

視聴覚刺激の音韻整合性が腹話術効果に与える影響

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

The ventriloquism effect, a shift of perceptual location of a sound source toward a synchronized visual stimulus, is known to be regulated only by physical, but not cognitive, factors. Though this conclusion is based upon simplified experiments which present single audio-visual stimuli, it may not fully explain our performances in a real world with many signals in each sensory modality. The hypothesis that cognitive factors, not only physical ones, modulate this effect in complex situations was examined. We compared simplified (Experiment 1) and complex (Experiment 2) experimental designs. Auditory stimulus was always a voice, but visual stimuli were a single movie of a face in Experiment 1, and two bilateral movies in Experiment 2. A cognitive factor, congruency of audio and visual speech information, was varied (same/different syllable). In Experiment 1, we found no impact of congruency, while in Experiment 2 audio-visual congruency elicited relatively large auditory localizations biases. A cognitive factor can affect the way audio-visual spatial information is integrated in a more complex, real world, situation.

Keywords: audio-visual integration, ventriloquism, speech perception.

問題・目的

視聴覚刺激を同時に提示した時に聴覚刺激の位置が視覚刺激の方向へ近づいて知覚される現象を、腹話術効果と呼び、感覚統合的な空間知覚を反映すると考えられている (Vroomen & de Gelder, 2004)。腹話術効果は従来、視聴覚刺激の時空間的な近接性または刺激の強度、顕著性といった物理的特性のみに規定され、刺激同士の意味的関連性や日常経験における共起関係といった認知的要因は関与しないと考えられてきた

(Radeau & Bertelson, 1977)。これらの知見は殆どが視覚または聴覚にそれぞれ一つずつの刺激を提示する単純な実験状況において得られた結果に基づくものだが、我々の日常世界では各モダリティに様々な刺激が存在しており、このような複雑な状況における本来の感覚融合過程ではより多くの手がかりが用いられる可能性が指摘されている (Kanaya & Yokosawa, in press)。本研究では、一つの聴覚刺激に対して複数の視覚刺激を提示した場合、物理的要因のみならず認知的要因も腹話術効果の大きさを規定するという仮説を検証する。そこで、実験1では従来研究のような単純な状況、実験2ではより複雑な状況を用いて音源定位課題を行った。用いた刺激は音声、および発話を行っている顔の動画である。どちらの実験においても認知的要因の一つとされる、音声と発話動画の音韻的整合性を操作した。仮説が正しければ、実験2においてのみ音韻の整合性が腹話術効果の大きさに影響すると考えられる。

方法

刺激

刺激は日本人女性による/ka/または/pa/の音声、/ka/または/pa/と発声している動画であった。

視覚刺激はCRTディスプレイの中央に現れる注視点の左右いずれかの側に一つ提示され、この動画と音声との組み合わせを操作した。どちらも同じ音韻を表す場合 (同音) と、異なる音韻を表す場合 (異音) があ

った。なお、実験1ではその反対側に何も提示されなかったが、実験2では反対側にもう一つ全く同じ動画が付加された。ただし、付加された動画は音韻情報を持たないよう、口元に白いマスクをかけられていた (figure 1)。聴覚刺激は左右2つのスピーカから発せられる音声信号の位相差 (1単位を $41.66 \mu\text{ms}$ とし、左チャンネルの先行を-、右チャンネルの先行を+で表す) によって仮想的に作られた13ヶ所の音源のうち、1ヶ所から提示された。これらの仮想音源は参加者の正面に対して左約24度 (位相差-6) から右約24度 (位相差+6) までの範囲に設定された。



figure 1. 実験1 (左) と実験2 (右) の視覚刺激例

手続き

課題はディスプレイの中央の注視点に対して、音源がその左/右のどちら側かを判断するものであった。実験中は注視点から目を逸らさないように、しかし回答は視覚情報によらず純粋な聴覚判断に基づいて行うように教示を行った。

実験1、2ともに、聴覚刺激の提示された13ヶ所の仮想音源の各所において、音源を右とする反応の比率を累積正規分布曲線に当てはめ、50%反応閾を求めた。これは主観的に正面と知覚される実際の音源位置 (主観的中心点) に相当する。これを左右チャンネルの位相差の単位 ($41.66 \mu\text{ms}$ /単位) で表したものを主観的中心点の値とした。

このような実験状況において聴覚的的定位は顔動画 (実験2においてはマスクをかけていない方の動画) により近い位置に錯覚されると考えられ、例えば動画が注視点の左側に提示される場合には定位は全体的に左

寄りになる。この時、主観的中心点は反対の右側に移動するため、視覚情報による定位バイアスの大きさは主観的中心点の移動量で表される。そこで、顔動画が左側に提示されていた場合（主観的中心点は相対的に右寄り）における主観的中心点の値から、顔動画が右側に提示されていた場合（主観的中心点は相対的に左寄り）における主観的中心点の値を引いたものを、各条件における腹話術効果の効果量とした。

実験参加者

実験1には18歳から29歳まで（平均年齢22.7歳）の12名、実験2には20歳から54歳まで（平均年齢27.9歳）の14名が参加した。全員が正常な視力、聴力を持つ日本語母語話者であった。

結果

実験1における腹話術効果の効果量をfigure 2に示す。エラーバーは標準誤差を表す。12名の参加者のうち、予備実験で、位相差が±6単位より大きい条件において音源の左右弁別課題の成績が75%以下であった3名、および課題を正しく理解していなかった1名を除く、8名のデータに関して分析を行った。

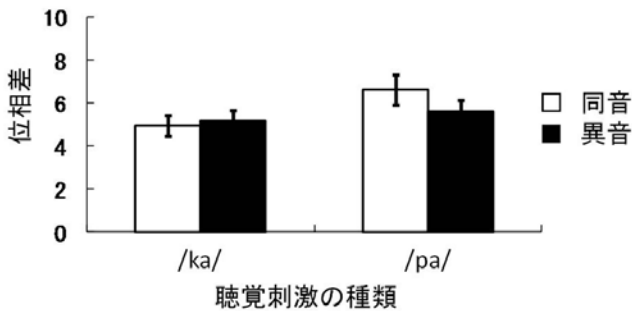


figure 2. 実験1における腹話術効果の効果量

実験2における腹話術効果の効果量をfigure 3に示す。エラーバーは標準誤差を表す。14名の参加者のうち、予備実験で、位相差が±6単位より大きい条件において音源の左右弁別課題の成績が75%以下であった4名、および課題を正しく理解していなかった1名を除く、8名のデータに関して分析を行った。

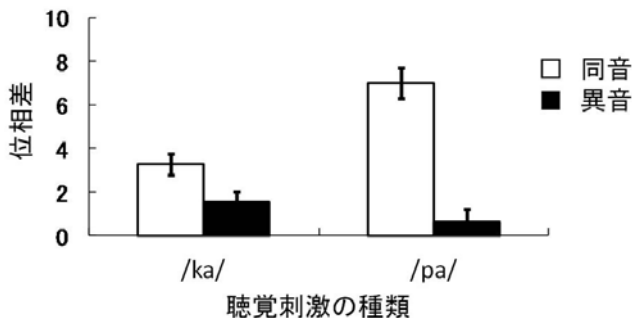


figure 3. 実験2における腹話術効果の効果量

実験1、2ともに聴覚刺激の種類 (/pa/・/ka/)によって音源定位の精度が異なっていたため、視聴覚刺激の組み合わせ（同音・異音）、聴覚刺激の種類 (/pa/・/ka/)を要因とする2要因分散分析を行った。実験1の結果についてはいずれの要因の効果および交互作用も見られなかった。一方、実験2では、組み合わせ、聴覚刺激の種類の主効果が共に有意であった ($F(1,8) = 52.98, p < .05, \eta^2 = .43, F(1,8) = 8.50, p < .05, \eta^2 = .05$)。同音条件では異音条件よりも、また聴覚刺激が/pa/の場合は/ka/の場合よりも、大きな効果が得られた。またこれらの交互作用も有意となった ($F(1,8) = 28.80, p < .05, \eta^2 = .14$)。視聴覚刺激の組み合わせの単純主効果はどちらの聴覚刺激条件においても有意であり、一方、聴覚刺激の種類の効果は同音条件においてのみ有意であった。

考察

顔動画（実験2においてはマスクをかけていない動画）の提示された方向への音源定位のずれ、すなわち腹話術効果を、音源の主観的中心点が顔動画と反対方向へ移動する量として定義した。本実験では腹話術効果の効果量が視聴覚刺激の表す音韻の整合性（同音・異音）に規定されるか否かを、2つの実験において検討した。注視点の片方に一つだけ動画を提示した実験1では、視聴覚的な音韻情報の組み合わせの異同は腹話術効果の効果量に影響を与えなかった。一方、注視点の反対側にもう一つの顔刺激を提示した実験2では、同音条件において異音条件よりも大きな腹話術効果が得られた。このことから、従来研究で用いられてきたような単純な実験状況と異なり、一つのモダリティに複数の刺激を提示するような複雑な実験状況においては、認知的要因が腹話術効果を規定するという可能性が示唆された。モダリティ内に複数の刺激が存在する時、多感覚情報の融合過程においては、まず一つのイベントおよびオブジェクトとして統合すべき刺激対を検出する必要がある。例えば本研究の実験2においては、一つの聴覚刺激が左右どちらの顔と統合されるのか、選択および判断を行うことになる。このような過程においては刺激の物理的要因のみならず、音韻整合性等の認知的要因も手がかりとして用いられるものと考えられる。

引用文献

Kanaya & Yokosawa (in press). Perceptual congruency of audio-visual speech affects ventriloquism with bilateral visual stimuli, *Psychonomic Bulletin & Review*.
 Radeau, M., & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics*, 22, 137-146.
 Vroomen, J., & Gelder, B. (2004). Perceptual Effects of Cross-modal Stimulation: Ventriloquism and the Freezing Phenomenon. In Calvert, G, Spence, C., and Stein, B. (Eds.), *The Handbook of Multisensory Processes* (141-150). Cambridge MA: MIT Press.