

顔と音声の情報統合過程：マガーク効果における顔の音声情報

神崎 利佳

ATR 人間情報通信研究所

〒619-02 京都府相楽郡精華町光台 2-2

1. はじめに

ノイズ下での音声知覚では、話者の顔の観察が音声の明瞭性を高める効果をもつ¹⁾。このことは、発話に伴う顔の動きの情報（以下、顔の音声情報とする）が、聴覚からの音声情報と統合されるためだと考えられる。一方、顔と音声から矛盾した情報が与えられた場合、ノイズの有無にかかわらず顔の音声情報が統合され、特定の音声を知覚することがある（マガーク効果）。本稿では、このマガーク効果を取り上げ、音声知覚においてどのような顔情報が統合されるかを検討する。また、筆者らが行ったマガーク効果の実験について述べ、顔と音声の情報の統合過程について考察を加える。

2. マガーク効果とは

1976年に McGurk と MacDonald²⁾は、次のような実験結果を報告した。ひとりの女性が /ga-ga/ と発話する正面顔の映像に、/ba-ba/ という音声をつけて被験者に提示する。単独提示した場合の音声の明瞭性は高かったが、被験者は 98% の試行で /da-da/ という音声を聞き、映像が /ba-ba/ で音声 /ga-ga/ のときには、81% の試行で /ba/ と /ga/ を組み合わせた音声（例えば /gabga/）を聞いた。

音声の明瞭性が高いときにも、顔情報との統合が起こるという点で、この結果は興味深い。音声知覚における顔の役割は、基本的には聴覚を補うことであると考えられてきた³⁾。

これに対し、マガークらの研究は、音声と一致しない顔からの情報も処理され、音声の知覚を変化させるように影響を与えうることを示したのである。

マガーク効果はたいへん頑健な現象で、この現象についてよく知っている場合でも、刺激を繰り返し提示される場合でも、効果が弱まることはない⁴⁾。聴覚により注意を向けるよう教示を与えても同様である⁵⁾。

マガーク効果は、音声知覚における視覚の影響の有無を知るだけでなく、視覚と聴覚からの情報がどのように統合されているかを知る手がかりにもなる。マガークらは実験結果を次のように説明している⁶⁾。視覚では、調音位置、すなわち音源を作る位置についての情報が得られる。読唇の研究によれば、調音が唇で起こる（唇音）のか、それ以外の部分で起こる（例えば、歯茎音、口蓋音）のかの弁別が視覚的に可能である⁷⁾。一方、聴覚では、どのように調音された音であるか（調音様式）についての情報が得られる（有声音か無声音か、あるいは破裂音か鼻音かなど）。音声 /ba/ で顔が /ga/ の場合、視覚的には音声は /ba/（唇音）ではないと判断され、/ba/ に聴覚的にもっとも近くて、調音の起こる位置が唇ではない音、すなわち /da/ が知覚される。

しかし、この説明は顔映像が /ba-ba/ で音声 /ga-ga/ のときに、/gabga/ を聞くケースには当てはまらない。また、その後の研究で

は、視覚からも調音様式に関する情報が得られていることが確かめられている⁷⁾。視覚と聴覚で異なる音声特徴が分析され、それらが統合されるといふマガークらの説明のみでは十分とはいえない。

3. マガーク効果における顔の情報

マガーク効果についての検討は多いが、統合される顔の音声情報について検討が始められたのは最近のことである。

Green ら⁸⁾は男性の顔に女性の声を組み合わせ、マガーク効果の実験を行っている。性別に関する認知的矛盾がマガーク効果を弱めるかどうかを調べたのである。視覚情報に音源位置が規定される腹話術効果では、認知的矛盾が視聴覚情報の統合性を弱めることが指摘されている⁹⁾。しかし顔と音声の性別の不一致は、マガーク効果の生起率を低下させることはなかった。顔から抽出された音声情報は、その性別にかかわらず、聴覚からの音声情報と統合されてしまうのである。

マガーク効果と顔の認知過程の関係については、倒立顔を用いた検討もある¹⁰⁾¹²⁾。顔を倒立させると、顔ボタンを同定しにくくなることが知られている¹³⁾。例えば、図1のように、倒立では左右の写真はほぼ同一に見えるが、正立させると違いが明らかになる（この現象は最初に英国の元首相サッチャー女史の写真を用いて発表されたことから「サッチャー錯視」と呼ばれている¹⁴⁾）。

もし、顔の認知過程が、顔からの音声情報の抽出に影響を与えるならば、倒立顔ではマガーク効果が起こりにくくなると予想される。これまでの検討では、倒立提示によって効果が弱まる顔と音声の組み合わせとそうでない組み合わせがあることが明らかにされている¹⁰⁾¹¹⁾。このことは顔の同定の困難さがマガーク効果を弱らせるというよりも、倒立提示された唇の動きがある音声の知覚には影響を与え、別の音声には影響を与えないことを示しているといえる。したがって、マガー

ク効果は顔の認知過程とは独立で、倒立によるマガーク効果の変化は、なんらかの唇の動きの情報が抽出されにくくなるために生じるものと推測される。

サッチャー錯視を用いた検討でも同様の結果が示されている¹²⁾。もし顔の認知過程が音声知覚に影響を与えるならば、顔ボタンを同定しにくい倒立顔に、正立した口がついている場合（図2(a)）にはマガーク効果は弱まり、正立顔に倒立した口（図2(b)）では弱まらないと考えられる。一方、倒立顔が口の運動情報を抽出をしにくくさせるのであれば、倒立した口に正立顔を組み合わせた場合（図2(b)）にはマガーク効果は弱まり、唇のみが正立しその他の部分が倒立している場合（図2(a)）には効果は弱まらないはずである。結果は后者であり、運動情報の抽出がマガーク効果に影響を与えることを示すものであった。

さらに、Green ら¹⁵⁾は光点で唇の動きを見せる方法を使って、この結果を確かめている。顔に蛍光塗料のついた光点シールを貼り、その動きだけを見せると、それを顔と認知しない場合にもマガーク効果が起こることが報告されている¹⁶⁾。もし、運動情報の抽出過程が関係しているならば、倒立提示は光点の動きを変化させ、マガーク効果に差が見られるはずである。予想通り、倒立提示はマガーク効果を弱めるという結果が得られている。

神経心理学的にも、顔の認知過程と顔からの音声情報の抽出過程は独立であると考えられる。Campbell ら¹⁷⁾が研究した相貌失認の患者は、健常人と同様のマガーク効果を示した。

以上の検討から、マガーク効果における顔からの音声情報の抽出は、口の部分の運動情報の抽出であり、顔の認知過程とは独立であることが推測される。しかしながら、これらの知見のほとんどが、限られた視聴覚刺激対のみを提示して得られたものであり、今後より多くの種類の顔と音声の組み合わせについ

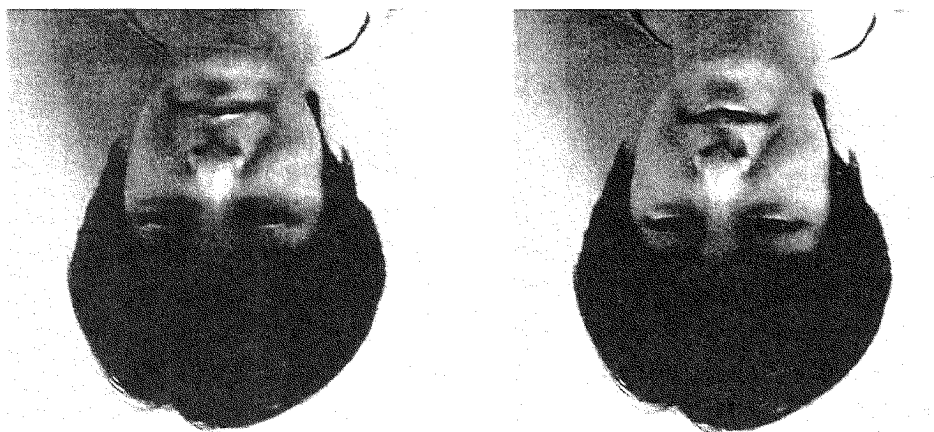


図1 サッチャー錯視の例

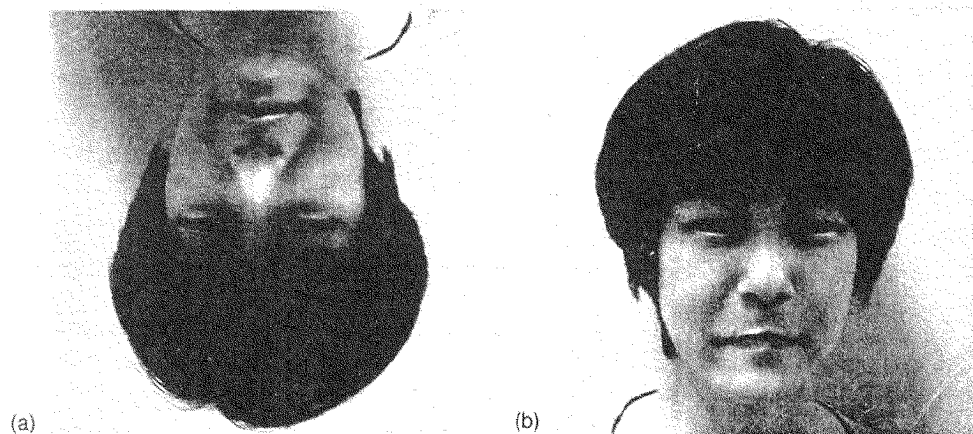


図2 (a) 倒立顔と正立口 (b) 正立顔と倒立口

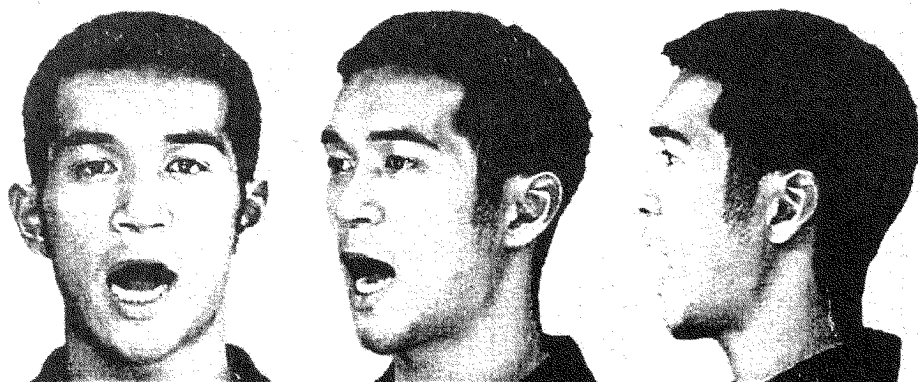


図3 実験で用いた顔映像

て、結果を確かめていく必要があると考えられる。

4. 顔の音声情報と統合過程

本節では、顔の音声情報について、筆者らが行った検討を中心に述べる。第2節でも述べたように、読唇研究によれば、我々は顔の動きから唇音か非唇音かを弁別することができるかとされている。果たして、我々はこの情報だけを聴覚情報と統合しているのだろうか。歯茎音や口蓋音の運動情報は統合されないのだろうか。

筆者らはこの問題について顔向き異なる映像を用いて検討した¹⁰⁾。もし、唇音か非唇音かだけの情報を得ているならば、顔向きにかかわらず、マガーク効果は生起するはずである。なぜならば、正面、斜め、横のどの向きでも唇音か非唇音かは観察できるからである。しかし、もし唇音か非唇音か以外の情報が抽出され、しかもそれが顔向きによって変化するものであれば、顔向きによってマガーク効果が変化するはずである。

刺激として用いたのは、9つの単音節(/ba/, /pa/, /ma/, /wa/, /da/, /ta/, /na/, /ga/, /ka/)とそれらを発話している顔映像(図3)である。正面、左45度、左90度の顔向き毎に、9つの音声

と9つの映像を組み合わせる81の視聴覚刺激を作成し、被験者に聞こえた音声を答えさせた。

ここでは、マガーク効果がより強くみられる破裂音の結果について述べる。まず音声のみを提示したときの正答率を、表1に示す。平均正答率は91%でやや明瞭性が低かったことがうかがえる。正答率が特に低かった/pa/の結果については後の分析から除外した。

図4に各音節の映像毎に、マガーク効果についての顔向きの影響を示す。/ga/や/ka/など口蓋音を発話する映像が提示される場合、マガーク効果はどの顔向きでも同じ強さで生じる傾向があった。この結果は、どの方向からも観察可能な顔の動きが音声知覚に影響を与えていることを示している。可能性としては、唇の開閉の情報が考えられる。

一方、唇音である/ba/を発話する映像の場

表1 音声課題の正答率

音声	/ba/	/pa/	/da/	/ta/	/ga/	/ka/
%	92	54	90	86	100	100

表2 音声課題の正答率(追試実験)

音声	/ba/	/pa/	/da/	/ta/	/ga/	/ka/
%	100	100	100	90	99	100

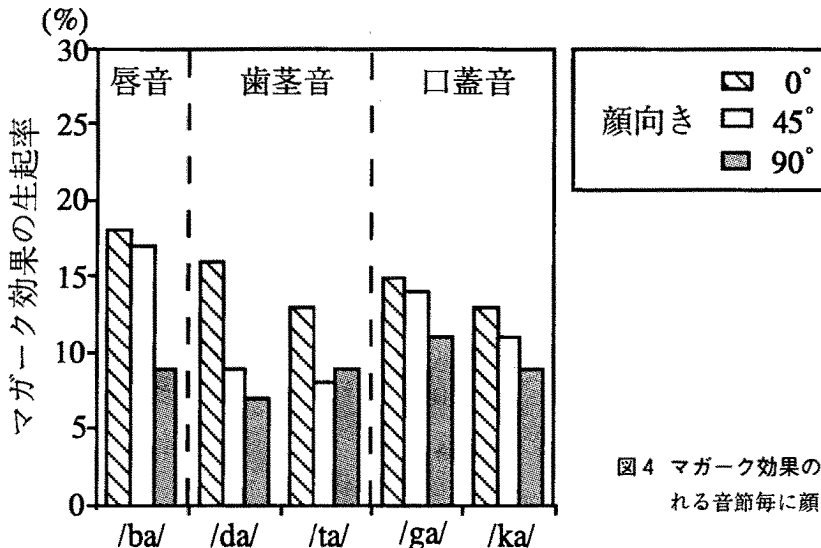


図4 マガーク効果の生起率。映像で提示される音節毎に顔向きの影響を示す。

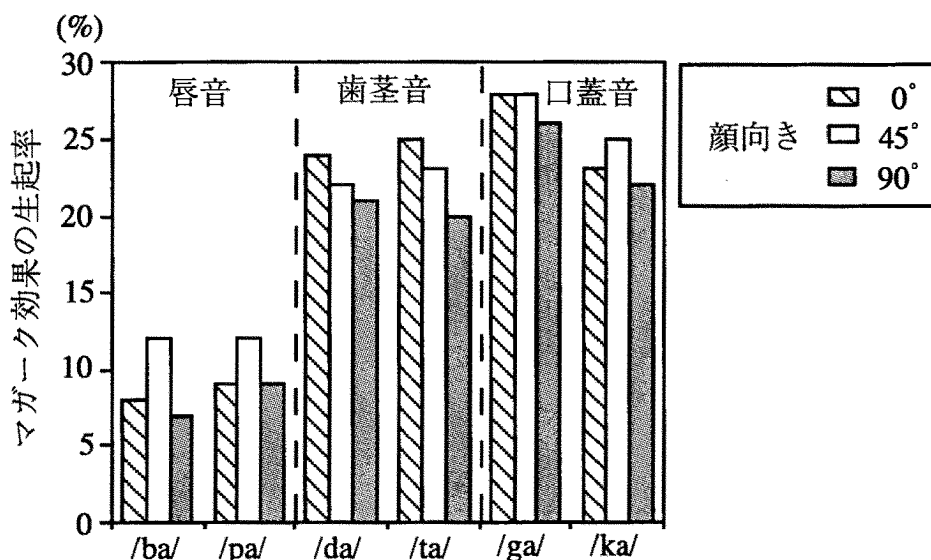


図5 マガーク効果の生起率（追試実験）。映像で提示される音節毎に顔向きの影響を示す。

合、正面顔と斜め45度でマガーク効果が生じやすく、90度で効果が弱くなった。正面と斜め45度からは観察されて、横顔では観察されにくい情報としては、唇や口腔の形状についての情報が考えられる。

また、/da/や/ta/などの歯茎音を発話する映像では、マガーク効果は正面顔で強く、斜め45度と横顔で弱くなった。正面からのみ観察可能な口腔内の情報が影響を与えていると考えられる。またこの結果が特に歯茎音で認められたことから、舌に関する情報の影響が考えられる（歯茎音は、舌尖を歯茎に打ちつけて発音する）。

筆者らの結果は、正面、斜め45度、90度のどの方向から観察しても、顔は音声知覚に影響を及ぼすことを示すものであるが、その情報とは唇音か非唇音かだけでなく、歯茎音の情報も含まれると考えられる。

筆者らは、この結果を確認するために、最近より明瞭性の高い音声を使って追試を行った。また、この追試では、口腔内の情報の抽出の有無を確かめるために、正面と斜め45度の顔刺激について、口腔内をグレーに塗りつぶした映像条件も加えた。

音声のみを提示した課題の正答率を表2に示す。平均正答率は、98%で前述の実験と比較して明瞭性が高いことがわかる。

図5にマガーク効果についての顔向きの影響を示す。追試実験では、先行実験とは異なり、ほとんど顔向きの影響は見られなかった。統計的に有意な顔向きの効果は/ta/の正面顔と横顔の間に見られるだけだった。また口腔内の情報の効果は、/ka/の顔についてのみ見られ、口腔内を塗りつぶした場合にマガーク効果は有意に弱まった。

追試実験で先行実験と同様の顔向きの効果が見られなかった理由として、音声の明瞭性が高かったことが考えられる。Sekiyamaら¹⁹⁾は日本人を対象に行ったマガーク効果の実験で、そのままの音声ではマガーク効果がほとんど生じないにもかかわらず、ノイズを加えて音を聞き難くすると強いマガーク効果が生じることを報告している。しかし筆者らの検討では、唇音を発話する映像ではマガーク効果が弱まる傾向があったが、その他の映像ではマガーク効果が強まる傾向があった（図5）。

追試実験では視聴覚課題の終了後、視覚刺

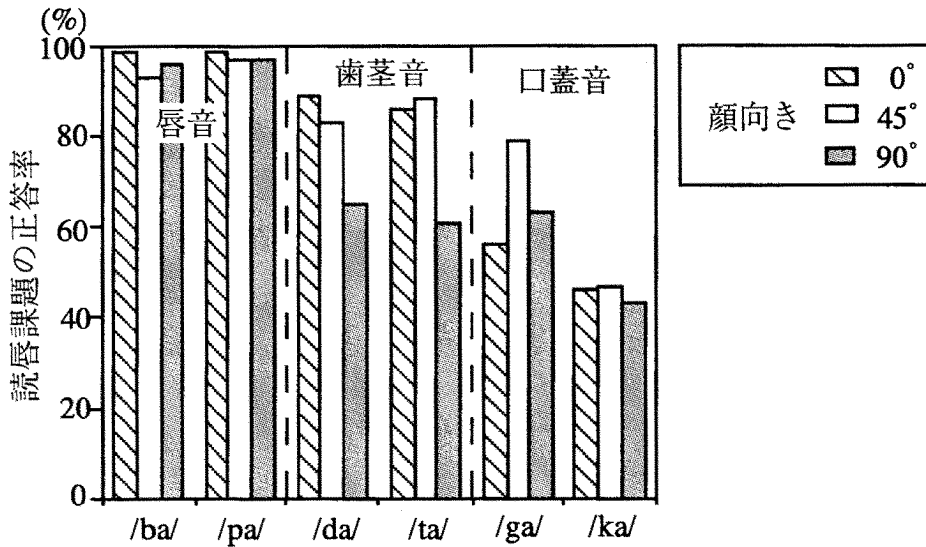


図6 読唇課題の正答率。映像で提示される音節毎に顔向きの影響を示す。

激のみを提示する読唇課題も行った。ある音声の発話について、調音位置が同じ音声を答えたときにはそれも正答として結果を分析した。結果を図6に示す。唇音についてはすべての顔向きで正答率は100%に近く、顔向きの効果は見られなかった。しかし、歯茎音(/da/と/ta/)を発話する顔では、横顔の読唇成績が正面顔よりも有意に低く、/ta/については斜め45度と横顔の間にも有意差が見られた。口蓋音の/ga/では斜め45度に比べて正面と横顔で読唇成績が有意に低下した。また歯茎音を発話する正面顔では、口腔内の情報が与えられた方が正答率が有意に高く、/ta/については斜め45度の顔でも、同様の差が見られた。これらの結果は口腔内の情報も読唇成績に影響を与えることを示唆するものである。

追試実験では、マガーク効果に前回と同様の顔向きの影響を認めることはできなかったが、読唇課題については、顔向きの影響が見られることが確認された。筆者らの2つの実験からは、どのような音声刺激が与えられるかによって、顔の音声情報と声の音声情報の統合の性質が異なることが示唆される。私た

ちは常に読唇から得られる情報を聴覚情報と統合しているのではなく、それぞれのモダリティから得られる情報の明瞭度に応じて、統合される情報の選択や重みづけが変化しているのではないかと考えられる。今後はこの予測を確かめるために、顔映像と音声の明瞭度の相互作用を系統的に操作し、顔と音声情報の統合について検討していく予定である。

5. 終わりに

最初にも述べたように、マガーク効果は今から20年以上前に発見され、この現象については多くの研究者が検討を重ねてきた。しかしながら、実際にどのような顔情報が聴覚からの音声情報と統合されるかについて詳細に調べ始められたのは最近のことである。それまでは、読唇に関する情報が、聴覚情報と統合されるという仮定の基で検討が進められてきた。

聴覚情報との統合性という観点から顔を検討すると、読唇研究だけではとらえられなかった顔の音声情報の処理過程が認められる。今後は、顔からどのような音声情報(運動情報)が抽出されているのかをさらに明確

にする必要がある。また、そこでの知見に基づいた視聴覚における音声知覚の統一的な理論の構築も今後の課題である。

文 献

- 1) W. H. Sumby and I. Pollack: Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215, 1954.
- 2) H. McGurk and J. MacDonald: Hearing lips and seeing voices. *Nature*, 264, 746-748, 1976.
- 3) Q. Summerfield: Some preliminaries to a comprehensive account of audio-visual speech perception. *B. Dodd and R. Campbell (eds): Hearing by eye: The psychology of lip-reading*. Erlbaum, London, 1987, 3-51.
- 4) Q. Summerfield and M. McGrath: Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology*, 36A, 51-74, 1984.
- 5) J. MacDonald, and H. McGurk: Visual influences on speech perception processes. *Perception and Psychophysics*, 24, 253-257, 1978.
- 6) 積山 薫, 城 和貴, 梅田三千雄: 読唇による単音節の認識. 電子情報通信学会技術報告, IE87-108, 33-40, 1988.
- 7) K. P. Green and P. K. Kuhl: The role of visual information in the processing of place and manner features in speech perception. *Perception and Psychophysics*, 45, 34-42, 1989.
- 8) K. P. Green, P. K. Kuhl, A. N. Meltzoff and E. B. Stevens: Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics*, 50, 524-536, 1991.
- 9) R. B. Welch and D. H. Warren: Intersensory interactions. *K. R. Boff, L. Kaufman and J. P. Thomas (eds): Handbook of Perception and Human Performance, Volume I: Sensory Process and Perception*. Wiley, New York, 1986, 25: 1-36.
- 10) K. P. Green: The influence of an inverted face on the McGurk effect. *Journal of the Acoustical Society of America*, 95, 3014, 1994.
- 11) D. W. Massaro and M. M. Cohen: Perceiving speech from inverted faces. *Perception and Psychophysics*, 58, 1047-1065, 1996.
- 12) D. A. Yakel, L. D. Rosenblum, K. P. Green, R. A. Vasquez and C. Bosley: Inversions on the McGurk effect. *Journal of the Acoustical Society of America*, 97, 3286, 1996.
- 13) T. Valentine: Up-side down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, 79, 471-491, 1988.
- 14) P. Thompson: Margaret Thatcher: A new illusion. *Perception*, 9, 483-484, 1980.
- 15) K. P. Green: The use of auditory and visual information in phonetic perception. *D. G. Stork and M. E. Hennecke (eds): Speech reading by Human and Machines*. Springer-Verlag, Berlin, 1996, 55-77.
- 16) L. D. Rosenblum and H. M. Saldana: An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 318-331, 1996.
- 17) R. Campbell, T. Landis and M. Regard: Face recognition and lip reading: a neurological dissociation. *Brain*, 109, 509-521, 1986.
- 18) R. Kanzaki, T. Kato and Y. Tohkura: Influence of facial views on the McGurk effect. *Journal of the Acoustical Society of Japan (E)*, in press.
- 19) K. Sekiyama and Y. Tohkura: McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, 90, 1797-1805, 1991.