

音声言語コミュニケーションに果す読唇情報の役割

Role of lip-read information in verbal communication

代表研究者 金沢大学文学部助手

Assist. Prof., Dept. of Phy., Kanazawa University
Kaoru SEKİYAMA

積山 薫

Some cultural difference in the effects of visual (lip-read) information on face-to-face speech perception was observed. When presented a video recording of a talking face in which audio signals were dubbed onto video signals for incongruent syllables, robust visual effects ('McGurk effect') have been reported in English speaking cultures. The present experiments, however, showed that Japanese listeners were much less influenced by visual cues than Americans. This was examined by testing 10 Japanese (Exp. 1) and 10 American (Exp. 2) subjects with 10 Japanese syllables, and 10 American subjects with 10 English syllables (Exp. 3). While Americans showed considerable visual influence for both Japanese and English syllables, Japanese listeners heard the syllables correctly with little visual influence.

研究目的

対面事態の音声言語コミュニケーションにおいて、人間は単に音声を耳で聞くだけでなく、目でも読み取っている（読唇）。自然なコミュニケーション事態においては、相手の顔が見えない場合よりも見える場合の方が、すなわち聴覚だけでなく視覚も利用可能な場合の方が、音声の聞き取り成績が良いことが知られている。一般に読唇情報は「聴覚情報の聞き取りをより正確にする」プラスの効果をもつ。

この読唇情報の効果を逆の形で示したのが、マガーク効果 (McGurk & MacDonald, 1976) と呼ばれる実験場面における現象である。この現象は、ビデオの吹き替えによって音声と映像を矛盾させた時に視覚が音声知覚に及ぼす妨害効果を示す。例えば、音声 [ba]+映像 [ga] で “da” と言っているように聞える。この例では、映像が提供する調音位置の情報 ([g]: “唇より奥”) が無視できないものであるため、それが聴覚的な調音位置情報 ([b]: “唇”) にとって代わると考えられる。その結果知覚される “da” は “唇より奥” で調音するという視覚情報と矛盾しないという条件を満たす

音のうち、聴覚が伝達する [ba] に音響的に似た音である。この現象では、矛盾した視覚情報に音声知覚が歪められると言え、「聴覚情報とは異なる音として聞える」という歪みの形で読唇情報の効果が示されている。

マガーク効果は、McGurk & MacDonald (1976) が初めて報告して以来、英語圏で多くの追試がなされ、頑健な効果であるとされてきた (Dodd, 1977; Green & Kuhl, 1989, 1991; MacDonald & McGurk, 1978; Massaro, 1983; Summerfield, 1987)。しかし、英語圏のみでの報告であったため、この現象が特定の言語ないし文化に依存するかどうかという観点から検討されたことはなかった。筆者らの先行研究では、言語に依存してマガーク効果の出方が変わることが示唆されていた。そこで、本研究の目的は、音声知覚における読唇依存性が文化ないし言語に左右される可能性を探ることであった。

研究経過

筆者らは、マガーク効果を日本語の単音節について日本人の被験者でテストしたところ、従来言われていたよりもずっと小さなマガーク効果しか

得られなかった (Sekiyama & Tohkura, 1991)。そこで、この読唇依存性の低さが日本語の刺激特性のためか日本人被験者の知覚特性のためかを調べるため、同一の日本語刺激・実験手続で日本人被験者 (実験 1) とアメリカ人被験者 (実験 2) の比較を行うことにした。また念のため、日本語刺激と同様の手順で作成した英語刺激をアメリカ人に提示する実験 (実験 3) も行い、英語圏で従来報告されてきたタイプのマガーク効果の追試とした。以下の方法で実験を行った。

方 法

被験者: 20~30 歳の、日本語を母国語とする人 10 名 (実験 1)、アメリカ英語を母国語とする人 (実験 2, 実験 3, 各 10 名)。聴力、視力は正常であった。

刺激: 材料は日本語および英語の単音節 10 個 (ba, da, ga, pa, ta, ka, ma, na, ra, wa)。日本語は日本人 1 名の、英語はアメリカ人 1 名の女性話者が発音した。その顔をビデオ録画し、それらの映像と別に録音した音声とを組合せて編集することにより、 $10 \times 10 = 100$ 個の視聴覚刺激を作成した。

手続: 100 個の視聴覚刺激をランダムに提示し、何と聞えるかを自由筆記で報告させた。視覚を無視することのないよう、見えと聞えに矛盾を感じるかどうかのチェックも求めた。

実験条件: 以上の視聴覚課題 (AV) の後に、映像なしで音声のみ聞き取る聴覚 (A) 課題も実施された。両課題は、雑音 (n) の付加された条件と付加されない条件で実施した。被験者は 2~4 日間

ですべての条件に参加し、各条件 6 回の繰返しがあった。

結 果

雑音の付加されない条件の実験結果の 1 部を、映像が唇音 (調音位置が唇: p, b, m, (w) がこれに当たる) であるか否かに分けて示す。表 1~3 のコンフュージョン・マトリクス中の数値は、各視聴覚刺激に対する当該反応の率 (%、 $n=60$) である。左端のセルの括弧内には聴覚課題の正答率を示す。マガーク効果 (視覚の影響) が生じれば唇音は非唇音として、非唇音は唇音として知覚されるはずであり、視覚の影響が無ければ聴覚課題の正答率と同じ数値が対角線のセルに並ぶはずである。

Table 1(A), (B) のように、日本人は日本語刺激に対して弱いマガーク効果しか示さなかった。ここでは、聴覚刺激 [pa] (“ta” 反応 33%), [wa] (“ra” 反応 20%), [ta] (“pa” 反応 17%) で視覚の影響が目につくのみであり、その他の七つの聴覚刺激に対してはほとんど聴覚的に正しい反応がなされている。

これに比べ、同一の日本語刺激に対してアメリカ人は強いマガーク効果を示した (表 2)。日本人とアメリカ人との最も明白な差は、唇音の聴覚刺激 [ba] が非唇音の視覚刺激 [ga] と組合せられたときにみられる。非唇音の視覚刺激に影響されて、聴覚刺激に [ba] に対しアメリカ人が 80% もの “da” 反応を生じ (Table 2(A)), マガーク効果が日本語でも起こることを示したのに対し、日本人の反応は 100% “ba” (聴覚的正答) であり (Table

Table 1. Stimulus-response confusions (%) in Exp. 1 (Japanese listeners, Japanese syllables). Asterisks indicate significant effects of vision (** $p < .001$, * $p < .01$, and * $p < .05$).

(A) vision = g (non-labial)

		response									
		b	p	m	w	d	t	n	r	g	k
audition	b (100)	100									
	p (98)		67				33				
	m (100)			96			4				
	w (95)				80				20		
	d (100)					100					
	t (95)						100				
	n (100)							100			
	r (100)								100		
	g (100)									100	
	k (100)										100

(B) vision = b (labial)

		response									
		b	p	m	w	d	t	n	r	g	k
audition	b (100)	100									
	p (98)		100								
	m (100)			100							
	w (95)				91				9		
	d (100)					100					
	t (95)						83				
	n (100)							100			
	r (100)								100		
	g (100)									100	
	k (100)										100

1(A)), 視覚の影響を全く受けなかった。同様の傾向が他の聴覚刺激についてもみられる。唇音の聴覚刺激が視覚刺激 [ga] と組合せられた時、および非唇音の聴覚刺激が視覚刺激 [ba] と組合せられた時に生じるマガーク効果の頻度は、アメリカ人対日本人で次のようであった。聴覚刺激 [pa]: 83% 対 33%, [ma]: 66% 対 4%, [wa]: 100% 対 20%, [da]: 40% 対 0% [ta]: 78% 対 17%, [na]: 52% 対 0%, [ga]: 12% 対 0%, [ka] 18% 対 0% これらの各々において、マガーク効果の頻度はアメリカ人と日本人とで統計的に有意な差があった。ここでは視覚刺激が [ga] と [ba] の場合の結果のみをあげたが、他の視覚刺激についても同様の結果であり、視覚刺激が唇音の場合は [ba] と、非唇音の場合は [ga] と同様の結果を示した。

実験 3 は、アメリカ人の読唇依存性が、今回の実験手続において外国語だけでなく母国語についてもみられるかどうかを検討する目的で行われた。Table 3 のように、アメリカ人は英語に対し

ても、視覚 [ga] と聴覚 [ba] で非唇音反応が 100% といった強いマガーク効果を示し、従来英語で報告されてきた現象が、今回の実験手続でも再現された。

なお、全体として、日本人は聴覚課題において正答が 100% を下回る音についてのみ視覚課題でマガーク効果がみられ、聴覚課題で 100% 正しく聴き取られた音ではマガーク効果が生じなかったのに対し、アメリカ人では聴覚課題で 100% 正しく聞き取られた音でもマガーク効果は生じた。

考察

以上三つの実験から、マガーク効果は日本語の刺激に対して起こりにくいのではなく、日本人の聞き手の場合に起こりにくいことが確認された。聞き手がアメリカ人の場合には、刺激が日本語でも英語でも容易に視覚に影響される。日本人とアメリカ人との間には、読唇依存性における顕著な知覚的差異がある。アメリカ人は音声言語コミュニケーションにおいて目に多くを負っているのに

Table 2. Stimulus-response confusions (%) in Exp. 2 (American listeners, Japanese syllables). Asterisks indicate significant effects of vision (***) $p < .001$, ** $p < .01$, and * $p < .05$.

		response											
		b	p	m	w	d	t	n	l	q	k	r	others
audition	b (100)	20				80							
	p (98)		17				83						
	m (100)			32			66						mn2
	w (38)			0				84	2		2		q5 d3 g2 h2
	d (90)					100							
	t (83)						100						
	n (96)							100					
	r (51)								51		13	gr18 dt10 nr8	
	g (98)									95			bg3 q2
	k (100)										100		

		response											
		b	p	m	w	d	t	n	l	q	k	r	others
audition	b (100)	100											
	p (98)		98				2						
	m (100)			100									
	w (38)	2		13	8				8			5	bl43 l m8 b5 a3 q2 nr2
	d (80)	40				35							bd2
	t (83)	78					17						ps
	n (96)			52				25					mn23
	r (51)	2		2					10			8	bl1 br10 ml10 n7 bd7 a2
	g (98)	12								76			ba12
	k (100)	18									74		bs pk3

Table 3. Stimulus-response confusions (%) in Exp. 3 (American listeners, English syllables). Asterisks indicate significant effects of vision (***) $p < .001$, ** $p < .01$, and * $p < .05$.

		response											
		b	p	m	w	d	t	n	r	q	k	l	others
audition	b (87)	0				13	2				55		lh20 s10
	p (88)		7				66				20		ps k2
	m (100)			22			32				46		
	w (98)				3						97		
	d (98)					98				2			
	t (100)						100						
	n (100)							100					
	r (100)								37			63	
	g (100)									100			
	k (100)										100		

		response											
		b	p	m	w	d	t	n	r	q	k	l	others
audition	b (87)	100											
	p (88)		100										
	m (100)			100									
	w (98)				13	38						37	bl12
	d (98)					100							
	t (100)					53			47				
	n (100)						13			74			mn13
	r (100)									90		2	br8
	g (100)										100		
	k (100)											2	98

対し、日本人はもっぱら耳で聞いているらしい。この結果は、これまで言語依存性をぬきに論じられてきた読唇情報の効果について、一石を投じることになろう。読唇情報が音声知覚に対していかに効果をもつかは、その人が育ってきた言語的・文化的背景にも影響されるのである。

日本人がアメリカ人よりも視覚に影響されにくいこと理由は今のところ特定することはできないが、日本人が相手の顔を見ないでしゃべる習慣をもつことは、一つの可能な説明であろう。また、日本語の音韻のレパートリーが英語より少ないことも理由として考えられる。日本語は、母音においても子音においても、その構成要素が英語より少ない。このため、読唇情報をわざわざ利用しなくても、それらを識別することが比較的容易であるといえる。

今後の課題と発展

今後の課題は、上記の可能な説明のどれが、今回見いだされた文化差の実際の原因であるかを検討することである。その一つの方法は、音韻構造が異なる種々の言語について同様の実験を重ねることであろう。それが、今後の研究の展開に重要であると考えられる。

引用文献

- Dodd, B. (1977): *Perception*, 6, 31~40.
Green, K. P. and Kuhl, P. K. (1989): *Perception & Psychophysics*, 45, 34~42.
Green K. P. and Kuhl, P. K. (1991): *Journal of Experimental Psychology: Human Perception and Performance*, 17, 278~288.
MacDonald, J. and McGurk, H. (1978): *Perception & Psychophysics*, 24, 253~257.
Massaro, D. W. and Cohen, M. M. (1983): *Journal of Experimental Psychology: Human Perception and Performance*, 9, 753~771.
McGurk, H. and MacDonald, J. (1976): *Nature*, 264, 746~748.
Sekiyama, K. and Tohkura, Y. (1991): *Journal of Acoustical Society of America*, 90, 1797~1805.
Summerfield, Q. (1987): In "Hearing by Eye," ed. by R. Campbell and B. Dodd, London: Erlbaum, pp. 3~51.

発表論文

(1) 論文

- 1) Cultural differences in dependence on visual cues in speech perception, *Journal of Phonetics* に投稿中.

(2) 学会発表

- 2) 音声知覚における読唇依存性の日米比較, 日本心理学会第55回大会(1991年10月).
- 3) Japanese listeners are less influenced by vision than Americans in speech perception, The 122nd Meeting of the Acoustical Society of America (1991年11月).