

# コミュニケーションにおける視聴覚統合

## —顔の向き・口内情報と McGurk 効果—

梶村 昌代（文学部人間学科心理学コース 4年）

指導教員

小島 治幸（人間社会研究域人間科学系 准教授）

### 1. 研究目的

人間は、環境情報の80%以上を眼から獲得しているといわれるほど、視覚に頼った生物である。例えば、テレビ上の登場人物が話しているとき、実際に音声が出ているのはスピーカーからであるが、私たちはその人物が話しているように感じている。このように、他の感覚情報に視覚情報が影響する事例が存在し、通常はその矛盾に気づかない。このような矛盾を実験的に発生させた一例として、McGurk 効果が挙げられる。これは、McGurk & MacDonald (1976) によって最初に報告され、視覚（画像）情報と聴覚（音声）情報とを矛盾させた場合に、情報の統合が生じて元の情報とは異なった音声として知覚されるというものであり、聴覚に対する視覚優位性を示した。例えば、視覚からの/ga-ga/という情報を、聴覚からの/ba-ba/という情報とともに呈示した場合、被験者は/da-da/という発音に聞こえたと報告した。McGurk らはこのような反応を融合反応と呼んだ。また、これとは逆の呈示条件、すなわち視覚情報が/ba-ba/で、聴覚情報が/ga-ga/のとき、被験者は/gab-ga/や/bag-ba/という子音の混ざったような発音に聞こえたと報告した。これは、複合反応と呼ばれる。

McGurk 効果は、顔の向きによって影響に差があることが分かっている。Kanzaki, Kato, Tohkura (1998) は、顔を正面顔・左斜め顔・左横顔の3種類に分けて被験者に呈示した。その結果、正面顔よりも、斜め顔、横顔の順でMcGurk効果が小さくなった。しかし、この研究において、斜め顔・横顔に関しては左向きのみの呈示である。顔の向きの影響を調べるためには、右向きの顔も呈示することが必要と考えられる。Kato (1978), 馬淵 (1999) の実験は、被験者に横顔の描画を求めると、多くの場合左向きの顔を描くという結果を得た。また Sackeim (1978) の実験において、向かって左側にある表情が強く感じられるという結果になるなど、顔認知における左右差の可能性が唆されていることにある。これらの報告は、顔情報の左右非対称性を示しており、McGurk 効果についても左右の顔で異なった効果をもたらす可能性が考えられる。

そこで本研究は、Kanzaki, Kato, Tohkura (1998) で用いられた3方向に右斜め顔と右横顔を加えた5つの方向を向いた顔刺激を用い、McGurk 効果における顔の向きの左右対称性を検討することを目的とした。

### 2. 方法

**実験参加者** 18歳から24歳の大学生50名（うち、女性27名、男性23名、平均20.53歳）が実験に参加した。

**呈示刺激** 刺激画像及び音声は、デジタルカメラ（Canon 製 IXY DIGITAL 2515）を用いて撮影した。それから、動画ソフト（ADOBE PREMIERE PRO CS3）によって、音声と映像の組み合わせを編集した。各刺激は発声前後の口を閉じた状態も含めて全体で3000msになるようにし、発声は3000msのちょうど中間で行われるようにした。また、各刺激において音声は一度のみ、語尾を延ばさずに発音された。背景は白色で統一

した。また、左向きの顔は刺激人物の右頬が被験者に見える状態、右向きの顔は刺激人物の左頬が被験者に見える状態であった。

**実験1 (音声—映像課題)** 音声刺激3条件 (/ba, ga, da/の音声) ×映像刺激3条件 (/ba, ga, da/の画像) ×顔の向き5方向 (正面, 左斜め, 左横, 右斜め, 右横) ×刺激人物10名=450の刺激を用いた。

**実験2 (音声課題)** 音声刺激3条件 (/ba, ga, da/の音声) ×顔の向き5方向 (正面, 左斜め, 左横, 右斜め, 右横) ×刺激人物10名=150の刺激を用いた。

**実験3 (映像課題)** 映像刺激3条件 (/ba, ga, da/の画像) ×顔の向き5方向 (正面, 左斜め, 左横, 右斜め, 右横) ×刺激人物10名=150の刺激を用いた。

また、それぞれの実験において各被験者に呈示される刺激は、刺激全体の半分 (刺激人物5名分) であった。したがって、刺激人物はあらかじめ5名ずつランダムに2つのグループに分けられた。

#### 実験手続き

**実験1** 刺激の呈示順序は、SuperLab4.0を用いてランダム化された。被験者は、講義室に設置されたプロジェクタからスクリーン上に投影された刺激を観察した。音声は、プロジェクタに接続したパーソナルコンピュータ (EPSON製 Endeavor NJ1000) にスピーカー (Diamond Audio Technology製 EMC2.0-USB) を接続し、被験者の前方から聞こえるように設定した。各試行は、ブザー音 (1000ms)、音声+映像 (3000ms)、白紙画面 (2000ms: 回答記入用) の順に呈示した。被験者は、ブザー音が鳴ったら画面に注目し、呈示される人物が発声し終わったら「聞こえたと思う音声」をマークシートに回答するように教示を受けた。マークシートを用いることによって、/ba, ga, da/のいずれかを強制的に選択させた。2000msの白紙画面の後に再びブザー音が鳴り、次の試行へと続いた。また、呈示刺激は、刺激人物5名分のもので、計225通り用意した。実験は、50試行ごとに1分間の休憩をはさんで行った。

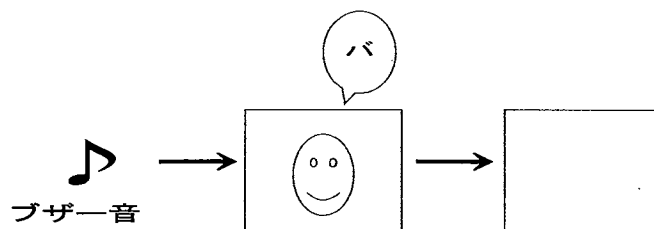


図1. 実験1の手続き

**実験2** 各試行は、ブザー音 (1000ms)、音声 (3000ms)、空白間隔 (2000ms: 回答記入用) の順に呈示した。被験者は、「聞こえたと思う音声」をマークシートに回答するよう教示を受けた。回答の記入方式と、次試行への移り方は実験1と同様であった。また、各被験者が呈示された刺激は、刺激人物5名分のもので、計75通り用意した。

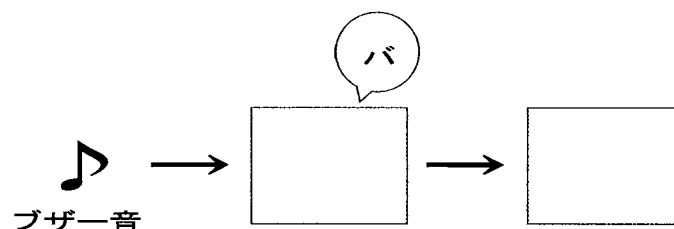


図2. 実験2の手続き

実験 3 各試行は、ブザー音 (1000ms)、映像 (3000ms)、白紙画面 (2000ms : 回答記入用) の順に呈示した。被験者は、ブザー音が鳴ったら画面に注目し、呈示される人物の「発音していると思う音」をマークシートに回答するように教示を受けた。刺激の呈示方法、回答の記入方式、次試行への移り方は実験 1 と同様であった。また、各被験が呈示された刺激は、刺激人物 5 名分のもので、計 75 通り用意した。

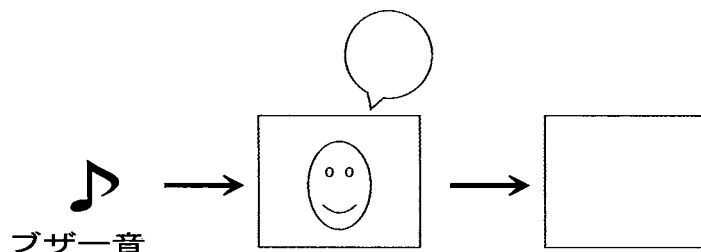


図3. 実験3の手続き

### 3. 結果と考察

まず、音声は/ba/であったとき、/ba/以外の音声に聞こえた割合 (エラー率) を示した (図 4)。そして、音声—映像課題に関して、音声×映像×顔の向きの 3 要因の分散分析を行った。要因は、全て被験者内要因であった。音声と映像の水準数は共に 3 (/ba, ga, da/) で、顔の向きの水準数は 5 (左横, 左斜め, 正面, 右斜め, 右横) であった。全方向において、音声/ba/-映像/ba/ (音声-映像一致) 条件と、音声/ba/-映像/ga/ (融合反応) 条件との間に有意な差が見られた。つまり、全方向で一般的な McGurk 効果の融合反応が生じ、そこには映像の効果があるということが示された。次に、音声—映像—顔の向きの交互作用が有意であった ( $F(16, 784) = 1.763, P < .05$ ) ため、多重比較を行った。

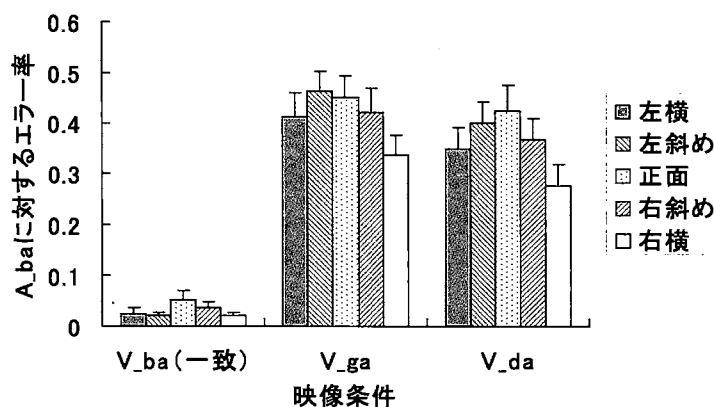


図4. 音声/ba/に対する、顔の向きごとの映像別エラー率

さらに、映像が/ga/であるときの、音声条件別エラー率を図 5 に示した。全方向において、音声/ba/-映像/ba/ (音声-映像一致) 条件と、音声/ba/-映像/ga/ (融合反応) 条件との間に、有意な差が見られた。つまり、融合反応には音声も影響を与えていることが示された。

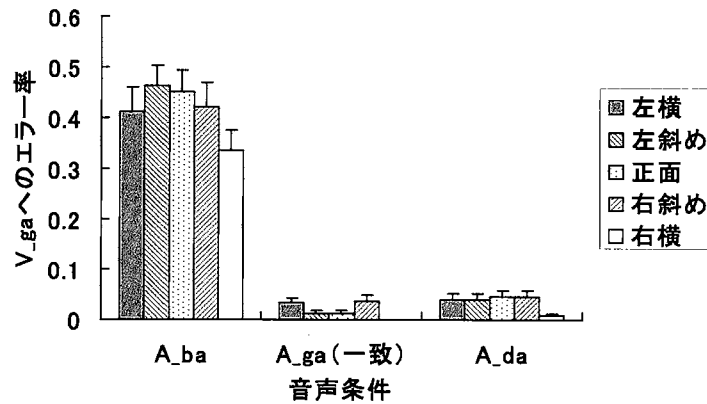


図 5. 映像/ga/に対する、顔の向きごとの音声別エラー率

複合反応について、どんな仮説が成り立つのか説明する。本実験では、複合反応は生じなかった。しかし、本実験の前に行った大学生と高校生を対象とした 2 回の McGurk 効果追試実験ではともに、複合反応の生じる条件は音声・映像一致条件よりも有意にエラー率が高かった。本実験で複合反応が生じなかった理由の可能性として、2 回の追試実験とは別に新しい刺激を作成したことが挙げられる。個人の発話の癖などの要因が入り込み、効果を阻害してしまったのかもしれない。また、音声に注意を向けるような教示をしたことが関係していたとも考えられる。McGurk & MacDonald (1976) の実験でも、もともと融合反応よりも複合反応が生じる割合は少ない。加えて、複合反応は日本人にとっては不慣れな子音の組み合わせ (b と g) が聞こえる条件であるため、生じにくかったと推察できる。さらに、聴覚を研ぎ澄ますことの多い、音楽系サークルの被験者が多数参加したことも関係しているかもしれない。

図 6a は、音声/ba/条件は他の 2 条件よりも有意にエラー率が高かったことを示している ( $F(2, 98) = 4.970$ ,  $P < .01$ )。ここから、/ba/という音声の知覚に、唇を一旦閉じて発声するという映像が影響しており、音声のみの課題では唇の動きという手がかりがなくなったためにエラー率の増加が生じたと推察できる。しかし全体としては、エラー率は音声ごとに 10%を下回っており、非常に低いと言える。これにより、音声—映像課題において、音声自体のあいまいさによる影響は小さいと考えられる。また、/ba, ga, da/は、音声のみであればほぼ弁別できるということが示された。

ちなみに、撮影は方向ごとに別々に行われたため、同じ音を発声していても、厳密には方向別に発音が異なっている可能性があった。本来であれば、各音で音声を一つに統一すべきであっただろうが、今回は例えば、正面向きの刺激では正面向きを撮影したときに録音された音声の中で吹き替えをし、左横向きの刺激では左横向きを撮影したときに録音された音声の中で吹き替えを行った。それに伴い、音声課題では、全方向からの音声を全て呈示した。しかし図 6a は、方向別の発音の違いを考慮していない。この影響を加味して検定を行ったところ、音声と方向の交互作用が有意となっており ( $F(8, 392) = 2.123$ ,  $P < .05$ )、音声/ba/においてエラー率が、右斜め向きが左斜めと右横向きよりもエラー率が有意に高く、左横向きが右横向きよりも有意にエラー率が高かった。このことから、音声—映像課題において方向別の音声明瞭度の差の影響があると考えられる。

図 6b は、映像/ba/条件は他の 2 条件よりも有意にエラー率が低かったことを示している ( $F(2, 98) = 75.140$ ,  $P < .001$ )。今回用いた音は、/ba/のみが唇を一度閉じて発音する唇音で、/ga, da/は共に唇を開いたまま発音する非唇音であった。このことから、唇音と非唇音を区別することはできるが、唇音同士の違いを読み取ることはチャンスレベル近くまで落ちることが分かった。

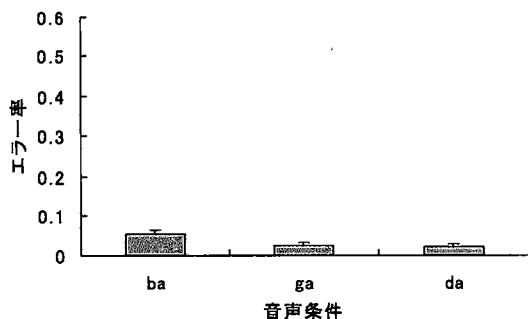


図 6a. 音声条件別エラー率

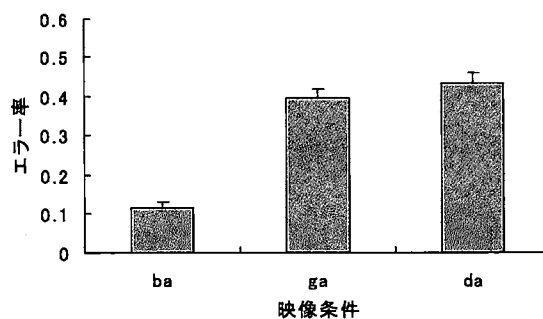


図 6b. 映像条件別エラー率

しかし、図 6b では顔の向きを無視している。映像刺激は、顔の向きによって形態が大きく異なるため、方向の条件を含めた分析が必要だと考えられる。そこで、映像課題に関して 2 要因の分散分析を行ったところ、映像と顔の向きとの交互作用が有意であった ( $F(8, 392) = 2.298, P < .05$ )。図 7 を見ると、全方向で、映像/ba/条件は他の 2 条件よりも有意にエラー率が低いことが分かる。また、映像/ga/条件と/da/条件の間では、左斜め向き条件以外では有意差は見られなかった。映像/da/についてのみ、正面向きが左横・左斜め向き・右横向きよりも有意にエラー率が低く、右斜め向きが左横向きよりも有意にエラー率が低かった。映像/da/に特有なこの結果は、/da/が上の歯茎に舌の端を接触させて発音する音声であることに関係しているかもしれない。つまり、正面顔で最もエラー率が低くなるのは、/ba/や/ga/の発音には見られない舌の動きという手がかりを利用することができたからではないかと推測できる。本実験ではスクリーン上で、普通の対話場面よりも拡大された人物を呈示したため、舌の動きが見やすかった可能性があると考えられる。

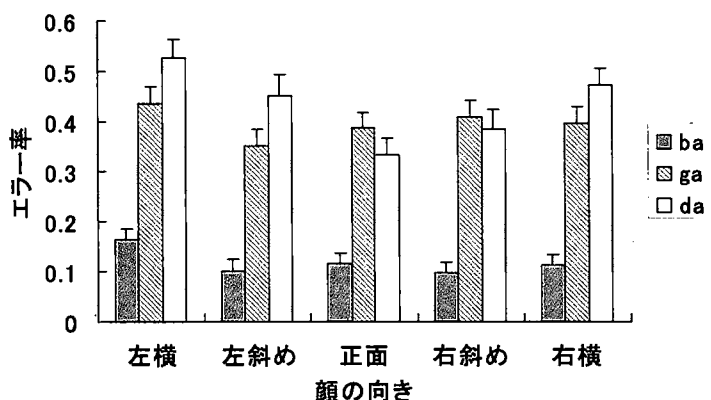


図 7. 映像課題における顔の向き別エラー率

最後に、本実験の結果を顔の向きの効果の非対称性という点から考察する。図 4, 5 において、融合反応が生じる音声/ba/・映像/ga/条件について、左斜め顔を頂点とする山型になっている。統計的には、右横顔を呈示したときのみ、他の 4 方向より有意にエラー率が低くなるという結果が得られた ( $F(4, 1764) = 10.46, P < .001$ )。つまりグラフは正面顔を中心として対称ではなく、左向きよりも右向きの顔に対するエラー率が低くなる傾向があると考えられる。先に、音声・映像課題における方向別の音声明瞭度の差の影響があると考えられると述べたが、明瞭度の差よりも非対称的な効果の方が大きいので、全体として顔の向きの効果の非対称性が現れていると推測できる。

これに関し、顔認知における顔の角度の効果を扱ったものとして、馬淵・片山・諸富 (2001) の実験がある。この実験で、右向き顔に対する左向き顔の人物識別における優位性が確認されている。この結果を本実験と合

わせて考察する。McGurk 効果が起こる条件を、視覚・聴覚情報間で拮抗が生じる条件として考えると、左向きの顔のときは視覚情報が聴覚情報よりも相対的に強い影響力を獲得し、McGurk 効果が生じやすくなったと推測することができる。McGurk 効果と人物識別の間に認知处理的な関連性があるかは、本実験のみでは分からないが、少なくとも右向きの顔は視覚優位性効果が薄れるということが示された。また、非対称的な影響は映像課題では見出せなかったことから、この影響は視覚と聴覚という 2 つのモダリティが競合する段階に作用しているのではないかと考えられる。つまり、ある程度の知覚的負荷がかかると非対称性が生じると推察できる。

今後、正面顔の右半分のみを呈示した場合と左半分のみを呈示した場合とで、McGurk 効果に差が現れるかどうかや、右半分と左半分で違う音声を発している顔を呈示した場合、どちらを優先するかという点について検討したい。もし差が現れれば、発話知覚レベルでの顔の左右差の影響がより明らかなものとなるだろう。

#### 4. 結論

1. 音声のみを聞かせた場合、/ba/と/ga/と/da/の 3 つの音は、高確率で聞き分けることができる。
2. 映像のみを見せた場合、唇音と非唇音の識別はできるが、非唇音同士の識別はほぼチャンスレベルである。また、これらの結果に顔の向きの左右差は現れなかった。
3. McGurk 効果の融合反応に関して、左向きの顔の方が視覚の影響を受けやすいということが明らかになり、顔の向きの左右差が示された。

#### 5. 引用文献

- Kanzaki, R., Kato, T., & Tohkura, Y. (1998). Influence of facial views on the McGurk effect. *The Acoustical Society of Japan(E)*, 19, 69-72.
- Kato, T. (1978). Dominance of left side faces in drawing profiles. *Artes Liberales*, 22, 47-53.
- 馬淵慶子 (1999). 顔の角度が顔の認識に及ぼす効果の ERP を用いた検討. 生理心理学と精神生理学, 17, 143.
- 馬淵慶子・片山順一・諸富 隆 (2001). 顔認知における顔の角度の効果 —ERP を指標にして— 電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎 pp. 39-46.
- McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices, *Nature*, 264, 746-748.
- Sackeim, H. (1978). Emotions are expressed more intensely on the left side of the face, *Science*, 202, 434-436.