

# 事象関連電位による視聴覚音声知覚の検討

## N1成分とP2成分による言語間比較

久永聡子・積山薫・伊賀崎伴彦・村山伸樹

### 1. はじめに

#### 1.1. 視聴覚音声知覚

話者の顔がみえる状況下での音声聴取は、そうでない場合よりも音声知覚が容易になる。このことは、特に騒がしい状況で顕在化し、調音運動に伴う視覚情報は雑音下で音声知覚に有用な情報として処理される (Sumbly & Pollack, 1954)。このような視覚情報の影響は、McGurk 効果のように視覚情報と聴覚情報に矛盾がある場合にも生じ、この場合、両者を融合した音韻が知覚されることが報告された (McGurk & MacDonald, 1976)。たとえば、“ba” という聴覚情報に “ga” という矛盾した視覚情報を同期させることで、多くの聴取者は “da” という音声を知覚するのである。この場合、音響レベルで “ba” と “da” が類似していること、調音レベルで “ga” と “da” が類似していることが聴取者の音声知覚を変容させたと示唆される。McGurk 効果は、現在までに数多くの研究で検討されており (e.g. Kuhl, Tsuzaki, Tohkura, & Meltzoff, 1994; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008; Sekiyama & Tohkura, 1991, 1993; van Wassenhove, Grant, & Poeppel, 2005)、音声知覚における視覚情報の影響が明らかにされてきた。

#### 1.2. 行動研究

McGurk 効果生起に例示されるような音声知覚における視覚情報利用の程度は、騒音などの環境要因のほか、被験者の年齢や母語によっても影響を受けることが報告されている。たとえば、日本語母語話者の大人は英語母語話者の大人に比べて McGurk 効果生起率が低いことが挙げられる (Kuhl et al., 1994; Sekiyama & Burnham, 2008; Sekiyama & Tohkura, 1991, 1993)。また、発達の変化に着目した研究から、英語母語話者は子どもから大人になるにつれて視覚情報の影響を強く受けるようになることが報告されている (McGurk & MacDonald, 1976; Massaro, Thompson, Barron, & Laren, 1986)。しかしながら、日本語母語話者では、このような発達のな変化がみられないことが Sekiyama & Burnham (2008) によって明らかにされた。

以上にみられた日本語母語話者と英語母語話者の視聴覚音声知覚の差異が何に起因するかを調査するために、Sekiyama & Burnham (2008) は聴覚のみによる聴取課題と視覚的な読唇課題を設定し、その反応時間を検討した。その結果、英語母語話者では6歳から8歳にかけて視覚による影響が大きくなり、大人では聴取課題よりも読唇課題で反応時間が短くなった。一方、日本語母語話者にはそのような傾向はみられず、大人においても聴取課題と読唇課題の反応時間に差はみられなかった。このこ

とから、視覚の影響が強い被験者では、聴取課題より読解課題の反応時間が速い傾向があり、処理の速い感覚が知覚内容の決定に大きな比重をもつことが示唆された。

### 1.3. 脳内活動

脳の神経活動の様子は、脳波形 (electroencephalography: EEG)、機能的核磁気共鳴画像法 (functional magnetic resonance imaging: f-MRI)、脳磁計 (magnetoencephalography: MEG) などの非侵襲性的手法で観察することができる。EEG に関しては、安静時に観察されるものと、外界からの刺激に対して誘発されるものがあり、後者は事象関連電位 (event-related potential: ERP) とよばれる。ERP や MEG は、時間分解能に非常に優れているのが特徴で、言語学や音知覚の研究においてひろく利用されている。

近年、音声知覚による視聴覚間の相互作用に関して、音声開始後少なくとも約 250ms まで観察されることが ERP によって明らかにされた (Basle, Fort, Delpuech, & Giard, 2004; Colin, Radeau, Soquet, Demolin, Colin, & Deltenre, 2002; van Wassenhove et al., 2005)。ERP では、縦軸に活動の大きさを示す振幅 ( $\mu\text{V}$ )、横軸に時間 (ms) が示され、上側を陰性極 (negative polarity: N)、下側を陽性極 (positive polarity: P) として表示される。表記の際は、出現した振幅を順に P1、N1、P2、N2 などのように略語で示すことが多い。N1、P2 に関して、英語圏の研究 (van Wassenhove et al., 2005) では、聴覚のみの聴取に比べて、聴覚と一致した視覚情報が付加された視聴覚条件で、振幅が減少し、潜時は短縮する傾向があることが報告された。さらに、MEG を用いたフィンランド母語話者での研究 (Davis, Kislyuk, Kim, & Sams, 2008) では、聴覚のみの聴取に比べて、視聴覚音声知覚で 100ms 付近にみられる磁場波形のピークが減少することが報告された。以上から、視覚情報が付加されたことによる音声知覚への影響が刺激呈示後 100ms から観察されることが示唆された。また、フランス語母語話者での研究 (Basle et al., 2004) では、視聴覚音声知覚では、聴取のみの判断より反応時間が短くなること、さらに ERP では van Wassenhove et al. (2005) の英語母語話者と同様に N1 振幅が減少することが報告されている。そこで、本研究では、行動研究から示唆された日本語母語話者と英語母語話者における音声知覚の視聴覚比重の差異が ERP ではどのように出現するのか、検討することをめざした。

### 1.4. 目的

本研究では、日本語母語話者と英語母語話者を対象に、音声知覚中の視覚情報の影響の言語差を検討した。実験 1 ではモダリティ間の処理速度の差に関する先行研究 (Basle et al., 2004; Sekiyama & Burnham, 2008) に着目し、反応時間の測定をおこない、実験 2 では ERP から脳内活動における時間的推移を検討した。また、前述したような日本語母語話者と英語母語話者間の視聴覚音声知覚の差異 (Kuhl et al., 1994; Sekiyama & Tohkura, 1991; 1993; Sekiyama & Burnham, 2008) に着目し、視覚情報の比重が相対的に小さい日本語母語話者の特有な音声処理過程を明らかにすることを目的とした。

## 2. 実験 1

実験 1 では、行動のデータとして反応時間 (RT) を指標に、音声知覚中の視覚情報の影響を検討した。

2.1. 方法

2.1.1. 実験参加者

日本語母語話者 (Japanese group: JG) 18名 (平均年齢、21.28歳; 範囲、20-26歳; 男性6名、女性12名) と英語母語話者 (English group: EG) 6名 (平均年齢、20.12歳; 範囲、19-22歳; 男性4名、女性2名) を対象とした。JGは熊本大学に通う者、EGは熊本大学に交換留学生として通うアメリカ人とイギリス人であった。全員が右手を利き手とし、視力(矯正含む)、聴力ともに本実験参加に十分なものであった。

2.1.2. 刺激

呈示法は、音声と凝視点で構成した聴覚呈示 (Audio-only: AO) と音声と映像が同期した視聴覚呈示 (Auditory-visual: AV) の2条件とした。AV条件に関しては、今回は視聴覚が一致したもののみを用いた。各条件は、2人の女性話者 [日本語刺激 (J)、英語刺激 (E)] がそれぞれ /ba/、/ga/ の発話をしたムービー・クリップを Adobe Premiere 5.1 を用いて編集した (映像: 640x480 pixel、29.97 フレーム、音声: 16bit、44kHz、ステレオフィールとして合成)。

音声は 65dB SPL で呈示し、さらにマシンノイズをマスクするためバンドノイズ (300-12000Hz: SN比 13dB SPL) を使用した。

Fig. 1 は、話者 J と E による /ba/、/ga/ を音声解析ソフト Praat (Boersma, & Weenink, 2010) を用いて検証したものである。スペクトログラムの低周波から高周波にかけて全体的に濃くなる時間点は、調音器官が閉鎖から開放されたことによる「破裂」を示している。/ba/、/ga/ は、いずれも有声音であり、通常、破裂前に声帯振動を伴う。話者 J に関して、音声波形、スペクトログラムから声帯振動が破裂に先行しておこなっているのがわかる。破裂から声帯振動の開始までの時間 (voice onset time: VOT) は、音声波形とスペクトログラムを参考に計測できる。声帯振動開始は、低周波数領域のエネルギーの立ち上がりで読み取ることが出来る。J の場合、声帯振動が破裂に先行しているため、VOT はマイナスの値で示され、J/ba/ では -78ms、J/ga/ では -98ms となる。清水 (1993) の研究では、日本語母語話者による有声破裂音に関する VOT の平均値は /ba/ が -89ms (範囲、-65 から -125ms)、/ga/ が -75ms (範囲、-35 から -125ms) であると報告している。一方、E/ba/、E/ga/ での VOT 値は、ほぼ 0 に近く破裂とほぼ同時に声帯振動が始まっていることがわかる。これは、英語母語話者の /ba/、/da/、/ga/ の VOT はゼロに近いことを示した Lisker (1964, p. 395) の報告と一致している。VOT

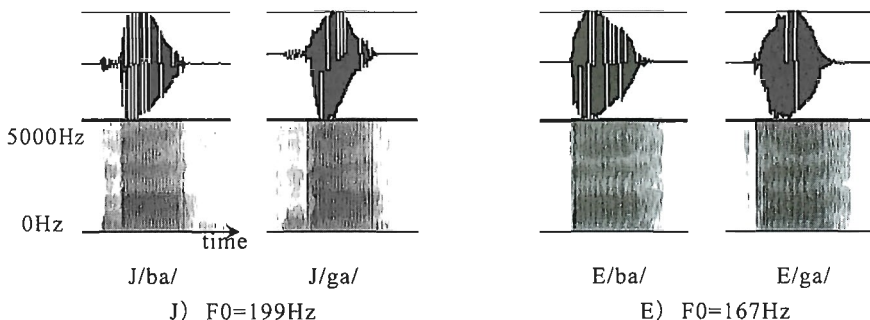


Fig. 1. Waveform & Spectrogram

は、有声破裂音と無声破裂音を弁別する知覚の手がかりとなることが知られており、VOT を操作した知覚実験では、英語においては、VOT がゼロ付近であれば有声破裂音として、+ 60ms 以上であれば無声破裂音として知覚されることが示されている (Lisker & Abramson, 1970)。

以上から、本研究で用いる刺激は、それぞれの言語の特徴を反映したものと考えられる。

### 2.1.3. 手続き

映像は液晶モニタ (SONY SDM - 51)、音声はその上方に置かれたスピーカ (AIWA SC-B10) を用いて呈示した。被験者は、液晶モニタの前に置かれた椅子に座り、開眼安静状態で実験に参加した。実験 1 は、実験 2 で計測する環境と同様にシールドルーム内でおこなった。(Fig. 2)。

実験構成は、AO 条件、AV 条件の 2 条件各 40 試行 1 セットであった。条件実施順は被験者の半数が AO 条件から、残り半数が AV 条件からとした。被験者には、視聴した音節が “ba”、“ga” のどちらであったかを判断し、ボタン押しで選択するよう教示した。Fig. 3 は、1 試行の流れである。

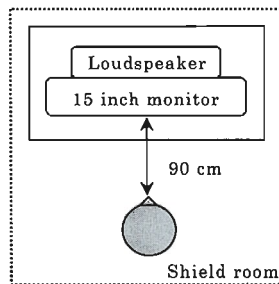


Fig. 2. Room setting

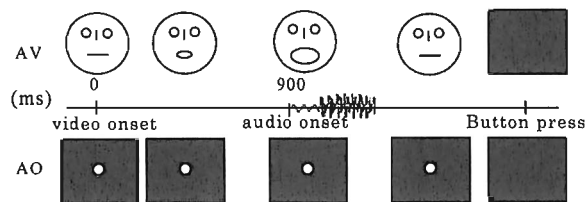


Fig. 3. A trial flow of movie: RTs were measured as the time from the audio onset to the button press. The onset of the next stimulus was 1500ms after the button press.

## 2.2. 結果

RT について、被験者の母語 (2: JG, EG) を被験者間要因、呈示条件 (2: AO, AV) × 刺激話者 (2: 母語話者、非母語話者) × 音節 (2: /ba/, /ga/) を被験者内要因とする 4 要因分散分析を行った。その結果、被験者の母語 × 呈示条件に関して交互作用が有意であった [ $F(1,22)=8.04, p<.001$ ]。JG では、AV 条件の RT は、AO 条件に比較して増大したが [ $F(1,22)=8.59, p<.01$ ] (Fig. 4a)、EG では短縮した [ $F(1,22)=9.46, p<.01$ ] (Fig. 4b)。さらに、AV 条件の RT は、JG より EG の方が短かった [ $F(1,44)=6.63, p<.05$ ]。

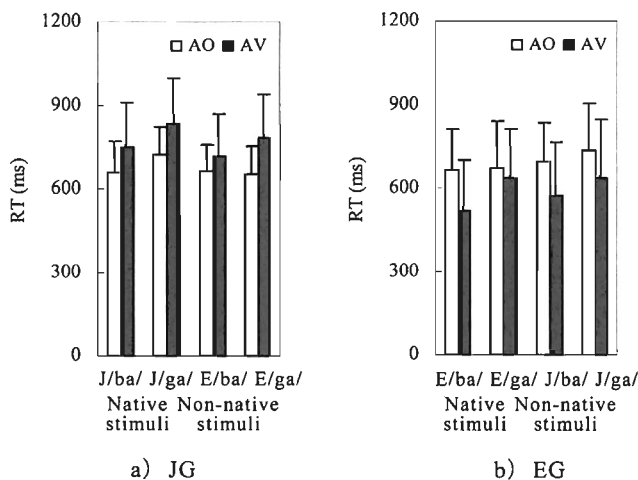


Fig. 4. RTs for two language groups

### 2.3. 考察

実験1では、JGとEGを対象にAO条件とAV条件中の音声知覚のRTを検討した。その結果、AO条件をベースラインとした視覚情報付加の効果は、言語間で異なっていた。EGは、AO条件に比べてAV条件でRTを短くしたが、JGにはそのような傾向はみられず、むしろAV条件でRTを延長した。以上から、音声知覚の際の視聴覚比重が母語によって異なることが示唆された。つまり、EGは積極的に視覚情報を利用することでAV条件でのRTを短縮したが、JGは視覚情報が付加したことによって妨害的な影響を受け、AV条件でRT増大となったと考えられる。この結果は、EGはAO条件より視覚条件（VO）でRTが速く、JGはVO条件とAO条件に差がみられないという先行研究 Sekiyama & Burnham (2008) と整合する。

これらの結果は、音声知覚において、JGでは聴覚情報が、EGでは視覚情報が相対的に重視されることを示唆している。従来、McGurk効果の生起率が母語によって異なることが報告されていたが (Kuhl, Tsuzaki, Tohkura, & Meltzoff, 1994; Sekiyama & Tohkura, 1991, 1993)、それはこのことと関連していると考えられる。

Basle et al. (2004) は、フランス語母語話者を対象に、AV呈示とAO呈示時のRTを比較し、AV呈示でRTが短縮したことを報告した。Basle et al. (2004) の実験は、AO、VO、AV呈示で構成された4音節 (/pa/, /pi/, /po/, /py/) を1ブロックとしてランダム呈示し、ブロック毎に4音節から1音節をtargetとした。課題はtargetの際にボタン押しをするというものであり、フランス語母語話者においてもAV情報が音声知覚に促進的に作用したことが示唆される。本実験結果から、AV情報の影響は、JGでは抑制的に作用することが示唆され、JGの特徴的な音声知覚形態があることを示したと考えることができる。

### 3. 実験2

実験2では、視聴覚音声処理における視覚情報の影響を、脳内活動の時間的推移から検討した。

### 3.1. 方法

#### 3.1.1. 実験参加者

実験1の参加者から、JG8名（平均年齢、22歳；範囲、21-26歳；男性3名、女性5名）とEG3名（平均年齢、19.7歳；範囲、19-22歳；男性2名、女性1名）を対象とした。

#### 3.1.2. 刺激

実験1と同様であった。

#### 3.1.3. 手続き

映像、音声の呈示は実験1と同様に行った。実験構成は、AO、AVの2条件各40試行10セット、計800試行であった。条件実施順は被験者の半数がAO条件から、残り半数がAV条件からとし、3日間に分けて脳波計測を行った。被験者には、視聴した音節が“ba”、“ga”のどちらであったかの内的な判断のみを求めた。刺激は1試行2000msで編集した。

#### 3.1.4. 脳波記録

脳波は、国際10-20法に準じたFp1、Fp2、F7、F3、Fz、F4、F8、T3、C3、Cz、C4、T4、T5、P3、Pz、P4、T6、O1、O2を探索電極（19ch）、両耳朶連結（A1 + A2）を基準電極（Fig. 5）、前頭極部中央および鼻根を接地電極として導出され、脳波計（日本光電 Neurofax EEG - 1100）により増幅およびフィルタリング（0.53 - 300Hz）されたのち、サンプリング周波数500HzでPC（DELL OptiPlex GX280）に記録された。トリガとして、ステレオフィールの一方に音声刺激と同期した30msのトーン信号を編集し、脳波の20chに記録した（Fig. 6）。

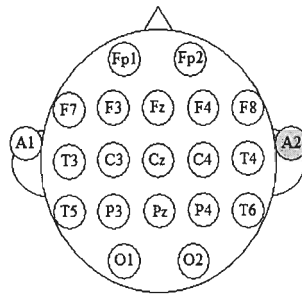


Fig. 5. The 10-20 system

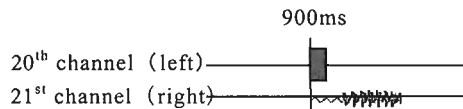


Fig. 6. The audio signals: The trigger signals were recorded on the 20th channel of the EEG and the speech signals were recorded on the 21st channel to make sure the synchronization.

3.1.5. 分析

解析は、audio onset 前 200ms をベースラインとし、条件 (AO、AV)、刺激 (母語話者、非母語話者)、音節 (/ba/、/ga/) 別に各被験者 100 回加算した。Appendix 1 は、JG、EG の総加算平均である。実験 2 の検討箇所としては、単音節に対する電位であるため、頭蓋頂 (Cz) で記録された N1、P2 の初期成分を分析した (Basle et al., 2004; Lewandowska, Bekisz, Szymaszek, Wrobel, & Szlag, 2008; van Wassenhove et al., 2005)。

3.2. 結果

N1、P2 について、被験者の言語 (2 : JG、EG) を被験者間要因、条件 (2 : AO、AV) × 刺激 (2 : 母語話者、非母語話者) × 音節 (2 : /ba/、/ga/) を被験者内要因とする 4 要因分散分析を行った。

3.2.1. 振幅

N1 振幅は、両被験者群において AO 条件に比べ AV 条件で減少し [ $F(1,9)=5.72, p<.05$ ]、被験者の言語 × 条件に交互作用はみられなかった。P2 振幅は、被験者の言語 × 条件 × 刺激に関して交互作用が有意であった [ $F(1,9)=6.22, p<.05$ ]。Fig. 7 に示すように、EG では、条件間において刺激の影響が異なる傾向がみられた [ $F(1,9)=8.42, p=.018$ ]が、JG にはそのような傾向はみられなかった。EG は、刺激が母語話者の際に AO 条件に比較して AV 条件で振幅を減少させた [ $F(1,18)=7.76, p=.012$ ] (Fig. 7a)。また、AO 条件で EG は、母語話者の刺激より非母語話者の刺激で振幅を減少させた [ $F(1,18)=12.33, p=.0025$ ] (Fig. 8)。

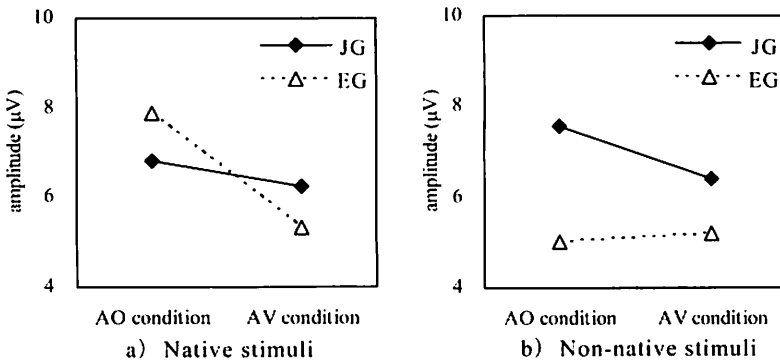


Fig. 7. P2 amplitude

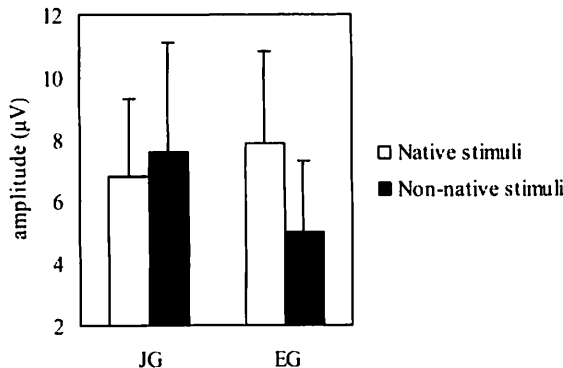


Fig. 8. AO condition on P2

### 3.2.2. 潜時

N1 潜時は、両被験者群において AO 条件に比べ AV 条件で減少した [ $F(1,9)=0.982, p<.01$ ] が、被験者の言語×条件に交互作用はみられなかった。P2 潜時に関しては、被験者の言語、条件による差はみられなかった。

### 3.3. 考察

実験 2 では、視聴覚音声知覚について ERP から検討した。N1 成分に関しては、JG と EG ともに AO 呈示に比較し、AV 呈示時で振幅は減少し潜時は短縮する傾向がみられた。van Wassenhove et al. (2005) は、EG を対象に母語話者の刺激を用いて AO、VO、AV の 3 条件による呈示方法を用いて実験をしたところ、聴覚情報に視覚情報が付加することによって、N1 - P2 振幅の減少、潜時の短縮が生じることを報告している。実験 2 で、EG は N1 振幅、潜時で van Wassenhove et al. (2005) と類似した傾向が確認されたといえる。しかし、本研究での視覚情報付加による P2 振幅の減少に関しては、EG においてのみみられ、JG に関しては、AO 呈示と AV 呈示の違いは N1 成分でのみしかみられなかった。また、EG でみられた P2 振幅減少は、母語話者の刺激に限られたものであった。

以上から、母語話者による刺激の際、言語間の差異が P2 振幅で確認された。EG が 200ms 付近まで AV 呈示による影響を受ける一方、JG はその影響が 100ms 付近までしか継続されなかったことが示された。

AO 呈示の際、P2 で EG は母語話者の刺激に比較し、非母語話者の刺激の際に振幅を減少させた。このことに関しては、2.1.2 の刺激の節で述べた言語の特性が関係しているのではないかと考える。

本実験結果から、視聴覚音声知覚に関して、視覚の影響が 100ms 付近にみられる JG と音声の影響が 200ms にみられる EG という音声知覚における言語差が示唆された。

## 4. 総合考察

本研究では、JG と EG を対象に音声知覚における視覚の影響を、RT (実験 1) と ERP (実験 2) から検討した。実験 1 では、視覚情報付加の影響により EG では RT が短縮したが、JG では RT が増大した。つまり、視覚の影響が EG では促進的であったのに対して、JG ではむしろ抑制的であり、逆効果であったことが示唆される。この結果は、EG は聴取課題に比べて視覚的な読唇課題で RT が短縮し、JG は聴取課題と読唇課題で RT に差がみられなかったという Sekiyama & Burnham (2008) の結果に類似していた。以上から、EG は音声聴取よりも調音情報の視覚的処理が速くできることによって、視覚的処理が聴覚的な音声知覚にプライミング効果を引き起こしたことが示唆される。しかし、JG ではそのような傾向はみられなかった。要するに、成人の EG で McGurk 効果が強く生起する (e.g. Kuhl et al., 1994; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008; Sekiyama & Tohkura, 1993) のは、このようなプライミング効果が起因していることが示唆される。

実験 2 において ERP で観察された視覚情報付加の影響は、EG では 200ms 付近までみられたが、JG では 100ms 付近までしかみられなかった。EG で McGurk 効果が強く生起するのは、視覚情報の処理が持続的であることが背景に考えられる。EG は、視覚情報の処理が持続的であることによって、音声知覚に視覚情報が有意な情報として処理され、その結果、AO 呈示に比較して AV 呈示で RT を短く



したと示唆される。

EGの視聴覚音声知覚中のERPの傾向は、N1振幅、N1潜時、P2振幅においてvan Wassenhove et al. (2005)の先行研究と整合しており、AO呈示に比較してAV呈示で振幅の減少、潜時が短縮する傾向が確認された。しかし、P2潜時については短縮傾向がみられず、van Wassenhove et al. (2005)とは異なった結果であった。だが、本研究のEGのERP被験者数が、小さなものであったことを考慮すると、P2潜時については被験者数を増加して再検討する必要がある。本実験では、単音節に対する電位であるため、N1、P2の初期成分を分析した(Basle et al., 2004; Lewandowska et al., 2008; van Wassenhove et al., 2005)が、実験1で示唆された視覚情報が音声知覚にプライミング効果を引き起こしたとするならば、N1、P2に先行するP1でAO、AV呈示による違いがみられることが考えられる。この点に関しても、検討の余地があるだろう。

本研究では、視聴覚音声知覚の言語差を検討するため、JGとEGを対象に、刺激話者についてもJとEを用いた。その結果、EGにおいてAO呈示時で、母語話者の刺激に比較し、非母語話者の刺激の際に振幅が減少したことが確認された。本実験で使用した刺激は、VOT値に関して、それぞれの言語の特性を示したものであった。そもそも、VOT値は音声の有声性、無声性を特徴づけるものである(Lisker, & Abramson, 1964; 1970)。清水(1993)はアジアの6言語(日本語、中国語、韓国語、ビルマ語、タイ語、ヒンディー語)を調査対象とし、VOTについて検証した結果、韓国語とヒンディー語を除く言語においてVOTは有声性、無声性の弁別に機能すると報告した。刺激言語による影響が200msに現れることが示されたEGに関して、このVOT値の言語差がどのように作用したのかについても再検討する必要があるだろう。

現在までに、視聴覚音声知覚の研究において母語話者を刺激話者として用いた研究はあるものの(Basle et al., 2004; Davis et al., 2008; van Wassenhove et al., 2005)、非母語話者を刺激話者とした際についての音声処理についてはまだ明確にされていない。しかし、本研究から、EGは刺激が母語話者であった場合と非母語話者の場合で視聴覚音声処理が異なる傾向がみられ、母語話者においてのみ視覚情報の付加が音声知覚に持続的にはたらく傾向が示唆された。

本研究ではP2成分においてのみ言語間の差異がみられたが、これはN1、P2のどのような特徴と関係しているのだろうか。N1成分、P2成分に関して、聴取課題の困難性を操作した研究でLewandowska et al. (2008)は、N1潜時、N1振幅、P2潜時には刺激の物理的な特性が反映され、P2振幅には課題の困難性が反映されることを示唆した。P2振幅は、困難性が低い課題で、困難性が高い課題よりも減少することを報告している。つまり、本研究結果を考慮すると、EGはAO呈示よりも視覚情報が付加したAV呈示で音声知覚が容易になったと考えることができる。この解釈は実験1でみられたRTの結果とも整合性がある。JGに関してはP2振幅に条件差はなく、音声知覚の際の視覚情報の付加が難易度を低下させることはなかったと考えることができる。N1成分に関して、高齢者と若年者を対象にした視聴覚音声知覚研究でWinneke & Phillips(2009)は、高齢者と若年者ともにAO呈示に比べてAV呈示時でN1振幅を減少させたことを報告している。若年者と高齢者では視聴覚の比重が異なることが示唆される(Thompson, 1995; 積山、坂本, 2007)が、その影響は100ms付近には報告されていない。また母語の異なる研究間において、N1成分にみられる傾向が類似している(Basle et al., 2004; van Wassenhove et al., 2005; Winneke & Phillips, 2009)ことから、N1成分は刺激に対する物理的処理を反映していることが示唆される。

今後、被験者数を増加し、RT と ERP 成分の関係について詳細に検討したいと考える。

## 文 献

- Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex, *European Journal of Neuroscience*, 20, 2225 - 2234.
- Boersma, P., & Weenink, D. (2010). Praat Ver.5.1.23 University of Amsterdam.
- Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clinical Neurophysiology*, 113, 495 - 506
- Davis, D., Kislyuk, D., Kim, J., & Sams, M. (2008). The effect of viewing speech on auditory speech processing is different in the left and right hemispheres, *Brain Research*, 1242, 151 - 161.
- Kuhl, P. K., Tsuzaki, M., Tohkura, Y., & Meltzoff, A. N. (1994). Human processing of auditory-visual information in speech perception : Potential for multimodal human-machine interfaces. *Proceedings of the International Conference of Spoken Language Processing*, Tokyo, 539 - 542
- Lewandowska, M., Bekisz, M., Szymaszek, A., Wrobel, A., & Szlag, E. (2008). Towards electrophysiological correlates of auditory perception of temporal order, *Neuroscience Letters*, 437, 139 - 143.
- Lisker, L., & Abramson, A. S. (1964). A crosslanguage study of voicing in initial stops: acoustical measurements. *Word*, 20, 384 - 422.
- Lisker, L., & Abramson, A.S. (1970). The voicing dimension: some experiments in comparative phonetics, in *Proceedings of the Sixth International Congress of Phonetic Sciences*, Prague, 1967. Prague: Academia, 563 - 567.
- Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, 41, 93 - 113.
- McGurk, H., & MacDonald, J. (1976). Hearing lips seeing voices, *Nature*, 264, 746 - 748.
- Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science*, 11, 306 - 320.
- 積山薫, 坂本真一 (2007). 耳の衰えは処理速度に出る - 視聴覚音声知覚の加齢変化 -. 日本音響学会聴覚研究会資料, 36 (4), 799 - 804.
- 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, 1797 - 1805.
- Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, 21, 427 - 444.
- 清水克正 (1993). 閉鎖子音の音声的特徴 - 有声性・無声性の言語間比較について -. 東京外国語大学アジアアフリカ言語文化研究, 45, 163 - 175.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212 - 215.
- Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology & Aging*, 10, 215 - 228.
- van Wassenhove, V., Grant, K., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory

speech. *Proceeding of the National Academy of Sciences*, 102, 4, 1181 - 1186.

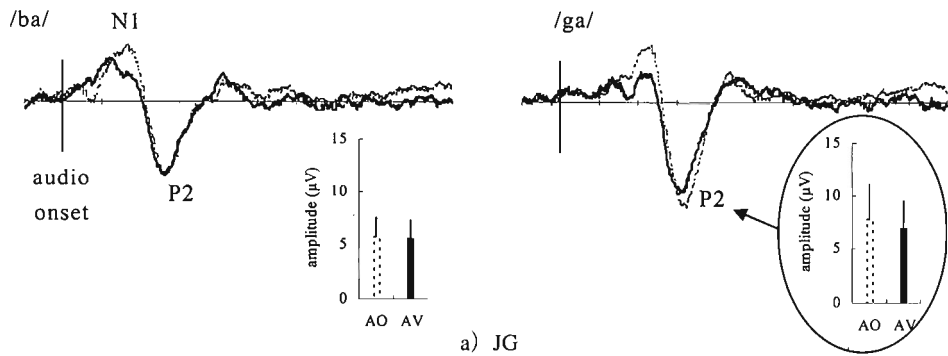
Winneke, A. H., & Phillips, N. Older and younger adults use fewer neural resources during audiovisual than during auditory speech perception. *Proceedings of the 2009 Conference on Audio Visual Speech Processing*, Norwich, 123 - 126.

## A Study of Audio-Visual Speech Perception by ERP Interlanguage Comparison of N1 and P2

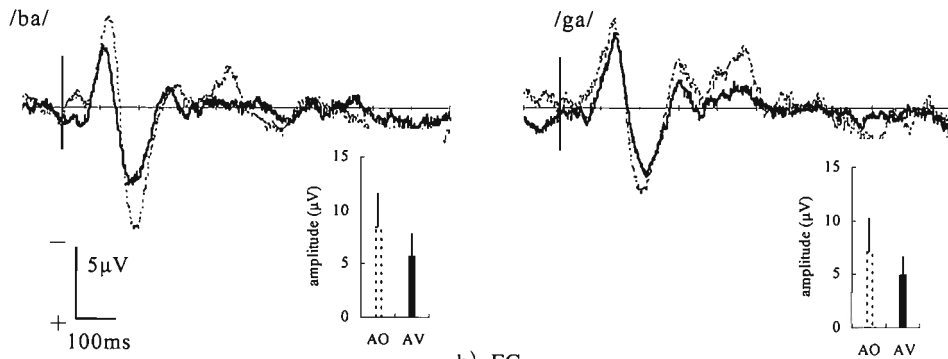
HISANAGA Satoko · SEKIYAMA Kaoru · IGASAKI Tomohiko and MURAYAMA Nobuki

By using event-related potentials (ERP, Experiment 2) and reaction times (RT, Experiment 1), the present study examined interlanguage differences between Japanese and English in audio-visual speech perception (Kuhl, Tsuzaki, Tohkura, & Meltzoff, 1994; Sekiyama & Burnham, 2008; Sekiyama & Tohkura, 1991, 1993). There were an auditory-only (AO) and a congruent auditory-visual (AV) conditions. In Experiment 1, RTs showed opposite tendencies between English-language (EG) and Japanese-language (JG) groups for the AO-AV relationship: the additional congruent visual information speeded up the speech perception processes for the EG, but it slowed down the processes for the JG. Thus, the visual influence was promoting for the EG but disturbing for the JG. In Experiment 2, different ERP patterns were found between the EG and JG: Whereas the visual influence was sustained (maintained from N1 to P2) in the EG, the influence was transient (limited only to N1) in the JG. The ERP and RT data were both consistent with the reported interlanguage differences that the JG perceivers use visual information to the less extent than the EG perceivers do.

Appendix 1

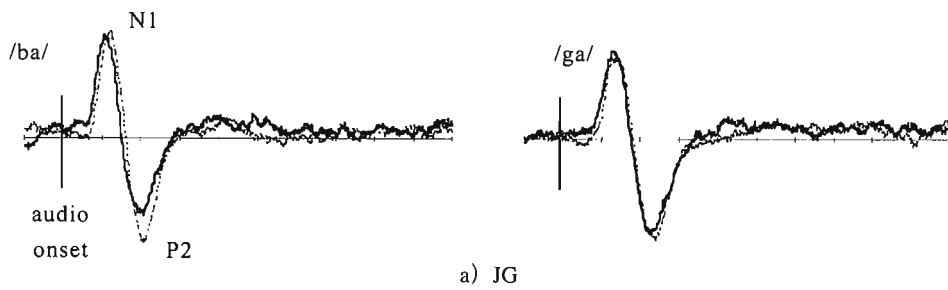


a) JG

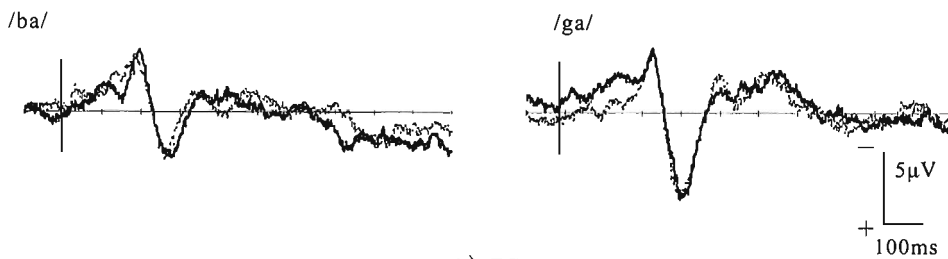


b) EG

ERP for the native stimuli



a) JG



b) EG

ERP for the non-native stimuli