音声知覚における視聴覚統合は知覚的同期と独立である

○北川智利¹・持田岳美¹ (非会員)・北村美穂¹

(¹NTT コミュニケーション科学基礎研究所)

キーワード: 視聴覚統合, 同期知覚, 音声知覚

Audiovisual integration in speech perception is independent from perceptual synchrony between audiovisual signals

Norimichi KITAGAWA¹, Takemi MOCHIDA^{1,*} and Miho KITAMURA¹

(¹NTT Communication Science Laboratories)

Key Words: Audiovisual integration, Perceptual synchrony, Speech perception

目 的

二つの感覚モダリティからの信号を統合する時, 信号間の時間的な同期が重要な手がかりとなる。しかしながら, 二つの信号の同期が知覚されることと, 二つの信号が統合されて知覚されること, この二つの知覚の依存関係は十分に明らかにされていない。マガーク錯覚を例にしてみよう。「パ」という音声と「カ」という発話の映像を組み合わせると, 両者が統合されて「ダ」と聞こえることが多い。この視聴覚統合は, 声と映像が同期していると知覚された時に生じるのだろうか, あるいは, 同期の知覚と, 視聴覚統合は独立なのだろうか。交差/反発錯覚を用いた最近の研究では, 視覚事象と聴覚事象が同時に知覚された時に (物理的な時間ずれによらず), 視聴覚統合が生じることが示されている (図1左; Kitagawa & Kitamura, 2014)。本研究では, 音声知覚でも同様に, 視聴覚統合が知覚的同期に基づいて生じかどうかを検討した。

方 法

実験参加者: 15名 (21~42歳の男性7名, 女性8名)。
刺激: 男性話者1名が発話した「パ」「タ」「カ」(各4回ずつ)を撮影, 録音した。このうち, マガーク錯覚が典型的に生じる「パ」の音声に「カ」の映像を組合せたもの (4×4=16組合せ) を検査刺激とした。また, 音声と映像のオリジナルの組合せ (パ, タ, カ, 各4種類) をダミー刺激として用いた。映像はディスプレイ (SONY, HMD-H200) に提示し, 音声はヘッドホン (Sennheiser, HDA200) から両耳に提示した。
手続き: 音声と映像は, 13種類の時間ずれ (Stimulus onset asynchrony (SOA): ±500, ±400, ±300, ±225, ±150, ±75, 0 ms: 負の値は音声が行先) を伴って提示された。実験参加者は2種類の課題を行った。**音声同定課題**では提示された音声は「パ」「タ」「カ」のいずれに聞こえたかを回答した。**同期判断課題**では音声と映像が同期していたか否かを回答した。実験では, まず二つの課題を個別の実験ブロックで行い (単一課題), その後の実験ブロックで, 各試行に対して二つの課題をランダムな順序で行った (二重課題)。二重課題では, 検査刺激の各SOAに対して64~88試行を行った。それとは別に, 分析対象とならないダミー試行が全試行の約30%提示された。

結 果

音声同定課題でのマガーク反応, 同期判断課題での同期反応のいずれかが極端に少なかった (全体の10%未満) 2名, および, 得られたデータへのガウス関数の当てはまりが悪かった2名 ($R^2 < 0.6$) の4名を分析から除外した (平均 R^2 は0.86)。得られたガウス関数のパラメータは, 音声同定課題においても, 同期判断課題においても, 単一課題条件と二重課題条件で有意な差はなかったことから, 二重課題でも単一課題と同じように課題を遂行できていることが示唆される。

マガーク錯覚の生起 (視聴覚統合の生起) が同期知覚に依

図1. SOAに依った反発錯覚生起率 (左) とマガーク錯覚生起率 (右) の平均値と標準誤差

存しているかどうかを検討するために, 二重課題の試行を同期判断に基づいて主観的**同期**試行と主観的**非同期**試行に分け, マガーク錯覚の生起率 (「タ」の回答率) をそれぞれ求めた (図1右)。時間ずれ (SOA) と主観的同期性を被験者内要因とする二元配置分散分析の結果, SOAの主効果のみが有意になった ($F(12,120)=9.368, p<0.0001$)。マガーク錯覚は, 音声と映像が物理的に同期しているほど生じやすく, 時間ずれ (SOA) が大きくなるほど生じにくいことが分かる。また, 主観的同期試行と主観的非同期試行では, マガーク錯覚の生起率に統計的に有意な差はなく ($F(1,10)=4.226, p=0.0669$), SOAに対して同じような関数となっている。

考 察

マガーク錯覚は, 音声と映像が同期していると知覚されたか否かに関わらず, 両者が物理的に同期しているほど良く生起した。つまり, 音声と映像が同期していないと感じたとしても, 物理的に同期していれば, マガーク錯覚は生じやすい。この結果は, 音声知覚における視聴覚統合が音声と映像の物理的な同期に依存していて, 知覚的な同期とは独立であることを示している。音声知覚においては, 音声と映像の同期を検出する処理と, 声と口の動きの映像を統合して音声を聞き取る処理とは独立であることが示唆される。

一方で, 交差/反発錯覚を対象にした研究 (Kitagawa & Kitamura, 2014) では, 反発錯覚 (視聴覚統合) は, 知覚的な同時性に依存して生じ, 物理的な時間ずれには依存しないことが示されている (図1左)。この違いがどのような視聴覚統合処理の違いを反映しているのか, 今後の検討が待たれる。

引用文献

Kitagawa, N. & Kitamura, M. (2014) The stream-bounce illusion depends on subjective audiovisual simultaneity. 15th International Multisensory Research Forum, Amsterdam, The Netherlands.