

# 視聴覚的な発話認知における位置、音韻知覚の共通性

金谷翔子  
横澤一彦

東京大学大学院人文社会系研究科  
東京大学大学院人文社会系研究科

Auditory-visual localization and identification processes, represented in ventriloquism and the McGurk effect, are assumed to involve independent mechanisms. We tested this hypothesis using an audio-visual matching task, in which participants localized one talker (from two faces) and identified his/her speech. Respectively, these judgments reflect ventriloquism and McGurk effect. Variables were spatial discrepancy between sound source and face of congruent stimuli (collocated or separated) and speech syllables (/pa/ versus /ka/). The results showed that spatial discrepancy affected both talker localization and speech identification. Moreover, the outcome of talker localization significantly altered perceived syllables. Certain common properties in perceiving "where" and "what" of a multisensory object were suggested.

Keywords: speech perception, McGurk effect, ventriloquism.

## 問題・目的

音源定位と音韻認知における視聴覚情報の利用を反映する腹話術効果とMcGurk効果は、類似の現象であるにも関わらず、相互独立な機構に基づくものと考えられてきた。Bertelson et al. (1994)は、発話している顔動画と音声刺激の提示位置の空間的距離、また顔動画の正立または倒立が、腹話術効果とMcGurk効果に与える影響を検討した。実験参加者の正面に置かれた一つのモニターと、半円状に配置された複数のスピーカから視聴覚刺激を提示し、知覚された音源位置と音韻の判断を求めたところ、定位の精度は刺激の空間的距離のみに、音韻の判断は顔動画の正倒立のみに規定された。しかしDriver(1996)では視聴覚刺激の空間的關係が音声の選択的聴取の効率に影響を与えていることから、視聴覚的な音源、音韻認知の関係について再検討する必要がある。

腹話術効果は多感覚的なオブジェクトの知覚を反映すると考えられており、よって音源定位と音韻判断を同時に行なうことはすなわち一つのオブジェクトとしての話者の位置、発話内容を認知することに等しい。そこで本研究では一種のマッチング課題を用い、二つの顔動画に加え、複数音源のうち一箇所から発せられる音声刺激を提示した上で、視聴覚情報の起源としての話者とその発話を同時に判断させた。独立変数は同じ音韻を表す視聴覚刺激の空間的位置関係、また音声刺激が表す音韻の種類であった。先行研究からは、それぞれが音源定位と音韻判断を規定することが示されている。一つのオブジェクトとしての話者と発話を認識する状況下で、腹話術効果とMcGurk効果が共通の基盤を持つとすれば、これらの二要因のうち一方または両方が音源と音韻の判断の双方に影響を与えることが予想される。

## 方法

### 実験参加者

20歳から29歳までの日本語母語話者（平均年齢22.9歳）が実験に参加した。

### 刺激と条件

視覚刺激は参加者正面のスクリーンに投影された左右二つの顔動画で、それぞれが/pa/と/ka/を発話しているものだった。聴覚刺激は、左右の視覚刺激のすぐ下に置かれた二つのスピーカのうち片方から発せられた、/pa/または/ka/の音声であった。要因は同じ音韻を表す視聴覚刺激の提示位置（同位置・逆位置）、聴覚刺激が表す音韻の種類（/pa/・/ka/）の二つであった

(Figure 1)。同位置条件では、例えば/pa/を発話している顔動画の下に置かれたスピーカから、/pa/音声提示された。逆位置条件では、例えば/pa/を発話している顔側のスピーカから/ka/音声提示された。聴覚刺激が/pa/の場合は、/pa/と/ka/を発話する二つの顔動画を観察することにより、音韻の錯覚(/ta/)が起こる可能性がある(McGurk効果)。これに対し聴覚刺激が/ka/の場合、日本語母語話者の被験者においては音韻の錯覚が生起しないことが先行研究より示されている(Sekiyama and Tohkura, 1993)。

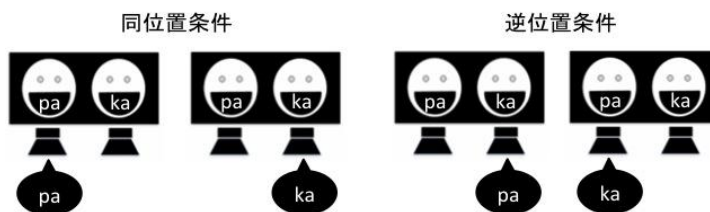


Figure 1. 刺激の提示条件

実験手続き

スクリーンの中央に注視点（100ミリ秒）が現れ、50ミリ秒後に視覚刺激と聴覚刺激が同時に提示された。刺激の提示時間は平均1500ミリ秒であった。参加者は、刺激提示中も注視点の位置（二つの顔動画の間）に視線を固定するよう教示され、これによって二つの顔動画の両方を中心視野で観察した。刺激提示後、どちらの顔刺激が話者と感じられたかについて、左右の二択で回答を求めた。続いて何という音韻を知覚したかについて、/pa/、/ka/、/ta/の三択により回答を求めた。条件は全て一ブロック内でランダム順に繰り返され、実験は全部で288試行であった。

結果

話者の定位において、実際の音源（スピーカ）側を回答した割合を Table 1 に示す。刺激の提示位置（同位置・逆位置）、提示音声（/pa/・/ka/）の二要因分散分析を行ったところ、提示位置の主効果が有意であった $[F(1, 23) = 35.40, p < .05]$ 。同位置条件に比べ、逆位置条件では非音源側の顔を話者として選択する割合が増加している。また、音韻判断において実際の提示音声と同じ音韻を回答した割合（正答率）を Table 2 に示す。刺激の提示位置（同位置・逆位置）、提示音声（/pa/・/ka/）の二要因分散分析を行ったところ、提示位置の主効果、提示音声の主効果および、それらの交互作用が有意であった $[F(1, 23) = 29.18, p < .05; F(1, 23) = 20.43, p < .05; F(1, 23) = 18.57, p < .05]$ 。音韻判断における錯覚は提示音声/pa/においてのみ、また同位置条件よりも逆位置条件において高い頻度で生じた。さらに音声/pa/が提示された試行における音韻判断の正答率について、話者として選択された顔（/pa/・/ka/）を要因としたウィルコクソンの符号順位検定を行った（Table 3）。同位置条件、逆位置条件の両方において、/ka/の顔が選択された場合は/pa/の顔が選択された場合よりも正答率が有意に低下した $[z = 2.56, p < .05; z = 3.28, p < .05]$ 。

Table 1. 音源側の顔を話者と判断した割合(%)

	/pa/	/ka/
同位置条件	0.91	0.92
逆位置条件	0.55	0.57

Table 2. 音韻正答率(%)

	/pa/	/ka/
同位置条件	0.84	0.98
逆位置条件	0.73	0.98

Table 3. 提示音声/pa/に対する音韻正答率

	/pa/(顔)	/ka/(顔)
同位置条件	89.41	55.82
逆位置条件	86.26	56.52

考察

逆位置条件よりも同位置条件において話者の定位がより正確であったこと、また逆位置条件では聴覚刺激と同じ音韻を発話している視覚刺激の側に誤定位する傾向が見られたことは、多感覚オブジェクトの定位における視覚情報の利用すなわち腹話術効果を反映している。しかしこの傾向は提示音声の影響を受けなかった。視覚刺激のみでは/ta/、/ka/の発音を弁別するのが困難であるため、提示音声/pa/に対し/ta/という錯覚が生じた場合は提示される二つの顔がどちらも聴覚刺激と整合するが、提示音声/ka/に対しては必ず/ka/の視覚的発話のみが整合する。このことから、話者定位は最終的に知覚された音韻ではなく物理的に提示された刺激音声と視覚刺激の対応に基づいて行われたと考えられる。音韻判断は聴覚刺激の種類によって規定されるだけでなく、刺激の提示位置条件によっても錯覚の起きやすさが変化した。これは本実験において視聴覚刺激の空間的位置関係が、話者の定位のみならず音韻の判断をも規定していたことを示唆するものである。さらに提示音声/pa/に対し、選択された顔によって知覚音声が有意に異なっていたことから、話者定位の結果が音韻判断に影響を与えたと考えられる。これらの結果は、一つのオブジェクトとしての話者とその発話を認識する状況下で、定位と音韻判断が共通の機構または要因に基づいて行われている可能性を示す。また定位と音韻判断を規定する要因の非対称性は、これらの処理が独立または並列に行われるのではなく、系列的に行われている可能性を示唆するものとも考えられる。

引用文献

Bertelson, P., Vroomen, J., Wiegand, G., & de Gelder, B. (1994). Exploring the relation between McGurk interference and ventriloquism. *Proceedings of the International Conference on Spoken Language Processing* (vol. 2, pp. 559-562). Yokohama.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, 381,

Sekiyama, K. & Tohkura, Y. (1993). Inter-language differences in the influences of visual cues in speech perception. *Journal of Phonetics*, 21, 427-444.