

視覚における情報統合：音声知覚時の視覚と聴覚の情報統合

積山 薫

公立はこだて未来大学 システム情報科学部

〒041-8655 北海道函館市亀田中野町 116-2

1. はじめに

日常事態において、ある事物が視覚情報と聴覚情報の両方を発していることはよくある。実験的に調べてみると、そのような場面では両方の感覚情報が補強しあって知覚の安定性をもたらしていることがわかる。われわれは、複数の同期した感覚情報を統合してとらえようとする傾向があるのである。もし、互いに矛盾する視覚情報と聴覚情報を同期させて与えると、独特の現象が生じ、統合のありさまが浮き彫りになる。

視覚情報が聴覚刺激の認知に影響を及ぼす例として、音声知覚の領域ではマガーク効果（矛盾した口の形が音声の聞こえを変える）¹⁾ など、音源定位の領域では腹話術効果（視覚的物体にひきずられた位置に音源が知覚される）が知られている。また、聴覚情報が視覚刺激の認知に影響を及ぼす例として、2物体の衝突・通過運動の知覚（カシャンという音が伴うと衝突したと知覚されるが、無音だと通過したと知覚される）²⁾、フラッシュ光の点滅の知覚（反復呈示されるピーブ音の回数にひきずられて光の点滅回数が知覚される）³⁾などが報告されている。ここでは、マガーク効果に焦点をあてて、統合のありさまについて考えてみる。

2. 視聴覚音声知覚の諸要因

対面場面での音声言語コミュニケーションにおいて、時間的に同期した声（聴覚）と口の動き（視覚）の情報は知覚的にきわめて統合されやすく、マガーク効果はそれを如実に

示す現象である⁴⁾。これは、ビデオの吹き替えによって話者の音声と映像とを矛盾させて呈示したときに、2つの情報が統合されて知覚される現象である（たとえば、audio /pa/ + video /na/="ta"）。この現象は、顔が女性で声が男性というような吹き替えであっても同性での吹き替えと同じくらい強く生じる⁵⁾、口が完全に見えていなくても、口のまわりにつけた幾つかの光点だけで生じる⁶⁾。また、顔が上下逆さにして呈示されると、マガーク効果すなわち視覚の影響は弱められる⁶⁾。発達的には、始語期以前の4ないし5ヶ月の乳児でも視覚と聴覚の両方によって音声を知覚しているという報告がある一方で^{7,8)}、幼児は視覚的な読唇の能力が大人よりもかなり劣っており、そのためマガーク効果も年少ほど弱いことが指摘されている^{9,10)}。この発達の問題と関連して、言語的ないしは文化的背景によってマガーク効果の強さが規定される部分もあり、日本語や中国語を母語とする被験者は英語を母語とする被験者よりもマガーク効果が弱いという¹¹⁾。

3. 統合のメカニズムをめざして

このように視聴覚音声知覚における統合の規定因がかなり詳細に調べられている一方で、統合のメカニズム自体はあまりよく分かっていない。わずかにみられる数理的なアプローチにおいて、代表的なモデルとしてファジィロジカル・モデルがあげられる¹²⁾。このモデルでは、視覚情報と聴覚情報の統合によって生じる反応の生起確率は単一感覚条件

で測定される正答率の機械的な積で与えられ、感覚モダリティに応じて注意や重み付けが異なるといったことはまったくないという。このような機械的なモデルは、統合をボトムアップ的な処理のみでとらえていることになるが、果たしてそれだけで説明できるのだろうか。

3.1 機能的脳画像による脳内活動部位の特定

筆者らは、視聴覚統合のメカニズムを調べることが目的に、発話の聴覚成分の聞き取りやすさ（聴覚明瞭度）を変化させた2種類の刺激セットを用いて、視聴覚音声知覚のさいの脳内活動をfMRIを用いて計測した。その結果、音が聞き取りにくいほど視覚情報がよく利用されてマガーク効果が生じやすい、という知覚の変化と対応する脳内活動の変化がみられた。この実験では、話者が発話した /ba/, /da/, /ga/ を録音・録画・デジタル編集した素材から、聴覚刺激(A)、視覚刺激(V)、視聴覚刺激(AV)の3通りを作成して呈示し、被験者には /ba/, /da/, /ga/ のいずれであるかの音韻判断を求めた。統制条件では、話者の顔の無音の静止画を呈示して、いずれの話者であるかの話者判断を求めた。これらの判断をおこなっているときの脳内活動をfMRIによって測定するとともに、被験者が報告した音韻判断・話者判断の行動的なデータも記録した。以上のすべての刺激条件は、音圧を変化させた2種類の聴覚明瞭度（High: S/N比 +7 dBおよびLow: -3 dB）で繰り返された。10人の被験者からデータを得た。

行動的データから、マガーク効果は聴覚明瞭度が低い方が高い時よりもずっと強く、先行研究と同様の結果であった¹³⁾。また、A条件でも、明瞭度が低い方が誤答が多く、明瞭度の低い条件では音が聞きにくかったことが裏付けられた。

脳内活動については、A、V、AV各条件の脳機能画像を統制条件との対比で解析したところ、単一モダリティ条件（AおよびV）では明瞭度による差がなかったのに対して、AV条

件では、聴覚明瞭度によって大きなちがいがみられた。A条件では、左側頭葉後部の聴覚領（BA 21/22）と前頭葉のブローカ野（BA 45）に活動がみられ、聴覚的な言語処理が生じていた。V条件では、ブローカ野（BA 44, 45）、頭頂葉の縁上回（BA 40）、後頭側頭回（BA 37/29, MT野ないしはV5を含む）に活動がみられた。これらは、視覚的な動きの処理（MT野）、内言的なりハーサル（ブローカ野）、音韻処理（縁上回）などに対応していると考えられ、読唇関連領域ということができるだろう。AV条件において、読唇にかかわる部位（MT野と縁上回）は、聴覚明瞭度が低いときには強く活動したが、明瞭度が高いときには活動がみられなかった。これらの結果から、口の動きに関する視覚情報は、音声聞き取りにくいときにはより集中的に脳内で処理されるが、音声が明瞭なときには聴覚情報の処理のウェイトが相対的に高くなるといえる。

3.2 反応時間による時間特性の検討

以上の脳内活動に関する「どの部位が」という空間的な情報は、「どのタイミングで」という時間的な情報と合わせて考える必要がある。そこで、脳内活動のタイミングを探る間接的な方法として、条件ごとの反応時間を測定する実験を追加した。防音室内で、一定音量のスピーチに対して音量を変化させたホワイトノイズを加え、上記のMRI室と同様のS/N比となる明瞭度を実現した（High: +7 dB, Low: -3 dB）。上記と同じ刺激（A、V、AV刺激）を2通りのノイズ強度のもとで呈示し、被験者が答えの音韻に対応したボタンを押すまでの反応時間を測定した。その結果、反応時間は無音で読唇を求めたV条件で最も短かった。これは、言語音を発音するさい、音響的な出力にかなり先行して口は音を作り出すための動きを始めていることによるのであろう。AV（マガーク）条件では、明瞭度が低いとき、反応時間はAのみ条件よりも短くなり、先行する視覚処理による促進効果がうか

がわれた。明瞭度が高い時，AV条件の反応時間はAのみ条件よりも遅かった。音が聞き取りやすい時は脳内で視覚的な動きの処理が生じにくかったMRIの結果と考え合わせると，遅い反応時間は視覚的な動きの処理を抑制する過程を反映しているのではないかと思われる。

3.3 結び

以上2つの実験結果から，視聴覚統合は，視覚と聴覚の各モダリティの活動強度を調節する過程を含むのではないかと考えられる。脳は，すべての感覚モダリティ入力を等しく処理するのではなく，状況に応じた最適な重み付けをおこなっているであろう。

文 献

- 1) H. McGurk and J. MacDonald: Hearing lips and seeing voices. *Nature*, 264, 746-748, 1976.
- 2) R. Sekuler, A. B. Sekuler and R. Lau: Sound alters visual motion perception. *Nature*, 385, 308, 1997.
- 3) L. Shams, Y. Kamitani and S. Shimojo: What you see is what you hear. *Nature*, 408, 788, 2000.
- 4) 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.
- 5) L. D. Rosenblum and H. M. Saldaña: An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 318-331, 1996.
- 6) D. W. Massaro and M. M. Cohen: Perceiving speech from inverted faces. *Perception and Psychophysics*, 58, 1047-1065, 1996.
- 7) P. K. Kuhl and A. N. Meltzoff: The bimodal perception of speech in infancy. *Science*, 218, 1138-1141, 1982.
- 8) L. D. Rosenblum, M. A. Schmuckler and J. A. Johnson: The McGurk effect in infants. *Perception and Psychophysics*, 59, 347-357, 1997.
- 9) D. W. Massaro, L. A. Thompson, B. Barron and L. Laren: Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, 41, 93-113, 1986.
- 10) N. S. Hockley and L. Polka: A developmental study of audiovisual speech perception using the McGurk paradigm. *Journal of the Acoustical Society of America*, 96, 3309 (4aSPb6), 1994.
- 11) K. Sekiyama: Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception and Psychophysics*, 59, 73-80, 1997.
- 12) D. W. Massaro: Perceiving Talking Faces: From Speech Perception to a Behavioral Principle. MIT Press, 1998.
- 13) 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.