

음성과 인상의 관계: 예비 연구

문 승 재

아주대학교 인문대학 영어영문학과
sjmoon@madang.ajou.ac.kr

요약

사람의 음성을 들으면, 낯익은 사람은 물론이지만 처음 듣는 목소리에 대해서도 그 목소리의 주인공에 대한 막연한 인상이 그려지게 된다. 본 논문은 이러한 현상이 얼마나 신빙성이 있는지, 즉, 음성만을 듣고 짐작한 그 음성의 주인공의 모습과 실제의 모습이 과연 얼마나 관계가 있는지를 알아보고자 하는 연구의 초기 단계에 대한 보고이다.

본 연구에서는 이처럼 음성이 전달하는 시각적 정보의 신빙성을 확인하기 위하여 남녀 각 8명의 사진을 찍고, 같은 내용의 짧은 문장을 녹취한 후, 100명 이상의 피실험자들에게 개별적으로 녹음을 듣고 가장 잘 어울릴 듯한 사진을 고르도록 한 것이다.

우선적으로 여성 8명의 녹음을 약간명(48명)의 피실험자에게 들려주어 살펴본 결과, 목소리의 주인공을 바로 찾는 경우는 드물었지만, 흥미로운 것은 비록 틀린 경우라도 어떤 특정한 목소리는 어느 특정한 사진과 집중적으로 연결되었다는 것이다. 이 결과를 source-filter theory와 연관시켜 생각해보고, 이를 바탕으로 좀 더 구체적인 앞으로의 연구방향을 제시한다.

1. 서론

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. 경우에 따라서는 그 그려진 모습이 매우 확실하게 여겨지기도 한다. 그래서 후에 목소리의 주인공을 만나게 되면 때로는 예상했던 바와 크게 달라서 놀라기도 하고 때로는 예상했던 모습과 비슷하여 당연하게 생각하게도 된다.

본 연구는 이러한 현상이 얼마나 신빙성이 있는지, 즉, 음성만을 듣고 짐작한 그 음성의 주인공의 모습과 실제의 모습이 과연 얼마나 관계가 있는지를 알아보고자 한다.

우선 음성을 들었을 때 머리 속에 그려지는 인상이 과연 신빙성이 있는가를 밝혀내야 할 것이나 만일 우리가 음성만을 듣고 그 주인공을 신빙성있게 알아낸다고 하면, 그것은 음성의 어떤 성질 때문인가 하는 의문을 낳게 된다.

Source-filter theory(Fant, 1970)에 의하면 성대에서 만들어지는 소리는 성대의 자질, 길이 등에 의하여 특정한 소리가 나게 되며, 그 소리는 체격에 의하여 저마다 다른 성도를 거치면서 여과되어 최종적인 말소리가 나게 된다. Filter에 해당하는 성도는 사람마다 그 크기가 달라서 같은 모음이라 하더라도 그 formant 값이 다른데도 불구하고 인간은 그 다른 formant 값을 갖는 모음을 한가지 모음으로 인식하는데 아무런 어려움이 없다. 이러한 현상은 인간이 두뇌에서 자동적으로

로 상대방의 체격 등에 대한 정보를 바탕으로 normalization을 하기 때문에 가능한 것이다. 즉, 키가 이 정도이면 성도의 길이가 그에 맞게 달라서 다른 formant를 갖는 모음이 나올 것이라는 것을 미리 예측한다는 것이다.

이러한 source-filter theory는 본 연구의 주제에 대하여 재미있는 가능성을 시사하고 있다. 즉, 사람에게 소리를 듣고 그 소리속에 있는 다른 음향학적인 정보(진력 다른 formant 값)를 한 소리로 알아들을 수 있는 능력이 있다면, formant 값 이외의 다른 정보도 추출해 낼 수 있는 능력이 있지 않을까, 혹시 상대방의 목소리를 듣고 반대로 그 소리로부터 체격을 추측해내는 것이 아닐까, 이러한 추측이 가능해진다. 바로 이러한 점을 알아보고자 하는 것이 본 연구의 목적이다.

본 연구에서 조사하고자 하는 것은 (1) 낯선 음성을 들었을 때 머리에 떠오르는 인상과 실제의 모습이 어느 정도 일치하는가, (2) 만일 일치하지 않는 경우에는 일치하지 않는 중에도 어떠한 일관된 양상이 나타나는가 하는 것이다.

그 뿐 아니라 우리가 일반적으로 가지고 있는 '좋은' 인상이 '좋은' 목소리와 관계가 있는지를 알아보는 것도 본 연구의 한 과제이다. 우리가 느끼는 주관적인 인상은 사람마다 얼마나 다른가? 만일 일관된 면이 있다면 과연 외모에서 느끼는 인상과 목소리에서 느끼는 인상이 비슷할까? 이러한 주관적인 면들은 얼핏 객관적으로 확증할 수 없을 듯 하지만 지금까지는 객관적으로 검증하고자 하는 노력조차 없어왔던 것 같다.

이러한 주제는 speech production 및 perception 연구와 깊은 관계가 있다고 할 수 있다. 이제는 음성합성 및 인식분야에서 정확한 발음만을 강조하던 추세에서 감성적인 면을 보다 중요시하는 쪽으로 연구 동향이 바뀌고 있다 (여양을 통하여 나타나는 감성에 대한 Leinonen & Hiltunen, 1997; Protopapas & Lieberman 1997 등). 본 연구는 이러한 추세에 맞추어서, 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.

2. 실험 방법

2.1 인지실험 자료 제작

위에서 밝힌 바와 같은 점들을 확인하기 위하여 인지실험을 실시하였다. 이 인지실험을 위하여 우선 남자 6명, 여자 8명의 목소리를 녹음하였다. 녹음 대상은 인지실험의 대상이 아주대학교 학생이 될 것을 고려하여 이들에게 낯선 목소리를 얻고 또한 연령층을 고르게 할 수 있기 때문에 아주대학교 교육대학원 학생들을 부적으로 골랐다.

녹음은 TASCOM PA-D1 DAT 녹음기(48kHz sampling rate)와 Shure SM 48 dynamic 마이크를 이

인지수 (%)	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)	

표 2. 인지실험 결과

용하여 조용한 저녁 시간에 보통 강의실에서 이루어졌다.

녹음의 내용은 말 자체의 내용에서 각 사람의 개인적 특징이 나타나는 것을 배제하기 위하여 다음과 같은 간단한 전화응답기 내용으로 통일하였다.

“안녕하세요, 지금은 전화를 받을 수 없으니 메시지를 남겨주세요, 감사합니다.”

녹음과 더불어 인상을 나타내기 위하여 사진을 찍었다. 이 때에는 각자의 재적 특성이 상대적으로 잘 나타나도록 강의실의 문을 배경으로 문이 모두 다 나오도록 찍었다. 또한 전체 체격이 인지에 영향을 미치는지 혹은 단순히 얼굴의 인상만이 영향을 미치는지를 비교하기 위하여 각각 전신 사진과 흉부 이상 얼굴 부분만을 확대한 사진, 두 장씩의 사진을 찍었다. 사진 촬영은 Minolta Maxum 7000에 35-105mm zoom lens와 Kodak 필름(ASA 400), Maxum 전용 flash를 이용하여 이루어졌다. 녹음을 하는 당사자들에게는 사진 및 녹음에 대하여 사진에 알리지 않고 녹음과 촬영 당일 날 이야기하였다.

이렇게 하여 얻은 녹음자료는 Kay Elemetrics의 CSL 5000B를 이용하여 8kHz low-pass filter를 통하여 20kHz sampling rate로 digitize하여 개인별로 소리 파일을 저장하였다. 사진은 보통 크기로 현상하여 8명씩 한꺼번에 볼 수 있도록 clear file 2면에 건치 배치하였다.

2.2 인지실험

본 연구에서는 우선 8명이 확보되어 있는 여성만을 대상으로 인지실험을 하여, 과연 목소리를 듣고 어떤 특별한 형태의 인지 양상을 보이는지 먼저 확인하고자 하였다. 만일 이러한 기본적인 실험의 결과, 전혀 공통적인 양상이 안 보인다면 그 다음 단계의 실험, 즉 성별간의 차이를 파악한다든가 혹은 전신 사진을 제시했을 때와 얼굴만을 제시했을 때의 차이를 살펴본다든가 하는 후속 실험을 진행한다는 것이 무의미한 것이기 때문이다. 따라서 1차적으로 본 연구에서는 45명(남자 19명, 여자 26명)을 대상으로 인지실험을 실시하였다.

본 연구의 인지실험은 사진과 목소리를 연계하는 실험인 만큼, 사진과 목소리를 같이 제시해야 하는데, 8장의 사진을 절의 저하 없이 여러 사람에게 한꺼번에

보일 수 있는 방법이 마땅치 않을 뿐 아니라, 소리 역시 피실험자들이 듣고 싶은 목소리를 제각기 반복하여 들을 수 있도록 하기 위하여, 다수의 피실험자들에게 한꺼번에 실험을 실시하지 않고 한 번에 한 명씩만 음성학 실험실로 불러서 실시하였다.

개별적인 실험을 위하여 CSL 5000B를 이용하여 A부터 H까지 8개의 창을 연 후, 각 창마다 8명의 여성의 소리를 무작위로 각각 불러들였으며 이 과정은 macro를 이용하여 매번 같은 순서로 나열이 되도록 하였다. 각 창의 파일 이름은 F1, F2 등으로 하여 그 이름을 직접적으로 알 수 없도록 하였다. 그리고 8명의 전신 사진은 두 장의 clear file에 끼우고 그 밑에 ①부터 ⑧까지 번호를 붙여서 나열하였다.

피실험자는 컴퓨터 화면 앞에 앉아서 CSL의 간단한 조작법 (창 이동 및 소리 듣기)에 대한 설명을 듣고, beyerdynamic DT211 헤드폰을 통하여 소리를 들었으며, 반복하여 듣는 횟수는 제한을 하지 않았다. 그 결과는 다음의 그림1과 같은 답안지에 기입하도록 하였다.

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

표 1. 인지실험의 응답지

각 창의 목소리에 맞는 사진의 번호를 해당 창 아래에 쓰는 이외에도, 각자 가장 좋다고 생각하는 소리를 A에서 H 중 하나, 그리고 가장 좋다고 생각하는 인상은 ①에서 ⑧ 중 하나를 고르도록 하였다. 이것은 혹시 가장 좋은 소리와 인상을 연관시키는 경향이 있지 않을까 하는 것을 파악하기 위함이다. 이 때 피실험자들에게는 “가장 좋은” 소리와 인상을 고르라는 것 이외에는 다른 특별한 설명을 부가하지 않았다.

3. 결과

통계학적으로 의미있는 결과를 얻기 위하여서는 상당한 수의 응답이 필요하다. 위에서 언급했듯이 대략적인 상황과 추세를 파악하기 위하여 본 연구에서는 8명

의 여성 화자에 대한 45명의 인지 실험 결과만을 보고하기로 하겠다.

여성 8명 화자에 대한 인지실험 결과는 다음의 표2와 같다. 45명의 피실험자가 각 화자의 목소리에 대하여 그 목소리의 주인공을 짐작한 결과를 응답수와 100분위(관호 속)로 나타내었다. 따라서 옳은 응답(즉 F1의 목소리를 F1의 사진과 제대로 인식한 경우)은 도표상에서 대각선으로 나타나는 셀들(빛줄을 그어서 표시했음)이며, 옳고 그름에 상관없이 가장 많이 연결이 된 셀은 진하게 나타내었다.

이 표를 보면, F8의 경우를 제외하고는, 처음에 예상했던 바와는 전혀 달리 대부분 목소리와 사진이 바로 짝지어진 경우가 없음을 알 수 있다. 목소리만을 듣고 그 목소리의 주인공의 모습을 제대로 인식한 경우는 불과 9-18%로 특별히 높은 인지율을 보이지 않았다.

그러나 이 표를 자세히 보면, 비록 목소리와 사진의 올바른 연결은 없었지만 재미있는 현상은 발견할 수 있다. 즉, 45명의 인지실험 대상자들은 각각의 화자의 목소리를, 비록 틀릴지라도, 어떤 특정한 모습과 연관지었다는 것이다. 물론 전체 응답 수가 그렇게 많지는 않았지만, 연결할 수 있는 가능한 사진이 8장이었던 것을 생각한다면 표에서 볼 수 있는 24% 이상의 일관된 응답은 우연이 아니라고 보는 데 큰 부리가 없을 것이다.

만일 이 소규모의 실험 결과에 바탕을 둔다면, 목소리를 듣고 머리 속에 그 목소리의 주인공을 연상할 때, 목소리로부터 짐작할 수 있는 인상과 연관성은 거의 없다고 볼 수 있을 것이다. 그러나 그렇다고 하여 그 연상되는 모습이 전혀 엉뚱한 것만은 할 수는 없다. 그것은 45명의 응답이 어떤 특정한 목소리를 나름대로 어떤 특정한 모습과 연관할 짓는 경향이 있다는대서 내릴 수 있는 잠정적인 결론이다.

이번 연구에서 또 한가지 알 수 있었던 것은, 소위 "좋은" 목소리를 반드시 "좋은" 인상과 연결시키지는 않는다는 것이었다. 표2에는 나타나 있지 않으나, 앞에서 밝혔듯이 본 인지실험의 대상자들은 막연히 어느 목소리가 가장 좋고 어느 인상이 가장 좋은지를 반드시 답하도록 강요를 받았는 바 그 결과로부터 좋은 소리와 좋은 인상을 연결시키려는지를 알아보았다. 이것은 응답자가 선택한 '좋은 목소리'에 대하여, 그 주인이라고 응답자가 연결한 사진이 과연 같은 응답자가 '좋은 인상'이라고 뽑은 사람인지를 비교함으로써 이루어졌다. 이 경우 역시 소리와 모습 연결의 객관적인 옳고 그름은 전혀 개의치 않았다. 그 결과 45명의 응답자 중 16명(36%)만이 좋은 목소리와 좋은 인상을 연계하였으며 나머지 29명(64%)은 좋은 인상으로 뽑은 사람과 좋은 목소리의 주인공으로 뽑은 사람이 서로 일치하지 않았다. 이는 곧 목소리의 주인에 대한 인상이 단순히 "좋다, 나쁘다"는 주관적인 판단에 의거한 것만은 아니라는 것을 나타낸다고 하겠다.

4. 앞으로의 연구 방향

본 연구에서 얻을 수 있는 잠정적인 결론은 (1) 목소리는 그 목소리 주인공의 모습을 객관적으로 정확히 전달하지는 않으나, (2) 듣는 사람들은 어떻게 해서인가 그 목소리의 주인공에 대하여 공통된 인상을 가지고 있지만 (3) 그 판단의 기준이 단순히 "좋고 나쁘다" 것에 의존하지는 않는다는 것이다.

위의 관찰 (2)로부터 비롯되는 당당한 질문은, 비록 주인공을 제대로 찾지는 못했지만 목소리의 어떤 상황

이 그러한 공통된 인상과 관련 정보를 듣는 이에게 제공하는 것일까 하는 것이다.

그러나 이러한 질문에 답하기에 앞서 과연 이러한 결론이 확실한 지를 확인해야 할 것이다. 물론 이번 초기 연구의 결과가 아무런 경향없이 나왔다면 더 이상 탐구할 여지가 없었으나 일단 일정한 경향이 관찰된 이상, 그 관찰이 여성 화자 8명에 대한 45명의 인지실험 응답에 바탕을 둔 것일 만큼 그 보편성을 확인하기 위하여 훨씬 더 큰 범위의 연구가 필요하다. 따라서 현재의 45명보다 더 큰 집단(100명 이상)을 대상으로 인지실험을 실시할 필요가 있다.

그리고 여성의 목소리뿐 아니라 남성의 목소리도 포함해야 할 것이다. 어느 정도 근거가 있는지는 알 수 없으나, 실험 도중 상당수의 사람들이 "남자라면 훨씬 더 자신있게 답할 수 있을텐데"라는 아쉬움(?)을 표한 것은 재미있는 일이었다. 과연 그들의 주장이 사실인지, 또한 남성들의 목소리에서도 위에서 문 현상들이 공통적으로 관찰되는지 확인해 보아야 할 것이다.

그리고 본 연구에서는 인지실험 참여자의 수가 부족하여 하지 않았으나, 실험 참여자의 성별에 따른 차이가 있는지도 분석에 있어서 하나의 변수가 될 수 있을 것이다.

이번 인지실험 참여자들의 또 다른 공통적인 반응은 화자 8명의 목소리가 서로 상당히 비슷했다는 것이다. 많은 사람들이 목소리가 "그게 그거" 같아서 배우 어려웠다고 평을 했는데, 그것이 과연 정말 우연히 비슷한 목소리들만 모아서 그런 것인지, 아니면 인지실험이라는 환경에서 오는 스트레스로 인한 것인지는 현재로서는 판단을 할 수 없다. 것처럼 "비슷한" 목소리인데도 위와 같이 일관된 결과가 나왔다는 것은 특기할 만한 일이라고 할 수 있지 않을까? 이를 밝혀고자 한다면 소위 특징있는 목소리만 모아서 인지실험을 실시하여 그 결과가 과연 다른지를 비교해 보면 좋을 것이다.

이 외에도, 각각 다른 집단에게 전산과 얼굴만을 보여주고 그 결과를 비교해 보는 것도 의미가 있을 것이다.

따라서 앞으로는 크게 5가지의 인지실험을 실시하여 그 결과를 비교하여야 할 것이다. 즉, 남성 전신, 남성 얼굴, 여성 전신, 여성 얼굴, 특별한 목소리만 모은 추가 집단을 각각 보여주는 실험이다.

그 실험들의 결과를 비교 분석하여 공통된 추세를 발견한다면 그 때에야 비로소 그것을 출발점으로 하여 어떤 특정한 인상을 연상하게 하는 목소리의 특성을 연구할 수 있을 것이다.

5. 참고문헌

- Fant, Gunar. (1970). *Acoustic Theory of Speech Production*, (2nd edition), The Hague: Mouton.
- Leinonen, L. & Hiltunen, T. (1997). "Expression of emotional motivational connotations with a one-word utterance," *Journal of the Acoustical Society of America*, Vol. 102 (3), 1853-1863.
- 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.