

한국어 말소리 지각에 미치는 조음에 관한 시각정보의 영향: 한국인의 McGurk 효과+

최양규* 남기춘**

*춘해대학 유아특수치료교육과 **고려대학교 심리학과

The influence of visual information on place of articulation in Korean speech perception: The McGurk effect in Korean subjects

Yang-Gyu Choi* Kichun Nam**

*Dept. of therapeutic & special education for children at Choonhae college

**Dept. of psychology at Korea university

ygchoi@choonhae.ac.kr

요 약

운동이론(motor theory)에 따르면 조음에 관한 정보는 말소리 지각에 중요한 역할을 한다고 한다. 조음에 관한 시각정보가 자음지각에 중요함을 시사하는 것이 바로 McGurk 효과이다. McGurk 효과는 말소리 지각에서 청각정보와 시각정보가 상충될 때 지각의 결과는 청각에 의한 조음정보와 시각에 의한 조음정보가 통합(integration)되어서 나타나는 것을 말한다. 예컨대, 시각적으로는 /ga/를 발음하는 모습을 보여주면서 동시에 청각적으로는 /ba/를 들려주면 그 결과로 /da/로 지각된다. 마찬가지로 시각적으로는 /ka/를, 청각적으로는 /ma/를 제시하면 /na/로 지각된다. 따라서 McGurk 효과는 시각적인 조음 정보가 자동적으로, 무의식적으로 말소리 지각과정에 통합됨은 보여준다. 한편 이러한 McGurk 효과는 문화마다 그 강도가 다르게 나타난다는 보고가 있다(Sekiyama, 1997). 예컨대, 일본과 중국 원어민의 경우 미국 원어민보다 McGurk 효과가 약하게 나타났다. 본 연구는 한국인에게는 McGurk 효과가 어떠한 양상으로 나타날지를 규명해 보고 아울러 기존의 미국, 일본 그리고 중국 원어민에 대한 연구결과와 비교 분석해 보았다.

서 론

말소리 지각은 순전히 청각적으로만 된다고 보통 생각하기 쉽지만 대면해서 말소리로 의사소통을 할 때 청자는 화자의 조음동작을 볼 수 있게 되는데 이때 시각적으로 주어지는 독순정보(lip-read informtin)는 말소리 지각에 중요한 역할을 한다는 증거들이 제시되어왔다(McGurk & MacDonald, 1976). 특히 말소리가 명료

하지 않을 때 화자의 얼굴을 바라보는 것은 말소리 지각을 개선시킨다(Sumby & Pollack, 1954). 청각정보만으로는 알기 어려운 조음장소정보를 시각정보를 통해서 쉽게 알 수 있기 때문이다(Binie, Montgomery, & Jackson, 1977). 말소리는 입술, 치조, 구개, 성문 등의 성도상의 여러 가지 장소들에서 조음된다. 독순에 관한 연구들에 따르면 인간의 시각은 양순음(/b/, /p/과 같은 입술소리)과 비양순음(/d/, /n/처럼 입술 안쪽에서 조음되는 소리)을 매우 예민하게 구분해 낼 수 있다.

사람들이 말소리를 지각할 때 어떻게 청각정보와 시각정보를 통합하는지를 알아보기 위해서 사용되는 실험 패라다임은 청각정보와 시각정보가 상충되는 시청각 자극을 이용하는 McGurk 효과 실험을 들 수 있겠다..

+ 본 연구는 정부통신부 위탁연구과제(계약번호: 1010-2002-0012)지원으로 수행되었음.

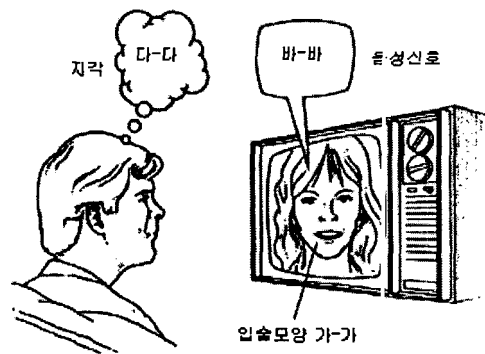


그림 1. McGurk효과: 여자의 입술이 마치 “가-가”라고 말하는 것처럼 보이지만 실제로는 “바-바”이다. 그러나 청자는 “다-다”소리를 들은 것으로 보고한다. 눈을 감고 더 이상 여자의 입술을 보지 않게 되면 “바-바”로 듣게 된다.

그림 1에서 처럼 McGurk와 MacDonald(1976)는 한 실험을 통해 이 사실을 극적으로 예시한 바 있다. [ga]를 발음하는 모습을 촬영한 비디오 장면에 [ba]를 발음한 소리를 더빙한 것을 제시하였을 때 사람들은 시행의 98%를 [da]라고 들었다. 이러한 McGurk 효과는 조음 장소에 관한 시각정보가 음소지각에 얼마나 쉽게 영향을 주는지를 잘 보여준다. 시각적인 강소정보([ga]: 비양순음)는 청각정보([ba]:양순음)를 왜곡시키는데 조음 장소정보는 시각을 통해서 조음장소에 관한 정보를 받아들이고 나머지는 청각을 통해서 정보를 받아들임으로써 결국에는 이 두 가지를 혼합한 것과 일치하는 [da]로 지각하게 된다. 보다 구체적으로 말해서 청각적 소리 [ba]는 양순음 정보를, 시각적 입모양 [ga]는 비양순음 정보를 갖고 있는데 지각된 소리인 [da]는 조음장소에서 시각정보와 일치하고 나머지 조음방식은 청각정보와 일치한다.

McGurk 효과는 말소리가 명료할 때조차도 독순정보가 중요한 역할을 한다는 것은 보여주는 예이다. 화자의 입술동작은 보고 관찰한다는 것이 청자가 듣는 것에 영향을 준다는 이 예는, 말 산출과 말 지각 사이의 연계를 보여주는 것이기 때문에 특별언어기제를 지지하는 증거로 인용되어왔다. 운동이론은 음성기들이 어떻게 음향신호를 생성해 내는지를 결정하는 모든 가능한 정보를 사용한다고 말함으로써 McGurk 효과를 설명하였다. McGurk 효과가 나타나는 상황에서 정보는 음성신호와 화자의 입술동작으로 이루어져 있다. 특히 입술동작은 음성신호가 어떻게 산출되는 방식에 관한 정보를 알려주기 때문에 중요하다.

McGurk 효과는 영어권 문화에서 매우 강력하게 나타난다. 그러나 일본인과 중국인, 미국간의 McGurk 효과를 비교한 Sekiyama등(1991, 1993, 1997)에 따르면 동양권인 일본인과 중국인의 경우 영어권인 미국인이

비해서 그 효과가 매우 약하게 나타남을 보고하였다. 일본인의 경우 말소리가 매우 분명할 때는 효과가 거의 나타나지 않았다. 그들은 시청각 통합 방식에 영향을 주는 요인으로 문화적 요인을 들고 있다. 일본인들은 옆사람이 말을 할 때 얼굴을 쳐다보는 것은 무례한 것으로 생각하기 때문에 말소리가 분명하게 알아들을 수 있을 때는 얼굴보기를 피하는 경향이 있다는 것이다. 이러한 문화적인 습관이 화자의 얼굴을 보고 있을 때조차도 시각정보와 청각정보를 통합하지 않는 식의 처리 방식을 개발시켰을 가능성이 있다. 안면회피(face-avoidance) 요인이 McGurk 효과의 두 언어간의 차이를 발생하게 만든 요인인지를 알아보기 위해 일본인과 같은 문화권인 중국인을 대상으로 동일한 실험을 한 결과 중국인도 일본인과 유사하게 약한 McGurk 효과를 보였다(Sekiyama, 1997).

한국인에게 있어서 McGurk 효과를 알아본 연구는 아직 없다. 본 연구는 한국인에게는 McGurk 효과가 어떤 양상으로 나타나는지를 알아보고 미국인과 일본인의 결과와 비교하고자 한다. 문화에 따라 McGurk 효과가 다르게 나타나게 하는 안면회피요인이 그 효과에 영향을 준다면 일본, 중국과 동일한 문화권인 한국인에게서도 비슷하게 그 효과가 미국인에 비해 약하게 나타날 것이다.

본 연구는 양순음 음절인 ‘마’, ‘바’와 비양순음 음절인 ‘가’, ‘나’, ‘다’, 이 5개의 음절의 발화모습(시각정보)과 발화소리(청각정보)를 조합하여 제작한 동영상 자극을 이용하여 한국인의 McGurk 효과를 알아보았다. 이를 위해 청각적으로는 양순음이고 시각적으로는 비양순음인 자극과 그 반대로 청각적으로는 비양순음이고 시각적으로는 양순음인 불일치 자극들 대한 지각을 실험을 통해 알아보았다.

방법

실험참가자. 부산과 울산지역에 거주하고 있으며 연령이 20-33세 사이이며 시력과 청력이 정상인 춘해대학 재학생 29명이 본 실험에 참가했다. 이들 중 남자는 2명, 여자는 27명이었다.

자극. 여성화자가 5개의 한국어 음절(/가/, /나/, /다/, /마/, /바/)를 발화하는 모습과 소리를 디지털 캠코더로 녹화한 것을 실험자극제작에 사용하였다. 이 음절들의 발화소리는 비교적 조용한 방에서 따로 녹음되어 녹화된 장면이 더빙되었다. 녹화된 장면을 편집하고 따로 녹음된 소리를 더빙하고 동기화(synchronization)하여 실험자극으로 사용될 동영상 제작과정에 Adobe Premiere 6.02가 사용되었다. 30개의 동영상이 파일로 제작되어 실험에 사용되었는데, 이중 5개의 발화모습

(동영상)과 5개의 발화소리를 모두 조합한 동영상 자극이 25개이고 5개의 발화소리와 얼굴없는 공백화면이 조합되어 시각정보가 없이 청각정보만 있는 5개의 동영상도 포함되어있다. 이 동영상의 해상도는 640x480 픽셀이고 소리는 44khz로 표집되었다. 1개의 동영상의 길이는 최초의 준비신호를 포함한 2초간의 백색공백화면과 다음의 1초간의 흑색공백화면, 이어지는 3초간의 얼굴모습과 마지막 6초간의 흑색공백화면으로 구성되어 모두 12초였다. 이 동영상은 LCD projector를 통해 제시되었고 소리가 제시되는 스피커는 영상이 제시되는 스크린과 가깝게 위치시켰다.

도구. 자극제작을 위한 녹화에는 JVC GR-DV3000KR 디지털 캠코더가 사용되었고 자극녹음에는 PC에 외장형으로 연결된 Creative사의 외장형 사운드카드인 Extigy와 콘덴서 마이크가 사용되었다. 자극제시통제는 IBM x20 노트북을, 영상제시에는 신도리코 SPR-2200 LCD projector를, 사운드제시에는 100w급의 PC 스피커를 사용하였다.

절차. 실험은 한번에 4, 5명씩 집단으로 실시하였다. 실험가자들을 스크린으로부터 1.5m 정도 떨어져 있는 책상 앞에 앉게 하고 자극이 정면 스크린에 제시되면 영상자극을 바라보면서 스크린 바로 밑에 위치한 스피커에서 나는 소리를 듣고 순서대로 응답지에 들은 음절이 무엇인지를 체크하도록 하였다. 이 때 반드시 스크린 영상을 보면서 음절소리를 듣도록 강조하였다. 이런 식으로 연습시행 10회와 실제시행 30시행을 실시하였다. 자극제시는 무선으로 하였다. 전체 1회 실험에 걸리는 시간은 약 30분이었다.

표1. 각 청각-시각 조건에서 혼동행렬(confusion matrices). 숫자는 반응의 백분율(%)이고, 음영처리된 칸은 McGurk 효과가 나타날 수 있는 청각-시각 불일치 자극 조건이다.

(a)시각: 가

반응

	가	나	다	마	바
가	100				
나		100			
다			100		
마		21		79	
바			52		48

(b)시각: 나

반응

	가	나	다	마	바
가	100				
나		100			
다			100		
마		35		65	
바			16		13

(c)시각: 다

반응

	가	나	다	마	바
가	100				
나		100			
다			100		
마		31		69	
바			62		38

(d)시각: 마

반응

	가	나	다	마	바
가	100				
나		59		41	
다			100		
마				100	
바					100

(e)시각: 바

반응

	가	나	다	마	바
가	100				
나		79		21	
다			100		
마				100	
바					100

표2. 청각적으로는 양순음이고 시각적으로는 비양순음인 자극과 그 반대로 청각적으로는 비양순음이고 시각적으로는 양순음인 불일치 자극들에 대한 분석결과.

반응

	가	나	다	마	바
가	100				
나		69		31	
다			100		
마		29		71	
바			56		47

결과 및 논의

표3. 한국인, 일본인, 중국인, 미국인간의 McGurk 효과 비교

구분	한국인	일본인	중국인	미국인
청각: 양순 시각: 비양순	42.5%	25%	35%	84%
청각: 비양순 시각: 양순	10.3%	5%	9%	13%
평균	26.4%	15%	22%	48.5%

실험참가자들은 소리만 제시된 조건에서는 그 소리로 100% 정확하게 반응하였다. 29명의 반응자료를 종합하여 각 자극조건별로 결과를 정리한 것이 표1이고 불일치 자극 전체를 종합한 결과가 표2이다. 표1, 2를 보면 McGurk 효과가 /가/, /다/에서는 나타나지 않지만 /나/, /마/, /바/에서는 나타나는 것을 볼 수 있다. 이 결과는 McGurk 효과가 문화 보편적으로 나타나는 현상임을 보여준다. 표2에서 청각적으로는 양순음이고 시각적으로는 비양순음인 경우 효과가 42.5%로서 청각적으로는 비양순음이고 시각적으로는 양순음인 경우 10.3%에 비해 더 크게 나타남을 볼 수 있다. 이러한 양상은 일본인, 중국인, 미국인의 결과에서도 마찬가지이다. 두 경우를 평균하면 26.4%로 전체적인 평균 McGurk 효과가 나타남을 보여준다.

표3에 볼 수 있듯이 전체 평균적인 McGurk 효과를 한국인, 일본인, 중국인, 미국인간의 비교를 해보았을 때 한국인은 동일한 동양문화권에 속하는 일본인과 중국인과 비슷한 결과를 보이나 다른 문화권인 미국인과 비교해보았을 때 상대적으로 매우 약하게 나타남을 볼 수 있다. 이러한 결과는 일본인의 경우 미국인의 경우보다 그 효과가 매우 약하게 나타나는 현상을 설명하기

위해 Sekiyama(1997)가 제안한 안면회피 가설을 지지하는 결과이다.

본 실험은 방법적으로 덜 통제된 집단실험을 실시했다는 점에서 제한점을 가지고 있으나 한국인에게 나타나는 McGurk 효과의 양상을 어느 정도 반영해 주고 있다. 그러나 향후 보다 정밀하게 통제된 상황에서 여러 가지 변수를 조작하여 실험한다면 McGurk 효과가 어떻게 발생하는지 구체적인 처리과정(process)을 밝힐 수 있을 것이다.

참고문헌

Binie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech & Hearing Research, 17*, 619-630.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics, 59(1)*, 73-80.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese of subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America, 90*, 1797-1805.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics, 21*, 427-444.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.