이라 하겠다. 하지만 source와 filter를 거의 별개로 취급한 것과는 달리, 본 연구의 결과는 source 자체가 이미 filter, 즉 체계에 대한 특성을 상당 부분 포함하고 있음을 보여준다고 할 수 있을 것이다.

그리고 단순히 "좋은" 목소리와 "좋은" 인상을 자동적으로 짝 짓는 것은 아님도 알 수 있었다. 그 뿐 아니라, "좋은" 목소리라는 것이 "좋은" 인상보다 훨씬 더 쉽게 주관적으로 동의할 수 있다는 것을 보았다.

이러한 결과들은 본 실험을 시작하기에 앞서 사람들에게 비공식적으로 물어보았던 것과 상반되는 것이 많았다. 많은 사람들이 자신들의 경험에 비추어 볼 때 자신들이 떠올리는 인상은 실제와 틀린 경우가 더 많았다고 기억했으며, 대부분 "좋은" 인상과 "좋은" 목소리가 함께 연결될 것이라고 생각했다. 그러나 이것은 아마도 예상이 틀려서 놀란 경우가 더 기억에 깊이 남기 때문이 아닐까 하는 추리를 가능케 한다.

본 실험은 그 실험 진행 방법에 있어서, 큰 제약을 가지고 있다. 시간 절약을 위하여 많은 사람을 한꺼번에 모아 놓고 빔 프로젝터를 이용하여 사진을 투사하고 실시하였기 때문에 이상적인 상태에서 실행한 실험은 아니라고 할 수 있을 것이다. 특히 빔 프로젝터로 비춘 그림의 선명도가 실제 사진보다 떨어졌다는 것이 본 연구의 제약점이라고 할 수 있을 것이다. 그러나 그러한 열악한 환경에서도 이와 같은 결과가 나왔다면 더 좋은 환경에서 할 때에는 이보다 더 나은 결과, 즉, 더 많은 목소리가 실제의 주인공과 짝이 지어지는 결과가 나오지 않을까 하고 생각해 본다. 물론 이것은 보장되어있는 것은 아니나, 앞으로 당연히 확인을 해야 할 과제라고 생각한다.

좀 더 확고부동한 결과를 얻기 위하여 앞으로는 개별적인 인지실험을 통하여 더 많은 응답을 얻을 계획이다. 그래서 화자의 성별, 전신과 얼굴, 응답자의 성별 등을 변수로 하는 통계학적인 분석을 해 보아야 할 것이다.

이렇게 얻어진 자료는 사람의 목소리와 외모와의 상관관계를 이해하는데 큰 도움을 줄 것이다. 즉, 음성의 어떤 특성이 그 주인공에 대한 시각적인 정보를 전달하는가

하는 질문에 대한 답을 구하는데 일익을 할 것이며, 그것은 다시, 어떤 소리가 어떤 시각적인 정보를 유도하는지를 밝히는데 자료를 제공할 것이다. 이러한 연구는 궁극적으로는 인간의 음성의 본질을 이해하고 규명하는데 기여

할 수 있을 뿐 아니라, 실용적인 측면에서는 정서적인 측면을 연구하는 최근의 추세[4]에 합하여 좀 더 자연스럽게 경우에 적합한 목소리를 합성하는데 도움을 줄 수 있을 것이다.

참 고 문 헌

1. 문승재, "음성과 인상의 관계 규명을 위한 실험적 시도," 말소리 35-36, 37-48, 1998.
2. J. Sundberg, *The Science of the Singing Voice* (Northern Illinois University Press, Dekalb, 1987). Chap.4, 7.
3. G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, 1970). Chap.1.1, pp.14-26.
4. L. Leinonen and T. Hiltunen, "Expression of emotional motivational connotations with a one-word utterance," J. Acoust. Soc. Am. 102(3), pp. 1853-1863, 1997

▲문 승 재(Moon, Seung-Jae) 1960년 10월 16일생



1983년 8월 : 서울대학교 인문대학
영어영문학과(학사)

1990년 12월 : University of Texas at
Austin(석사 : 언어학)

1991년 12월 : University of Texas at
Austin(박사 : 언어학/음
성학)

1992년 1월 ~ 1993년 8월 : University of Texas at Austin-
(Post-Doctoral Research Associate)

1999년 9월 ~ 2000년 8월 : Visiting Scholar in the Depart-
ment of Linguistics, University of Texas at
Austin

※주관심분야 : 음성학, 음성처리, 음성DB, 판소리발성법