



Kobe Shoin Women's University Repository

Title	時間知覚における視聴覚情報の相互作用
Author(s)	Mark Scott Katrin Dohlus Gabor Pinter
<i>Citation</i>	Theoretical and applied linguistics at Kobe Shoin, No.8 : 133-142
Issue Date	2005
Resource Type	Bulletin Paper / 紀要論文
Resource Version	
URL	
Right	
Additional Information	

時間知覚における視聴覚情報の相互作用*

Mark Scott[†], Katrin Dohlus[‡], and Gábor Pintér[§]

Seeing geminates and hearing singletons: A pilot study

Abstract

This paper presents the results of a pilot study on the McGurk effect for consonantal length distinctions. Subjects saw a video of a speaker pronouncing a geminate consonant and simultaneously heard an audio consonant that varied in length from trial to trial. Despite the fact that subjects showed a fusion of visual and auditory information with respect to place-of-articulation, no such fusion occurred for phonological-length information. This suggests that different forms of visual information are integrated into speech perception differently. This is perhaps because duration-perception is more reliable in the auditory modality and so any disagreement between audition and vision with respect to the duration of a stimulus results in the visual information being disregarded.

1. 序

言語知覚に際して脳は、耳を通して入ってくる聴覚的な情報のみならず、目を通して入ってくる情報にも頼る。つまり、脳は一つのモダリティからだけでなく、両方のモダリティから得た情報を自動的に統合しているわけである。言語知覚において視覚情報が重要であることは、特に雑音が多いといった聴覚情報の取得が困難な場면을例にすると明白である (Schwartz et al. 2004)。

耳と目から入ってくる情報は、普通日常場面においては一致するが、実験的に、この二種類の情報流れを分離したり、一致しないように編集したりすることが可能である。このような実験は、脳が視覚か聴覚のどちらか一方のモダリティを優先させるのではなく、折衷の音を知覚するという興味深い結果をもたらす。例えば、[ba] を発音する顔面

*本研究を取り込むにあたって、松井理直先生、松田謙次郎先生、榎園久美子氏にご協力をいただきましたことを心より感謝いたします。

[†] 神戸松養女子学院大学大学院研究生

[‡] 神戸大学大学院研究生

[§] 神戸大学大学院生

の動画を、時間的に一致する [ga] という音声と組み合わせると、[da] という音声の知覚が生じる。この知覚される [da] の子音の調音点は、視覚的な動画情報である [ba] の両唇音や、音声の [ga] の軟口蓋音ではなく、その間の歯茎音になるわけである。この現象は McGurk 効果として広く知られている (McGurk and MacDonald 1976)。以下は調音点における McGurk 効果を「伝統的な McGurk」として参照することにする。

これまで、言語知覚に視覚情報が寄与する現象としては、特に調音点の知覚に注目した研究が行われてきた (Colin and Radeau 2003)。しかし、子音の長さを弁別する場合にも、視覚情報が影響を与えるのかについては、現在まで研究が行われていない。その一つの理由として、従来の実験では、被験者が英語母語話者であったことが考えられる。英語には、子音の長短の区別がないため、子音の長短は実験対象とされてこなかった。しかし、日本語においては子音の長短が意味を区別する機能を持つ。このことから、日本語は視覚情報が子音の長さの知覚に影響を与えるのかを確認するのに適切な言語だと言える。

調音点の知覚に関わる視覚的情報と子音の長さの知覚に関わる視覚的情報とは、異なる性質を持ったものである。異なる音の調音点は異なる顔面の形状で表現され、それらの知覚は「形認識」という範疇に入る。しかし、音の長さの弁別においては、顔面の形状の違いではなく、顔面の形状の継続時間の違いで表現される。つまり、視覚情報の観点から観ると、調音点の知覚は「形認識」の問題であり、音の長さの知覚は基本的に「継続時間」認識の問題だと言える。過去の研究は言語知覚に「形認識」が寄与することについては十分に証明してきた。しかし、視覚的な継続時間の認識が、言語知覚に寄与するのかについては未だ研究が行われていない。

しかしながら、本研究では必ずしも調音点における McGurk 効果と同時に、音韻の長短における McGurk 効果が現れるとは仮定していない。伝統的な調音点の McGurk 効果が現れるかどうかに関わらず、視覚的情報を長短弁別の知覚に際して聴覚的な情報と統合する日本人話者もいると予測する。言い換えると、McGurk 効果は音声の長短においても働くと思われる。

この仮説を確認するために、母音間の子音の長短知覚に関して以下のような実験条件を設定した。一つ目は聴覚情報のみを被験者に与える条件 (AO: audio-only) であり、二つ目は聴覚情報と同時に視覚情報を被験者に与える条件 (AV: audio-visual) であった。もし、聴覚情報のみ AO 条件と、視覚情報のある AV 条件との間に、被験者の子音の継続時間知覚に違いがあるならば、視覚的な情報が音声の長短の知覚にも影響するという証拠になる。

2. 実験の設定

ターゲットワードとして、長子音の場合も、短子音の場合にも日本語として無意味の単語を用意した。音と動画のずれが子音の長短の知覚において変化をもたらすかを調べるために、二種類の刺激のセットを用意した。一つ目は、母音間の長子音の閉鎖時間を 4ms きざみで短くした音声のセットであった (AO 条件)。二つ目の刺激セットでは、音声

と動画が同時に呈示され、音声は一つ目の音声セットと同様であり、また動画はすべて長子音のターゲットワードを発声した画像を用いた（AV条件）。これらの刺激は、関西出身の25歳女性が「マップ」というターゲットワードを「昨日行った学会のモットーは<マップ>」という文で発話するのをビデオカメラで撮影し作成した。

2.1 刺激

刺激は防音室で作成し、ターゲット人物がターゲット文を三回繰り返す場面を撮影した。発話の速さは厳密にコントロールしなかったが、ターゲット人物にできるだけ同じスピードで話すように指示した。録画した三つの動画の中で、もっとも自然に聞こえる刺激のみを使用した。

撮影した刺激をビデオ編集ソフトウェア（AviUtl98d）を使用して、音声と動画に分離した。分離後、音声部分は、ストリームのターゲット刺激（[mappa]）という部分のみを変更した。なお、動画刺激には変更を加えなかった。変更する前の長子音の閉鎖時間は180msであった。また、閉鎖時間は、音声のスペクトグラムにおいても無声の部分に決定した。長子音の閉鎖時間を、スピーチ編集ソフトウェア（Praat）を使用して、160msから76msまで4ms刻み間隔で縮め、22個の刺激を作成した。予備調査において、180msから28msの刺激を4ms刻みで39個作成し、被験者に短子音と長子音の識別テストを行った。その結果、未編集のターゲット刺激の閉鎖時間180msと160msのターゲット刺激の閉鎖時間は、同様に促音と判断されたため、160msのターゲット刺激を最長閉鎖時間とした。短子音についても、30msのターゲット刺激と76msのターゲット刺激の短子音は同様と判断されたため、76msの刺激を最短閉鎖音とした。したがって、長子音を含むターゲット刺激（[mappa]）から短子音を含むターゲット刺激（[mapa]）までの22段階の刺激のセットとなった。

視聴覚の刺激（AV条件）を作成するために、22個の音声刺激を未編集の動画と組み合わせた。22個の動画は全て同一のもので、音声刺激は長子音の[mappa]から短子音の[mapa]に段階的に縮めた刺激であった。

ターゲット刺激は、無意味単語[mappa]と[mapa]を用いた¹。その理由は3つある。一つ目は、語彙頻度などの効果を避けるためであった。二つ目は、視覚的な情報を被験者が容易に受け取れるように、子音の中で視覚的に一番目立つものである両唇音にした（[m, p]）。三つ目も同様の理由で、母音の中で顎開きがもっとも大きい[a]にした。日本語には、母音間に短子音の両唇無声閉鎖音が稀にしか現れないという反論はあるかもしれないが、外来語及び擬態擬声語に母音間の[p]は頻繁に現れるので、日本語音韻体系の一つであるといえる。

更に、予備調査において、一般に[mappa]と[mapa]のアクセントパターンに差がなかったため、長子音と短子音のターゲット刺激の区別の際、アクセント情報を手がかりにする恐れがないと分かった。

¹ 「マバ」は日本語に存在しない単語である。「マップ」（末波）は辞書に記載されているが、非常に稀な単語で実験被験者も知らなかった

2.2 実験手続き

実験は三人のボランティアの被験者（NK, GA, TK）に対して行われた。いずれも関西地方の母語話者で、視力は健全で（または眼鏡で矯正済み）聴力も異常がなかった。

全ての被験者に対して、約 20 分かかる 6 回のセッションの実験を行った。六つのセッションのいずれにおいても、22 個の視聴覚的な刺激 (AV 条件) と 22 個の聴覚のみの刺激 (AO 条件) を 4-6 回（被験者のスケジュールに対応して）提示した。従って、各刺激は全てのセッションにおいて合計約 30 回流されたことになる。AO で始まるセッションと AV で始まるセッションを交互に行った。

実験は防音加工がなされた部屋で行われた。視聴覚の刺激の場合は、刺激を表示するコンピュータが良く見える距離で画面の前に被験者を座らせ、スピーカーを画面左右の上に被験者に向けて設置した。被験者には、画面に現れてくる人物の顔面に集中し、コンピュータのマウスを使って、文の最後に聞こえてくるターゲットワードに応じて、用意した [mappa] または [mapa] のボタンをクリックように指示を与えた。

最初のセッションの前には、伝統的な McGurk 効果及び、被験者の読唇術を確認するための 2 つの簡単なテストも含まれていた（参照：Kubozono 2002）。まず、調音点の視聴覚的な知覚を観察するために、動画の [aba] と音声の [aga] を組み合わせた刺激を使用した。この刺激を 4 回被験者に流し、解答用紙にある “aga”、“aba”、“ada” の中から聞こえるものに丸を付けるように指示を与えた。又、[aba] という音声のみの刺激を使用した実験も行った。もし、聴覚のみの刺激の場合において被験者が [aba] だと知覚し、編集した視聴覚刺激の場合において、[ada] であると解答すれば、McGurk 効果が確認されたことになる。

本実験の前のもう一つのテストは、被験者の読唇術能力を調べる実験であった。子音の長短が視覚的に区別できるのかを確認するために、ターゲットワードの [mappa] と [mapa] の無声動画を 6 回流し、被験者に解答用紙に記述した “mappa” と “mapa” の中、画面に現れる人物が発音していると思われる方に丸を付けるように指示を与えた。

3. 結果

3.1 調音点の McGurk 実験の結果

被験者 NK 及び TK において伝統的な強い McGurk 効果が確認できた。被験者 GA では、調音点の知覚にはある程度視覚的な情報も影響を与えたが、視覚的な情報と聴覚的な情報を統合した回答はなかった。具体的には、回答の一部は [aga] であり、他の回答は [aba] というもので、伝統的な McGurk 実験に現れる聴覚的な情報と視覚的な情報を統合した [ada] という回答はなかった。更に AV 条件において、被験者 NK と GA は違和感を報告したが、被験者 TK は内観報告しなかった。

3.2 読唇術テストの結果

全ての被験者において、無声動画の [mappa] と [mapa] の区別が正確にでき、長短の弁別は視覚的な情報だけに頼って可能になるということが明らかになった。

3.3 継続時間の McGurk 効果の実験結果

被験者から得たデータは、聴覚刺激のみの AO 条件に対する回答と視聴覚刺激の AV 条件に対する回答とを、対応する刺激毎に合わせ 22 個のペアに、有意水準 α を 0.05 に設定し、 χ^2 検定にかけた。その結果、長短子音の閉鎖時間の知覚に関して、聴覚情報の AO 条件と視聴覚条件の AV 条件での回答との間に有意な差はなかった。つまり、全ての被験者は短子音と長子音の間の閉鎖時間の境界線を聴覚のみの場合と視聴覚刺激の場合と、どちらも同様に判断したというわけである。

各被験者の回答にはセッション毎に異なる傾向があり、あるセッションでは、子音の長短の境界線が視聴覚的な刺激の際、視覚のみの刺激と比べて一つの方向に移動したり、他のセッションでは反対の方向に移動したりすることもあった。しかし、全てのデータを合わせると、全体的に子音の長短境界線を移動させる傾向がないと分かった。提示した刺激の順序（最初が AO か AV か）も回答に影響を及ぼさなかったと考えられる。以下の図は、各被験者から得たデータをまとめたものである。

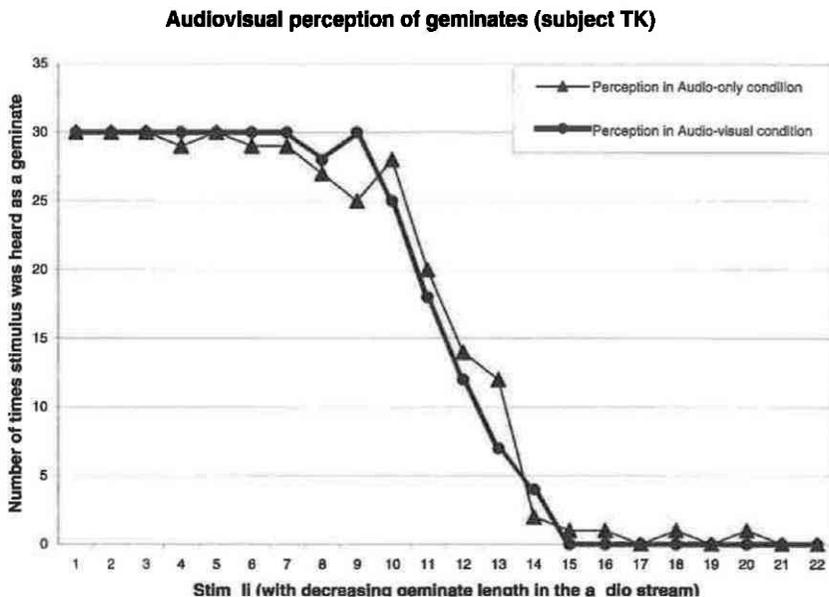


図 1: TK 被験者の実験結果

4. 議論

上述の結果で分かるように、予想に反して時間の弁別における McGurk 効果が確認できなかった。三人の被験者中二人において、調音点に関する伝統的な McGurk 効果を確認したにも関わらず、すべての被験者において、子音の閉鎖時間に関しての McGurk 効果は確認できなかった。いくつか考えられる理由を以下で述べる。

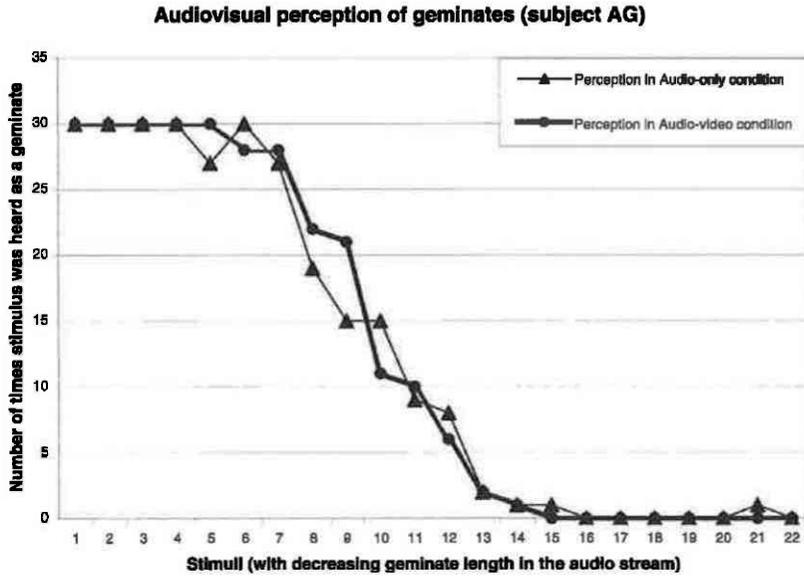


図 2: AG 被験者の実験結果

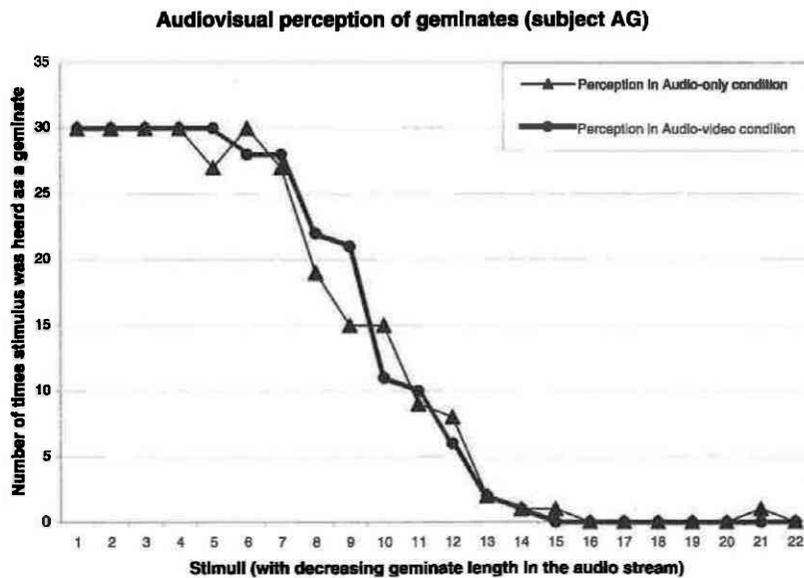


図 3: KN 被験者の実験結果

4.1 視覚対聴覚知覚

視覚的情報は、調音点の知覚には影響を与えたが、音素の長さ知覚には影響を与えなかった。このような違いから、言語知覚においていずれの視覚的情報も同様に扱われているわけではないことがわかる。調音点は、顔面と口の異なる形状によって視覚的に区別され、基本的に「形認識」の過程として解釈できる。一方、音韻的な長さは顔面の形状の継続時間によって視覚的に知覚され、ここでは形認識が役割を果たさないわけである。

視覚的情報は多量のニューラルプロセッシングを必要とするため (Levine and Shefner 2000:347)、視覚的情報を得るためには、聴覚的情報を得る場合よりも時間がかかるわけである (Welch and Warren 1986)。視覚情報の処理が相対的に遅い理由として、「杆体細胞と錐体細胞の光化学的なプロセスが比較的遅い」(Welch and Warren 1986) ことが考えられる。即ち、言語知覚過程において、視覚情報を処理する生理学的なプロセスは、聴覚情報を処理するプロセスに比べて遅いと言える。

このように視覚プロセスは聴覚プロセスと比べ遅いことから、視覚は時間的な情報に関して、聴覚ほど信頼性が高くないと言える。従って、視覚と聴覚の間に時間的な不一致が起こると、脳は自動的に信頼性の高い聴覚の方を優先し、一致しない視覚情報を無視するのである。このような時間的矛盾が起こった場合に聴覚が優先されるという事実は、既に Droit-volet et al. (2004)、Wada (2003)、Wearden et al. (1998)、Goldstone and Lhamon (1974) などによって明らかにされている。ただし、時間知覚における優先性は絶対的ではない。Wada et al. (1998) が示したように聴覚知覚が困難といった場合、視覚情報が聴覚知覚に影響を及ぼすこともある。

4.2 日本語と英語を対照とする McGurk 実験

日本語話者において継続時間の McGurk 効果が観察されなかった理由としてもう一つの要素が考えられる。日本語母語話者では、調音点の McGurk 効果が弱いことから、継続時間においても McGurk 効果が弱い可能性があるかと推論できる。

過去の研究において、日本語母語話者は、例えば英語話やスペイン語母語話者と比べると、さほど伝統的な McGurk 効果を見せないことが指摘されている (Sekiyama and Tohkura 1993)。我々は、Sekiyama (1993) の文化論の説明と異なり、この理由として日本語の音韻目録が英語やスペイン語と比べ、視覚的に目立つ音に乏しいからだと考える。英語などの場合は、視覚的にも容易に知覚できる音が多く存在している (Mac Eachern 2000)。例えば、英語には両唇音 ([m, p]) や唇歯音 ([f, v])、歯茎音 ([θ, ð]) もあるのに対して日本語には両唇音のみしかない。更に、英語の [ʃ, ʒ] の発音にも円唇性があるのに、日本語にはない。その上、英語の後舌の母音が円唇であるのに対して、日本語の後舌の母音 (特に [ɯ]) は典型的に円唇性が弱い。つまり、日本語母語話者の聞き手には、英語と比べ発話の視覚情報が少なく、言語知覚における視覚モダリティの重要性が低いと言える。

さらに、日本母国語話者において McGurk 効果のある人が少ない事の音韻的な理由として、日本語の音節構造の比較的単純さと音韻目録の短小さも考えられる。日本語母語話

者の聞き手が、話の流れから音素（または音節）を聞き取るのは、選択肢が比較的少ないため、英語の場合より容易である。要するに、英語母語話者の言語知覚においては視覚情報が相当頼りになる。一方、日本語母語話者は視覚情報を重要とせず、言語知覚において視覚情報が頼りにも妨げにもならず聴覚情報が優先されるわけである。英語母語話者では全体の98%で McGurk 効果が確認できるのに対して (McGurk and McDonald 1976)、日本語母語話者では、25%と少ない割合でしか McGurk 効果を確認できない (Yamanaka 2004)。

4.3 技術的な限界

実験結果に技術的な要素が影響した可能性もある。現在最も一般的に使用されているビデオカメラのフレームレートは30Hzである。つまり、刺激用に撮影した画像は、フレーム毎に $33\frac{1}{3}ms$ の時間間隔があることを意味している。これに対して、撮影した音声のサンプリングレートは32kHzであり、音声は $0.03125ms$ 毎に区切られているわけである。このように、動画における情報は、唇の閉鎖と開放といった発音イベントの正確なタイミングと比べて、 $30ms$ 程遅れている可能性もある。従って、この技術的な限界が引き起こす時間の揺れは、視覚システムの内的不確かさに加わるという可能性もあると思われる。

5. 結論

本実験は、調音点の視聴覚知覚と音素の長さの視聴覚知覚とは性質が異なるものであることを示した。調音点の視覚情報は規定通りに言語知覚に統合されているのに対して、音素の継続時間についての視覚情報も同様に言語知覚に統合されている証拠は見出せなかった。

長短に関わる視覚情報が言語知覚に統合されているかどうかを確認するために、聴覚のみの刺激 (AO 条件) と視聴覚的の刺激 (AV 条件) の知覚を比較した。もし、この二種類の刺激に対する反応に有意差があり、視聴覚的の刺激において、子音の長短境界線が長子音の方に近づけば、それは長子音を表す動画の影響と考えられ、継続時間の知覚に視覚的な要素も働くといえる。しかし、実験結果からは、二種類の刺激において有意差は観察されず、被験者が子音の継続時間の知覚において視覚情報を統合しないと明らかになった。

結論として、全ての視覚的情報が言語知覚に同様に寄与するというわけではないといえる。調音点の視覚的知覚は、「形状認識」の過程で、伝統的に言われている McGurk 効果ではこのような情報は一般的に言語知覚に統合されているということを立証しているが、我々の実験では、継続時間に関わる情報は一般的に言語知覚に統合されていないということを指摘する。

恐らく、言語の視聴覚知覚において形状に関わる情報と持続時間に関わる情報に対して2つの異なるメカニズムが採用されているということは、驚くべき結果であろう。その理由として、時間的な知覚において聴覚モダリティが優先されることや、日本語にお

ける言語知覚において、視覚情報の重要性が低いことなどが考えられる。

5.1 今後の課題

現在、刺激を改変し、被験者数を増やすといった更なる実験を計画している。この新しい実験の目的は、日本語を第二言語として話している学習者の言語知覚に日本語の子音の長短に関する視覚的情報が統合されているのかを確認することである。言語学習者は、外国語である日本語を知覚するに際して、母語と比べて聴覚はもちろん、視覚的情報などにもより大きく頼り、長さに関わる McGurk 効果がより強く働くと予測される。日本語学習者を、母語に子音の長短の区別のある言語（韓国語、ハンガリー語など）、及びそのような区別のない言語（中国語、スペイン語など）に二分し、この2つの相違点も考慮して実験を行う予定である。

参考文献

- Colin, Cécile & Radeau, Monique (2003). The McGurk illusions in speech: 25 years of research/Les illusions mcgurk dans la parole: 25 ans de recherches. *L'annee Psychologique*, **103** (3), 497–542.
- Droit-Volet, Sylvie, Stéphanie Turrett & Wearden, John (2004). Perception of the duration of auditory and visual stimuli in children and adults.. *The Quarterly Journal of Experimental psychology*, **57A** (5), 797–818.
- Goldstone, S. & Lhamon, W. T. (1974). Studies of auditory-visual differences in human timing judgment: 1. Sounds are judged longer than lights. *Perceptual and Motor Skills*, **39**, 63–82.
- Kubozono, Haruo (2002). Temporal neutralization in Japanese. In Gossenhoven, Carlos & Natasha, Warner (Eds.), *Laboratory Phonology*, Vol. 7, pp. 171–201. Berlin: Mouton de Gruyter.
- Levine, Michael W. (2000). *Levine and Shefner's Fundamentals of Sensation and Perception*. Oxford University Press, New York.
- MacEachern, Margaret R. (2000). On the visual distinctiveness of words in the English lexicon. *Journal of Phonetics*, **28**, 367–376.
- Massaro, D. W., M.M. Cohen & Smeele, P.M.T. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America*, **100**, 1777–1786.
- McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264**.

- Paré, Martin, Richler, Rebecca C., Hove, Martin Ten, & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception and Psychophysics*, **65** (4), 553–567.
- Schwartz, Jean-Luc, Berthommier, Frederic, & Savariaux, Christophe (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, **93**, B69–B78.
- Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception and Psychophysics*, **59**, 73–80.
- Sekiyama, K. & Tohkura, Y. (1991). 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** (4), 1797–1805.
- Sekiyama, K. & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, **21**, 33–36.
- Wada, Yuji, Norimichi Kitagawa & Noguchi, Kaoru (2003). Audio-visual integration in temporal perception. *International Journal of Psychophysiology Special Issue: Multisensory Research*, **50** (1-2), 117–124.
- Wearden, J., Edwards, H., Fakhri, M., & Percival, A. (1998). Why sounds are judged longer than lights: Application of a model of the internal clock in humans. *Quarterly Journal of Experimental Psychology*, **51B** (2), 97–120.
- 山中雅史 (2004). 自閉症児の音韻処理過程における視聴覚情報の統合. 『調音結合の生成と知覚に関する発達的研究 (研究課題番号 136101138)』, pp. 48–66.

Author's E-mail Address: shark_scott@hotmail.com