【非公開】

No. 19

TR－A－0041

単音節の受聴における読唇情報の役割

積山　薫　　東倉　洋一

Kaoru Sekiyama　Yoh'ichi Tohkura

1988. 12. 21

ＡＴＲ視聴覚機構研究所

単音節の受聴における読唇情報の役割

Effects of lip-read information on auditory perception of syllables

積山　薫　*Kaoru Sekiyama*
東倉洋一　*Yoh'ichi Tohkura*

ＡＴＲ視聴覚機構研究所

ATR Auditory and Visual Perception Research Laboratories

Abstract

The McGurk effect is a phenomenon that demonstrates a perceptual fusion between auditory and visual(lip-read) information in speech perception under visual-auditory discrepancy condition (using dubbed video tapes). This paper examined the relation between the "McGurk effects" and the intelligibility of auditory stimuli. A female narrator's speech was video taped for ten Japanese syllables(/ba/,/pa/, /ma/, /wa/, /da/, /ta/, /na/, /ra/, /ga/, /ka/). The video and audio signals for these ten syllables were combined, resulting in 100 audio-visual stimuli. These stimuli were presented to ten subjects who were required to identify the stimuli as heard speech in both noisy and noise-free conditions. For both conditions, the intelligibility of the auditory stimuli was measured, presenting the auditory stimuli alone. In the noise-free condition, the McGurk effect was small and found only in conditions in which the intelligibility of the auditory stimuli was not 100%. In the noisy condition, the McGurk effect was very strong and widespread. These results suggest that incomplete intelligibility of auditory stimuli is necessary for the McGurk effect.

1．はじめに

対面事態での音声知覚において、話者の唇などの動きがもたらす視覚情報（lip-read information；以下、読唇情報とよぶ）が音声知覚に影響を及ぼすことが知られている。しばしば指摘されるのは、騒音の中で音声情報を受容するさいの読唇情報の有用性である。読唇情報を伴う音声情報は聴覚のみの情報に比べ、S/N比が低くても正しく知覚されやすく、雑音に頑健である（e.g.,Erber,1975, Fig. 1-a）。我々の音声知覚を成立させているものは、聴覚情報だけではないと言える。最近では、人工内耳を埋め込まれた患者にとって、読唇情報を加えてやると音声知覚が容易になることが注目されている（福田、城間、船坂, 1988, Fig. 1-b）。

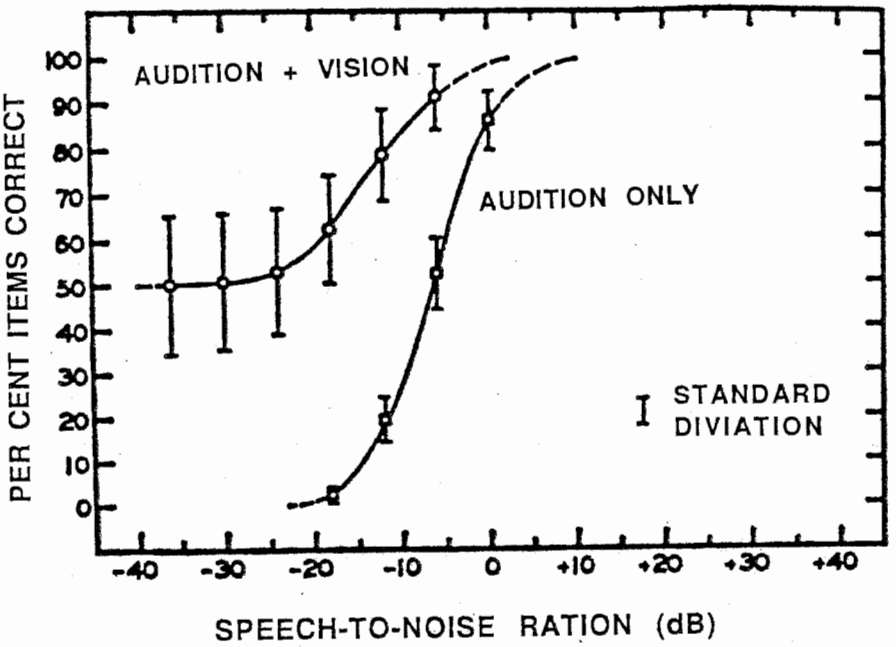

Fig.1-a Auditory and auditory-visual recognition of 250 spondaic words in broad-band noise by five adults with normal hearing. (From Erber, 1975)

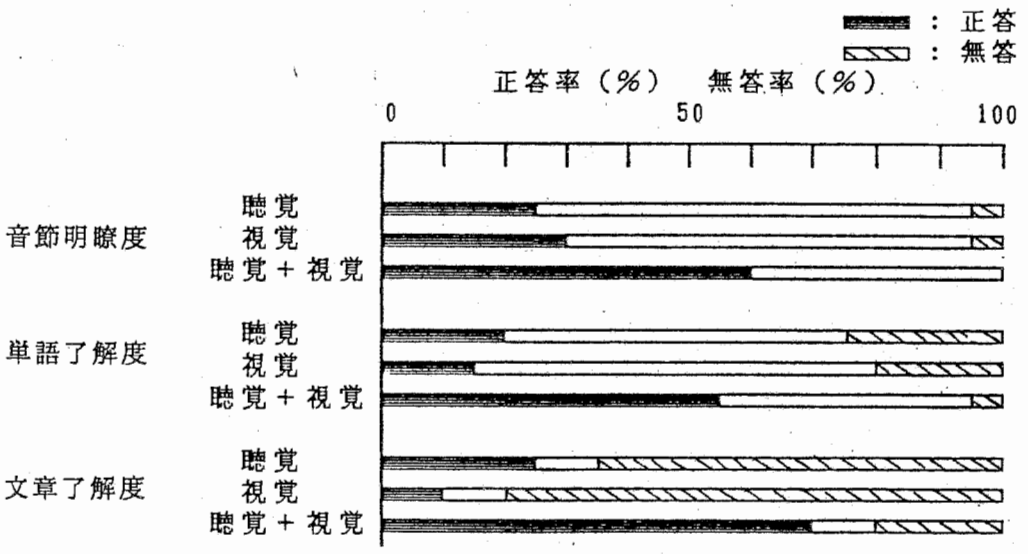

Fig.1-b　人工内耳埋め込み患者の音声知覚　　福田、城間、船坂（1988）より

聴覚のみ、視覚のみ、聴覚＋視覚での比較

McGurk effect　読唇情報の有用性を言わば逆手に取ったのが McGurk効果である（McGurk & McDonald,1976; McDonald & McGurk, 1978）。これはビデオを用い、例えばビデオ音声／ba／に対して話者の映像／ga／をダビングしたものを提示すると（Fig. 2参照）、聴覚情報と視覚情報が融合して／da／などとして知覚されるという現象である。彼らによると、この現象は音声が唇音で映像が非唇音であるときに顕著で、その逆の組合せでは起こりにくいと言う（Table 1）。この現象は、次のように説明されている。すなわち、知覚される音は、聴覚から調音様式（この例では破裂音）と有声性（ここでは有声音）の情報を、視覚から調音位置（この例では唇が閉じない音）の情報を取りこんだ音であるという説明である。

[McGurk & McDonald, 1976]

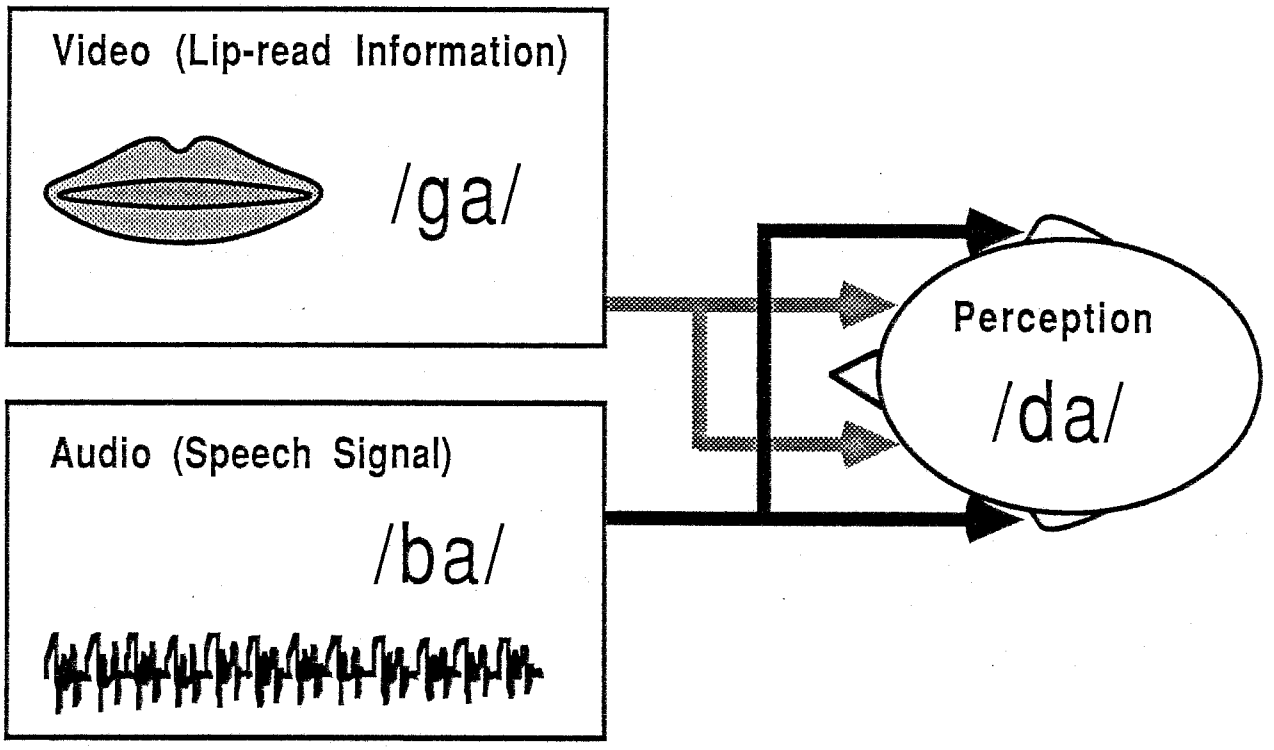

Fig. 2 The McGurk Effect

Table 1 The "original McGUrk effect".

(from McGurk & McDonald, 1976, Nature, p747)

1-1 Stimulus conditions and definition of response categories

Stimuli		Response Categories				
auditory component	visual component	Auditory	Visual	Fused	Combi-nation	Others
ba-ba	ga-ga	ba-ba	ga-ga	da-da		
ga-ga	ba-ba	ga-ga	ba-ba	da-da	gabga bagba baga gaba	dabda gagla etc.
pa-pa	ka-ka	pa-pa	ka-ka	ta-ta		tapa pta kafta etc.
ka-ka	pa-pa	ka-ka	pa-pa		kapka pakpa paka kapa	kat kafa kakpat etc.

1-2 Percentage of responses in each category

Stimuli		Subjects	Response Categories				
auditory	visual		Auditory	Visual	Fused	Combi	Others
ba-ba	ga-ga	3-5yr	19	0	81	0	0
		7-8yr	36	0	64	0	0
		18-40yr	2	0	98	0	0
ga-ga	ba-ba	3-5yr	57	10	0	19	14
		7-8yr	36	21	11	32	0
		18-40yr	11	31	0	54	4
pa-pa	ka-ka	3-5yr	24	0	52	0	24
		7-8yr	50	0	50	0	0
		18-40yr	6	7	81	0	6
ka-ka	pa-pa	3-5yr	62	9	0	5	24
		7-8yr	68	0	0	32	0
		18-40yr	13	37	0	44	6

Function of lip-read information　人間が読唇情報を調音の位置に関する情報という形で処理していることは、読唇の研究でも指摘されるところである（Fukuda & Hiki, 1982; Summerfield, 1987）。実のところ、読唇情報から母音は弁別できても、子音については非常に大まかに調音位置を弁別する情報しか得ることができないので、読唇は非常に難しい。しかし、かなり明確に弁別できる音節のクラスターが存在することも事実である。筆者らは、日本語 100音節を発話している話者の映像をビデオで提示し、普通成人に読唇させる実験を行った（積山，城，梅田，1988）。その結果を子音に関して多次元尺度法で分析すると、我々が抽出し得る視覚的な特徴の重要なものは、唇の動きの際立ちや唇がすぼまるかどうかという特徴であることが示唆された。そして26個の子音は、大きく分けて３つのクラスターを形成しており、それらは／w／、／m，b，p／、およびそれ以外であった。より詳細に見れば、もっと多くのクラスターに分けることはできるが、誰でも確実に区別できるものはこの３つであろう。このように３つに分けると、結局、特別な訓練経験の無い者にとっての読唇情報は、調音位置がはっきり見える唇であるか、あまりよく見えない口の中であるかを区別するものと言える。

読唇情報のどのような側面が音声知覚に働くのかを検討した Summerfield（1979）の実験によれば、連続的な散文を背景妨害として提示される文章の報告において、ビデオで話者の顔全体、唇、唇の中央と端の４点を見せると、成績がそれぞれ43%、31%、 8%上昇した。しかし、オシロスコープの光点の動きで音節のタイミングを見せるだけでは、全く上昇しなかった。このことからも、読唇情報の音声知覚への寄与は、時間情報よりも調音運動に伴う空間情報にあることが示唆される。

McGurk effect in words　単語の場合でもMcGurk効果は見出される。 Dodd（1977）は白色雑音を聴かせながら、例えば聴覚で “tough”、視覚で“hole”と提示すると“towel” として知覚されるというように、どちらかが無視されるのではなく、融合が生じることを報告している。 一方、Easton & Basala（1982）の単語を用いた実験では（特に雑音の提示なし）、音声と映像の矛盾状態で「聴覚情報に注意を払うよう」教示された統制群の被験者は、99%正しく聴覚提示語を同定し、McGurk効果は見出されなかった。逆に、「視覚情報に注意を払うよう」教示された実験群はほとんど視覚提示語を報告できず、しばしば聴覚提示語を報告し、McGurk効果に相当する反応（視覚 “next”、聴覚“chime”で知覚“chest”など）は１割ほどし

かなかったという。したがって、単語のように文脈情報のある材料では、明瞭な音であれば視覚情報に影響されることは少ないのかもしれない。

Noise and McGurk effect　　雑音を同時提示するとMcGurk効果が生じ（Dodd, 1977）、提示しないと生じない（Easton & Basala, 1982）のだとすれば、McGurk効果は音声の明瞭性が完全でない状況で生じるのではないかと考えられる。McGurkらの実験（McGurk & McDonald,1976）では、聴覚提示のみの統制条件で測定された音声の明瞭度は、幼児、児童、大人の被験者においてそれぞれ91%、97%、99%であった。これは、用いられた刺激セット（／ba-ba／／ga-ga／／pa-pa／／ka-ka／の４つ）の大きさから言って、それほど高い明瞭度とは言えない。我々は予備的に14個の日本語単音節セット（／ba／／da／／ga／／pa／／ta／／ka／／ma／／na／／ra／／wa／／ha／／ya／／sa／／za／）で実験を行ない、McGurk効果を見出した（積山、東倉、1988, Fig. 3）。このとき、聴覚提示のみの条件で得られた平均明瞭度は80%と低かった（Table 2)。明瞭度が低かった理由として、話者に「誇張せ

Table 2　Inteligibility of auditory stimuli in the experiment of Sekiyama & Tohkura (1988)

response (%)

auditory stimulus	z	s	d	t	g	k	h	y	w	n	r	b	p	m	other
z	95				3							3			
s		92		8											
d			90								5				5
t			10	85							3				3
g					100										
k						100									
h						3	97								
y								100							
w									100						
n					56			3		5	13				23
r			8		43						15				34
b			1		18						13	56	1		10
p		1	1	3								9	81		3
m														100	

ず自然に発話するよう」教示したこと、また、音声と映像を同時に収録し、刺激作成の段階で編集を繰り返したため、最終的に音声の S/N比がかなり低下したことが考えられる。

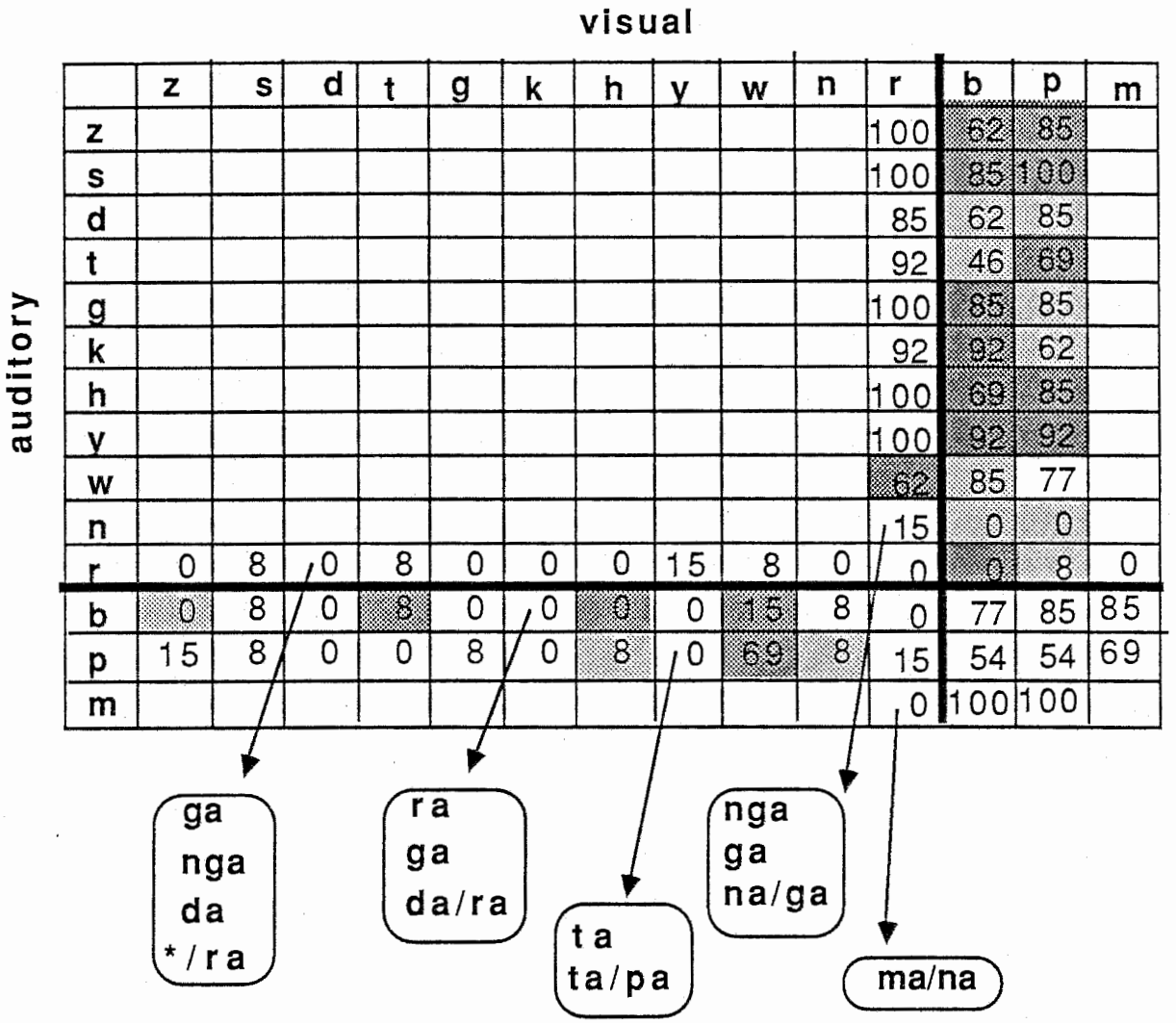

Fig.3 % of correctly heard response (numeric) and incongruently perceived response (shade)

(Sekiyama & Tohkura, 1988)

Incongruity □ *0-30%* ▒ *45-55%* ■ *62-85%*

以上のことから、McGurk効果が報告された実験ではいずれも音声の明瞭度が完全でなかったものと考えられる（Table 3）。そこで、本稿では付加雑音を用いて明瞭度を操作し、雑音のMcGurk効果生起に及ぼす影響、McGurk効果と明瞭度との関係について検討する。

Table 3　McGurk効果の必要条件は雑音の存在か？

	材料	McGurk効果	条件
McGurk & McDonald (1976)	単音節 4個	あり	明瞭度 幼児91% 児童97% 大人99%
積山，東倉 (1988)	単音節 14個	あり	明瞭度 80%
Dodd (1977)	単語	あり	白色雑音提示
Basala & Easton (1982)	単語	なし	雑音なし

2．実験の方法

本実験では、10個の音節について音声（聴覚情報）と映像（視覚情報）を組み合わせた刺激を作成し、これを雑音のない明瞭な条件と雑音のある条件で被験者に提示し、どんな音に知覚されるかを検討した。

【刺激材料】日本語単音節１０個、／ba／／da／／ga／／pa／／ta／／ka／／ma／／na／／ra／／wa／。

【刺激の作成】話者は女性アナウンサー１名。発話は１音節につき１回ずつ。映像は、正面から放送用カメラで首から上を撮影し、 1/2インチＶＴＲに収録した。音声は、「はっきりしゃべるよう」教示し、音声品質劣化を伴わないようＶＴＲの音声録音チャンネルとは別にディジタル録音した。録音条件はサンプリング周波数20 kHz 、16ビット。各音節の持続時間は約250msec であった。また、刺激提示用として各音節についてシグナル音を含むFig. 4のような刺激を作成し、これをＶＴＲの音声チャンネルにダビング録音して提示した。ダビングの際の音声と映像の同期は、実験者にとって最も良く同期して知覚されるタイミングを映像フレーム単位（33.3 msec）で決定した。10個の音節について、対応する音声と映像（／ba／の音声なら／ba／の映像）との同期タイミングを決定したのち、同じタイミングで音声と映像のすべての組み合せを作成した（Fig.6）。 組み合せ数は、10 × 10＝100 である。

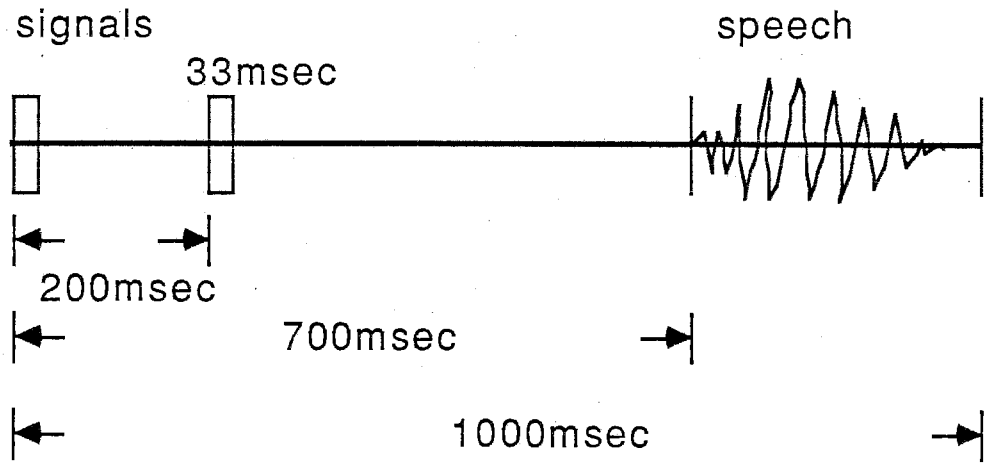

Fig.4 Time course of auditory stimuli

【刺激の提示】Fig. 5に刺激提示の概略を示す。ＶＴＲ（SONY BVW-40）の出力を、映像は20インチ・カラーモニタ（Shibasoku CMM20-11）に提示した。話者はほぼ実物大。音声はスピーカより提示した。スピーカはモニタの両側に取り付けられた。被験者の観察距離は1m。刺激は Fig. 6 のように７秒に１音節のペースでランダム順に提示された。

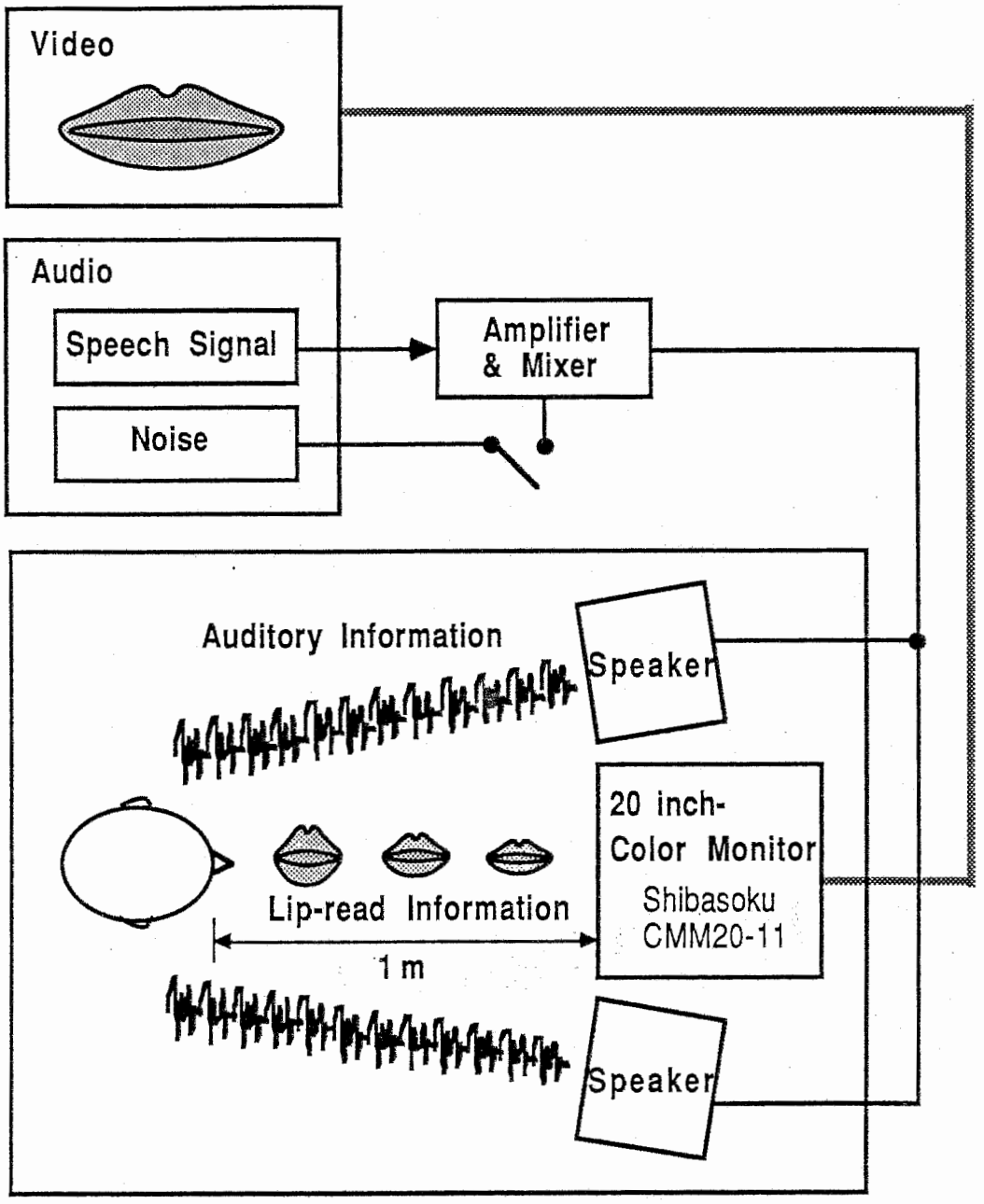

Fig.5 Blockdiagram of Lip-read and Auditory Information Perception Experiment

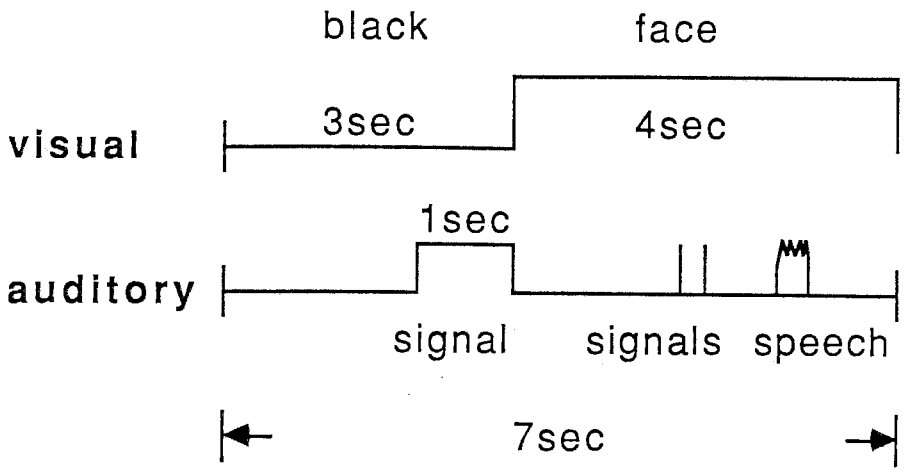

Fig.6　Time course of audio-visual stimuli

【実験条件】実験条件はすべて実験者内変数とした。実験は４つのフェーズからなる。雑音の有無で２条件、これに統制条件として、音声のみの提示か音声＋映像の同時提示かの２条件が加わり、計４つの実験条件を選んだ。各被験者は、次の４つの条件において話者の発話音節として「聞こえた音」をそのまま報告した。

①雑音なしＡ条件　：音声のみ提示される。

②雑音なしＡＶ条件：音声と映像の同時提示。被験者は映像を見ながら音声を受聴する。

③雑音を伴うＡ条件：音声に連続的なガウス雑音を付加し聴覚提示する。音声と雑音の S/N比は被験者の受聴位置で 0dBとなるよう調整される。

④雑音を伴うＡＶ条件：上記③の条件に加えて映像を同時提示する。

各条件とも同一刺激について６回の繰り返しがあり、ＡＶ条件は 600試行が２日、Ａ条件は60試行が１日で実施された。

【教示】反応は自由反応。基本的には日本語のア段の音節であるが、聞こえたままを報告するよう強調した。複合音として聞き取れる場合、例えば、 /bda/のように通常の日本語には存在しない音として聞こえる可能性もあることを教示した。これは、McGurkらの実験でこのような複合反応が報告されているため。映像への注意を確実にするため、音声と映像の間に顕著な矛盾が感じられるときは、矛盾感を記録するよう求めた。反応は筆記でローマ字による。
【被験者】正常な聴力をもつ20才台の女性10名（平均年齢24才）。いずれも研究所職員。

3．実験の結果

結果は、映像の種類ごとに音声刺激ー反応マトリクス（以下、異聴表とよぶ）を作成し、各反応カテゴリーの反応数を集計した。複数回答のあった反応については、教示で「より確かなものから先に記入するよう」指示していたので、例えば回答が2つの場合には、始めの反応に 0.6、後の反応に 0.4の重みをかけるといった処理をした。これらの異聴表を実験条件ごとに検討する。

雑音がないとき

①雑音なしA条件： Table 4 に異聴表を示す。数値は%で、分母は60である。この条件は刺激音声のスピーカ受聴における通常の明瞭度の測定と等価である。Table 4 によると、大部分の音節は 100%の明瞭度を示しているが、 /pa/（98%）、/wa/（95%）、/ta/（95%）の3つの音節はわずかに明瞭性を欠いた。異聴は、/ta/と/pa/の間で双方向に起こり、/wa/は一方向的に/ra/に異聴される。

②雑音なしAV条件： Table 5 は、①で測定された明瞭度の高い音声に映像が加わった場合の結果である。映像の種類（音節）ごとに異聴表を作成した。

Table 5 の10個の異聴表を全体的に見ると、総じて正答率が高く、Table 4 と比較しても、視覚情報が加わったことによる影響はそれほど大きいとは言えない。し

かし、音節個別にみると、主として／ta／、／pa／、／wa／の受聴においてMcGurk効果が認められる。　まず、Table 5-a に示されるように、映像が唇音のときは、４つの異聴表に共通して刺激音／ta／から知覚／pa／への異聴が20%ほどみられる。映像が非唇音のとき、Table 5-b の６つの異聴表に共通して刺激音／pa／が／ta／に、刺激音／wa／が／ra／に異聴される率が約20%みられる。これらの刺激音／pa，ta，wa／は、①の雑音なしＡ条件で明瞭度が 100%を欠いた音声である。これに比較して、音声の明瞭度が完全な場合には映像の影響はほとんど認められないことから、McGurk効果が顕著に生じるのは音声の明瞭度が完全でない場合に限られることが示唆される。

Table 4. The intelligibility of the auditory stimuli for the noise-free condition(%).

response

audition	b	p	m	w	d	t	n	r	g	k	others
b	100										
p		98				2					
m			100								
w				95				5			
d					100						
t	5					95					
n							100				
r								100			
g									100		
k										100	

For the noise-free condition, the intelligibility of the auditory stimuli was high. Most of the stimuli were always identified correctly. Three syllables, however, show incomplete intelligibility: the rates of correct responses for /pa/, /ta/, and /wa/ were 98%, 95%, and 95%, respectively.
The auditory stimuli were presented by speakers throughout the experiment.
The values here indecate the rates to 60 observations (10 subjects x 6 repitions).

Table 5-a Confusion matrices for the sitmuli for which visual components were labials . Noise-free condition. Indicated in %.

vision = b

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	100										
p		100									
m			100								
w				91				9			
d					100						
t		17				83					
n							100				
r								100			
g									100		
k										100	

vision = p

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	98				2						
p	2	97				2					
m			100								
w				91				9			
d	1				99						
t	1	33				66					
n							100				
r								100			
g		1							99		
k		1							2	98	

vision = m

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	100										
p		100									
m			100								
w				90				10			
d					100						
t		22				78					
n			2				98				
r								100			
g									100		
k			2							98	h1

vision = w

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	98	2									
p		97				2				2	
m			100								
w				97				3			
d	1			1	97				2		
t		14				86					
n							100				
r								100			
g									100		
k		2		2						97	

Table 5-b Confusion matrices for the stimuli for which visual components were nonlabials . Noise-free condition.

vision = d

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	100										
p		76				24					
m			98				2				
w				81				19			
d					100						
t		2				98					
n							100				
r						2		98			
g						2			98		
k										100	

vision = t

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	100										
p		76				24					
m			93				7				
w				78				22			
d					100						
t						100					
n							100				
r								100			
g									100		
k										100	

vision = n

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	100										
p		78				22					
m			94				6				
w				82				18			
d					100						
t		2				98					
n							98	2			
r								100			
g								2	98		
									2	98	

vision = r

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	100										
p		72				28					
m			96				4				
w				77				23			
d					100						
t						100					
n							100				
r								100			
g								2	97	2	
k										100	

Table 5-b (2)

vision = g

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	100										
p		67				33					
m			96				4				
w				80				20			
d					100						
t						100					
n							100				
r								100			
g									100		
k										100	

vision = k

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	98				2						
p		78				22					
m			88				12				
w				68				32	1		
d					100						
t						100					
n							100				
r								100			
g									100		
k										100	

Under noise-free conditions, visual effects were weak. For visually presented labials, i.e., /b, p, m, w/ (Table 5-a), it was only with auditory /ta/s that the McGurk effectoccurred. Presented with visual labials, about 20% of auditory /ta/'s were perceived as /pa/. This response occurred 17% of the time for visual /ba/'s, 33% for visual /pa/'s, 22% for visual /ma/'s and 14% for visual /wa/'s.
Compared to Table 4 (auditory alone condition), these /pa/-responses are interpreted as involving visual effects as well as auditory errors.

For the visual non-labials, i.e., /d, t, n, r, g, k/ (Table 5-b), it was only with auditory /pa/'s and /wa/'s that the McGurk effec occurred. When auditory /pa/'s were combined with any visual non-labial, a reponse of /ta/ occurred to some extent. Such erroneous /ta/ responses were 24%, 24%, 22%, 28%, 33%, and 22% for visual /da/, /ta/, /na/, /ra/, /ga/, and /ka/ stimuli. When auditory /wa/'s were combined with any visual non-labial, a response of /ra/ occurred to some extent. Such erroneous /ra/-responses were 19%, 22%, 18%, 23%, 20%, and 32% for visual /da/, /ta/, /na/, /ra/, /ga/, and /ka/ stimuli. These erroneous responses are interpreted as involving both auditory errors and visual effects.

In both Table 5-a and Table 5-b, it seems that the McGurk effect occours only for auditory stimuli of incomplete intelligibility. Note that the auditory simuli in which the McGurk effect occurred, i.e., /ta, pa, wa/ stimuli, are the stimuli that are not completely intelligible as shown in Table 4.

雑音があるとき

③雑音を伴うＡ条件：　Table 6 に、雑音があるときの音声の異聴表を示す。また、Fig.7 に、雑音の有無による明瞭度は相違を音節別に比較図示する。雑音によって、一般に明瞭度は低下するが、このうち低下の著しい音節が／pa，ta，ra，ga，ka／、あまり低下しない音節が／ba，ma，wa，da，na／となる。低下の著しいものは、明瞭度が50～60％まで低下し、とくに／ga／は24％まで低下する。一方、雑音がないときに明瞭度が完全でなかった音節のうち、／wa／ではほとんど低下がみられない。

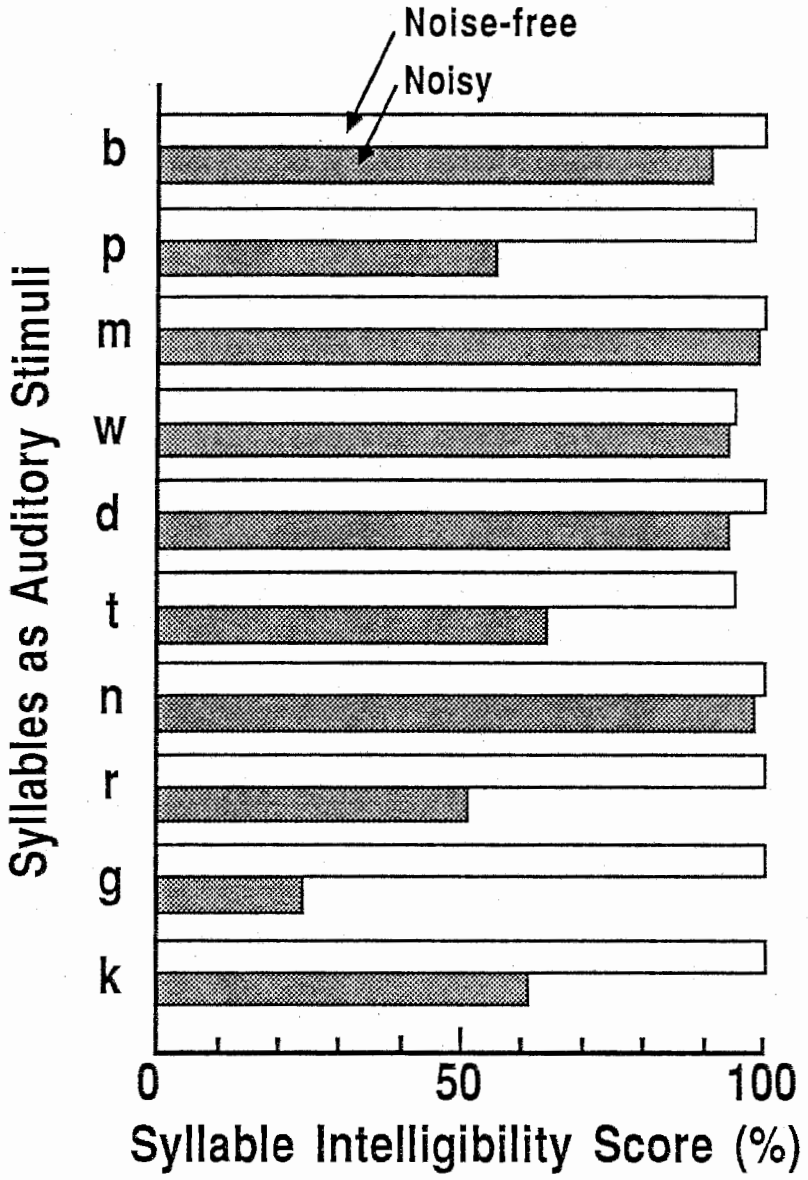

Fig.7　Comparison between noise-free and noisy speech intelligibility scores

主な異聴は、／pa／→／ta／37％、／ta／→／pa／27％、／ra／→／wa／33％、／ga／→／da／41％、／ga／→／ba／35％、／ka／→／ta／11％、／ka／→／pa／14％である。

ここで注意しておきたいのは、異聴表内のカテゴリーが調音位置の順に並べられていることである。すなわち、／ba，pa，ma，wa／は調音位置が唇、／da，ta，na，ra／は歯茎、／ga，ka／は口蓋にあり、調音位置で口の先から奥へとカテゴリーが並べられている。音だけで知覚する場合には、調音位置の視覚情報は被験者にはほとんど利用することができないので、Table 6 でも異聴のパターンに特別な偏りは見られない。つまり、唇音が非唇音になる異聴もその逆もある。しかし、読唇情報が加わると、次に述べるように、異聴は調音位置に関して生じる。

Table 6. The intelligibility of the auditory stimuli for the noisy condition (%).

response

audition	b	p	m	w	d	t	n	r	g	k	others
b	91				1				8		
p		56	1			37				6	
m			99				1				
w				94			1	4	1		
d	3				94				3		
t		27		1		64				8	
n			2				98				
r			1	33			4	51	2		y8
g	35				41				24		
k		14				11				61	h7 a7

With the addition of gaussian noise, the intelligibility of the auditory stimuli decreased compared to that for the noise-free condition. For five syllables, i.e., /pa/, /ta/, /ra/, /ga/, and /ka/, the intelligibility is poor: responses for these stimuli are only 20-60% correct. The other syllables preserve relatively high intelligibility: more than 90% of responses are correct . The S/N ratio was 0.dB.

④雑音を伴うＡＶ条件：　雑音があるとき、映像が加わると、　知覚される音声は読唇情報の影響を著しく受け、McGurk効果が顕著となる。Table 7 に雑音下における音声と映像の知覚結果を示す。まず、音声と映像が同一内容の場合には、従来から指摘されているように、読唇情報が聴覚情報の有用な補助となり、雑音の影響で低下した音声の明瞭度を回復するよう働くことが予想される。この様子をFig.8 に

示す。この図は、雑音を伴うＡＶ条件における各音節の明瞭度（音声と映像が同一内容の場合）をFig.7 に追加図示したものである。これによると、読唇情報の働く唇音において、総じて明瞭度の回復がみられ、とくに／pa／では56％から98％へと、著しい明瞭度向上を示す。また、これに次いで／ta／も64％から98％まで回復する。これは、調音位置（歯茎）の視覚情報が働き、雑音下での明瞭度低下の原因であった／pa／、／ka／への異聴（27 + 8％）が全く生じなかったためである。

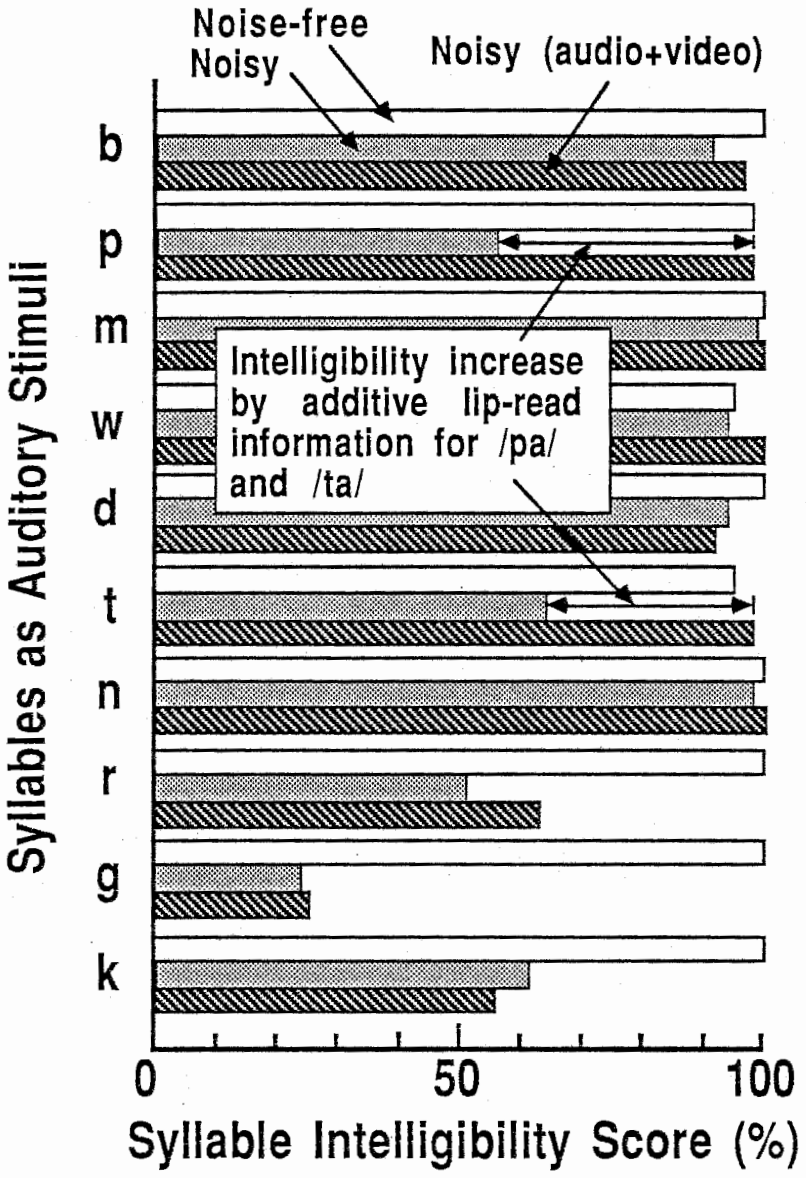

Fig.8 Speech intelligibility increase by additive visual (lip-read) information

Table 7-a は、映像が唇音の場合の結果である。異聴表中、反応／wa／と／da／の間に縦に引かれた太線は、知覚された音声の調音位置が唇か口の中（歯茎、口蓋）かの境を示している。この観点からみると、映像が唇音の場合、反応の主な部分が唇音の側にシフトしている。すなわち、読唇情報は音声の調音位置の知覚に対して極めて強い影響を与える。これに対し調音様式の知覚には、読唇情報の影響はほとんど認められない。実際には、／da，ga／→／ba／、／ta，ka／→／pa／、／na／→／ma／のように、それぞれ調音様式の等しい唇音に異聴される。とくに、／ta／、／ka／、／ga／の異聴は著しく90％を越える異聴率を示す。また、／ra／は／wa／または／ma／に知覚されている。

Table 7-b は、映像が非唇音の場合である。ここでは、上記とは逆に、読唇情報の影響で反応が非唇音の側に大きくシフトしている。とくに、／pa／は90％以上／ta／に異聴され、他の音節についても／ba／→／da，ga／、／ma／→／na／、／wa／→／ra，ga／といったように、それぞれ調音様式の等しい非唇音に異聴される。

ここで、Ａ条件における各音節の明瞭度と、これに映像が加わったＡＶ条件における異聴との関係をみてみよう。Fig.9 にこの関係が示される。図中、縦軸の異聴率は次のように定義された。まず、実験に用いた10音節を唇音／ba，pa，ma，wa／とこれ以外の非唇音に分類する。そして、音声が唇音で映像が非唇音、およびその逆の組み合せだけに着目して、各々の音節に関する異聴率を算出する。この異聴率は読唇情報の影響の強さ、いいかえれば、McGurk効果の強さを表す。Fig.9 によれば、読唇情報の影響による異聴の程度は、提示された音声の明瞭度（読唇情報がない場合の）に依存することを示す。すなわち、音声の明瞭度が 100％と完全な場合には、異聴は起こりにくいが、少しでも明瞭性を欠いた音声の知覚には、読唇情報の影響がただちに現れる。この影響は、音声の明瞭度95～99％で既に強く現れ、70％を下まわる条件では、完全に視覚優位の知覚が示される。

Table 7 より、さらに以下のことが指摘される。読唇情報の影響は、もともとMcGurkらが報告した条件（音声／pa／と映像／ka／で知覚／ta／、あるいは音声／ba／と映像／ga／で知覚／da／）だけにとどまらない。すなわち、McGurk効果は、音声と映像とが、唇音－非唇音、非唇音－唇音の全ての組み合せ場合に生じ得る。具体的には、雑音下での音声／pa／は、映像／ka／との組み合せによってのみ／ta／に知覚されるのではなく、／ka／以外のどの非唇音の映像と組み合わされた場合にも／ta／に知覚される。

Table 7-a Confusion matrices for the stimuli for which visual components were labials. Noisy condition. Indicated in %.

vision = b

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	97	3									h1
p		98				2					
m			100								
w	2		6	86				5			h2
d	60				39						bd2
t	2	94				3				1	
n			74			2	24				
r	8	1	27	45				19			y1
g	88				7				5		
k	1	94				2				2	bp2

vision = p

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	98								2		
p		98				2					
m			100								
w	2		7	77				11			y3
d	45				51			2	3		
t		98				2					
n		1	67				32				
r	9		29	34				23			y5
g	93				5				2		
k	2	98									

vision = m

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	96				2				2		h1
p		98				2					
m			100								
w	5	1	7	79				8			
d	64				34		2	1			
t		97				2				2	
n			84				14				h2
r	8		40	29			2	14			y5 wm2
g	93				4				2	2	
k		98				2					h1

vision = w

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	88			2	1	2			5		h3
p		93				3					h1 f2
m			100								
w				100							
d	33			2	61				3		z1
t		90		1		3				1	f4
n			35				65				
r			2	84				13			a1
g	72			1	17			1	9		
k		92		1		2				2	pk2

Table 7-b Confusion matrices for the stimuli for which visual components were nonlabials. Noisy condition.

vision = d

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	25				38	3			29		h4 z1
p		7				92		1			h1
m			20				77	4			
w				31	2			37	22		y9
d	1				92				8		
t	1					93		2		3	pk2
n			1				96		3		
r				4			8	77	7		y4
g		2			64				34		z1
k		7				47			3	35	a4 h2 pt2

vision = t

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	11				57	6			25		h1 z1
p		2				97					h2
m			18				80	2			
w				20			6	45	13		
d					97				3		
t						98			1	2	
n			2				98				
r				7			12	63	5		
g	1				82				16		z2
k		4				42			3	39	a5 h3 s2 pt2

vision = n

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	24		2		33	2			36		h3
p		6				90				1	pt2
m			21				77		2		
w				32	1			40	19		y8
d	2				95				2		
t		4				81				14	h1
n							100				
r				5	1	2	6	63	13		y10
g	2				61			1	34		z2
k		2				36			2	53	a3 h2

vision = r

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	13				33	1		3	46		h4
p		2				93				3	h1 s1
m			15				82	3			
w				21			2	46	21		y10
d					82				17		h1
t						90			9		h1
n							93		7		
r				3			3	63	12		y20
g	2				52			1	46		
k		2				35			2	47	h6 s2 pt2

Table 7-b (2)

vision = g

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	13				48				35		h3
p		1				98				1	
m			17				78	4			
w				22				47	18		y13
d	1				98				1		
t		1				98				2	
n							98		2		
r				1	2		7	64	13		y13
g	2				71				25	2	h1
k						45			2	47	a3 h2

vision = k

audition \ response	b	p	m	w	d	t	n	r	g	k	others
b	10				30	2		1	46		h6 z2 dg2
p		4				87				5	a2 h1 pt2
m			11				86	3			
w				20			3	43	21		y13
d					91				9		
t		1				90				9	
n							100				
r					5		6	63	12		y14
g					59				40		z1
k		1				38				56	h2 a3

Under noisy conditions, visual effects were very strong and widespread. When the visual stimulus was labial (Table 7-a), every auditory non-labial stimulus was to a large extent perceived as labial. For visual /ba/'s, for example, auditory /da/'s were perceived as /ba/ 60% of the time, while 94% of these auditory /da/'s were correctly perceived by audition alone. Similarly, erroneous shifts to labials are as follows: /ta/'s were perceived as /pa/'s 94% of the time, auditory /na/'s were perceived as /ma/'s(74%), auditory /ra/'s were perceived as /wa/'s(45%) or /ma/'s(27%), auditory /ga/'s were perceived as /ba/'s(88%) and auditory /ka/'s were perceived as /pa/'s(94%). This response pattern is also true with visual /pa/, /ma/ and /wa/ stimuli.

When the visual stimulus was non-labial (Table 7-b), every auditory labial stimulus was to large extent perceived as non-labial. When the visual stimulus was /da/, auditory /ba/ was perceived as /da/(38%) or /ga/(29%), while this auditory /ba/ was 91% correctly perceived by audition alone. Similarly, the erroneous shift to non-labial is as follows: auditory /pa/ was perceived as /ta/(92%), auditory /ma/ wa perceived as /na/(77%), and auditory /wa/ was perceived as /ra/(37%) or /ga/(22%). This response pattern is true with the other visual non-labial stimuli.

The original McGurk effect was found only for pairs of auditory /pa/'s and visual /ka/'s or auditory /ba/'s and visual /ga/'s. Our results, however, revealed that under noisy conditions, the visual effect occurs in any visual-auditory incongruent pair, i.e., both the visual-labial and auditory-nonlabial pair and the visual-non-labial and auditory-labial pair. In these incongruent pairs, auditory stimuli are perceived as the stimuli having the same manner of articulation and voicing, but different place of articulation..

In our experiment, there were very few "combination errors" (e.g., visual /ba/ and auditory /ga/ make perceived /bga/).

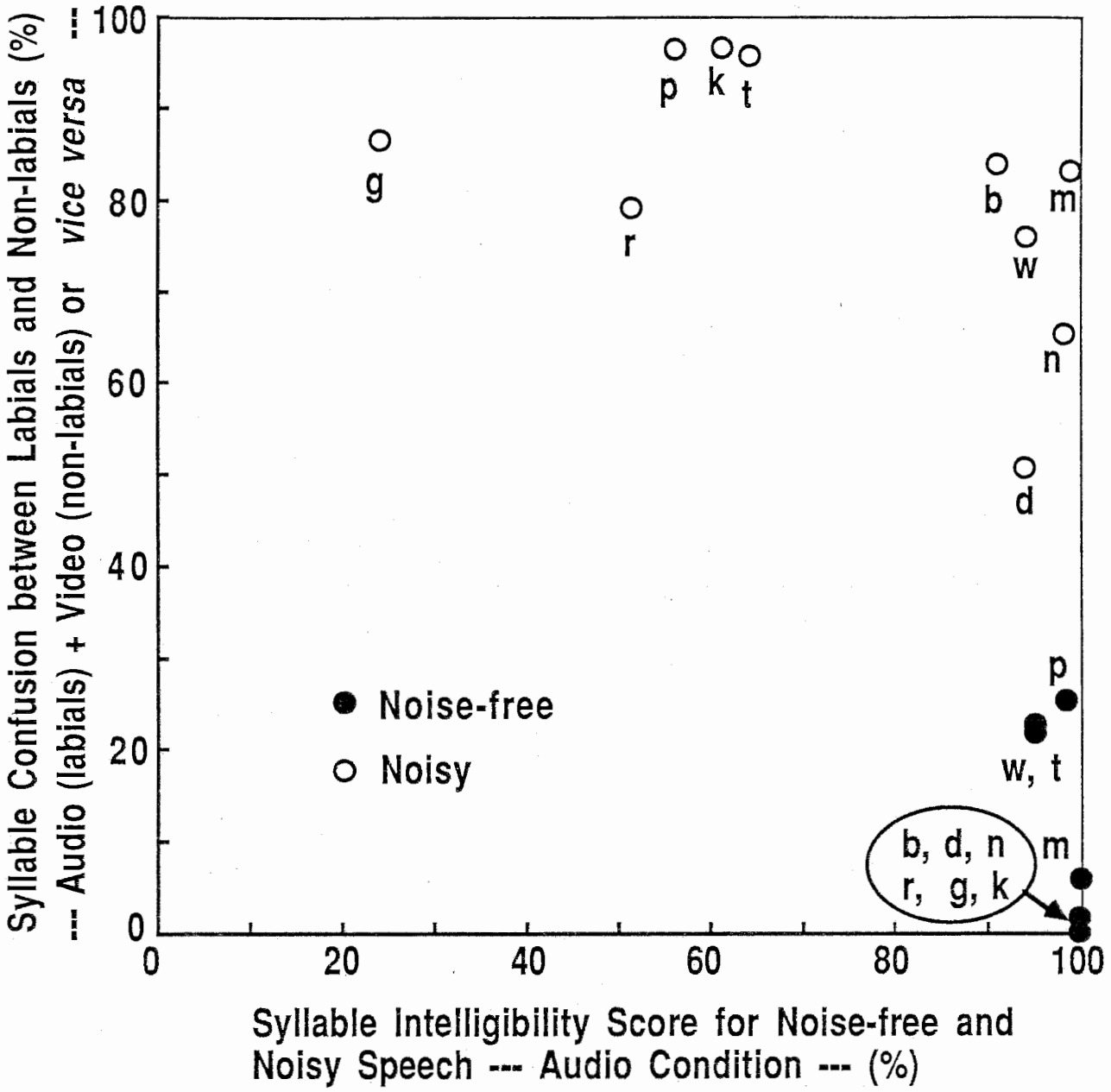

Fig.9 Relationships between syllable intelligibility in audio-only conditions and syllable confusion caused by misleading visual information

4．まとめと考察

本実験において、以下のことが明らかになった。

①雑音のない条件でもMcGurk効果がある程度生じたが、それは明瞭度が 100％ではない音節／pa，wa，ta／の知覚に限られていた。このとき、McGurk効果として反応の約20％が他の音節に異聴された。
②雑音のある条件における音声知覚では、非常に強いMcGurk効果が認められた。音節の異聴率でみた場合、この効果は音節／pa，ta，ka，ga／に対して90％を越える強いものであった。
③McGurk効果の強さは、刺激音である音節の明瞭度と負の連関をもつ。完全に明瞭な音声に対してはMcGurk効果は極めて生じにくい。しかし、わずかでも明瞭性を欠いた音声の受聴では、McGurk効果が生じ、雑音で明瞭度が低下した場合には、極めて強いMcGurk効果が容易に現れる。
④McGurkらは、音声が唇音で映像が非唇音のときに、視覚情報の効果が最も大きく、その逆の組み合せでは効果が小さいとしている。しかし、本実験の結果からは、そのような非対称性は認められず、効果は、むしろ対称的であった。
⑤McGurkらの言う複合反応（例えば、音声／ba／+映像／da／で知覚／bda／）は、教示で特に注意を促したにもかかわらず、ほとんど見出されなかった。これは、刺激作成における音声と映像の同期に十分な注意を払ったこと、日本語にこのような複合音節が存在しないことと関係しているものと思われる。
⑥McGurk効果による音節の異聴は以下のようであった。McGurk効果は、明瞭度が完全でない音節の知覚において調音位置の異なる音声と映像の組み合せの場合に生じる。このとき、各音節の知覚は映像刺激の調音位置に強く影響され、調音様式が等しいが調音位置の異なる音節に異聴される。

以上、雑音によって音声の明瞭度を操作した実験から、音声知覚における読唇情報の様々な効果が明らかになった。人間は、聴覚情報が完全でない場合には読唇情報を併せて利用し、聴覚情報が不明瞭な場合には、視覚がもたらす調音位置の情報に頼った視覚優位の知覚を行う。我々が音声によるコミュニケーションを交わす日常環境は、多かれ少なかれ雑音を伴う環境であるから、文脈情報と共に、このよう

な読唇情報がしばしば音声知覚に影響を及ぼすにちがいない。

【引用文献】

[Erber, N. 1975] Auditory-visual perception of speech. Journal of Speech and Hearing Dis-orders, XL481-492

[McGurk, H. & McDonald, J. 1976] Hearing lips and seeing voices. Nature, vol.264, pp.746-748

[McDonald, J. & McGurk, H. 1978] Visual influences on speech perception processes. Perception & Psychophysics, vol.24, pp.253-257

[福田、城間、船坂 1988] 人工内耳埋め込み患者の音声の知覚：聴覚と視覚の併用について 電子情報通信学会技術研究報告, SP88-91

[Fukuda, Y. & Hiki, S. 1982] Characteristics of the mouth shape in the production of Japanese: Stroboscopic observation. Journal of Acoustical Society of Japan(E), vol.3, pp.75-91

[Summerfield, Q. 1987] Some preliminaries to a comprehensive account of audio-visual speech perception. In B.Dodd and R.Cambell(eds.), Hearing by Eye: The Psychologyof Lip-reading. Lawrence Erlbaum Associates: London. pp.3-51

[Summerfield, Q. 1979] Use of visual information for phonetic perception. Phonetica, vol.36, pp.314-331

[Dodd, B. 1977] The role of vision in the perception of speech. Perception, vol.6, pp.31-40

[Easton, R. & Basala, M. 1982] Perceptual dominance during lipreading. Perception & Psychophysics, vol.32, pp.562-570

[積山、城、梅田 1988] 単音節の読唇における混同行列の分析：多次元尺度法による知覚属性の検討 電子情報通信学会技術研究報告, PRU87-113

[積山、東倉 1988] 単音節の受聴におよぼす唇の動きの影響：日本語におけるMcGurk効果 関西心理学会第100回大会論文集