

## 聴覚情報および調音運動情報に基づく音声知覚処理の役割に関する研究

田村, 俊介

<https://doi.org/10.15017/4060183>

---

出版情報 : 九州大学, 2019, 博士 (学術), 課程博士  
バージョン :  
権利関係 :

聴覚情報および調音運動情報に基づく音声知覚  
処理の役割に関する研究

田村俊介

## 概要

本論文は、主要な音韻弁別素性である有声性の知覚を題材としてヒトの音声知覚メカニズムを検討したものである。有声性知覚を通じた音声知覚メカニズムの検討は英語母語話者を対象とした数多くの先行研究で行われており、主に有声性知覚のために聴覚系での情報処理で音声信号からどのような情報が抽出されるのかについて調べられている。しかしながら、英語圏以外の話者を対象とした研究では、英語母語話者と有声性の知覚特性に大きな違いがあることが示されているもののそのメカニズムに関する検討はほとんど行われていなかった。また、近年の研究では、聴覚系での情報処理で抽出した情報に基づく音声知覚処理だけでなく、音声信号からその信号作り出す原因となった調音運動を推定し、その情報を基にした音声知覚処理の存在が盛んに議論されている。しかしながら、有声性知覚における後者の役割についてはこれまでにほとんど検討がなされていなかった。

本論文では、有声性の知覚処理過程で聴覚情報と調音運動情報がどのように利用されているのかについて検討を行った。研究1では、言語間で有声性の知覚特性の違いが見られることに注目しながら、日本語母語話者の有声性知覚にどのような聴覚情報が利用されているのかを調べる行動実験を行った。その結果、日本語母語話者は逆向性マスキングに係る聴覚処理特性を利用して抽出した情報を基に有声性のカテゴリー知覚を行っており、英語母語話者とは異なる聴覚処理特性を用いて有声性の知覚を行っている可能性が示唆された。研究2では、有声性知覚において調音運動情報がどのように利用されているかを調べるために、有声性の生成を変化させるための聴覚フィードバック課題を行った前後で有声性の知覚にどのような変化が見られるかを調べる行動実験を行なった。その結果、有声性の生成の変化が有声性知覚に及ぼす影響は大きくはないが、その明確性に影響を及ぼすことが分かったことから、調音運動情報

に基づく音声知覚処理が有声性知覚の明確性に関与している可能性が示唆された。研究3では、機能的磁気共鳴画像法と脳磁図を用いて有声性知覚時の聴覚皮質や音声の生成に関与する脳領域の神経活動を調べることで聴覚情報と調音運動情報の利用に係わる脳内メカニズムについて検討した。その結果、聴覚皮質と音声の生成に関与する脳領域の間で機能的な結合が見られるとともに両方の領域において有声性知覚と相関した神経活動が生じていることが分かった。また、音声生成に関与する脳領域では、有声性の生成に密接に係ると思われる **subcentral area** で有声性知覚と相関する神経活動が生じていることが分かった。研究3の結果から、聴覚情報と調音運動情報の両方が有声性知覚に利用されていることが研究1,2の行動実験だけではなく脳機能計測実験によっても実証された。

研究1~3での聴覚情報および調音運動情報に基づいた音声知覚処理の役割の検討を通して、聴覚系で音声信号から抽出された情報は音声知覚の主要な手がかりとなるが、その具体的な処理方策は言語圏によって異なる可能性があることが示唆された。また、調音運動情報に基づいた音声知覚処理が聴覚情報に基づいた音声知覚処理を補完する形で音声知覚の明確性を高めるという仕組みが日本語母語話者に備わっていることが明らかになった。調音運動情報に基づく音声知覚処理はヒトの頑健な音声知覚システムを支える機能であり、工学的な音声情報処理（音声認識）でも考慮すべき処理方略だと考えられている。そのため、本研究で得られた成果はヒトの音声コミュニケーション原理を理解する上で非常に重要だと考えられるとともに工学的な音声研究の進展を促す知見にもなるであろう。



# 目次

第1章 序論	9
第2章 研究背景	18
2.1 音声知覚メカニズムに関する仮説.....	18
2.2 行動実験による音声知覚メカニズムの検討.....	20
2.2.1 聴覚情報に基づく音声知覚処理の役割を検討した先行研究.....	20
2.2.2 調音運動情報に基づく音声知覚処理の役割を検討した先行研究.....	22
2.3 脳機能計測実験による音声知覚メカニズムの検討.....	25
2.3.1 脳機能計測手法.....	25
2.3.2 聴覚情報に基づく音声知覚処理の脳内メカニズムを調べた先行研究.....	28
2.3.3 調音運動情報に基づく音声知覚処理の脳内メカニズムを調べた先行研究...	30
2.4 先行研究での未解決点.....	33
2.5 本論文での具体的な検討課題.....	35
第3章 研究 1: 日本語母語話者の有声性知覚に係る聴覚処理特性の検討	40
3.1 目的.....	40
3.2 実験.....	41
3.2.1 実験参加者.....	41
3.2.2 実験装置および刺激.....	41
3.3 実験手続き.....	43
3.3.4 データ分析.....	44
3.3.5 結果.....	45

3.3 考察.....	47
<b>第4章 研究 2：有声性の生成処理の変化が知覚処理に与える影響の検討</b>	<b>51</b>
4.1 目的.....	51
4.2 実験.....	53
4.2.1 実験参加者.....	53
4.2.2 実験装置および刺激.....	53
4.2.3 実験手続き.....	54
4.2.4 データ分析.....	55
4.2.5 結果.....	55
4.3 考察.....	58
<b>第5章 研究 3: 有声性知覚時の聴覚情報および調音運動情報の利用に係る脳内 メカニズム：MEG および fMRI を用いた脳機能計測による検討</b>	<b>63</b>
5.1 目的.....	63
5.2 実験 1.....	64
5.2.1 実験参加者.....	64
5.2.2 実験装置および刺激.....	64
5.2.3 実験手続き.....	65
5.2.4 データ分析.....	66
5.2.5 結果.....	68
5.3 実験 1 考察.....	70
5.4 実験 2.....	72
5.4.1 実験参加者.....	72

5.4.2 実験装置および刺激.....	72
5.4.3 実験手続き.....	73
5.4.4 データ分析 行動データ.....	73
5.4.5 データ分析 MEG データ.....	74
5.4.6 結果.....	78
<b>5.5 実験 2 考察.....</b>	<b>84</b>
<b>第6章 本論文のまとめ</b>	<b>88</b>
<b>謝辞</b>	<b>92</b>
<b>文献</b>	<b>95</b>



## 第1章 序論

音声は我々の日常生活において欠かすことの出来ないコミュニケーションツールである。そのため、ヒトの音声コミュニケーションの仕組みを解明することは、ヒトの認知情報処理メカニズムの一端を明らかにするという学術的意義だけでなく、聴覚補償や音声合成・認識などを扱うヒトの情報処理を工学的に実現する研究を進展させる上でも非常に重要である。本論文ではヒトが音声信号から言語情報を知覚するまでの情報処理メカニズムを聴覚系での情報処理のみならず音声生成処理との相互作用にも注目しながら検討を行った。音声生成処理と音声知覚処理の相互作用に関する研究は近年注目を集めており、そのメカニズムの解明は高齢者などの聴覚機能低下に対する新たな音声コミュニケーション補償法の開発に繋がると考えられている (Du et al., 2016)。また、ヒトの音声生成と音声知覚における情報処理の共通性は、機械による音声合成技術や音声認識技術では考慮されていない機能であり、その解明が工学的な音声情報処理技術を発展させる指針になるとも考えられている (廣谷, 2017)。

speech chain と称される音声コミュニケーションの枠組み (Denes & Pinson, 1993) や2章で詳しく解説する音声知覚メカニズムに関する仮説を参考に、ヒトの音声生成および音声知覚の情報処理過程を表す概念図を図1に示す。まず、ヒトが音韻情報を音響信号に変換して聴取者に伝える際には、音声の生成に伴って生じる聴覚情報を予測した上で調音器官の制御が行われ、その後に予測した通りに発声が行われたかどうかを自らの発話音声を聴取することで確認する。そして、もし両者の間に誤差が生じた場合には音韻情報や調音運動と聴覚情報の対応関係を修正することで音声生成処理に変更を加えるという処理が行われている (図1 実線)。音声信号から音韻情報を知覚する情報処理過程 (図1 点線) については、長年に渡って主に2つの仮説が議論さ

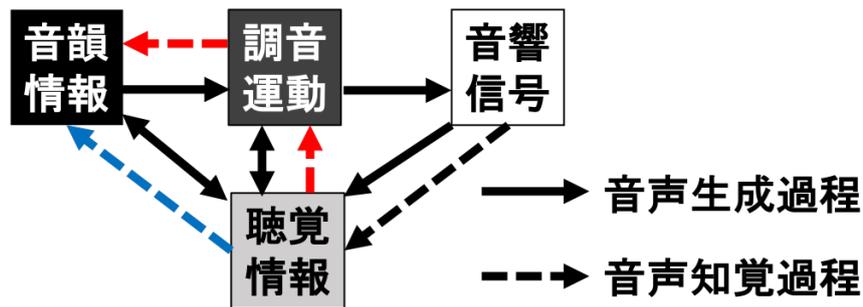


図 1. 音声生成及び音声知覚における情報処理過程の概念図。

れている。1 つは、聴覚系での音響信号処理で抽出した情報を基に音韻情報が知覚されると考える説（図 1 青線）（Blumstein & Stevens, 1979, 1980; Diehl et al., 2007; Holt et al., 2004; Kuhl & Miller et al., 1978; Stevens, 1989）である。もう 1 つは音声生成系を伴った情報処理が音声知覚において重要な役割を担っており、聴覚系での情報処理を介することなく音響信号から直接的にその信号を作り出した調音運動を推定し、その情報を基に音韻情報を知覚するという説（Fowler, 1986, 2016; Liberman et al., 1967; Liberman & Mattingly, 1985）である。後者の仮説と関連付けて近年提案されている仮説では、音響信号を作り出す原因となった調音運動情報は聴覚情報と調音運動の対応関係に基づいて求められると考えられている（Barnaud et al., 2016; Laurent et al., 2017; Schwartz et al., 2012）（図 1 赤線）。さらに、音声から抽出された聴覚情報と調音運動情報の両方が音韻情報の知覚に利用されると考えられており、基本的には前者が主要な手がかりとして利用され、後者は雑音環境下など音声の持つ音響情報が不明瞭で聴覚情報が有用でない状況で利用されることも示唆されている。音声知覚の脳内メカニズムについても脳機能計測技術の発展に伴って検討が進んでおり（Hickok & Poeppel, 2004, 2007; Rauschecker et al., 2009, 2011; Sepulcre, 2015; 廣谷, 2017），聴覚腹側経路（聴覚皮質→下前頭回）が聴覚情報、聴覚背側経路（聴覚皮質→縁上回→運動皮質）や operculum

parietal (OP4) を介した聴覚皮質と運動野・運動前野の処理経路が調音運動情報を基にした音声知覚処理過程に対応すると考えられている (図 2)。

以上のように、近年の研究では聴覚情報や調音運動情報に基づく音韻情報の知覚処理の両方がヒトの音声知覚に貢献すると考えられているが、実際にこれら両方の処理の役割を具体的に検討し、その脳内メカニズムを調べるといった包括的な研究は数少なく、特に英語圏以外では存在しないと言っても過言ではない。そこで本論文では、主要な音韻弁別素性の 1 つである有声性の知覚において聴覚情報と調音運動情報がどのように利用されているかを検討する 2 つの行動実験を行うとともに、聴覚情報と調音運動情報を基にした有声性知覚処理に係る脳内メカニズムを調べる脳機能計測実験を行うことで音声知覚のメカニズムに関する包括的な検討を行った。

有声性とは、音声信号から音韻情報を知覚するまでの情報処理過程において、/d/と/t/、/b/と/p/、/g/と/k/などの閉鎖子音における有声音と無声音の音韻対を弁別するための素性のことであり、全世界に存在する言語の約 6 割で用いられている重要な音韻弁別素性である (Diehl, 2008)。また、有声性は音響信号から子音を知覚するまでの情報処理過程で最初に処理される弁別素性とされており、その処理の正確性は他の音韻弁別

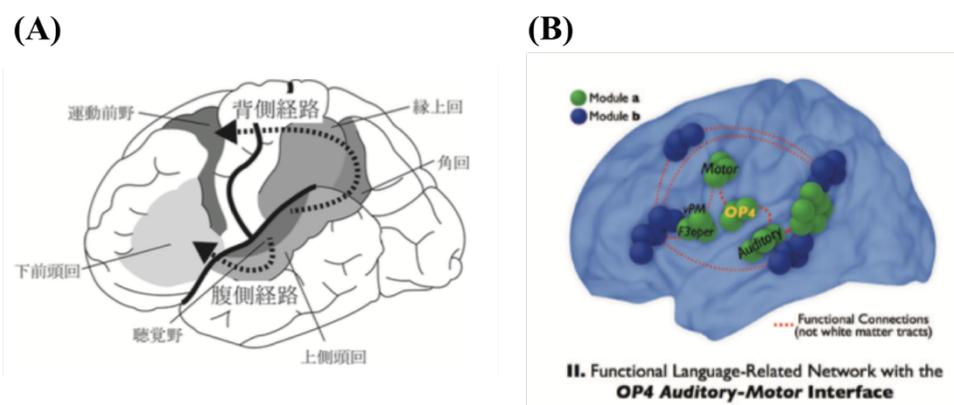


図 2. 音声知覚に係る脳内ネットワーク. (A) 二重経路モデル (廣谷, 2017, p.513, 図 2). (B) OP4 を介して聴覚皮質と運動野・運動前野を接続する経路 (Sepulcre, 2015, p.662, Fig.3B の一部) .

素性（例えば、両唇音（/b/, /p/）、歯茎音（/d/, /t/）、軟口蓋音（/g/, /k/）など調音位置の異なる音韻対を弁別する調音位置素性など）の処理の正確性にも影響を与えると考えられている（Greenberg & Chistiansen, 2019）。有声性は主に有声開始時間（voice-onset time, 以下 VOT）の違いによって特徴付けられる（Lisker & Abramson, 1964）。VOT とは、声門の閉鎖解放から声帯振動の開始までの時間（子音部の開始から母音部の開始までの時間）で無声音の VOT は有声音の VOT よりも長くなることが知られている（図 3A）。有声音と無声音の VOT 長は多くの先行研究で調べられているが、言語圏によって異なることが知られている。例えば、英語母語話者の有声音は 10-40 ms, 無声音は 50 ms 以上の VOT 長を持つ一方（Bailey & Haggard, 1980; Fledge & Eefting, 1986, 1987; Lisker & Abramson, 1970; Newman, 2003）、日本語母語話者やスペイン語母語話者では有声音は 0-20 ms の VOT 長、無声音は 20-50 ms の VOT 長を持つ（日本語母語話者: Ogasawara, 2011; Riney et al., 2007; Tamura et al., 2019, スペイン語母語話者: Fledge & Eefting, 1986; Lisker & Abramson, 1970）。

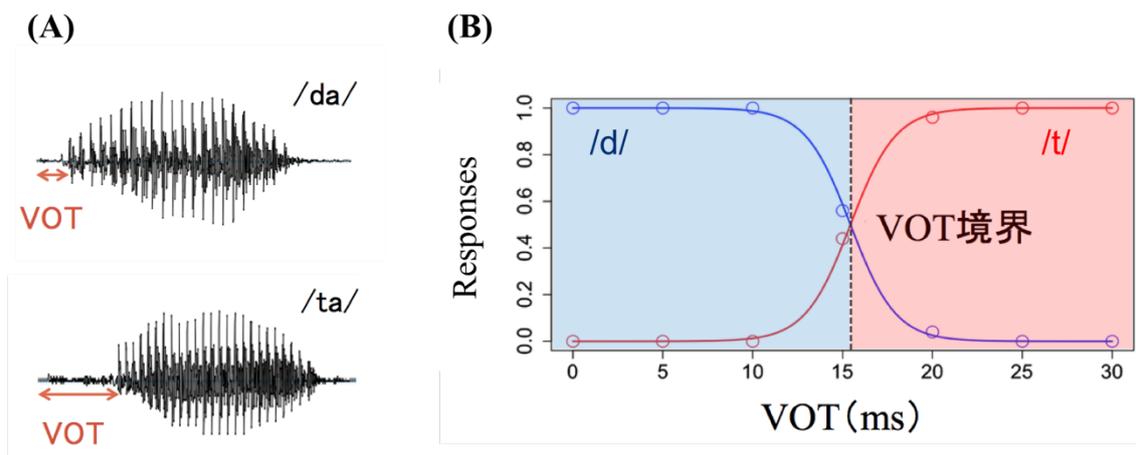


図 3. 有声性の生成と知覚. (A) 有声音/d/と無声音/t/の音声波形とその VOT. (B) VOT を手掛かりとして行われる有声性 (/d/と/t/) の知覚.

有声性の知覚に関しては, VOT を等間隔で段階的に操作した刺激を用いた知覚実験が多く、多くの言語圏の話者を対象として行われており、例えば日本語母語話者の有声音/d/と無声音/t/の知覚判断では 10~20 ms の VOT の長さを境として急激に知覚が変化し、その長さよりも短い場合には/d/, 長い場合には/t/が知覚されることが知られている (図 3B)。このような刺激パラメータに対する非線形的な知覚の変化 (カテゴリー知覚) は他の音韻対を用いた知覚実験でも示されており、ヒトが効率よく音声信号から音韻情報を知覚する上で重要な機能だと考えられている。カテゴリー知覚の特徴としてよく調べられるのは、音韻対の知覚率がともに 50%になる点である音韻境界 (有声性の知覚では VOT 境界と呼ばれる) や知覚曲線の傾きの鋭さであるカテゴリー知覚の明確性の 2 つである (Medina et al., 2010)。有声性の生成と同様に有声性の知覚にも言語差があることが知られており、日本語母語話者における有声音と無声音の VOT 境界は 10-30 ms 程度でスペイン語母語話者と近い値を取る一方 (日本語母語話者: Shimizu, 1977; Tamura et al., 2018, 2019, スペイン語母語話者: Elangovan & Stuart, 2011; Hay & Garcia-Sierra, 2005; Lisker & Abramson, 1970; Williams, 1977), 英語母語話者ではそれよりも凡そ 10 ms 以上長くなることが知られている (英語母語話者: Blumstein et al., 2005; Brandmeyer et al., 2012; Elangovan & Stuart, 2008, 2011; Flege & Eefting, 1987; Hay & Garcia-Sierra, 2005; Lisker, 1975; Lisker & Abramson, 1970; Lisker et al., 1977; Kuhl & Miller, 1975, 1978; Miller et al., 1983; Parker, 1988; Pisoni, 1974, Repp, 1979, Samuel, 1982; Sharma & Dorman, 1998; Simos et al., 1998a; Steinschneider et al., 1999, 2004; Toscano et al., 2018)。

これまでに閉鎖子音の有声性知覚を通して音声知覚メカニズムを検討した先行研究は英語母語話者を対象としたものがほとんどである。さらに、その大半は聴覚情報に基づいた音声知覚処理に注目したものであるが、その契機となった研究として Kuhl and Miller (1975, 1978) の研究が挙げられる。この研究では、言語を持たないげっ歯類

のチンチラが短期の聴覚訓練によって VOT を操作した音声刺激を英語母語話者と同じように弁別出来ることが発見された。また、有声性の音韻対立を特徴付ける数ミリから数十ミリ秒単位の VOT の違いは調音運動の違いとして視覚的に捉えにくいこともあつてか、調音運動情報を基にした音声知覚処理のメカニズムを探る研究では有声性の知覚が題材として扱われることがほとんどない（詳しくは 2 章で述べるが、主に舌で調音されるか唇で調音されるかなどの視覚的に捉えやすい調音運動の違いに特徴付けられる音韻対の知覚が扱われている）。そのため、これまでに有声性の知覚に係る調音運動情報の利用についてはほとんど検討が進んでおらず、英語母語話者の有声性知覚では聴覚情報への依存度が非常に高いという考え方が主流である。

本論文では、音声知覚のメカニズムに関する包括的な検討を行うことを目的として、日本語母語話者が有声性を知覚する過程で聴覚情報と調音運動情報をどのように利用しているかを調べるための行動実験とその脳内メカニズムを調べるための脳機能計測実験を行った（図 4）。聴覚情報の利用については、日本語母語話者がどのような聴覚処理特性を利用して有声性を知覚するために必要な情報を抽出しているのかを Repp（1979）が提案した逆向性マスクング特性が有声性知覚に利用されているという仮説に注目して具体的に検討した（研究 1）。ここでは、その結果と英語母語話者が有声性の知覚に用いている聴覚処理特性について検討した先行研究の結果を比較しながら言語圏によって有声性知覚の VOT 境界が異なる原因について考察を行なった。続いて、言語圏を問わずこれまでほとんど検討が進んでいなかった有声性知覚時の調音運動情報の利用について検討を行うための行動実験を行った（研究 2）。先行研究では有声性の生成と知覚の相関関係を調べることで検討が行われていたが、本研究では聴覚フィードバック課題を用いて有声性の生成を変化させた場合にそれらの知覚にどのような影響が生じるかを調べることで有声性の知覚と生成の相互作用を明らかにするとともに調音運動情報が有声性知覚時にどのように利用されているかについて議論を行な

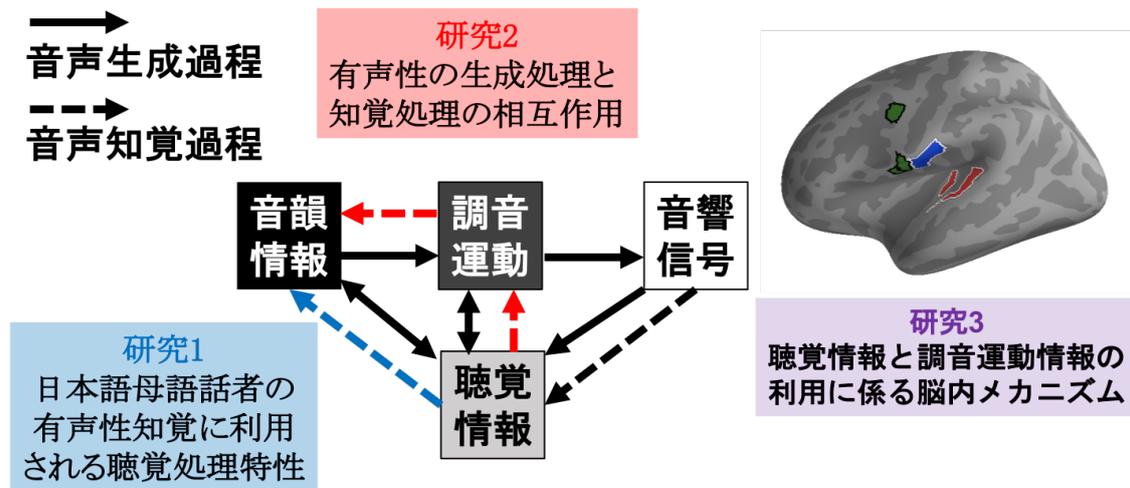


図 4. 本論文を構成する 3 つの研究の目的

った。さらに、有声性知覚時の聴覚情報と調音運動情報の利用に係る脳内メカニズムを明らかにするために脳磁図 (magnetoencephalography, 以下 MEG) と機能的磁気共鳴画像法 (functional magnetic resonance imaging, 以下 fMRI) を用いた脳機能計測実験を行った (研究 3)。有声性知覚時の聴覚情報の利用に係る脳内メカニズムを調べることを目的とした研究では、数ミリから数十ミリ秒単位の時間情報 (VOT) に対する聴覚皮質の応答を調べられるように時間解像度の高い脳機能計測手法を用いる必要がある。一方で、調音運動情報の利用に係る脳内メカニズムを調べることを目的とした研究では、聴覚皮質に比べて活動強度の低い音声生成に関与する脳領域の活動を精度よく捉えるために空間解像度が高い脳機能計測手法を用いる必要がある。有声性知覚時の聴覚情報と調音運動情報の利用に係る脳内メカニズムの両方を同時に調べるのは脳機能計測手法の制約を考慮すると難しい課題であるが、本研究では高い時間解像度と空間解像度を合わせ持つ MEG を用いた実験を行うことでこれを実現した。さらに、MEG よりも空間解像度の高い fMRI で有声性知覚時や有声性生成時に活動する脳領域を特定してその結果を基に MEG の解析を行うことで、MEG 計測実験で得られた結果

の信頼性を高めるとともに音声知覚処理の脳内メカニズムを生成処理の脳内メカニズムと関連付けて詳細に議論を行なった。



## 第2章 研究背景

### 2.1 音声知覚メカニズムに関する仮説

聴覚情報に基づく音声知覚処理を重要視する仮説では、特定の音韻あるいは音韻弁別素性に不変的な特徴が聴覚系での情報処理を通して容易に取り出されるように言語システムが成り立っていると考えられている。Blumstein and Stevens (1979, 1980) が提唱した *Acoustic invariance theory* では、音声信号自体に音韻弁別素性に不変的な特徴が含まれていると考えられている。この研究では、調音位置の異なる3種類の閉鎖子音（両唇音 (/b/, /p/) , 歯茎音 (/d/, /t/) , 軟口蓋音 (/g/, /k/)）のスペクトル特徴から調音位置の異なる音韻を弁別するための不変的な特徴の探索が行われた。その結果から、広い周波数帯域にパワーが拡散するか狭い帯域にパワーが集約するか、また、広い周波数帯域にパワーが拡散する場合にパワーが高域で強くなるか低域で強くなるかに注目することで調音位置素性に関する不変的な特徴を見出すことが出来ると主張がされている。ただし、これらの特徴はあくまで物理的な特徴であるため聴覚系での情報処理を経た後にも不変であるかどうかは定かではないという問題点が存在する。他にも、Liljencrants and Lindblom (1972) の *Dispersion theory* では各言語の母音体系の成り立ちについて議論がなされており、母音体系を構成する複数の母音が聴覚系での情報処理を通して容易に聴き分けられるように、それらが出来る限り音響的な違い（距離）を持つようにデザインされていると考えられている。Diehl et al. (2003) は、母音の音響的な距離だけではなく聴覚的な距離に注目をして母音体系の成り立ちについて議論するために聴覚モデルを用いた母音の分析を行っている。また、音声信号自体に特定の音韻や音韻弁別素性に不変的な特徴が存在しない場合であっても、音響信号の持つ音響的特徴と聴覚情報処理の間の非線形的な性質によって作り出される聴知覚の不連続点を利用することで不変的な特徴が抽出されるという仮説も存在する (Diehl, 2007;

Holt et al., 2004; Kuhl & Miller et al., 1978; Stevens, 1989)。後節で詳しく解説するが、有声性の知覚における不変特徴の探索については、この仮説に基づいて数多くの先行研究（Miller et al., 1976; Parker, 1988; Pisoni, 1977; Repp, 1979; Simos et al., 1998a, b; Steinschneider et al., 2004）で検討がなされている。

調音運動情報に基づく音韻情報の知覚処理に関する仮説として最も有名なものは *motor theory of speech perception* (Liberman et al., 1967; Liberman & Mattingly, 1985) である。この仮説では聴覚系での情報処理を介することなく音声信号からその信号を生成した調音運動を推定し、その情報を基に音韻情報の知覚が行われると考えられている。*Motor theory of speech perception* で想定されている情報処理過程は音声知覚特有のものだと考えられていたが、Fowler (1986) の *Action theory* では音響信号からそれを作り出した運動情報を推定するメカニズムは音声知覚特有のものではなく、非音声の処理にも共通するものだと述べている。この考えは、音声信号ではなくても聴取した音からその原因となった運動を想起できるものであればそれらの運動に係る脳領域の活動が生じるという知見からも支持される (Agnew et al., 2011; Chen et al., 2008; Patel & Iverson, 2014; Schaefer et al., 2014)。近年、Fowler et al. (2016) は上記の2つの仮説の考えの違いと共通点をまとめて *Gesture theory* と呼んでいる。また、音声知覚時の音声生成系の働きに注目した他の仮説として、*Analysis-by-Synthesis theory* (Stevens & Halle, 1967; Poeppel et al., 2008) が存在する。この仮説では、まず始めに、聴取された音声信号が聴覚的に分析されることでその信号に含まれる音韻あるいは音韻弁別素性について仮説が立てられる。そして、その仮説を検証するために脳内で音声を生成して入力音声との照合が行われるという処理メカニズムが考えられている。*Analysis-by-Synthesis theory* が *Gesture theory* と異なる点は、音声生成系の働きだけでなく聴覚情報の利用も重要視している点だと考えられる (筧, 1995)。近年提案されている仮説でも、音響信号を作り出す原因となった調音運動情報の推定は聴覚情報と調音運動の対応関

係に基づいて行われると考えられている (Barnaud et al., 2016; Laurent et al., 2017; Schwartz et al., 2012)。

## 2.2 行動実験による音声知覚メカニズムの検討

### 2.2.1 聴覚情報に基づく音声知覚処理の役割を検討した先行研究

2.1 章で取り上げたように、聴覚情報に基づく音韻情報の知覚処理を重要視する仮説では音声信号から聴覚系での情報処理を通して特定の音韻あるいは音韻弁別素性に不変的な特徴が容易に取り出されるように言語システムが成り立っていると考えられている。有声性の知覚に関しては、VOT と聴覚情報処理の間の非線形的な性質が利用されることで有声音と無声音を弁別するための不変的な特徴が取り出されると考えられており (Diehl et al., 2007; Kuhl & Miller, 1978) , 先行研究では具体的にどのような聴覚処理特性が有声性知覚に利用されているのかが検討されている。英語母語話者の有声性知覚に用いられる聴覚処理特性として挙げられているのは、2つの音イベントの同時性知覚 (時間順序知覚) に係る処理である (Pisoni, 1977; Simos et al., 1998a; Steinschneider et al., 2004)。同時性知覚に係る聴覚処理特性がどのようにして有声性知覚に利用されていると考えられているのかについて具体的に説明をしていく。まず、2つの音イベントの同時性知覚は2つの音の時間間隔に比例するのではなくある長さを境に非線形的に変化し、それよりも短い場合には2つの音イベントは同時に始まったと知覚され、また、長い場合には最初に始まった音と遅れて始まった音の時間順序が正しく知覚される (Pisoni, 1977)。そして、2音の開始の時間順序が正しく知覚されるための時間間隔は、2つの音がどのような音響特徴を持つかにも依存するが、英語母語話者の VOT 境界と比較的近いことが明らかにされている (Pisoni, 1977; Parker, 1988; Summerfield, 1982)。このような処理特性を有声性の知覚に当てはめると (図 5) , 雑音成分 (子音部) の開始と周期成分 (母音部) の開始が同時に始まったと知覚される

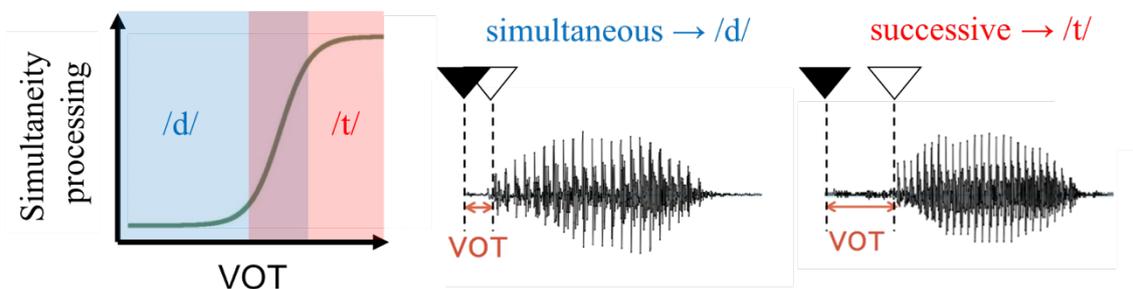


図 5. 同時性知覚に係る聴覚処理特性を利用した有声性知覚メカニズムの説明図.

VOT 長の刺激は実質的に VOT が 0 ms の刺激と同じように知覚されるため有声音が知覚される, また, 雑音成分の開始よりも周期成分の開始が遅れていると知覚される VOT 長では無声音が知覚されると考えられる。この仮説は有声性知覚と同時性知覚の類似性が示された行動実験の結果から支持されるだけでなく, 後節で詳しく解説する Simos et al. や Steinschneider et al. の脳機能計測研究で, 2 音の開始の同時性知覚を行っている際に見られる一次聴覚野の応答が有声性知覚時のものと類似することが報告されていることから支持されている。

英語母語話者を対象とした実験では, 有声性知覚に關与する具体的な聴覚処理特性について検討がなされているが, その他の言語圏の話者ではほとんど検討がなされていない。Hay and Garcia-Sierra (2005) は, 英語母語話者とスペイン語母語話者を対象として有声性知覚課題および同時性知覚課題を行い, 有声性知覚については言語間で明確な違いが見られたのに対し, 同時性知覚課題については言語間で結果に違いが見られなかったことを報告している。この結果は, 英語母語話者とスペイン語母語話者で有声性の知覚に違いが出るのは同時性知覚に係る聴覚処理特性に違いがあるからではないことやスペイン語母語話者は同時性知覚に係る聴覚処理を有声性の知覚に利用していない可能性を示唆している。Elangovan and Stuart (2008) や Kuhl and Miller (1978) では, 言語によって有声音と無声音の知覚が切り替わる VOT の長さが異なる

ため、有声性の知覚に用いられる聴覚特性も言語によって異なる可能性も示唆されているが、検討は全く行われていない。

## 2.2.2 調音運動情報に基づく音声知覚処理の役割を検討した先行研究

調音運動情報に基づく音韻情報の知覚処理の役割を重視した仮説では、音声知覚処理に音声生成に関する知識が用いられると考えられている。そのため、音韻の生成とそのカテゴリ知覚の間に有意な相関が見られること報告した先行研究 (Bailey & Haggard, 1980; Chao et al., 2019; Perkell et al., 2004) は調音運動情報に基づいた音韻情報の知覚処理の役割の重要性を支持するものだと考えられている。有声性の生成と知覚の関係についてはいくつかの先行研究で検討がなされており, Bailey and Haggard

(1980) の研究では、有声音と無声音の発声においてそれらの VOT の違いが大きな参加者ほど VOT を操作した刺激を用いた知覚課題でのカテゴリ知覚が明確になる傾向があることが示されている。また、音声生成に関与する脳領域に損傷が見られる患者を対象として有声性の生成や知覚を調べた研究では、有声音と無声音の生成においてそれらの VOT の違いが健常者と比べると小さいこと (Blumstein, 1980; Ivry & Gopal, 1993) や VOT を操作した刺激を用いた知覚実験におけるカテゴリ知覚の明確性が健常者と比べて低いことが示されている (Ackermann et al., 1997; Basso et al., 1977)。母音の生成と知覚の関係性を調べた研究でも有声性知覚と同様に母音対の持つ音響的な距離の関係とカテゴリ知覚の明確性の間に相関があることが報告されている。Perkell et al.の研究では4つの母音 (/a/, /ɛ/, /u/, /o/) の生成課題と第1, 2, 3フォルマント周波数を等間隔で段階的に変化させて/a/から/ɛ/, または,/u/から/o/へと知覚が変化する刺激を用いた知覚課題を行なった。その結果,/a/と/ɛ/, /u/と/o/の発声の違いが大きい参加者ほどそれぞれ/a/と/ɛ/, /u/と/o/の音韻対のカテゴリ知覚の明確性が

高くなることが示された。以上の先行研究から、調音運動情報に基づく音声知覚処理は音韻対のカテゴリー知覚の明確性に関与することが示唆される。

上記の先行研究では、音韻の生成とそのカテゴリー知覚の間の相関関係が示されているが、その結果から両者の相互作用や因果関係を示すことは出来ない。そこで、近年の研究では、変形聴覚フィードバック (transformed auditory feedback, 以下 TAF) 課題を用いて音韻の生成を変化させることで、それらのカテゴリー知覚にどのような影響が生じるかを調べる研究が行われている (Lametti et al., 2014; Shiller et al., 2009; Schuerman et al., 2017a, b)。TAF 課題とは、発話音声の音響的特徴を実時間で変形して発話者に聴覚フィードバックする課題のことである (Houde & Jordan, 1998; Kawahara, 1994)。1章で述べたように、音声生成過程では発声に伴って生じる聴覚情報をフィードバックとして受け取ることで音韻情報や調音運動と聴覚情報の対応関係の確認が行われると考えられている。TAF 課題では音韻情報や調音運動から予測される聴覚情報とは異なる聴覚情報がフィードバックされるため、音韻情報や調音運動と聴覚情報の対応関係が修正されることで音声生成処理に変更が加えられる。具体的には、TAF 課題である特定の音響的特徴が変形されると、その変形とは逆方向に音響的特徴が修正された発声が行われるようになること (補償応答) が多くの先行研究で明らかにされている (Houde & Jordan, 1998, 2002; Jones and Munhall, 2000, 2005; Villacorta et al., 2007)。Shiller et al.は、/s/の音韻を含む単語を発声する際に、発話音声のスペクトル重心を/j/に近づける方向にわずかに変形した音声を発話者にフィードバックする課題を行い、その課題の前後で/s/と/j/のカテゴリー知覚に変化が見られるかどうかを調べた。その結果、/s/の生成に補償応答が見られるとともに、/s/と/j/のカテゴリー知覚の音韻境界も変化することを示した。このような知覚の変化は、TAF 課題で用いた聴覚刺激を聴くのみでは起こらないこと、また、フィードバック音声を変形しない通常の発声課題の前後に確かめられた知覚の変化とは変化の仕方が異なることから、音韻の生成処理の変容がそのカテゴリー知

覚に影響を及ぼしたと考察されている。Lametti et al. では、/ɛ/の音韻を含む単語を発声する際にその第1フォルマント周波数を/æ/に近づけるようにわずかに下げる、または、/ɪ/に近づけるようにわずかに上げた音声を発話者に聴かせる課題を行い、その課題の前後で/ɛ/と/æ/や/ɛ/と/ɪ/のカテゴリー知覚に影響が生じるかどうかを調べた。その結果、/æ/の音響的特徴に近づけた音声がフィードバックされることで/ɛ/の発声が/ɪ/に近い音響的特徴に変化するという補償応答が起きた前後では、/ɛ/と/ɪ/のカテゴリー知覚の音韻境界が変化するが/ɛ/と/æ/のカテゴリー知覚には変化が見られないことが明らかにされた。それとは逆に、/ɪ/の音響的特徴に近づけた音声がフィードバックされることで/ɛ/の発声が/æ/に近い音響的特徴に変化するという補償応答が起きた前後では、/ɛ/と/æ/のカテゴリー知覚の音韻境界が変化するが/ɛ/と/ɪ/のカテゴリー知覚には変化が見られないことが明らかにされている。つまり、特定の音韻間でそれらの生成に近接が起きた場合のみそれらのカテゴリー知覚にも影響が生じることが示された。また、このような知覚の変化は聴覚フィードバックに変形は加えず通常通りに発声を行う前後では起きないことが示されたことから、TAFによる音韻生成処理の変容がそのカテゴリー知覚にも影響を与えたことが示唆されている。

Patri et al. (2018) では、Lametti et al.の研究でTAF課題の前後でなぜ知覚に変化が生じたのかについてより詳細に議論がなされている。Patri et al.は、TAF課題が音声生成・知覚の情報処理過程における音韻情報、調音運動、聴覚情報の対応関係にどのような影響を及ぼすのかを考察した上で、それらの対応関係の変化が音韻の生成や知覚にどのような影響を及ぼすのかを検討している。まず、/ɛ/の音韻を含む単語の生成時に音響特徴がわずかに変形された聴覚情報がフィードバックされると、音韻情報や調音運動から予測される聴覚情報とは一致しない情報が入力される。そのため、音韻情報 (/ɛ/) に対応する聴覚情報の知識や音韻情報 (/ɛ/) を生成する際に起こす調音運動情報と聴覚情報の対応関係に関する知識が更新されると考えられている (図6 橙枠)。

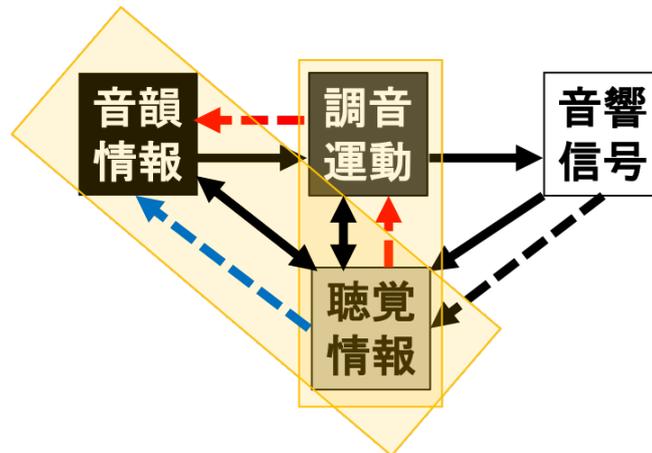


図 6. 変形聴覚フィードバック課題が音韻情報・調音運動・聴覚情報の対応関係に与える影響の予測 (Patri et al., 2018 を基に作成) .

このような更新が起きると、音韻情報を音響信号として生成する処理も更新されて補償応答が生じる。次に、聴覚情報と調音運動情報の両方が音声知覚に利用される仮定のもとで、上記の2つの更新が音声知覚処理にどのような影響を及ぼすのかについて議論しており、音韻情報と聴覚情報の対応関係の変化の影響についてはその音韻を知覚する際の聴覚情報に基づく音韻知覚処理に影響を及ぼすと考えられている (図 6 青点線)。また、調音運動情報と聴覚情報の対応関係の変化については、音声知覚時に聴覚情報と調音運動情報の対応付けを行う処理を行う調音運動情報に基づく音韻知覚の処理経路に影響を及ぼすとされている (図 6 赤点線)。つまり、TAF 課題は音声知覚における聴覚情報と調音運動情報の利用の両方に変容をもたらすと考えられており、Lametti et al. (2014) や Shiller et al. (2009) の研究では調音運動情報に基づく音韻知覚処理の役割を直接的に探ることが出来ていない可能性が指摘されている。

## 2.3 脳機能計測実験による音声知覚メカニズムの検討

### 2.3.1 脳機能計測手法

脳機能計測手法の内、音声知覚のメカニズムを検討した関連研究でよく用いられている脳波 (electroencephalography, 以下 EEG) や皮質脳波 (electrocorticogram, 以下 ECoG), 本研究で用いる脳磁図 (magnetoencephalography, 以下 MEG) や機能的磁気共鳴画像法 (functional magnetic resonance imaging, 以下 fMRI) についてそれらの計測方法や特徴について簡潔に述べる。

EEG, ECoG はどちらも脳内の神経細胞で生じた電気活動を計測する脳機能計測手法で, EEG では頭皮上に設置した電極で, ECoG では脳の表面に設置した電極で計測される。MEG では, 脳内の電気活動に伴って生じる磁場の変化を超伝導量子干渉計という磁束感度が非常に高いコイルを用いて計測する。上記の脳機能計測手法の特徴は時間分解能が高い点で, 脳内で生じた電気活動に関連した信号をミリ秒単位で計測することが出来る (図 7)。一方で, EEG や MEG は空間分解能が低く, 計測されたデータからその信号源となる脳領域を特定することは容易ではない。特に, EEG の場合には, 電

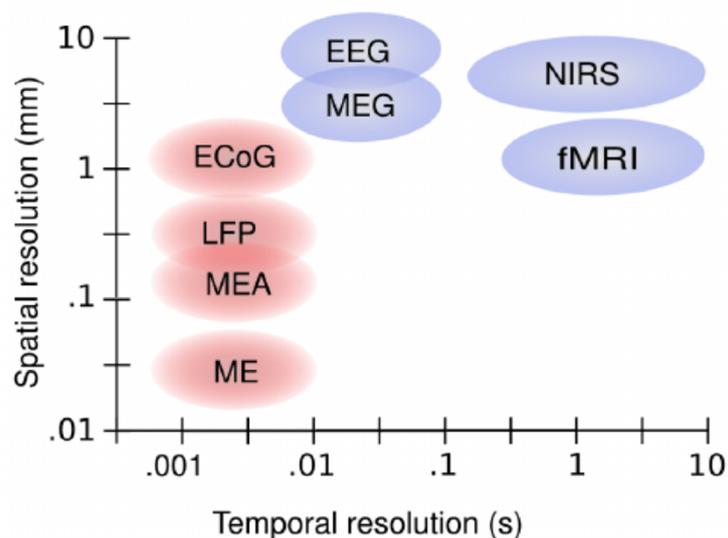


図 7. 各種脳機能計測手法の時間分解能と空間分解能 (van Gerven et al., 2009, Figure2) .

ECoG = 皮質脳波, LFP = 局所フィールド電位記録, MEA = マイクロ電極アレイ記録, ME = マイクロ電極記録, EEG = 脳波, MEG = 脳磁図, NIRS = 近赤外分光分析法, fMRI = 機能的磁器共鳴画像法

気活動が生じた神経細胞からデータを計測する頭皮上までの間に誘電率の異なる物質が複数存在し、その間の電流伝搬を完全にモデル化することは非常に難しい。そのため、計測データから信号源での電気活動を求める逆推定の精度は必然的に低くなる。MEG については、頭部の透磁率がほぼ均一であることから EEG に比べて信号源の推定精度が高いとされている。しかしながら、その推定精度は実験方法やデータの解析方法によって大きく異なってくるため注意が必要である (Gross et al., 2013)。実際に MEG の空間解像度は後述する fMRI と同等と考えられることもあれば EEG と同程度と考えられることもある (図 7)。ECoG を用いた実験では、脳の表面に電極を設置するため特定の脳領域の活動を精度良く評価することが出来るが、主に脳外科手術の一環で行われることが多いこともあって実験が麻酔下で行われるという欠点がある。さらに、一部の脳領域に対象を絞って計測が行われる場合が多く、脳内のネットワーク活動を調べるのが難しいなどの欠点もある。

fMRI は電磁気的変化を測定している点は EEG, ECoG, MEG と同じだが、計測対象とするものは神経細胞の電気活動ではなく、それに伴って生じる血流動態および代謝活動に起因する Blood Oxygenation Level Dependent (以下 BOLD) 信号の変化である (Ogawa et al., 1990, 1992)。fMRI の長所は空間解像度が高いことで、全脳をミリメートル単位のボクセルに分割して活動を調べることが出来る (図 7)。一方で短所として挙げられるのは、特定の知覚や認知に応じて生じる神経活動を直接捉えているわけではない点と時間解像度が低く各脳領域の活動の時間的变化を十分に調べる事が出来ない点である。また、BOLD 信号の計測時には非常に大きなスキャン音が鳴ることから聴覚刺激を用いた実験を行う場合にはスパースサンプリング撮像法を用いるなど実験デザインをする上で工夫が必要である (Hall et al., 1999; Peele, 2014; Perrachione & Ghosh, 2013)。

### 2.3.2 聴覚情報に基づく音声知覚処理の脳内メカニズムを検討した先行研究

聴覚情報に基づく音声知覚メカニズムの脳内メカニズムを検討した研究の中で最も代表的なものは ECoG を用いて音声刺激聴取時の聴覚皮質（ヘッシェル回や上側頭回）の活動を調べた研究である（Chang et al., 2010; Steinschneider et al., 1999, 2004）。有声音の知覚に関しては, Steinschneider et al. が VOT を等間隔で段階的に操作して, 有声音 /d/ から無声音 /t/ へと知覚が変化する刺激を作成し, それらの刺激を聴取している際のヘッシェル回の活動を ECoG で調べた。その結果, /t/ と知覚される刺激（VOT 40, 60, 80 ms）の聴取時には子音部の始まりに対して生じる神経応答とは分離して母音部の始まりに対して生じる神経応答が現れることが明らかにされた（図 8）。一方, /d/ と知覚される刺激（VOT 0, 20 ms）の聴取時には子音部と母音部の始まりに対する神経応答は分離せず単一のピークが見られることが明らかにされた。これと類似した結果は, EEG を用いた実験でも示されている（Sharma & Dorman, 1999）。また, Chang et al. では調音位置素性の知覚時の聴覚情報の利用に関する脳内メカニズムが検討されている。この

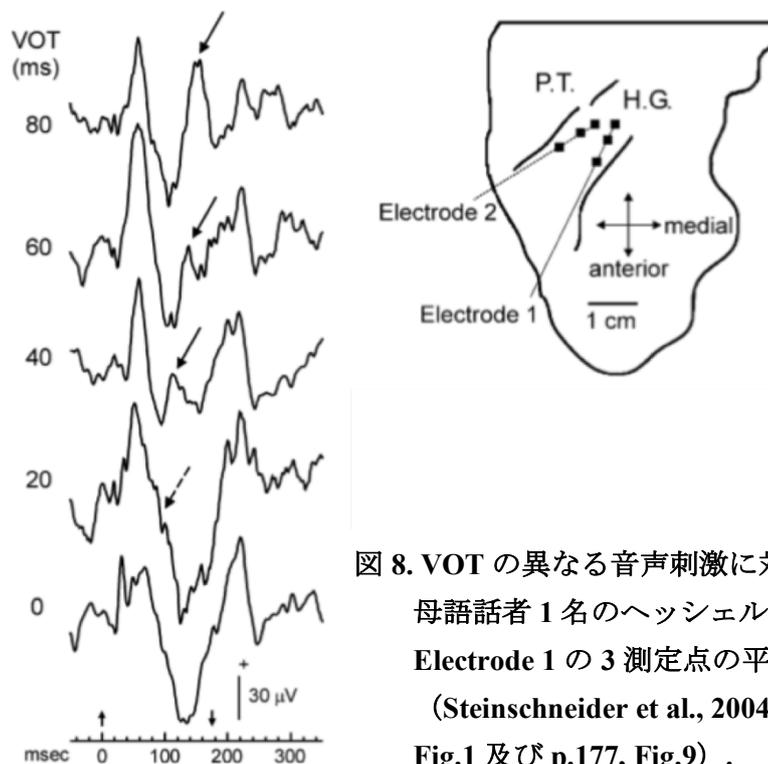


図 8. VOT の異なる音声刺激に対する英語母語話者 1 名のヘッシェル回（右図 Electrode 1 の 3 測定点の平均）の応答（Steinschneider et al., 2004, p.171, Fig.1 及び p.177, Fig.9）。

研究では、第2フォルマントの遷移開始周波数を等間隔で段階的に操作して、両唇音/b/から歯茎音/d/、歯茎音/d/から軟口蓋音/g/へと知覚が変化する刺激が作成され、それらの刺激を聴取している際の上側頭回の活動計測が行われた。音声刺激呈示後110~150msの潜時で見られる上側頭回での脳活動分布から各刺激間での聴覚皮質応答の類似度が調べられた結果、刺激間の音響パラメータの違いに依存せず、それぞれ/b/、/d/、/g/と知覚される刺激同士の類似度が高いことが明らかにされた。上記の研究は英語母語話者を対象としたものだが、刺激の音響パラメータに対する聴覚応答パターンに音韻知覚（有声性や調音位置素性の知覚）と相関するパターンが見られることが示されている。

Simos et al. (1998a, b) のMEG研究では、英語母語話者を対象として有声性知覚に係る聴覚情報の利用に係る脳内メカニズムを調べるだけでなく2つの音イベントの同時性知覚に関連する脳活動も調べられている。Simos et al. (1998a) では、VOTを0msから60msまで20ms間隔で操作して有声音/g/から無声音/k/へと知覚が変化する刺激を作成し、それらの刺激を聴取している際のMEG計測が行われた。この研究では、どのVOT長の刺激に対しても単一のピークを持つ聴覚反応が確かめられたが、そのピーク振幅が/k/と知覚される刺激（VOT 40, 60ms）を聴いている時よりも/g/と知覚される刺激（VOT 0, 20ms）を聴いている時の方が大きくなるのに対し、同じ音韻が知覚される刺激同士ではほとんど振幅が変わらないことが明らかにされた。また、Simos et al.

(1998b) では、2種類の純音の開始時間差を0msから60msまで20ms間隔で操作した刺激を作成し、それらの刺激を聴取している際のMEG計測が行われた。その結果、聴覚反応のピーク振幅が2音のどちらか一方が先に始まったと知覚される刺激（開始時間差 40, 60ms）を聴いている時よりも2音が同時に始まったと知覚される刺激（開始時間差 0, 20ms）を聴取している時に大きく異なるのに対し、同じ知覚結果が生じる刺激同士では振幅にほとんど差が見られないことが示されている。2つの実験結果

を比較すると、有声音が知覚される時と無声音が知覚される時の聴覚反応の違いが2音の開始が同時と知覚される時と一方が先に始まったと知覚される時の聴覚反応の違いと酷似していることが分かる。英語母語話者の有声性知覚に2つの音イベントの同時性知覚に係る聴覚処理特性が密接に関与しているという仮説は行動実験だけでなく脳機能計測実験の結果からも支持されている。

有声性知覚に相関する聴覚応答パターンは英語母語話者以外を対象とした実験では見つかっていない。例えば, Elangovan and Stuart (2011) は英語母語話者およびスペイン語母語話者を対象とした EEG 計測を行い, VOT を操作した刺激に対する聴覚反応を調べた。その結果, スペイン語母語話者でも英語母語話者と同様に VOT 30 ms より長い場合には母音部の始まりに対する神経応答が刺激の始まりに対する神経応答と分離して見られるのに対し, 30 ms より短い場合には単一のピークが得られることを示した。しかしながら, スペイン語母語話者の有声音と無声音の知覚は 10-20 ms 付近で切り替わることからピークが分離するか否かの聴覚応答パターンはスペイン語母語話者の知覚結果と一致しないことを示している。

### 2.3.3 調音運動情報に基づく音声知覚処理の脳内メカニズムを検討した先行研究

調音運動情報に基づく音声知覚の脳内メカニズムを検討した研究は, ミラーニューロンの発見 (Rizzolatti et al., 1996) や Wilson et al. (2004) による fMRI 研究で音声知覚時に音声生成に関与する脳領域が活動することが報告されて以降, 音声の生成に関与する脳領域の活動を調べたものが数多く存在する。Pulvermüller et al. (2006) の fMRI 研究では, 唇で調音される/p/の音韻を含む音節や舌で調音される/t/の音韻を含む音節の生成及び知覚を行なっている際の脳活動が調べられた。その結果, 唇や舌で調音される音韻を知覚する際にはそれぞれ唇や舌を動かす脳領域が賦活することが示された (図 9)。この結果は, 調音位置素性の知覚に調音位置の制御に係る脳内メカニズムが

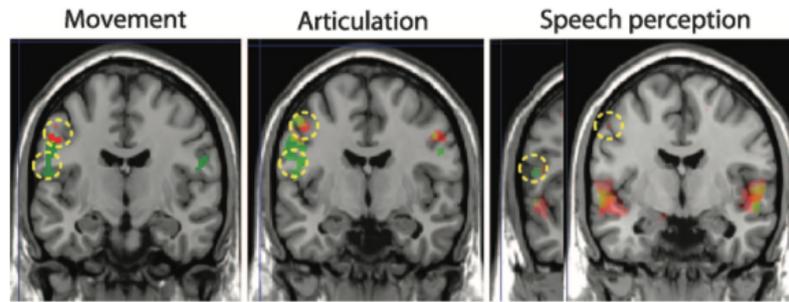


図 9. (左) 唇を動かした際に賦活した脳領域 (赤) と舌を動かした際に賦活した脳領域 (緑) . (中央) /p/を生成した際に賦活した脳領域 (赤) , /t/を生成した際に賦活した脳領域 (緑) . (右) /p/を聴取した際に賦活した脳領域 (赤) , /t/を聴取した際に賦活した脳領域 (緑) (Pulvermüller et al., 2006, p.7867, Fig.3) .

関与していることを示しており、音声知覚時にその音声信号を生成した調音運動に関する情報を脳内で推定している (Barnaud et al., 2016; Fowler, 1986; Laurent et al., 2017; Liberman et al., 1967; Liberman & Mattingly, 1985; Schwartz et al., 2012) , あるいは、入力音声と脳内で生成した音声の照合をしている (Poeppl et al., 2008; Stevens & Halle, 1967) という音声知覚における音声生成系の役割を重要視した仮説を支持するものだと考えられる。さらに、Chevillet et al. (2013) は、fMRI で調べた音声生成に関与する脳領域の活動が音韻カテゴリー知覚に関与することを示唆している。この研究では、第 2 フォルマント周波数を等間隔で段階的に操作することで歯茎音/d/から軟口蓋音/g/へと知覚が変化する刺激が作成され、それらの刺激を用いた知覚実験と脳機能計測実験が行われた。fMRI のデータの分析では、音韻情報を知覚する際の音声生成に関与する脳領域の活動には大きな個人差が見られるという報告 (Szenkovitz et al., 2012) に基づいて、運動前野の活動の個人差が調べられ、その個人差が調音位置の違いに特徴付けられる音韻対のカテゴリー知覚の明確性に反映されることが明らかにされた。この結果は、調音運動情報に基づく音声知覚処理が音韻のカテゴリー知覚の明確性に関与することを示唆する行動研究の知見とも一致するものである。

有声性知覚時の音声生成に関与する脳領域の活動を報告した研究は英語母語話者を対象とした研究でも数少ないが, Toscano et al. (2018) が fast optical imaging という脳機能計測手法を用いて VOT を操作した刺激群を聴取している際の脳活動を計測し, 音声刺激呈示後 200 ms 付近で運動前野の活動が見られることを報告している。しかしながら, この研究では運動前野は刺激 (VOT) 依存性の活動を起こすことは示されているが, その活動と知覚成績の関連性については議論されていない。その他にも, 有声性知覚時の脳活動を探索的に調べた研究は, 他にも Blumstein et al. (2005) や Myers et al. (2009) の fMRI 研究があるが, これらの研究では音声生成に関与する脳領域の活動は報告されていない。

脳機能計測とは異なるが, いくつかの研究で, 経頭蓋磁気刺激 (transcranial magnetic stimulation, 以下 TMS) を用いて音声の生成に関与する脳領域の活動を一時的に抑制することで音韻の知覚にどのような影響が生じるのかが調べられている。D'Ausilio et al. (2009) の実験では, Pulvermüller et al. (2006) の fMRI 研究の結果を受けて, 唇や舌の動きを制御する脳領域の活動を一時的に抑制することで, 雑音環境下でそれぞれ唇や舌で調音される音韻の知覚が困難になることを示した。Möttönen and Watkins (2009, 2012) は, 音声生成に関与する脳領域の活動抑制がカテゴリー知覚に影響を及ぼすことを明らかにした。この研究では, 第 2 フォルマントを操作して両唇音から歯茎音 (/b/から/d/や/p/から/t/) へと知覚が変化する刺激が作成され, それらの知覚成績が TMS で運動野の唇を制御する領域の活動を抑制する前後でどのように変化するのかが調べられた。その結果, TMS を施行する前よりも後の方が有意にカテゴリー知覚の明確性が低くなることが明らかにされた。この結果は, Chevillet et al. (2013) の fMRI 研究で示された, 音声の生成に関与する脳領域の活動が調音位置の違いに特徴付けられる音韻対のカテゴリー知覚の明確性に関与するという知見と一致する結果である。一方, Möttönen and Watkins の研究では VOT を段階的に操作した刺激を用いた有声

性知覚課題も行われているが、唇の運動制御に関与する脳領域の活動を抑制してもその知覚結果に影響を与えないことが示されている。この結果は、有声性の知覚には調音運動情報が利用されていない、あるいは、利用されていたとしても調音位置の制御を行う脳内メカニズムは調音の時間的制御によって特徴付けられる有声性の知覚には関与していない可能性を示唆している

## 2.4 先行研究での未解決点と本論文での具体的な検討課題

ヒトの効率的な音声情報処理を支える機能である音韻カテゴリー知覚において、聴覚情報および調音運動情報がどの程度利用されているかを調べた先行研究の結果から、聴覚情報とその主要な手がかりになるとともに、調音運動情報が聴覚情報を補完してカテゴリー知覚を明確にするというメカニズムの存在が示唆される。ただし、本論文で題材に扱う有声性のカテゴリー知覚における聴覚情報および調音運動情報の利用については未解決点が多い。英語母語話者を対象として有声性の知覚メカニズムを調べた研究では、聴覚情報に基づく音韻知覚処理の役割を検討したものが数多く存在し、行動実験と脳機能計測実験の両方の結果から、2つの音イベントの同時性知覚に係る聴覚処理特性が有声性の知覚に密接に関わっていることが明らかにされてきた。一方で、英語以外を母語とする参加者では、有声性知覚のVOT境界が英語母語話者と大きく異なることから、言語によって有声性知覚に利用される聴覚処理特性が異なることが示唆されている (Elangovan & Stuart, 2008; Kuhl & Miller, 1978)。しかしながら、英語以外を母語とする話者がどのような聴覚処理特性を利用して有声性のカテゴリー知覚を実現しているのかについては全く知見がなく、VOTに対する聴覚応答パターンを調べた研究でも英語母語話者で見られたような有声性知覚との相関は発見されていない。

有声性知覚時の調音運動情報の利用については、英語圏でもその他の言語圏でもほとんど検討が進んでいない。Bailey and Haggard (1980) は、有声性の生成とカテゴリ

一知覚の明確性の間の関連性を報告しているが、それらの中に強い相関が見られているわけではなく因果関係が存在するかどうかについても検討されていない。調音運動情報に基づく音声知覚処理のメカニズムを調べるための行動研究では、TAF 課題を利用して音声生成と知覚の相互作用を調べることで両者の因果関係を探る研究が近年行われているが (Lametti et al., 2014; Schuerman et al., 2017a, b; Shiller et al., 2009) , 有声性の生成と知覚の相互作用については検討されていない。その理由としては考えられるのは、VOT という数ミリから数十ミリ秒単位の時間情報を実時間で変形をしてフィードバックすることは困難であるためである (Mitsuya et al., 2014)。さらに、Patri et al.

(2018) によると、TAF 課題は聴覚情報と調音運動情報を利用した音声知覚処理の両方に影響を及ぼす可能性が考えられるため、TAF 課題が音声知覚に与える影響について検討した先行研究では調音運動情報に基づく音声知覚処理のメカニズムを直接的に検討できていない可能性が考えられる。

有声性知覚時の調音運動情報の利用に係る脳内メカニズムについては、Toscano et al. (2018) によって音声刺激呈示後 200 ms 付近で音声の生成に関与する運動前野の活動が生じることが報告されているが、その活動と有声性知覚の関連性については十分に検討がなされていない。また、この研究では VOT を 5 ms 間隔で変化させた刺激が用いられ、聴覚皮質や運動前野が刺激 (VOT) 依存性の活動を示すことが報告されたが、彼らが用いた fast optical imaging の時間分解度は 24 ms と VOT の刻み幅よりも大幅に低かった。そのため、有声性知覚に係る脳内メカニズムが十分に検討されているとは言いがたい。脳機能計測によって調音運動情報の利用に係る脳領域の活動を調べるためには高い空間分解能を持つ脳機能計測が必須であるが、ミリ秒単位の時間情報を手掛かりとして行われる有声性知覚に係る脳内メカニズムを検討するためには高い時間分解能も必要であろう。

## 2.5 本論文での具体的な検討課題

本論文では、上述した先行研究での未解決点を解決するために、日本語母語話者を対象として有声性知覚に聴覚情報および調音運動情報がどの程度利用されているのかを検討するとともにそれらに対応する脳内メカニズムを明らかにすることを目的とした。聴覚情報に基づく有声性知覚処理の役割を検討した行動実験（研究1）では、日本語母語話者の有声性知覚に係る具体的な聴覚処理特性について検討を行うために、逆向性マスキングという聴覚現象が有声性の知覚に密接に関与する可能性について言及した先行研究（Repp, 1989）での仮説に注目し、有声性知覚との関連性を探る行動実験を行なった。逆向性マスキングとは後続する音が先行する音の聴こえを阻害する現象であり、阻害の程度は先行音と後続音が持つ音響的特徴の関係性に依存することが知られている。例えば、後続音に対して先行音が弱いパワー（エネルギー）を持つ場合、先行音と後続音が同じ周波数帯域にパワーを持つ場合、先行音と後続音の開始の時間差が短い場合などでは後続音によって先行音の聴こえが阻害されやすくなる（Elliott, 1971; Massaro, 1973）。具体的に、逆向性マスキングの特性がどのように有声性知覚に関与すると考えられているかを説明すると（図10）、まずVOT区間に存在する雑音成分（子音部）は後続する周期成分（母音部）と同じ周波数帯域にパワーを持つため逆向性のマスキングを受けるが、VOTが短い場合には先行する雑音成分と後続する周期成分の開始時間差が小さく、雑音成分のエネルギーも小さいため逆向性マスキングの程度が大きくなる。そのため、VOTが短く雑音成分の存在が知覚されにくい場合にはVOT区間は存在しないと見なされるため有声音が知覚される。一方、VOTが長い場合には先行する雑音成分と後続する周期成分の開始時間差が大きく、雑音成分のエネルギーも大きくなるため逆向性マスキングの程度が小さくなり、雑音成分の存在が明確に知覚されるため無声音が知覚されると考えられている。研究1ではこの可能性を直

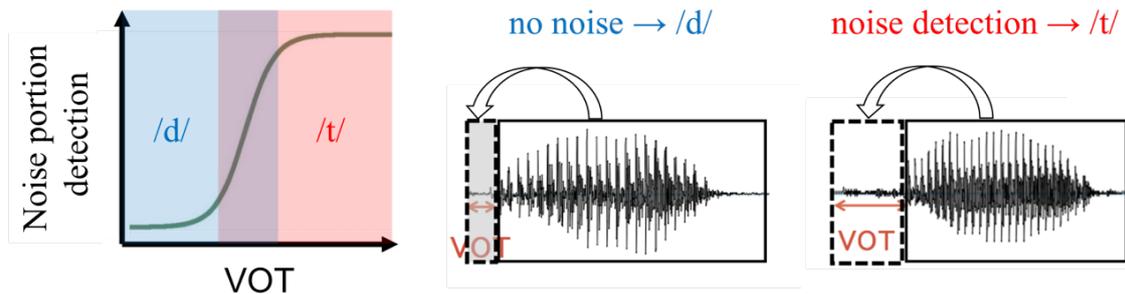


図 10. 逆向性マスクングに係る聴覚処理特性を利用した有声性知覚メカニズムの説明図.

接的に検討するために、VOT を操作した音声刺激を用いて、有声性知覚課題及び雑音成分の検出課題を行って両課題の成績を比較する実験を行った。

調音運動情報に基づく有声性知覚処理を調べるための行動実験（研究 2）では、Mitsuya et al. (2014) が行なったカテゴリー間聴覚フィードバック（Cross-categorical auditory feedback, 以下 CAF）実験を行って有声性の生成処理を変容させることでその知覚処理にどのような影響が生じるかを調べた。Mitsuya et al. の CAF 課題では、有声音の生成時にそれよりも VOT の長い無声音をフィードバックされる場合と無声音の発声時にそれよりも VOT の短い有声音をフィードバックされる場合の CAF 課題が行われ、前者の実験では有声音の VOT が短く、後者の実験では無声音の VOT が長くなり、TAF 課題と同様に補償応答が起こることが明らかにされた。ただし、TAF 課題では発話音声の音響的特徴がわずかに変形された音声でフィードバックされるのに対して、CAF 課題では自らが発声した音韻とは異なる音韻がフィードバックされるため状況が大きく異なる。Patri et al. (2018) によると、TAF 課題では発声した音韻（言語情報）と聴覚情報の対応関係の知識が更新されることで聴覚情報に基づく音声知覚処理に影響が生じ、また、発声時の調音運動情報と聴覚情報の対応関係の知識が更新されることで調音運動情報に基づく音声知覚に影響が生じると考えられている。しかしながら、CAF 課題では発声した音韻とは異なる音韻が、また、調音時に予測される聴覚情報とは明らか

に異なる聴覚情報がフィードバックされるため、発声した音韻（言語情報）や調音運動情報と聴覚情報の対応関係の知識が更新されることはないと考えられる。そのため、CAF 課題が有声性の生成や知覚に影響を与えるメカニズムは TAF 課題とは大きく異なると考えられる。Mitsuya et al.は、CAF 課題では発声と同時に呈示される聴覚刺激が自らが発したものではないことは容易に分かるが、その聴覚刺激が呈示されることで自らが発した音声のフィードバックを受け取ることができないことが原因となって有声性の生成に変化が生じた可能性があるとして述べている。現時点で CAF 課題が有声性の生成に影響を及ぼすメカニズムは断定できないが、CAF 課題が有声性の生成に影響を及ぼすということは有声性の生成処理過程における音韻情報と調音運動情報の対応関係を変化させることと同義だと考えられる（図 11）。そのため、その変化が音声知覚に与える影響を検討することで調音運動情報に基づく音声知覚処理の役割を直接的に検討できる可能性が考えられる。Patri et al.では、TAF 課題が音声生成・処理過程における音韻情報、調音運動、聴覚情報の対応関係に与える影響について、音韻情報と聴覚情報、調音運動と聴覚情報の対応関係のみに注目しているが、TAF 課題や CAF 課題によって音韻の生成に変化が生じていることを考慮すると、音韻情報と調音運動の対応関係も変化すると考えて音声知覚処理への影響を検討すべきだろう。

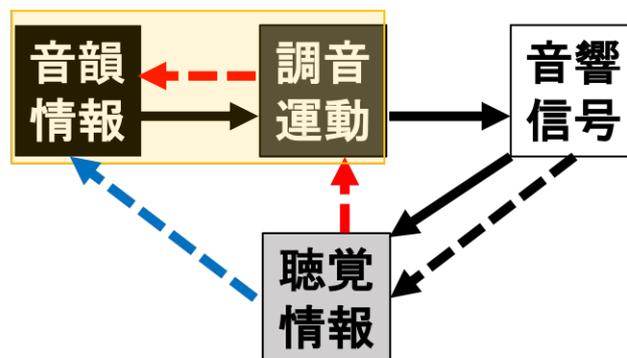


図 11. カテゴリー間フィードバック課題が音韻情報・調音運動・聴覚情報の対応関係に与える影響の予測。

有声性知覚における聴覚情報と調音運動情報の利用に係る脳内メカニズムを調べるための脳機能計測実験（研究3）では、VOTを操作して有声音/d/から無声音/t/へと知覚が変化する刺激を作成し、MEGを用いてそれらの刺激を聴取している際の脳活動を調べた。MEGは時間分解能と空間分解能の両方に優れた脳機能計測手法であるため、ミリ秒単位の時間情報（VOT）に対する脳活動を精度良く調べる、且つ、音声生成に関与する脳領域の活動を精度良く検出出来ると考えられる。VOTに対する聴覚応答については、英語母語話者の有声性知覚に密接に関与すると考えられている、子音部と母音部の始まりに対する神経応答が分離するか融合するかという応答パターンに注目するのではなく、ウェーブレット変換を用いた時間周波数解析によって聴覚皮質の神経振動を詳細に分析することで日本語母語話者のVOTに対する聴覚応答パターンが有声性の知覚と相関するかどうかについて検討を行なった。有声性知覚時の調音運動情報の利用に係る脳内メカニズムについては、先行研究でよく調べられている運動野や運動前野だけでなくその他の脳領域、特に調音の時間制御の役割を担い、有声性の生成に密接に関与すると考えられる脳領域の活動についても有声性知覚との間に関連が見られるかどうかを調べる。そこで、本研究ではMEG計測実験の前段階として、fMRIを用いた脳機能計測実験で有声性知覚時に有声性生成時と共通して活動が見られる脳領域を調べることで、調音の時間制御に係る脳領域が有声性の生成時だけでなく知覚時にも生じるかどうかを調べる。そして、fMRI計測実験で有声性生成と知覚時に共通して活動が見られた脳領域について、MEG計測実験で機能的結合や神経振動の時間変化を調べて有声性知覚との関連性を詳細に検討する。



## 第3章 研究1：日本語母語話者の有声性知覚に係る聴覚処理特性の検討

### 3.1 目的

日本語母語話者の有声性知覚に係る聴覚処理特性を具体的に明らかにするために、VOT を操作した刺激を用いて有声音/d/と無声音/t/の知覚課題（有声性知覚課題）を行うとともに2種類の非音声課題を行ってそれらの共通点を調べた。1つ目の非音声課題では、Repp（1979）が提案した逆向性マスクングの聴覚処理特性が有声性知覚に利用されている可能性を検討するために、周期成分（母音部）から逆向性のマスクングを受ける雑音成分（子音部）の存在に気づくか否かを判断する逆向性マスクング課題を行った。ここでは、雑音成分の知覚が有声性の知覚と同様にVOTの増加に伴って非線形的に変化するかどうか、また、雑音成分の知覚の有無が切り替わるVOT境界が有声音/d/と無声音/t/のVOT境界と一致するかどうかを調べることで逆向性マスクングに係る聴覚処理特性と有声性知覚の処理に共通性が見られるかを検討した。2つ目の非音声課題では、英語母語話者の有声性知覚に密接に関与するとされている2つの音イベントの同時性知覚処理が日本語母語話者の有声性知覚にも密接に関与している可能性について検討をするために、閉鎖子音における雑音成分（子音部）の開始が周期成分（母音部）の開始と同時であったか、雑音成分が周期成分よりも先に始まったかを判断する同時性知覚課題を行った。そして、同時性の知覚が有声性知覚と同様にVOTの増加に伴って非線形的に変化するかどうか、また、同時性の知覚が切り替わるVOT境界が有声音/d/と無声音/t/のVOT境界と一致するかどうかを調べることで2音の同時性知覚に係る聴覚特性と有声性知覚の共通性を検討した。また、有声性の知覚は主にVOTを手がかりに行われることが知られているが、第1フォルマント周波数（F1）や

雑音成分の音圧といった音響的特徴もその副次的な手がかりとなり、F1 が低い場合に比べて高い方があるいは雑音成分の音圧が小さい場合に比べて大きい方がより短い VOT 長で無声音が知覚されやすい（VOT 境界が短くなる）ことが知られている

（Lisker, 1975, Lisker et al., 1977; Repp, 1979）。そこで本研究では、これらの音響手がかりが有声性知覚の VOT 境界だけでなく雑音成分と周期成分の同時性の知覚が切り替わる VOT 境界や雑音成分の知覚の有無が切り替わる VOT 境界にも影響を及ぼすかどうかについても調べることで有声性知覚と 2 つの非音声課題の間の処理の共通性について更なる検討を行なった。

## 3.2 実験

### 3.2.1 実験参加者

実験参加者は日本語母語話者 12 名（女性 4 名、男性 8 名）で平均年齢は 22.5 歳であった。どの参加者も聴覚障害を有しておらず、本実験で用いた刺激の聴取に問題はなかった。この実験は、九州大学システム情報科学府の実験倫理審査で承認を受けた上で行われ、参加者は実験前に実験の説明を受けた上で実験参加の同意書に署名を行った。

### 3.2.2 実験装置および刺激

実験は防音室内で行われ、刺激の呈示には、パーソナルコンピュータ（DELL, Latitude E5530）、聴覚実験システム（Tucker-Davis Technologies, TDT System 3）、ヘッドホンアンプ（STAX, SRM-006tS）、ヘッドフォン（STAX, SR-303）を使用した。参加者の回答の収集には、聴覚実験システムに接続された反応ボックス（Tucker-Davis Technologies, RBOX）を用いた。刺激の呈示音圧の調整には、騒音計（Brüel & Kjær,

2260) , 人工耳 (Brüel & Kjær, 4153) , マイクロフォン (Brüel & Kjær, 4192) を用いた。

VOT を操作して /da/ から /ta/ へと知覚が変化する合成音声刺激を, F1 あるいは雑音成分の音圧が異なる Standard, High-F1, High-amplitude の 3 つの刺激条件毎に作成した

(図 12) 。合成音声刺激は, 帯気雑音から成る雑音成分と 100 Hz の基本周波数を持つ周期成分 (/a/) から構成され, 雑音成分の音圧レベルは Standard 条件と High-F1 条件では 45 dBA, High-amplitude 条件では 60 dBA に設定した (図 12, 橙枠) 。周期成分の音圧レベルは 85 dBA であった。フォルマントは雑音成分の開始から生起させ, F1 周波数については Standard 条件と High-amplitude 条件では 600 Hz, High-F1 条件では 800 Hz

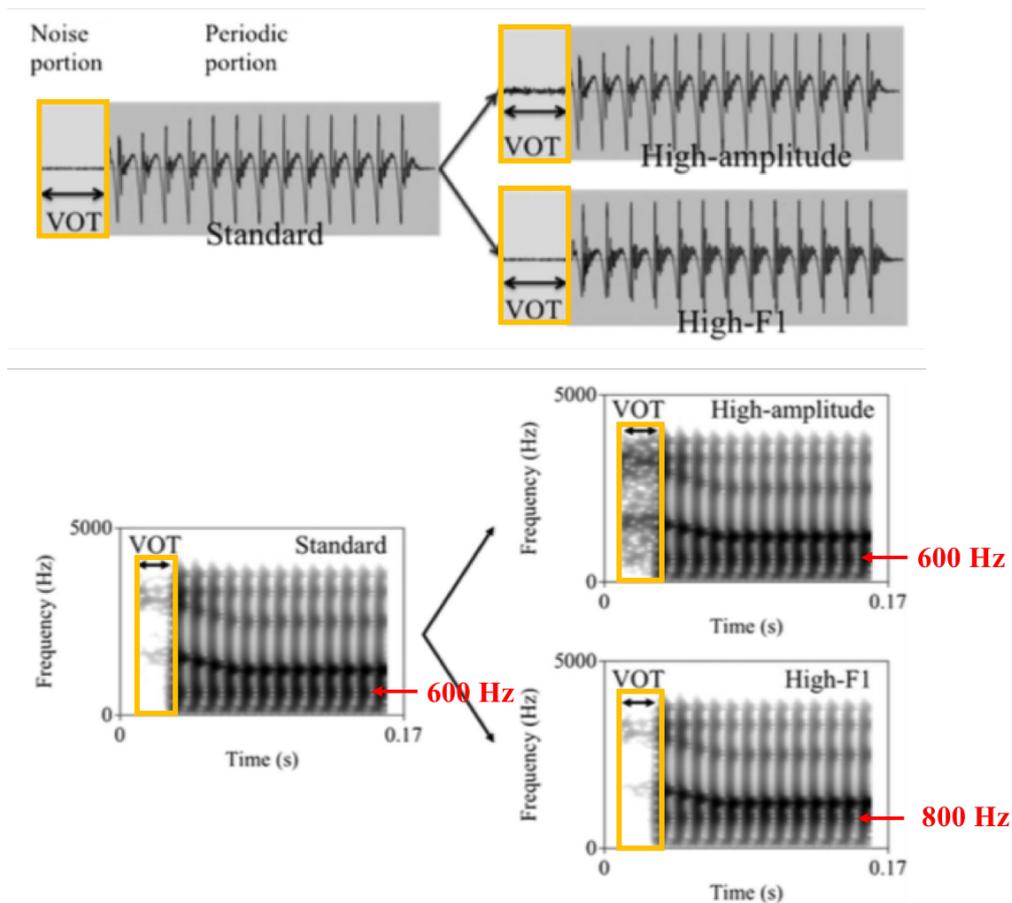


図 12. Standard, High-amplitude, High-F1 条件の刺激の波形とスペクトログラム (Tamura et al., 2018, p.791, Fig.1 及び p.792, Fig.2 を改変) .

に設定した（図 12, 赤矢印）。F2, F3 周波数については子音の調音位置素性を決める重要な音響的特徴であるため、歯茎音である/da/や/ta/が知覚される刺激を作成するために予備実験の結果に基づいて設定を行った。具体的には、F2, F3 周波数は雑音成分ではそれぞれ 1600 Hz, 3000 Hz と一定で、周期成分の開始から 40 ms の間にそれぞれ 1600 Hz から 1200 Hz, 3000 Hz から 2500 Hz へ一定の割合で変化させ、その後は一定に設定した。通常、F2, F3 の遷移は雑音成分の開始時から始まるが、本実験では雑音成分と周期成分のフォルマント特徴が VOT に依存して変化しないようにするために上記のような設定にした。F4, F5 は F1 と同様、雑音成分の開始から一定でそれぞれ 3300 Hz と 3750 Hz に設定した。VOT は雑音成分の長さを変えることで操作し、有声性知覚課題及び逆向性マスクング課題では VOT を 5 ms 刻みで 0 ms から 30 ms まで変化させた 7 つの刺激を用いた。同時性知覚課題での VOT の設定は、逆向性マスクング課題の結果に基づいて決定した。具体的には、雑音成分が知覚されるために必要な VOT 長（50 % の確率で知覚される VOT の長さ）の小数点以下を切り上げた値を求め、その長さから 30 ms 長い VOT まで 5 ms 刻みで VOT を操作した 7 つの刺激を用いた。

### 3.2.3 実験手続き

VOT を操作した刺激を用いて、まず始めに 2 つの非音声課題を行い、その後には有声性知覚課題を行った。有声性知覚課題を行う前に非音声課題を行った理由は、特に逆向性マスクング課題において呈示された刺激が有声音に聴こえるか無声音に聴こえるかの判断が非音声課題の成績に影響を及ぼす可能性を考慮したためである。同時性知覚課題では、逆向性マスクング課題の成績をもとにして VOT のパラメータを決める必要があるため、非音声課題は逆向性マスクング課題から行った。また、各課題における 3 つの刺激条件（Standard, High-F1, High-amplitude）の実施順序については実験参加者内でカウンターバランスを取った。

逆向性マスクング課題では、課題前に雑音成分を持たない VOT 0 ms の刺激と VOT30 ms に設定したもののうち雑音成分のエネルギーが最も大きい刺激を実験参加者に聴いてもらい、前者では雑音成分が知覚出来ないことと後者では明確に雑音成分が知覚出来ることを確かめた。VOT 30 ms の刺激については、それを構成する雑音成分と周期成分を分離した刺激も聴いてもらった。課題では、VOT の異なる 7 つの刺激をランダムな順番で 25 回ずつ呈示し、実験参加者には呈示された刺激に雑音成分が含まれているか否かを答えてもらった。同時性知覚課題では、VOT が最も短い刺激と長い刺激を聞いてもらい、前者では雑音成分が周期成分と同時に始まっていると知覚され、後者では雑音成分が周期成分よりも先に始まっていると知覚出来ることを確かめた。課題では、VOT の異なる 7 つの刺激をランダムな順番で 25 回ずつ呈示し、実験参加者には呈示された刺激において雑音成分の始まりが周期成分の始まりと比較して、同時であったか速かったかを答えてもらった。有声性知覚課題でも同様に、VOT 0 ms の刺激が明確に /da/、VOT 30 ms の刺激が明確に /ta/ と知覚されることを確かめた上で、VOT の異なる 7 つの刺激をランダムな順番で 25 回ずつ呈示し、実験参加者には呈示された刺激が /da/ に聴こえたか /ta/ に聴こえたかを答えてもらった。

### 3.2.4 データ分析

各課題各刺激条件で VOT の異なる 7 つの刺激に対する回答データを参加者ごとにまとめた。有声性知覚では無声音 /t/ が知覚された回答率、同時性知覚課題では雑音成分が周期成分よりも先に始まったと答えた回答率、逆向性マスクングでは雑音成分が知覚される回答率をまとめた。各課題各刺激条件でまとめた参加者ごとの回答率データに以下に示すロジスティック曲線（式 1）を当てはめ（図 13）、VOT 境界（回答率が 50% になる VOT 長）と曲線の勾配を求めた。

$$P(\text{VOT}) = \frac{1}{1 + e^{-\beta(\text{VOT} - \alpha)}} \quad (1)$$

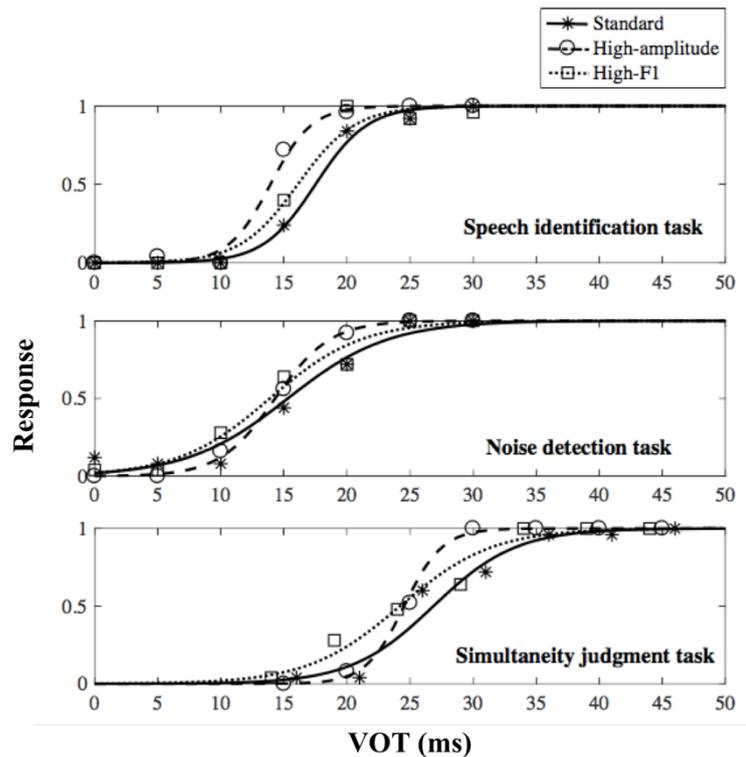


図 13.1 名の参加者の各課題各刺激条件の回答率とロジスティック補完曲線 (Tamura et al., 2018, p.793, Fig.3) .

式 1 における  $P(\text{VOT})$  は刺激の VOT に対する回答率の予測値を表し、 $\alpha$  と  $\beta$  のパラメータがそれぞれ VOT 境界と曲線の勾配を表す。曲線の勾配について、1 名の参加者の  $\beta$  が他と比べて異常に大きな値であったため（参加者の全体平均 + 3 × 標準偏差よりも大きかった）、外れ値と見なしてデータを分析から除外した。

### 3.2.5 結果

各課題各刺激条件で求めた VOT 境界について参加者全体での平均を求めたところ (図 14) , 雑音成分の知覚の有無が切り替わる VOT 境界が /d/ と /t/ の VOT 境界と近い値を取ることが分かった。一方で、雑音成分と周期成分の同時性の知覚が切り替わる VOT 境界はどの刺激条件でも /d/ と /t/ の VOT 境界より 10 ms 以上長くなった。さらに、刺激条件による VOT 境界の違いはどの課題でも同じ傾向が見られ、Standard 条件に比

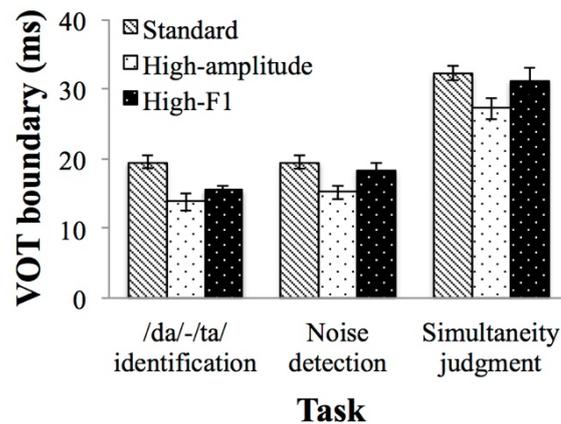


図 14. 各課題各刺激条件における VOT 境界の参加者平均。エラーバーは標準誤差を示す (Tamura et al., 2018, p.794, Fig.4) .

べて High-amplitude 条件で VOT 境界が短くなった。High-F1 条件でも Standard 条件に比べて VOT 境界が短くなる傾向は見られたが、High-amplitude 条件と Standard 条件の差に比べると明瞭ではなかった。曲線の勾配についても VOT 境界と同様に各課題各刺激条件で全参加者の平均を求めたところ (図 15) , 有声性知覚課題や逆向性マスクング課題で得られた曲線の傾きが同時性知覚課題と比べてやや急峻な傾向は見られたが、どの課題でもあまり大きな差は見られなかった。また、刺激条件による曲線の勾配の違いについては、どの課題でも似た傾向が見られ、High-amplitude 条件では Standard 条件や High-F1 条件に比べて勾配が鋭くなった。High-F1 条件と Standard 条件の勾配は同程度になることが分かった。

VOT 境界や曲線の勾配について、課題 (3 水準: 有声性知覚課題, 雑音成分の検出課題, 同時性知覚課題) と刺激条件 (3 水準: standard, High-F1, High-amplitude) の 2 要因での反復測定分散分析を行った。VOT 境界については、課題と刺激条件の主効果が有意であったが (課題:  $F(2, 20) = 93.78, p < .01$ , 刺激条件:  $F(2, 20) = 8.48, p < .01$ ) , 課題と刺激条件の交互作用は有意ではなかった ( $F(4, 40) = 0.74, p = 0.57$ ) 。曲線の勾配については、刺激条件の主効果は有意であったが ( $F(2, 18) = 11.62, p < .01$ ) , 課題の主効果や課題と刺激条件の交互作用は有意ではなかった (課題:  $F(2, 18) = 2.67, p = 0.10$ ,

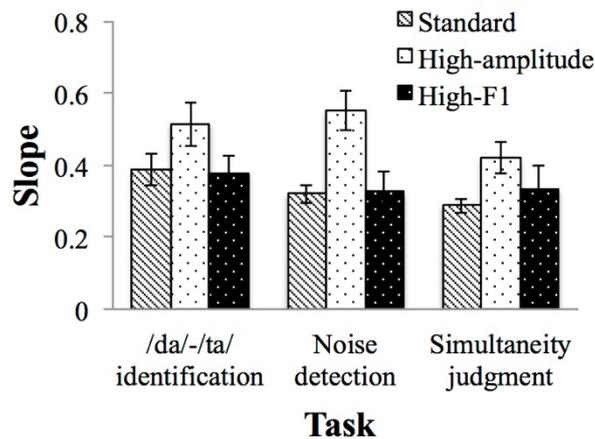


図 15. 各課題各刺激条件における曲線の傾きの参加者平均。エラーバーは標準誤差を示す (Tamura et al., 2018, p.794, Fig.5) .

交互作用:  $F(4, 36) = 1.56, p = 0.21$  )。続いて, VOT 境界について課題の 3 水準間と刺激条件の 3 水準間で多重比較検定 (Bonferroni 法) を行ったところ, 課題の要因では雑音成分の知覚の有無が切り替わる VOT 境界と /d/ と /t/ の VOT 境界の間で有意な差が見られないのに対して, 雑音成分と周期成分の同時性の知覚が切り替わる VOT 境界はその他の 2 つの課題で得られた VOT 境界よりも有意に長くなることが分かった。刺激条件の要因では Standard 条件と High-F1 条件の VOT 境界に比べて High-amplitude 条件の VOT 境界が有意に短くなることが分かった。曲線の勾配については刺激条件の 3 水準間で多重比較検定を行ったところ, Standard 条件や High-F1 条件よりも High-amplitude 条件の方が曲線の勾配が有意に鋭くなることが分かった。Standard 条件と High-F1 条件については有意な差は見られなかった。

### 3.3 考察

研究 1 では, 日本語母語話者の有声性知覚に利用される聴覚処理特性を調べるために, 同時性知覚や逆向性マスクングの聴覚特性と有声性知覚のカテゴリー知覚の関連性を調べた。VOT を操作した刺激を用いて, 2 音 (雑音成分と周期成分) の同時性知覚

課題と雑音成分の知覚課題を行ったところ、どの刺激条件においても有声性知覚と同様に VOT の増加に伴う非線形的な知覚の変化が見られた。このことは、これらの聴覚処理特性が有声性のカテゴリ知覚を生み出すために有用であることを示唆している。VOT 境界については、雑音成分と周期成分の同時性知覚が切り替わる VOT 境界が有声性の知覚が切り替わる VOT 境界よりも長くなるのに比べて、雑音成分の知覚の有無が切り替わる VOT 境界がどの刺激条件でも有声性知覚の VOT 境界と非常に近い値を取ることが分かった。この結果は、日本語母語話者は、逆向性マスクングに係る聴覚処理の非線形特性を有声性知覚に用いており、雑音成分が知覚されるか否かが有声音と無声音を知覚する上での不変的な特徴になり得ることが示唆される。

英語母語話者を対象とした実験では、2つの音イベントの同時性知覚に係る聴覚処理特性が有声性知覚に利用され、雑音成分と周期成分が同時に始まったと知覚されるか否かが有声音と無声音の知覚を決定する不変的な特徴だとされているが (Pisoni, 1977; Simos et al., 1998 a, b; Steinschneider et al., 2004) , 本研究の結果から日本語母語話者はこの特徴を有声性知覚に用いていないと考えられる。また, Hay and Garcia-Sierra (2005) の研究でも、同時性知覚に係る聴覚処理特性がスペイン語母語話者の有声性知覚には用いられていないことが示唆されている。本研究や Hay and Garcia-Sierra の研究結果を考慮すると, Kuhl and Miller (1978) や Elangovan and Stuart (2008) で述べられているように有声性知覚に利用される聴覚処理特性とそれによって生み出される知覚の不連続点は言語によって異なるのであろう。ただし、日本語母語話者を対象とした本研究で発見した逆向性マスクングに係る聴覚情報処理特性と有声性知覚の関係性については他の言語圏の話者を対象とした研究では検討されていない。そのため、英語母語話者やスペイン語話者も含め複数あるいは全ての言語圏の有声性カテゴリ知覚を統一して説明できる可能性について今後検討を進めるべきであろう。スペイン語母語話者の VOT 境界が日本語母語話者と近い値になることを考慮すると、スペイン

語母語話者は日本語母語話者と同じように逆向性マスクングの聴覚処理特性を有声性のカテゴリー知覚に利用している可能性が考えられる。

研究1では、音声刺激を用いて逆向性マスクング課題や同時性知覚課題を行うことで日本語母語話者の有声性知覚に利用される聴覚処理特性を直接的に検討した。その結果、英語母語話者を対象とした研究以外で初めて聴覚情報に基づく音声知覚処理の役割を重要視する仮説（Diehl et al., 2007; Holt et al., 2004; Kuhl & Miller et al., 1978; Stevens, 1989）を支持する結果が得られた。ただし、通常非音声刺激を用いて行われる逆向性マスクング課題や同時性知覚課題に音声刺激を用いたため、これらの課題の結果に有声音に聞こえるか無声音に聞こえるかという音声としての知覚結果が影響を及ぼした可能性も考えられる。そのため、今後は有声音や無声音を模擬した非音声刺激を作成して逆向性マスクング課題を行った場合にも同様の結果が得られるかどうかを検討する実験を行う必要がある。



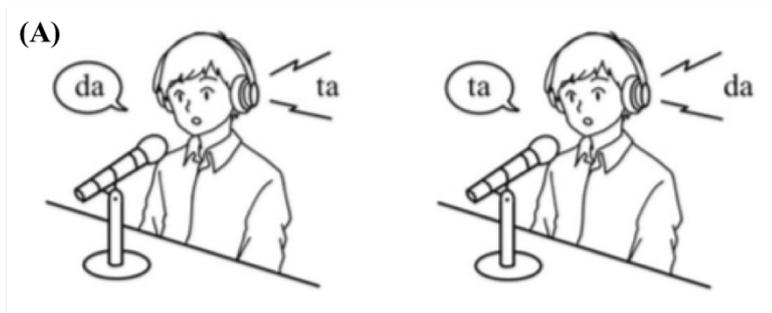
## 第4章 研究2：有声性の生成処理の変化が知覚処理に

### 与える影響の検討

#### 4.1 目的

調音の時間的性質（VOT）の違いに特徴付けられる有声性の音韻対における音声生成・知覚の相互作用を検討するために、Mitsuya et al. (2014) の研究で用いられた CAF 課題を用いて有声性の生成を変化させ、その前後で有声性のカテゴリー知覚にどのような影響が生じるかを調べた。本研究の CAF 課題では、実験参加者に /da/ と /ta/ の発声をランダムな順番で行なってもらい、/da/ の発声時には事前に録音しておいた /ta/ の音声を、/ta/ の発声時には事前に録音しておいた /da/ の音声を実時間でフィードバックする課題（図 16A）を行った。

実験では、有声性の生成処理の変化が知覚処理に与える影響を調べる聴覚フィードバック（auditory feedback, 以下 AF）条件とその対照条件として聴覚順応（auditory adaptation, 以下 AA）条件を行なった（図 16B）。AF 条件では、まず始めに聴覚刺激は与えず通常通り発声を行う課題（Non-perturbed auditory feedback 課題, 以下 NP 課題）を行なった。続いて、VOT を操作した刺激を用いて有声性知覚課題を、CAF 課題の間に挟んで 2 回行った。有声性の生成の変化については NP 課題と CAF 課題の発声を比較することで、知覚の変化については CAF 課題の前後で行う知覚課題の成績を比較することで調べ、両者の関連性についても検討を行った。AA 条件では、ランダムな順番で呈示される /da/ と /ta/ の音声を受動的に聴取してもらう課題（Passive listening 課題, 以下 PL 課題）を行ってもらい、その前後で有声性の知覚課題を行った。AA 条件を対照条件として設けた理由は、CAF 課題で特定の聴覚刺激を繰り返し聴取すること自体が知覚を変化させる原因になっていないことを確かめるためである。先行研究にて、



(B)

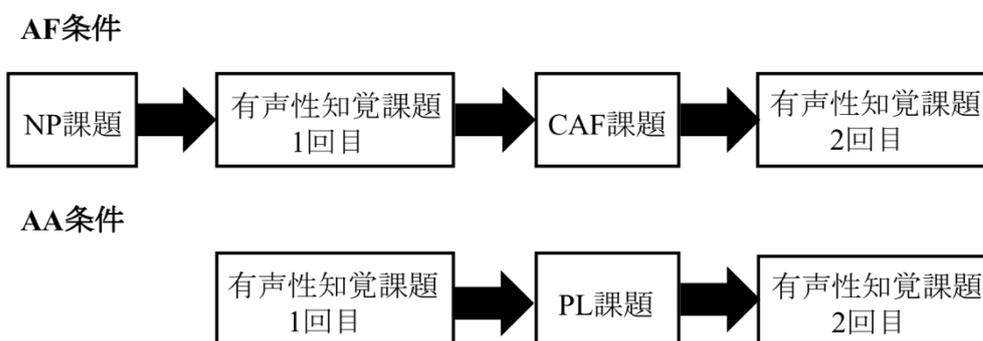


図 16. (A) カテゴリー間フィードバック (CAF) 課題の説明図. (B)聴覚フィードバック (AF) 条件と聴覚順応 (AA) 条件の実験の流れ (Tamura et al., 2019, p.2199, Fig.1 を改変) .

特定の音韻を聴き続けることでその音韻が知覚されにくくなる現象が知られており、例えば、有声音や無声音を聴き続けた後に、VOT を段階的に操作した刺激を用いた有声性知覚課題を行うと、有声音を聴き続けた場合には有声音の知覚率が落ちて VOT 境界が短くなる、無声音を聴き続けた場合には無声音の知覚率が低下して VOT 境界が長くなることが知られている (Eimas & Corbit, 1973; Miller et al., 1983; Samuel, 1982) 。ただし、本研究の CAF 課題では発声させる音韻を 1 つに限定せずランダムな順番で有声音と無声音の発声を行わせることで、ランダムな順番で有声音と無声音を聴く状況を作るため、CAF 課題で聴覚刺激を聴くこと自体が知覚に及ぼす影響は少ないと考えられる。

## 4.2 実験

### 4.2.1 実験参加者

実験参加者は日本語母語話者 17 名（女性 6 名，男性 11 名）で平均年齢は 21.6 歳であった。どの参加者も正常聴力で，本実験で用いた刺激の聴取に問題はなかった。この実験は，九州大学システム情報科学府の実験倫理審査で承認を受けた上で行われ，参加者は実験前に実験の説明を受けた上で実験参加の同意書に署名を行った。実験参加者は AF 条件及び AA 条件の両方に参加した。条件の実施順はランダムで，2 つの条件の間には 1 週間以上の期間を設けた。

### 4.2.2 実験装置および刺激

実験は防音室内で行われた。CAF 課題及び NP 課題での発声指示は，液晶ディスプレイ（EIZO, FlexScan S2000）を用いて行われ（PL 課題でも発声はしないが文字は呈示された），発話音声はマイクロホン（AKG, D7S），USB オーディオインターフェイス（EDIROL, UA-3D）を介してパーソナルコンピュータ（DELL, Latitude E5530）で記録された。CAF 課題での聴覚フィードバック刺激は，録音音声は USB オーディオインターフェイスを介してパーソナルコンピュータ（DELL, Latitude）に入力され，発話の開始が検出されたと同時に，聴覚実験システム（Tucker-Davis Technologies, TDT System 3），ヘッドホンアンプ（STAX, SRM-006tS），ヘッドフォン（STAX, SR-303）を用いて呈示された。また，発声時の骨伝導によるフィードバックの影響を軽減するために聴覚フィードバック刺激と同時に白色雑音の呈示も行った。聴覚フィードバックの実験制御には MATLAB の Data Acquisition toolbox を用い，発話の開始の検出に対する聴覚刺激呈示の遅延は 10 ms 以下であった。PL 課題や有声性知覚課題の刺激の呈示には，CAF 課題で聴覚刺激を呈示するために用いた機器と同じものを使用した。有声性知覚課題での実験参加者の回答収集には，聴覚実験システムシステムに接続された反応ボ

ックス (Tucker-Davis Technologies, RBOX) を用いた。刺激の呈示音圧の調整には、騒音計 (Brüel & Kjær, 2260) , 人工耳 (Brüel & Kjær, 4153) , マイクロフォン (Brüel & Kjær, 4192) を用いた。

CAF 課題及び PL 課題で用いたフィードバック音声刺激は参加者ごとに 1 回目の実験参加 (AF 条件あるいは AA 条件) に先立って、録音された。録音では、実験参加者に /da/ と /ta/ を 30 回ずつランダムな順番で発声してもらった。/da/ と /ta/ それぞれで 30 回の発声の VOT を求め、その平均に最も近い VOT 長を持つ刺激を選定し CAF 課題や PL 課題で用いた。有声性知覚課題では、VOT を 3 ms 刻みで 3~27 ms まで操作して作成して作成した合成音声刺激を用いた。VOT 長以外の設定は研究 1 の Standard 条件の刺激設定と同じである。CAF 課題での聴覚フィードバック刺激、有声性知覚課題の合成音声刺激は 85 dBA で前者については両耳に、後者については右耳に呈示した。聴覚フィードバック課題における雑音の呈示音圧は 50 dB に設定した。

#### 4.2.3 実験手続き

AF 条件と AA 条件で行う課題の順序は、図 15 に示した通りである。AF 条件における CAF, NP 課題では、実験参加者に /da/ と /ta/ をランダムな順番で 100 回ずつ、合計 200 回発声してもらった。CAF 課題では、/da/ の発声時には /ta/ の録音音声を、/ta/ の発声時には /da/ の録音音声を発声と同時に呈示した。実験前に、実験参加者に自分が発声した声が聴こえないかどうかを確認した。NP 課題では、録音音声の呈示は行わず自然に /da/ と /ta/ の発声を行なってもらった。有声性知覚課題では、9 つの刺激をランダムな順番で 10 回ずつ呈示し、実験参加者は呈示された刺激が /da/ と聴こえたか、/ta/ と聴こえたかを回答した。AA 条件の PL 課題では、/da/ と /ta/ の録音音声をランダムな順番で 100 回ずつ呈示した。この際、発声は行わなかったが、CAF 課題と同様に聴覚呈示される音韻と

は異なる文字をディスプレイで視覚呈示した。AA 条件での有声性知覚課題は AF 条件と全く同じ方法で行った。

#### 4.2.4 データ分析

有声性の生成の変化を調べるために、CAF 課題での発声を NP 課題での発声と比較した。参加者ごとに CAF 課題と NP 課題それぞれで /da/ と /ta/ の 100 回の発声の VOT を調べた。/da/ の発声では、声門の閉鎖解放よりも声帯振動が先に始まり VOT が負の値を持つ場合がある。そこで、まず正の VOT と負の VOT を持つ /da/ の発声率を求めた。負の VOT についてはその長さを精確に測ることが難しいため、Mitsuya et al. (2014) と同様、正の VOT を持つ /da/ と /ta/ の発声のみ詳細な分析を行い、CAF 課題と NP 課題それぞれで /da/ と /ta/ の VOT の平均値を求めた。

有声性知覚課題のデータについては、VOT の異なる 9 つの刺激に対する回答データを参加者ごとにまとめ、3.2.4 と同様の方法を用いてロジスティック関数を当てはめることで VOT 境界と曲線の勾配を求めた。AF 条件の CAF 課題の前後に行った知覚課題の成績を比較することで有声性の生成処理の変化がその知覚処理に与える影響を検討した。AA 条件でも同様に PL 課題前後の成績を比較した。

#### 4.2.5 結果

##### 生成の変化

正の VOT を持つ /da/ の発声率を全参加者で平均したところ、CAF 課題では約 70%、NP 課題では約 80%であった（課題間で有意差は見られなかった）。続いて、/da/ 及び /ta/ の正の VOT の平均値を全参加者で平均したところ、/da/ の発声に関しては CAF 課題と NP 課題で違いは見られなかったものの、/ta/ に関しては NP 課題よりも CAF 課題で VOT が長くなることが分かった（図 17）。また、/da/ と /ta/ の VOT 差について、NP 課題と CAF

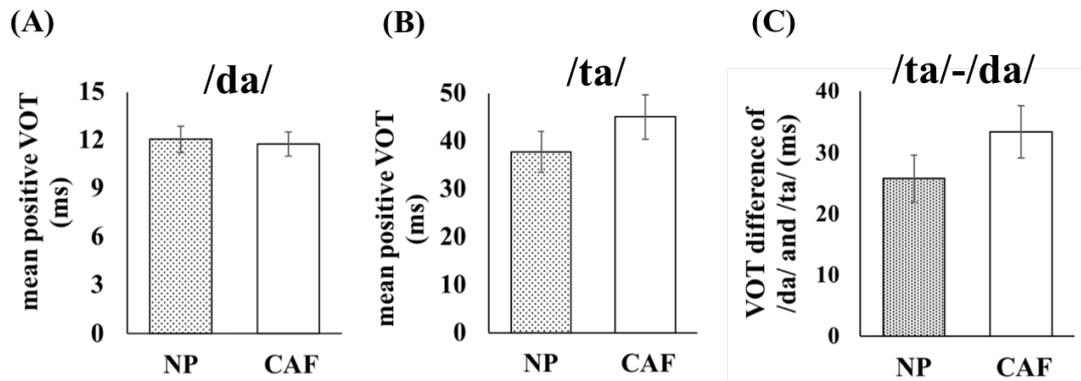


図 17. (A) NP 課題及び CAF 課題での /da/ の発声における正の VOT の参加者平均. (B) /ta/ の発声における正の VOT の参加者平均. (C) /da/ と /ta/ の VOT 差の参加者平均. エラーバーは標準誤差を示す. (Tamura et al., 2019, p.2201, Fig.2)

課題で比較したところ, NP 課題よりも CAF 課題で /da/ と /ta/ の VOT 差が大きくなることが分かった。/da/ と /ta/ の VOT, /da/ と /ta/ の VOT 差のそれぞれについて, 課題間で  $t$  検定を行なったところ, /da/ の VOT については, 課題間で有意差は見られなかった ( $t(16) = 0.90, p = 0.38$ )。/ta/ の VOT と /da/ と /ta/ の VOT 差については, 課題間で有意差が見られた (/ta/ の VOT:  $t(16) = 3.49, p < 0.01$ , /da/ と /ta/ の VOT 差:  $t(16) = 3.45, p < 0.01$ )。

### 知覚の変化

CAF 課題 (AF 条件) と PL 課題 (AA 条件) の前後に行なった有声性知覚課題ごとに求めた VOT 境界と曲線の傾きについて, 参加者全体で平均値を求めるとともにその平均値からロジスティック曲線を作成したところ (図 18), CAF 課題前に比べて後の方が曲線の勾配が急峻になること, PL 課題前に比べて PL 課題後に VOT 境界が短くなる傾向があることが分かった。VOT 境界や曲線の勾配について, 課題 (CAF 課題と PL 課題の 2 水準) とセッション (課題前後の 2 水準) の 2 要因での反復測定分散分析を行った。VOT 境界については, 課題とセッションの主効果ともに有意ではなく (課題:  $F(1, 15) = 0.01, p = 0.97$ , セッション:  $F(1, 15) = 2.83, p = 0.11$ ), 課題とセッションの

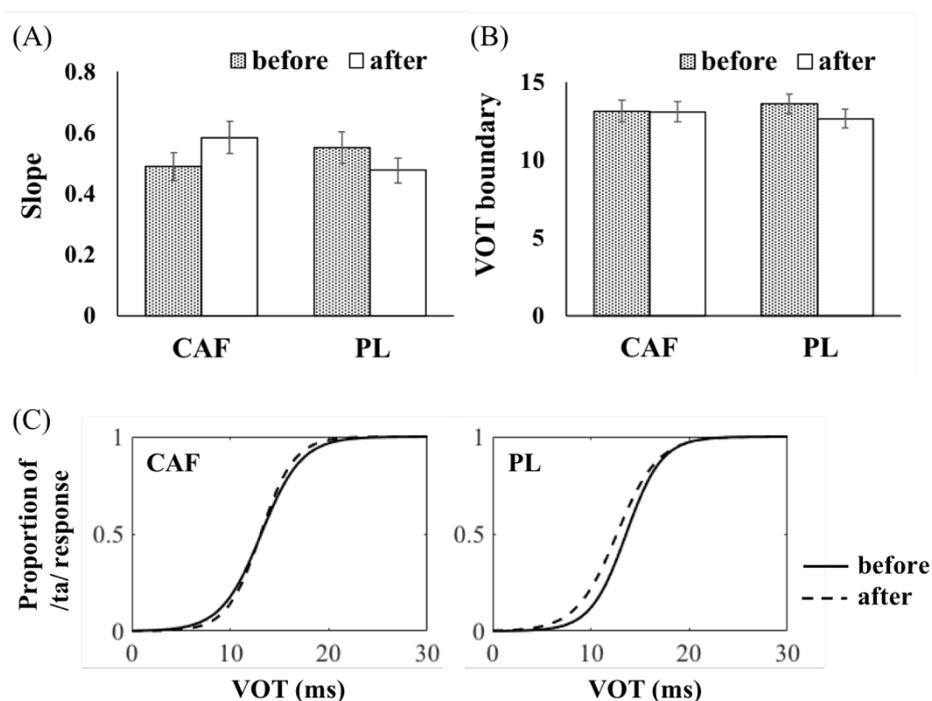


図 18. (A) CAF 課題及び PL 課題前後での有声性知覚課題における曲線の勾配の参加者平均. (B) VOT 境界の参加者平均. (C) 曲線の勾配, VOT 境界の参加者平均をパラメータとして作成したロジスティック曲線. (Tamura et al., 2019, p.2202, Fig.3 を改変)

交互作用についても有意ではなかった ( $F(1, 15) = 3.06, p = 0.14$ )。CAF 課題でも PL 課題両方とも VOT 境界には影響を与えないという結果が得られた。曲線の勾配については、課題とセッションの主効果ともに有意ではなかったが (課題:  $F(1, 15) = 0.26, p = 0.61$ , セッション:  $F(2, 18) = 0.11, p = 0.74$ ) , 課題と時点の交互作用は有意だった ( $F(1, 15) = 5.14, p = 0.04$ )。続いて、単純主効果検定を行ったところ, CAF 課題前に比べて後の方が曲線の傾きが鋭くなることが分かった ( $F(1, 15) = 5.30, p = 0.04$ )。一方で, PL 課題では課題前後で有意差は見られなかった ( $F(1, 15) = 1.90, p = 0.19$ )。続いて, CAF 課題での知覚の変化と生成の変化の関係性を調べるために, CAF 課題前後での曲線の傾きの変化量と NP 課題に対する CAF 課題の間の相関分析 (スピアマンの順位相関係数) を行ったところ, 両者の間には有意な正の相関が見られた ( $r(14) = 0.64, p < 0.01$ ) (図 19)。

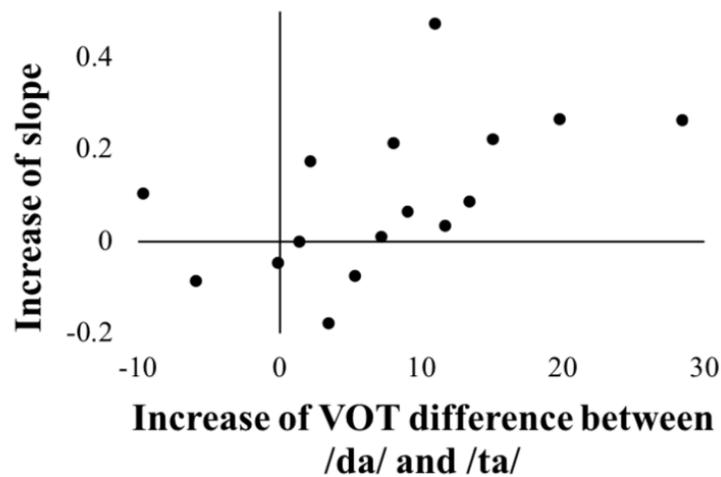


図 19. /da/と/ta/の発声における VOT 差の変化と知覚曲線の勾配の変化の散布図 (Tamura et al., 2019, p.2202, Fig.4) .

### 4.3 考察

有声性の知覚と生成の間の相互作用を検討するために、CAF 課題を用いて有声性の生成を変化させ、その前後で知覚にどのような変化が見られるかを調べたところ、有声音と無声音の VOT の違いが大きく明確になるのに伴って有声性知覚のカテゴリー知覚判断が明確になることが分かった。さらに、生成における有声音と無声音の VOT 差の変化とカテゴリー知覚の明確性の変化の間に有意な正の相関が見られた。一方で、CAF 課題で聴取した音声刺激を聴取するだけの PL 課題の前後ではカテゴリー知覚の明確性に変化が見られないことが分かった。上記の結果は、有声音と無声音の VOT 差とカテゴリー知覚の明確性の間に有意な正の相関があることを報告した先行研究 (Bailey & Haggard, 1981) の結果を補強するものであり、調音運動情報を用いた音声知覚処理が音韻対のカテゴリー知覚の明確性に関与することを示唆するものである。

CAF 課題が有声性知覚を変化させた理由について詳細に議論するために、まず CAF 課題が音声生成・処理過程における音韻情報、調音情報、聴覚情報の対応付けに

どのような変化をもたらしたのかを議論する。そして、それらの変化が有声性の知覚にどのように影響したのかを, Patri et al. (2018) の TAF 課題に関する議論と比較しながら考察する。Patri et al.は, TAF 実験が音韻情報と聴覚情報, 調音情報と聴覚情報の対応関係を変化させることで聴覚情報と調音運動情報に基づく音声知覚処理の両方を変化させたと述べている (図 6)。Lammeti et al. (2014) の TAF 課題を用いた実験で生成の変化と知覚の変化の間に相関が見られなかったことについても, TAF 課題が 2 つの処理両方に影響を与えたことが原因だと考えられている。一方で, CAF 課題ではフィードバックされる聴覚情報が調音情報から予測される聴覚情報と大きく異なるため, それらの対応付けに変化は生じないと考えられる (Mitsuya et al., 2014)。さらに, CAF 課題では発声した音韻とは異なる音韻がフィードバックされるため音韻情報と聴覚情報の対応関係についても変化は起きないと考えられる。そのため, CAF 課題が音声生成・知覚処理過程における音韻情報, 調音情報, 聴覚情報の対応付けに及ぼす影響は TAF 課題が及ぼす影響とは全く異なると考えられる。それでは, CAF 課題はどのようなメカニズムで有声性の生成や知覚に変化を及ぼしたのだろうか。生成の変化については Mitsuya et al.によって議論がなされており, CAF 課題で呈示される聴覚刺激は自らが発したものではないことが容易に分かるが, 実際に自らが発声したことで得られるはずの聴覚フィードバックが聴覚刺激によって遮られることが原因で生成に変化が生じた可能性があるとして述べられている。本研究で生じた有声性知覚の変化については, 有声性の生成が変化したこと自体が知覚に影響を及ぼした可能性が考えられる。音韻の生成が変化することは音韻情報と調音情報の対応関係を変化させることと同義であり (図 11), もし音声知覚時に聴覚情報から音韻情報を知覚する際に調音情報を介在させる処理が働くのであれば音韻情報と調音情報の対応関係の変化は知覚を変化させる要因となり得るだろう。

本研究では、CAF 課題を用いた実験を行うことで有声性の生成処理の変化が有声性のカテゴリー知覚に影響を及ぼすことを示すとともに、その相互作用のメカニズムについて音声生成・知覚処理過程における音韻情報、調音情報、聴覚情報の対応付けに注目して議論を行った。CAF 課題を用いた本実験では音韻情報と調音運動の対応付けの変化が調音運動情報に基づく音声知覚処理に影響を及ぼした可能性を示唆したが、更なる追加実験を行うことで TAF 課題を用いた先行研究で議論されている聴覚情報と調音運動の対応付けの変化が調音運動情報に基づく音声知覚処理に与える影響についても検討を行うことが可能だと考えられる。追加の実験案について具体的に説明をすると、/da/の生成時に比較的長い VOT を持つ/da/の音声刺激をフィードバックし、/ta/の生成時には比較的短い VOT を持つ/ta/の音声刺激をフィードバックするといった擬似的な TAF 課題を行うことで、聴覚情報と調音運動の対応関係を変化させるとともに補償応答（CAF 課題と同じ発話の変化）を引き起こすことが出来ると考えられる。そのため、このような擬似 TAF 課題の前後での有声性知覚の変化を CAF 課題の前後での変化と比較すれば音韻情報と調音運動の対応付けの変化だけでなく聴覚情報と調音運動の対応付けの変化が音声知覚における調音運動情報に基づく音声知覚処理に与える影響についても議論が可能である。

音声生成と知覚の間で見られた相互作用のメカニズムを探る他の方法として、TAF 課題や CAF 課題と脳機能計測実験を組み合わせたアプローチも考えられる。例えば、Schuerman et al. (2017b) は、/ɛ/と/æ/の音韻対を用いて TAF 課題前後にカテゴリー知覚にどのような変化が生じるかを調べる実験を行ったが、TAF 課題時には EEG の計測も行った。この実験では、TAF 課題での発声の変化と TAF 課題前後での音韻カテゴリー知覚の変化の間に有意な相関が見られること、さらに、TAF 課題時の発声の変化と EEG で計測した聴覚応答（自らの発話音声を聴くことで生じた応答）の間に有意な相関が見られることが示された。この結果から、TAF 課題時に生じた聴覚情報と調音運動の

対応関係の変化が音声の生成および知覚の変化を引き起こした要因であると考えられている。TAF 課題や本研究で用いた CAF 課題を行っている際の脳活動に加えてその後で行う音声知覚課題時の脳活動を調べ、音声の生成や知覚の変化に関連する脳活動を特定して両者の共通性について検討することが出来れば音声生成と知覚の相互作用について更なる議論が可能となるであろう。



## 第5章 研究3：有声性知覚時の聴覚情報および調音運動情報の利用に係る脳内メカニズム：MEG および fMRI を用いた脳機能計測による検討

### 5.1 目的

有声性知覚時の聴覚情報及び調音運動情報の利用に係る脳内メカニズムを検討するために、VOT を操作した刺激を用いた知覚課題を行なっている際の脳活動を調べた。序論でも述べたように、有声性知覚時の聴覚情報の利用に係る脳内メカニズムを調べるためには数ミリから数十ミリ秒単位の時間情報（VOT）に対する聴覚皮質の応答を調べることが出来るように時間解像度の高い脳機能計測手法を用いる必要がある。一方で、調音運動情報の利用に係る脳内メカニズムを調べるためには音声生成に関与する脳領域の活動を精確に検出するために空間解像度が高い脳機能計測手法を用いる必要がある。そこで本研究では、高い時間解像度と空間解像度を合わせ持つ MEG を用いて有声性知覚時の脳活動を計測することで聴覚情報及び調音運動情報の利用に係る脳内メカニズムを同時に調べることを試みた。ただし、MEG の空間解像度については、実験方法やデータの分析方法に依存する面があるため必ずしも信頼性が高いとは言えない。そこで、MEG 実験の前段階として MEG よりも空間解像度の高い fMRI を用いて有声性知覚時に活動する脳領域を特定する実験を行い、その結果に基づいて MEG データの分析を行うことで空間解像度に関する信頼性を高めた。さらに、fMRI 計測実験では MEG では発声時に生じる体動ノイズの影響で調べるのが困難な有声性生成時の脳活動についても調べることで、有声性知覚時の脳内メカニズムを有声性生成時の脳内メカニズムと照らし合わせながら議論することを可能とした。MEG データの分析では、聴覚皮質や音声生成に関与する脳領域における神経振動が VOT の違いや知覚結果

の違いによってどのように異なるのかを調べることで聴覚情報及び調音運動情報が有声性知覚にどのように利用されているかを検討した。さらに、聴覚皮質や運動皮質における神経活動を独立に調べるだけでなく、それらの脳領域の間の機能的結合についても調べることで有声性知覚時に働く脳内ネットワークについて先行研究で提案されている音声知覚時の脳内ネットワークとの比較を行った。

## 5.2 実験 1

### 5.2.1 実験参加者

有声性の知覚課題と生成課題それぞれで 13 名（知覚課題: 女性 8 名, 男性 5 名, 平均年齢 23.5 歳, 生成課題: 女性 6 名, 男性 7 名, 平均年齢 28.9 歳）が実験に参加した。上記の実験参加者の内, 1 名は知覚課題と生成課題の両方に参加した。全ての参加者は右利きで, 聴覚障害を有しておらず, 本実験で用いた刺激の聴取に問題はなかった。この実験は, 九州大学システム情報科学府の実験倫理審査で承認を受けた上で行われ, 参加者は実験前に実験の説明を受けた上で実験参加の同意書に署名を行った。

### 5.2.2 実験装置及び刺激

fMRI 計測実験では, 大分市の医療法人 至誠会 帆秋病院に設置された 3T MRI 装置 (Siemens, MAGNETOM Spectra) が用いられた。機能画像の撮像ではグラディエントエコー型エコープレナー法を用い (マトリックスサイズ:  $64 \times 64$ , TR: 10 s, TE: 30 ms, FA:  $90^\circ$ , FoV: 192 mm, スライス厚: 3.0 mm, スライスギャップ 0.75 mm, スライス枚数: 38 枚, ボクセルサイズ:  $3.0 \times 3.0 \times 3.0$  mm), 構造画像の撮像では 3D 超高速撮像法 (MPRAGE) を用いた (マトリックスサイズ:  $240 \times 256$ , TR: 2.3 s, TE: 3.33 ms, FA:  $9^\circ$ , FoV: 192 mm, スライス厚: 1.2 mm, スライスギャップ: 0 mm, スライス枚数: 176 枚, ボ

クセルサイズ:  $1.0 \times 1.0 \times 1.0$  mm)。知覚課題での刺激は、パーソナルコンピュータ (DELL, Precision Tower 5810) , ヘッドホンアンプ (Marantz, HD-DAC1) を用いて、fMRI 装置に備え付けられたヘッドホンから呈示した。参加者の回答の収集には MEG あるいは MRI 装置専用のレスポンスパッド (Physio-Tech, HHSC-1 x5-N4) を用いた。生成課題での発話の指示は、プロジェクタを用いて MRI 装置のヘッドコイルに備え付けた鏡にスクリーンを通して投射することで視覚呈示した。また、fMRI のスキヤンのタイミングと知覚課題での聴覚刺激呈示や生成課題での発話タイミングの指示の同期を取るために MRI 同期装置 (Physio-Tech, MS-06C) を用いた。

知覚課題では、VOT を 10 ms 間隔で 0 ms から 40 ms まで操作して、/d/から/t/へと知覚が変化する音声刺激 (後続母音/a/) を用いた。フォルマントについては、第 2, 3 フォルマント周波数を刺激の開始から 40 ms の間にそれぞれ 1600 Hz から 1200 Hz, 3000 Hz から 2500 Hz へと変化させてその後は一定とした。第 1, 4, 5 フォルマント周波数は常に一定で、それぞれ 600, 3300, 3750 Hz に設定した。後続母音の F0 は常に一定で 100 Hz に設定し、持続時間は VOT によって異なり、全体長さが 250 ms になるように調整した。音声刺激は 85 dBA の音圧に調整し、右耳から呈示した。刺激の呈示音圧の調整には、騒音計 (Brüel & Kjær, 2260) , 人工耳 (Brüel & Kjær, 4153) , マイクロフォン (Brüel & Kjær, 4192) を用いた。

### 5.2.3 実験手続き

知覚課題はスパースサンプリング撮像を用いた実験デザインで行い、2.5 s のスキヤンを 10 s 間隔で行った。スキヤンが終了した 1 s 後から 1.2 s 間隔で 5 回音声刺激が呈示される 20 試行と、音声刺激が呈示されない 10 試行がランダムな順番で行われた。刺激は全体を通して 100 回呈示されたが、VOT の異なる 5 つの刺激がそれぞれ 20 回ずつランダムな順番で呈示された。実験参加者は、音声刺激が呈示される試行では、左手

に持ったボタンで、/da/に聴こえたか/ta/に聴こえたかを回答した。生成課題での fMRI の計測方法は知覚課題と同じで課題内容のみが異なった。2.5 s のスキャンが終了した 1 s 後から /da/, /ta/, /a/, x のいずれかがスクリーン上に 6 s 間呈示された。文字が呈示された場合にはその文字が消えるまで繰り返し発声するように、x が呈示された際には口は動かさず安静を保つように教示した。それぞれの文字はランダムな順番で 10 試行ずつ呈示された。両課題は 1 回のセッションで行われ、約 6 分で 37 回のスキャンを行なった。最初の 5 スキャンは磁場が安定しない状態で撮像が行われている可能性があるため課題は行わず解析からも除外した。

#### 5.2.4 データ分析

fMRI データの解析には脳画像解析用のソフトウェアである FreeSurfer を用いた。知覚課題と生成課題のデータともに以下の手順でデータの前処理を行った。まず、機能画像を回転及び平行移動させることで最初の画像の位置に後続の画像の位置を合わせ、頭部の動きを補正する処理を行った。その次に機能画像と構造画像の位置合わせを行った上で個人脳データを標準脳上に変換した。最後に、ガウスフィルターを用いて半値幅 8mm の設定で空間的平滑化を行った。

知覚課題及び生成課題の前処理後の各参加者のデータについて、一般線型モデルを用いた分析を行った。知覚課題では各ボクセルにおける BOLD 信号の時系列データを従属変数、音声刺激が呈示される場合とされない場合の試行系列に基づいて作成した BOLD 信号の時系列変化モデルを独立変数としてボクセルごとに線型重回帰分析を行った。生成課題では /da/ を発声する場合、/ta/ を発声する場合、/a/ を発声する場合、発声を行わない場合の試行系列に基づいて作成した BOLD 信号の時系列変化モデルを独立変数として知覚課題と同様の分析を行った。BOLD 信号の時系列変化のモデルは、刺激の呈示や発声の開始から 4~7 秒で脳血流が増えることを想定した標準的な血流動態反

応関数を用いて作成した。各参加者で求められたボクセルごとの重回帰分析のパラメータに基づいて統計的有意性検定を行うことで、知覚課題の音声刺激が呈示される条件と呈示されない条件、生成課題の/da/、/ta/の発声を行う条件と発声を行わない条件、また、/da/、/ta/の発声を行う条件と/a/の発声を行う条件の間で活動に差が見られる脳領域を調べた。有意性検定では、まず始めに比較する条件間で重回帰分析のパラメータの平均値を比較するための  $t$  検定を行って  $p$  値が 0.001 以下になるボクセルを特定した。次に、特定したボクセルの隣接性に基づいてクラスターを形成した後にクラスターサイズに関する有意性検定を行った。クラスターサイズが 0 になるという帰無仮説が棄却された場合にはそのクラスターにおいて条件間で差があると見なした。クラスターサイズの検定における  $p$  値は確率場理論に基づいて導出された (式 2)。

$$p = e^{\left(-b \left[\frac{s}{r} h (h^2 - 1)\right]^{\frac{2}{3}}\right)} \quad (2)$$

式 2 において  $b$  は定数 (0.5334942...) で、 $h$  はクラスター作成のための閾値 ( $p = 0.001$  となる  $t$  値)、 $r$  はクラスター内での  $t$  分布における  $x, y, z$  方向の半値幅の積である。有意性の検定は、複数のクラスターで繰り返し検定を行うことで生じる多重性の問題を解決するためにファミリーワイズエラー (Family Wise Error, FWE) 率の補正を行った上で行った。

有声性知覚時と生成時で共通して活動が見られた脳領域については、知覚課題の音声刺激が呈示されない条件よりも呈示される条件で有意に活動が大きくなった脳領域と生成課題の発声を行わない条件よりも/da/、/ta/の発声を行う条件で有意に活動が大きくなった脳領域を比較することで特定した。また、/da/、/ta/の発声を行う条件と/a/の発声を行う条件を比較して有意な差が見られた脳領域を調べることで、上記で特定した脳領域の内、どの脳領域が主に子音部分あるいは母音部分の生成のどちらに係るのかについても検討を行なった。

### 5.2.5 結果

知覚課題において、音声刺激が呈示される条件の活動とされない条件の活動で有意な差が見られた脳領域を調べたところ、主に上側頭部、中心前部、上前頭部、島皮質などの6つのクラスターで有意な差が見られることが分かった（表1, 図20）。これらのクラスターは聴覚情報に基づく音声知覚処理の役割を担うとされる聴覚腹側経路や調音運動情報に基づく音声知覚処理の役割を担うとされる聴覚背側経路に当たると考えられる脳領域と概ね一致した。生成課題において、/da/と/ta/の発声を行う条件の活動と発声を行わない条件の活動で有意な差が見られた脳領域を調べたところ、知覚課題と同様に上側頭部、中心前部、上前頭部のクラスターで有意な差が見られることが分か

表 1. 知覚課題において音声刺激が呈示される条件と呈示されない条件の間で有意差が見られた脳領域（脳領域名, MNI 座標, ボクセル数）

Cluster No.	Annotation	MNI coordinates (mm)			Cluster size (mm <sup>2</sup> )	t-value
		X	Y	Z		
1	superior temporal	-60.5	-15.1	2.1	6195.76	9.11
2	precentral	-54.3	1.8	33.2	2381.76	6.35
3	superior frontal	-9.2	10.2	50.2	1053.83	7.36
4	insula	-28	23.4	5.9	843.38	6.77
5	lingual	-5.1	-90.1	-9.5	278.93	5.61
6	rostral middle frontal	-25.7	45.9	16.2	186.04	3.95

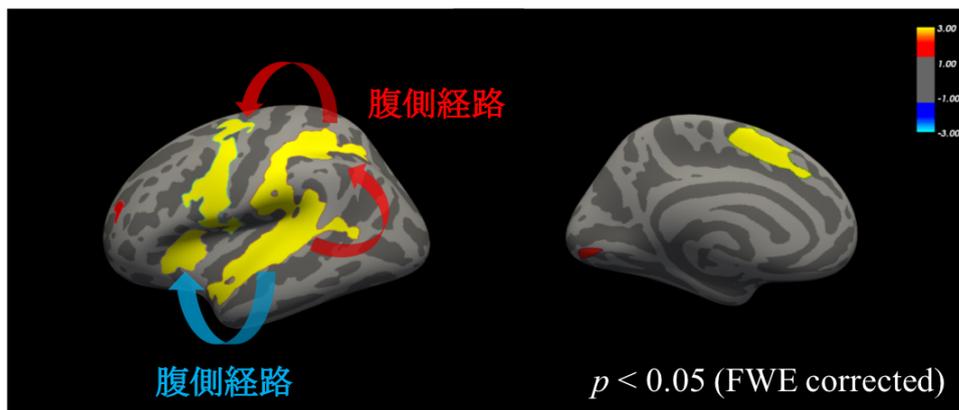


図 20. 知覚課題において音声刺激が呈示される条件と呈示されない条件の間で有意差が見られた脳領域。

った（表 2, 図 21A）。続いて, /da/, /ta/の発声を行う条件と/a/の発声を行う条件を比較して有意な差が見られた脳領域を調べたところ, 主に上側頭部, 中心前部など 10 個のクラスターで有意な差が見られることが分かった（表 3, 図 21B）。

上側頭部, 中心前部, 上前頭部のクラスター内で知覚課題と生成課題の両方で活動が認められた脳領域を Glasser et al. (2016) の大脳皮質区画に基づいて詳細に調べたと

表 2. 生成課題において /da/, /ta/を発声する条件と発声をしない条件の間で有意差が見られた脳領域（脳領域名, MNI 座標, ボクセル数）。

Cluster No.	Annotation	MNI coordinates (mm)			Cluster size (mm <sup>2</sup> )	t-value
		X	Y	Z		
1	superior temporal	-64.1	-19.8	0.5	1789.45	5.86
2	precentral	-58.7	3.1	25.3	1628.39	6.29
3	superior frontal	-6.8	0.4	60.1	339.94	5.35

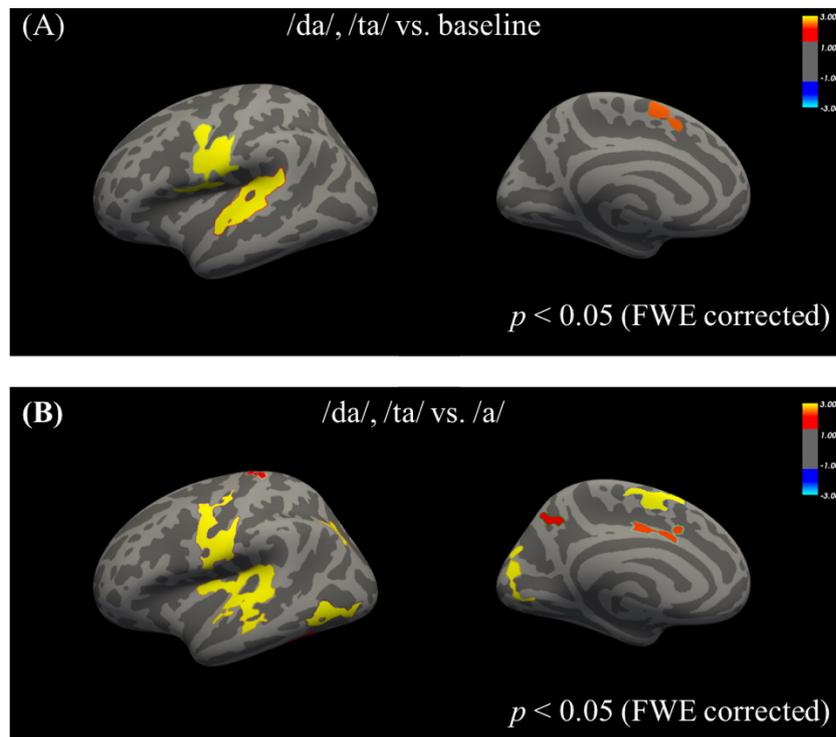


図 21. 生成課題において(A)/da/, /ta/を発声する条件と発声をしない条件の間で有意差が見られた脳領域. (B)/da/, /ta/を発声する条件と/a/を発声する条件の間で有意差が見られた脳領域.

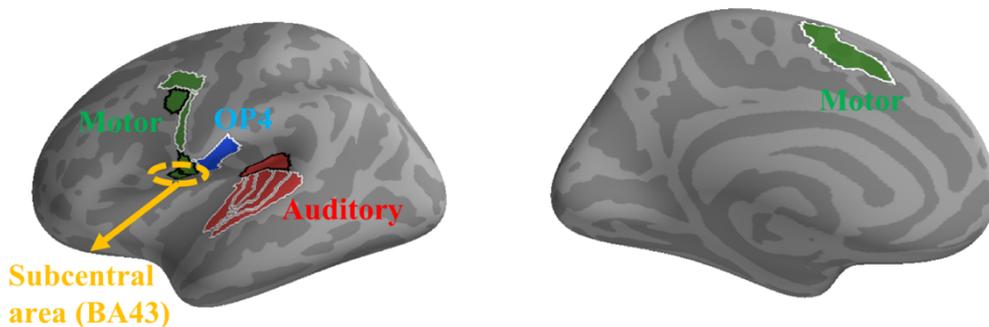
表 3. 生成課題において /da/, /ta/ を発声する条件と /a/ を発声する条件の間で有意差が見られた脳領域（脳領域名, MNI 座標, ボクセル数）。

Cluster No.	Annotation	MNI coordinates (mm)			Cluster size (mm <sup>2</sup> )	t-value
		X	Y	Z		
1	superior temporal	-55.5	-15.6	3.5	2117.4	5.67
2	precentral	-52.3	-6.9	39.8	1483.31	10.33
3	cuneus	-4.5	-77.8	17.8	731.31	4.83
4	inferior temporal	-49.2	-64.5	-4.2	668.94	4.67
5	superior frontal	-7.4	-1.7	55.1	625.53	7.11
6	inferior parietal	-31.9	-76.8	38.6	451.99	5.32
7	poterior cingulate	-5.6	-6.9	40.2	262.84	5.61
8	postcentral	-25.8	-30.5	66.6	220.06	4.42
9	precuneus	-6.5	-65.9	47.1	197.47	4.3
10	inferior temporal	-45.7	-43.3	-15.7	165.43	4.31

ころ、聴覚皮質や音声生成に関与する脳領域と両者との間に相互接続があるとされる OP4 でも共通した活動が見られることが分かった（図 22）。運動皮質については、声帯振動の準備や開始など調音の時間的な制御の役割を担い、VOT の制御にも密接に関与すると考えられる subcentral area (SC) の活動が見られることが分かった。さらに、知覚課題と生成課題で共通して活動が見られた脳領域の内、/da/, /ta/ の発声を行う条件と /a/ の発声を行う条件を比較して有意な差が見られた脳領域を調べたところ、聴覚皮質と運動皮質の一部を除いてほとんどの脳領域で有意な差が見られたが（図 22 白枠, 太文字）、SC では有意な差は見られなかった。

### 5.3 実験 1 考察

有声性の知覚実験と生成実験で共通して活動が生じる脳領域を調べたところ、Sepulcre (2015) が提唱した OP4 を介した聴覚皮質と運動皮質のネットワークと一致する脳領域で活動が見られることが分かった。調音運動情報に基づく音声知覚処理の役割を担うとされる聴覚背側経路 (Hickok & Poeppel, 2004, 2007; Rauschecker et al., 2009, 2011) における縁上回やシルビウス裂後部の頭頂・側頭接合部の付近では有声性知覚時には活動が見られたが、有声性生成時には活動が見られなかった。音韻情報の



● superior temporal	● precentral
<b>Primary Auditory Cortex (A1)</b>	<b>Ventral Area 6 (v6)</b>
<b>Medial Belt Complex (MBelt)</b>	<b>Area 55b (55b)</b>
<b>Lateral Belt Complex (LBelt)</b>	Premotor Eye Field (PEF)
<b>Parabelt Complex (PBelt)</b>	Frontal Opercular Area1 (FOP1)
<b>Auditory 4 Complex (A4)</b>	Area 43 (43) = subcentral area
RetroInsular Cortex (RI)	● superior frontal
<b>Area OP4/PV (OP4)</b>	<b>Supplementary and Cingulate Eye Field (SCEF)</b>

図 22. 有声性の知覚課題と生成課題で共通して活動が見られた脳領域（白枠、黒枠はそれぞれ生成課題において /da/, /ta/ を発声する条件と /a/ を発声する条件で有意差が見られた脳領域と見られなかった脳領域）と脳領域の具体的な名称（太文字は生成課題において /da/, /ta/ を発声する条件と /a/ を発声する条件で有意差が見られた脳領域）。

知覚時と生成時の脳活動を fMRI で計測して比較した幾つかの先行研究（Du et al., 2014; Evans & Davis, 2015）でも同様に音韻情報の知覚時には聴覚背側経路の活動が検出されている一方で、音韻情報の生成時には活動が検出されていない。このことを考慮すると、聴覚背側経路が音声生成および知覚において感覚情報と運動情報の情報交換を行うという脳内メカニズムについては再考の余地があり、OP4 がその役割を担う可能性についても議論を進めていく必要があるだろう。

さらに、運動皮質の中には、先行研究で音韻対の知覚との関連が示唆されている運動野・運動前野のみならず SC でも共通した活動が見られた。SC は声帯振動の準備や開始に関わるとされるため（Scott & Johnsrude, 2003; Toyoda et al., 2014）、有声性生成時の VOT の制御にも密接に関わると考えられる。このことは、調音の時間的制御を行う脳領域の活動が、調音の時間的性質の違いに特徴付けられる音韻弁別素性（有声

性) の知覚に關与することを示唆するものである。SC では、/da/と/ta/を發声している際の活動と/a/を發声している際の活動に有意な差は見られなかったが、この結果はどちらの發声条件でも声帯振動を繰り返して行なっていたためだと考えられる。その他の運動關連の脳領域については、一部の領域 (PEF) を除いて/dɑ/と/tɑ/を發声している際と/a/を發声している際で活動に差が見られたため、有声性生成時には主に舌の運動制御など子音部の生成に主に關与すると考えられる。

## 5.4 実験 2

### 5.4.1 実験参加者

実験参加者は日本語母語話者 20 名 (女性 8 名, 男性 12 名) で平均年齢は 27.0 歳であった。20 名の参加者の内, 1 名は実験 1 の知覚実験と發声実験の両方に, 4 名は發声実験に参加した。どの参加者も正常聴力で, 本実験で用いた刺激の聴取に問題はなかった。この実験は, 九州大学システム情報科学府の実験倫理審査で承認を受けた上で行われ, 参加者は実験前に実験の説明を受けた上で実験参加の同意書に署名を行った。

### 5.4.2 実験装置と刺激

実験で用いた刺激や刺激の呈示耳, 音圧設定は fMRI 計測実験と同じであった。刺激の呈示には, パーソナルコンピュータ (DELL, Latitude E5470), ヘッドホンアンプ (Marantz, PS3001), インサートイヤホン (Etymotic Research, ER3A) を用いた。MEG 計測は, 九州大学病院ブレインセンターに設置されている全頭型 306 ch (マグネトメータ 102 ch, グラジオメータ 204 ch) 脳磁計 (Elekta, Neuromag) を用い, サンプルング周波数 1000 Hz, 帯域通過フィルタ 0.1-300 Hz の設定で計測を行った。MEG センサーと脳の位置關係を把握するために, MEG 計測実験の前に 4 つの Head point indicator

(HPI) コイルを頭部に配置して MEG センサーと HPI コイルの位置関係を取得するとともに 3D カメラ (Konica Minolta, VIVID 9i) を用いて頭部形状と頭部に設置した HPI コイルを撮影した。これらのデータと各参加者で撮像した MRI 構造画像データから復元した頭部形状のデータを用いて MRI 構造画像の座標と MEG データの座標の位置合わせを行った (Hironaga et al., 2014)。各参加者の MRI 構造画像の撮像は、大分市の医療法人 至誠会 帆秋病院に設置された 3T MRI 装置 (Siemens, MAGNETOM Spectra), または、高知工科大学に設置された 3T MRI 装置 (Siemens, MAGNETOM Verio) が用いられた。

#### 5.4.3 実験手続き

5 つの刺激をランダムな順番でそれぞれ 100 回ずつ呈示し、実験参加者には呈示された音声が /da/ に聴こえるか /ta/ に聴こえるかを回答してもらった。1 試行の流れを具体的に説明すると (図 23A), 始めに注視点が見れ、その 1~2 秒後に 5 つの刺激のいずれかが呈示された。刺激呈示の 1 秒後に回答画面に移り、注視点の左右どちらか一方に /da/, もう一方に /ta/ の文字が呈示された。実験参加者はマウスを使って、音声刺激の聴

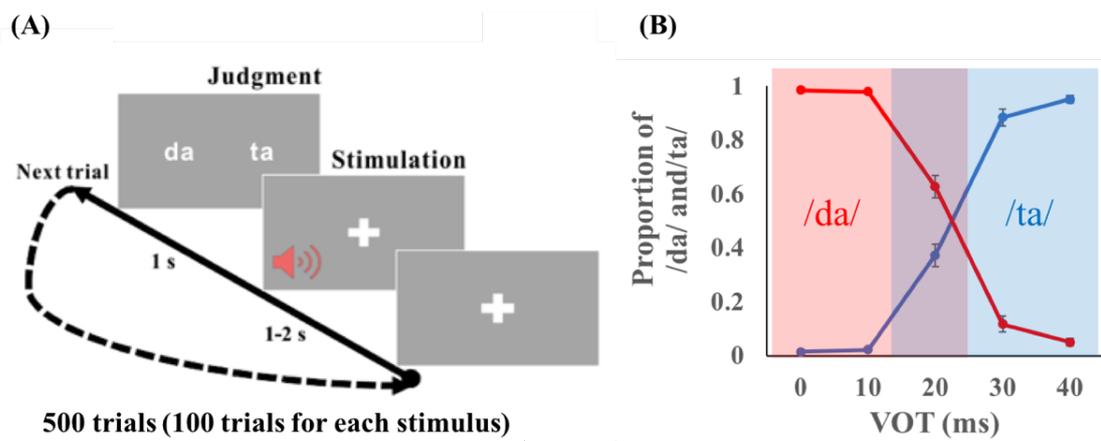


図 23. (A) MEG 計測中の知覚課題における 1 試行の流れ. (B) 知覚課題における /da/ と /ta/ の回答率の参加者平均. エラーバーは標準誤差を示す.

こえと一致する文字を、左クリックまたは右クリックで選んだ。実験は全 500 試行の 100 試行ごとに休憩を挟んで行い、1 セッション毎の所用時間は約 6 分であった。

#### 5.4.4 データ分析 行動データ

VOT の異なる 5 つの刺激に対する /da/ と /ta/ の回答率を参加者ごとに求めた後、参加者の全体平均を求めた (図 23B)。VOT 0 ms と 10 ms の刺激は明確に /da/、VOT 30 ms と 40 ms の刺激は明確に /ta/ と知覚された。VOT 20 ms の刺激については、ほとんど全ての参加者で /da/ と /ta/ のどちらかが明確に知覚されることはなかった。

#### 5.4.5 データ分析 MEG データ

##### 前処理及び信号源解析

MEG データの解析には MNE (Gramfort et al., 2013, 2014) 及び FreeSurfer を用いた。データの前処理では、Elekta 社提供の Maxfilter (Taulu et al., 2006) を計測データに適用して外部ノイズを除去した後、帯域通過フィルタ (1-100 Hz) 及びノッチフィルタ (60 Hz) を適用した。続いて、刺激呈示の 100 ms 前から刺激呈示の 500 ms 後の 306 チャンネルのセンサー波形を抽出した。この際、マグネトメータについては 8000 fT 以上、グラジオメータについては 5000 fT/cm 以上の振幅を持つ波形は除外した。

信号源解析では、まず始めに各参加者の MRI 構造画像に基づいて脳内に 20484 点の信号源を設定し、各点で特定の電流が生じた場合に MEG センサーでどのような磁場が得られるかを予測するモデルを作成した (式 3) (Hamalainen & Sarvas, 1989)。

$$B_k = \sum_{p=1}^{20484} g_{kp} (\vec{r}_{kp}, \hat{e}_{kp}) \mathbf{q}_p^T + N \quad (3)$$

式 3 において、 $B_k$  は特定の MEG センサーで得られる磁場、 $\mathbf{q}_p$  は特定の信号源で生じる電流を表している。 $g_{kp}$  は、特定の MEG センサーに対する特定の信号源の位置  $\vec{r}_{kp}$  や電

流の向き  $\widehat{e}_{kp}$  を変数としてその信号源で生じた単位電流が特定の MEG のセンサーにもたらす磁場を予測する方程式である。N はセンサーに混入するノイズを表している。

次に、作成した順モデルに基づいて、1 試行ごとに抽出したセンサー波形や VOT の異なる 5 つの刺激ごとに求めた加算平均波形から信号源分布の時間的変化を求める逆推定を行った。fMRI 計測実験の結果から複数の脳領域において電気活動が生じていることが予想されるため、極端に限局した信号源分布が解として求められないようにするために逆問題の解法には最小ノルム法 (Hamalainen & Ilmoniemi, 1994; Molins et al., 2008) を用いた。最後に、刺激呈示前 (-100 ~ 0 ms) の計測データを用いて dynamical statistical parametric mapping (dSPM) 法 (Dale et al., 2000) によるデータの正規化を行った。

#### 関心領域 (Region of Interest, 以下 ROI) の設定

ROI の設定を行うために、まず始めに fMRI 計測実験の知覚課題と生成課題で共通して活動が見られた脳領域において、MEG 計測実験でも同様に活動が推定されたかどうかを確かめた。参加者ごとに加算平均波形データを用いた信号源解析を行うことで 100 ~ 500 ms の潜時における各刺激聴取時の信号源分布を求めた後、時間分解度が高い MEG のデータを時間分解度が低い fMRI のデータと比較するために MEG の信号源分布のデータを時間平均した。さらに、各刺激で求めた信号源分布の時間平均データを標準脳上に変換した上で全て平均した。図 24A に示す全参加者の平均と fMRI 計測実験における知覚課題の結果 (図 20) と比較すると活動の限局性が低い傾向があるものの、活動が見られる脳領域は類似することが分かった。fMRI の知覚課題と生成課題で共通して活動が見られた領域 (図 24A 黒枠) で活動が見られたかどうかを調べたところ、聴覚皮質付近の脳領域では全ての領域で明瞭な活動が見られた。その中でも一次聴覚野 (A1) とパラベルト領域 (PBelt) において特に強い活動が見られたため (図

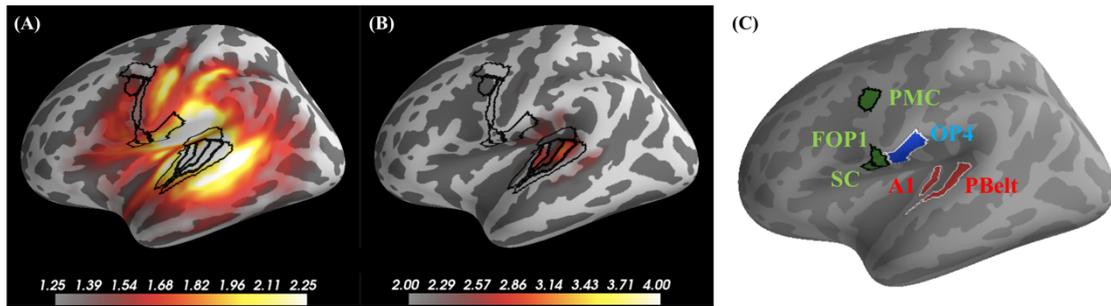


図 24. (A) 潜時-100-500 ms における信号源分布の時間平均データの全刺激全参加者平均. 黒枠→fMRI 計測実験の知覚課題と生成課題で共通して活動が見られた脳領域. (B) (A)の表示閾値を高め設定. (C) (A) (B)の結果を基に設定した関心領域(ROI).

24B), この2つの領域を ROI に設定して詳細な分析を行った (図 24C)。調音運動情報に基づく音声知覚に係る脳領域については, PEF や SC, FOP1, OP4 で活動が見られたため ROI に設定して詳細な分析を行った (図 24C)。

#### 時間周波数解析

上述のように設定した6つのROIにおける神経振動が聴取した刺激のVOTや知覚結果によってどのように異なるのかを検討するために, それぞれのROIから求めた各参加者の活動波形に時間周波数解析を適用し, その結果を5つの刺激条件間で比較した。解析の手順を具体的に説明すると, まず始めに各参加者で1試行ごとに抽出した306チャンネルのセンサー波形データを用いて信号源解析を行い, 各ROIでROI内にあるグリッド点の活動を平均することで活動波形を抽出した。次に, 1試行ごとに求められた各ROIでの活動波形にモーレットウェーブレット変換を利用した時間周波数解析を適用した後, VOTが異なる5つの刺激それぞれで各潜時(-100~500 ms)・周波数(1~40 Hz)における試行間位相同期度(Inter-trial coherence, 以下ITC)を求めた。ITCは0から1の値を取り, 値が大きい方が試行毎に生じる神経活動の同期性が高いことを表す。

各 ROI で ITC に有意差が見られる潜時・周波数範囲を全ての刺激ペア（0-10, 0-20, 0-30, 0-40, 10-20, 10-30, 10-40, 20-30, 20-40, 30-40 ms）で調べて、同じ知覚結果になる刺激ペア（0-10, 30-40 ms）と知覚に違いがある刺激ペア（0-30, 0-40, 10-30, 10-40 ms）で結果を比較することで ROI に設定した 6 つの脳領域において有声性の知覚に依存した神経振動が生じているのかについて検討した。有意性の検定は、cluster-based permutation 検定（Maris & Oostenveld, 2007）を用いて行なった。まず始めに、10 個の刺激ペアそれぞれで  $t$  検定を行って  $p$  値が 0.05 以下になる潜時・周波数範囲を特定した後、特定された潜時・周波数範囲の近接性に基づいて複数のクラスターを作成した。Permutation 検定を用いた有意性の検定では、比較する刺激ペア間で ITC のデータをランダムに入れ替えたデータを用いて各潜時・周波数の  $t$  値を求めて、上述の手順で作成したクラスター内での  $t$  値の合計値と全てのクラスターで計算された  $t$  値の合計値の中の最大値を求めた。この計算を 1000 回繰り返し行って全てのクラスターで計算された  $t$  値の合計値の中の最大値の分布を作成し、オリジナルの ITC データから計算した各クラスター内の  $t$  値の合計が分布の上位 0.5%以上の値を取るかどうかを確かめた。上位 0.5%以上の値取るクラスターについてはクラスター内の潜時・周波数において刺激ペア間で ITC に有意差があると見なした（多重比較の補正を行った有意水準を設定した）。

### 機能的結合解析

聴覚皮質や音声生成に関与する脳領域の間で機能的な相互作用が見られるかどうかを調べるために、各参加者各刺激で聴覚皮質や音声生成に関与する脳領域の両方と相互接続があるとされる OP4 と脳内に設定した全てのグリッド点との間で Phase slope index (PSI) (Nolte et al., 2004, 2008) を求めた。PSI の計算では、まず始めに OP4 の活動波形と各グリッド点での活動波形にウェーブレット変換を利用した時間周波数分

析を適用し、両者の間のコヒーレンス（クロススペクトル）を求めた。続いて、各潜時・周波数におけるコヒーレンスの虚数部を計算し、関心のある周波数帯域の計算結果を加算することで OP4 と各グリッド点の間の PSI の時間変化を求めた（式 4）。

$$PSI(t) = \Im \left( \sum_{f \in F} C_{ij}^*(t, f) C_{ij}(t, f + \delta f) \right) \quad (4)$$

式 4 において、 $t$  は潜時、 $f$  は周波数、 $\delta f$  は周波数分解能、 $F$  は周波数帯域の設定（ $\theta$  帯域（4–8 Hz）、 $\alpha$  帯域（8–13 Hz）、 $\beta$  帯域（13–30 Hz）の 3 つの周波数帯域で計算）である。 $C_{ij}$  は OP4 の活動波形（ $i$ ）と各グリッド点の活動波形（ $j$ ）から計算した複素コヒーレンスで、 $C_{ij}^*$  は  $C_{ij}$  の複素共役である。 $\Im(z)$  は  $z$  の虚数部を取り出すことを意味しており、式 3 は、 $PSI(t) = \sum_{f \in F} |C_{ij}(t, f)| |C_{ij}(t, f + \delta f)| [\sin[\Phi(t, f + \delta f) - \Phi(t, f)]]$  のように展開できる（ $\Phi$  はコヒーレンスの位相成分）。また、 $[\sin[\Phi(t, f + \delta f) - \Phi(t, f)]]$  の部分は  $[\Phi(t, f + \delta f) - \Phi(t, f)]$  と近似出来る。

A1, PBelt, PMC, SC と OP4 の間の PSI の時間的な変化を調べるために、各潜時で A1, PBelt, PMC, SC の ROI 内にあるグリッド点と OP4 の間の PSI の平均値を求めた。PSI は -1 から 1 の値を取り、負の値を持つ場合は OP4 へ情報を伝達する、正の値を取る場合は OP4 から情報が伝達される脳領域だと解釈される。

#### 5.4.6 結果

全ての刺激ペアで A1, PBelt の ROI における ITC に ROI における ITC に有意な差が見られた潜時・周波数帯を調べたところ（図 25, 26）、A1 では 0–10, 20–30 ms の刺激ペア、PBelt では 20–30 ms の刺激ペアで有意差が認められる潜時・周波数帯は見つからなかったが VOT 10 ms 差ペアの内 10–20, 30–40 ms のペアでは A1, PBelt 共に有意な差が認められる潜時・周波数帯が見られた。VOT 20 ms 差のペアについては、A1 では 0–20 ms の刺激ペアで有意な差が認められた潜時・周波数帯は見つからなかったが、PBelt では全ての刺激ペアで有意な差が認められた潜時・周波数帯が見られた。続

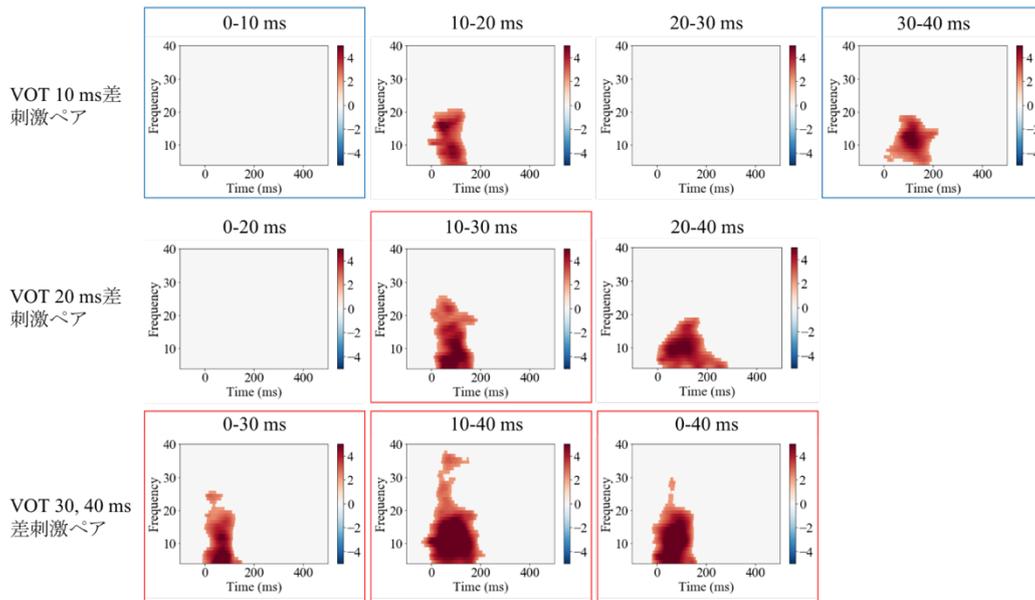


図 25. A1 の ROI において各刺激ペアで有意 ( $p < 0.001$ ) な差が認められた  
 潜時・周波数帯. 青枠は知覚に違いがない刺激ペア. 赤枠は知覚に違いが  
 ある刺激ペア.

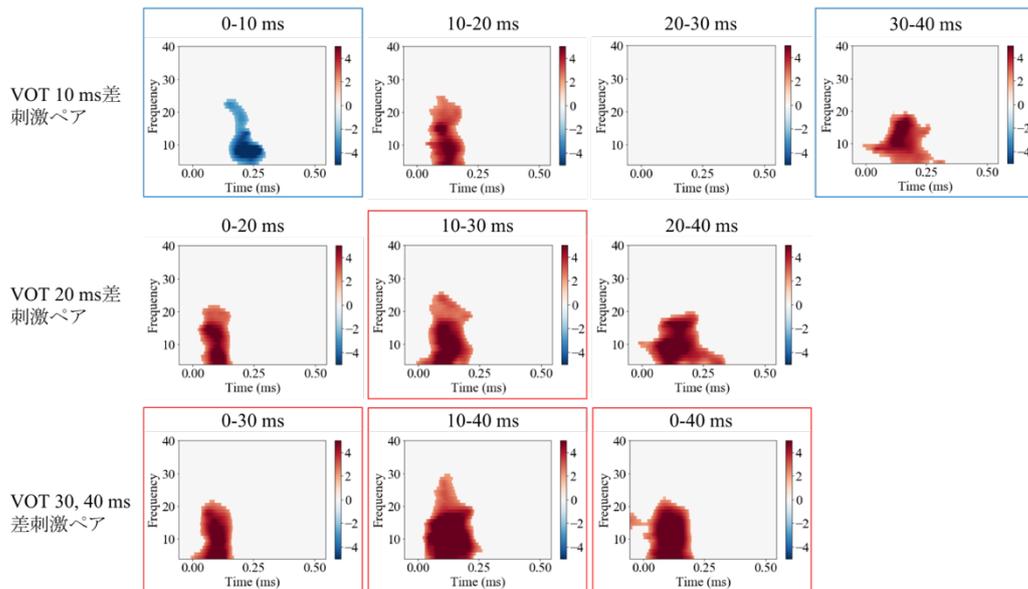


図 26. PBelt の ROI において各刺激ペアで有意 ( $p < 0.001$ ) な差が認められた  
 潜時・周波数帯. 青枠は知覚に違いがない刺激ペア. 赤枠は知覚に違いが  
 ある刺激ペア.

いて, A1 や PBelt などの聴覚皮質において有声性知覚に関連した神経振動が生じてい  
 るかどうかを調べるために, 知覚に違いがある刺激ペア全て (0-30, 0-40, 10-30, 10

-40 ms) で有意な差が認められる潜時・周波数帯を特定して、その潜時・周波数帯において知覚に違いがない刺激ペア (0-10, 30-40 ms) で有意な差が見られるかどうかを確認した。その結果、主に潜時 100 ms よりも早い潜時帯の  $\beta$  帯域の活動において知覚に違いがある刺激ペア全てで有意な差が認められる且つ知覚に違いがない刺激ペアで有意な差が見つからない神経振動が見つかった (図 27)。

調音運動情報に基づく音声知覚処理に関与すると考えられる OP4, PMC, SC でも A1 や PBelt と同じ方法で刺激ペアごとに有意差が見られる潜時・周波数帯を調べて (図 28-31), 有声性知覚との関連性を検討した。その結果, OP4 では知覚に違いがある刺激ペア全てで有意差が認められた潜時・周波数帯は見られたもののその大部分は知覚に違いがない刺激でも有意な差が見られた (図 32)。PMC や FOP1 では知覚に違いがある刺激ペア全てで有意差が見られる潜時・周波数帯は見つからなかった (図 32)。SC では知覚に違いがある刺激ペア全てで有意差が認められた潜時や周波数帯域が見られた一方で、知覚に違いがない刺激ペアでは有意差が認められなかった (図 32)。この結果は、A1 や PBelt といった聴覚皮質における神経振動だけではなく SC における神経振動が有声性知覚により密接に関連することを示している。

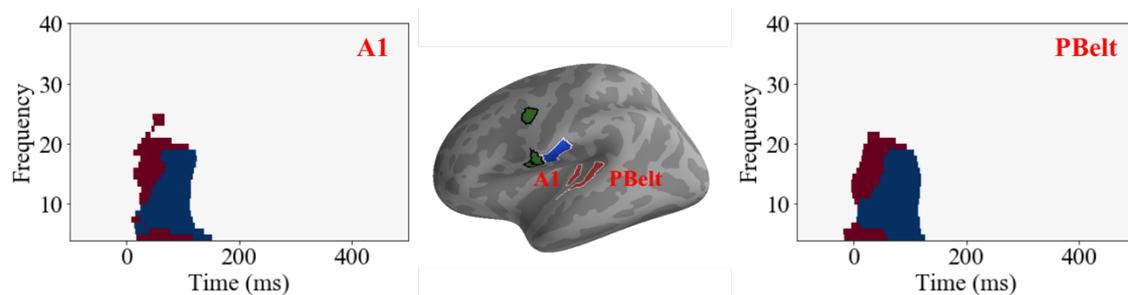


図 27. A1, PBelt の ROI において知覚に違いがある全ての刺激ペアで有意差が確かめられる且つ知覚に違いがない刺激ペアで有意差が確かめられなかった潜時・周波数帯 (赤)。知覚に違いがある全ての刺激ペアで有意差が確かめられる且つ知覚に違いがない刺激でも有意差が確かめられた潜時・周波数帯 (青)

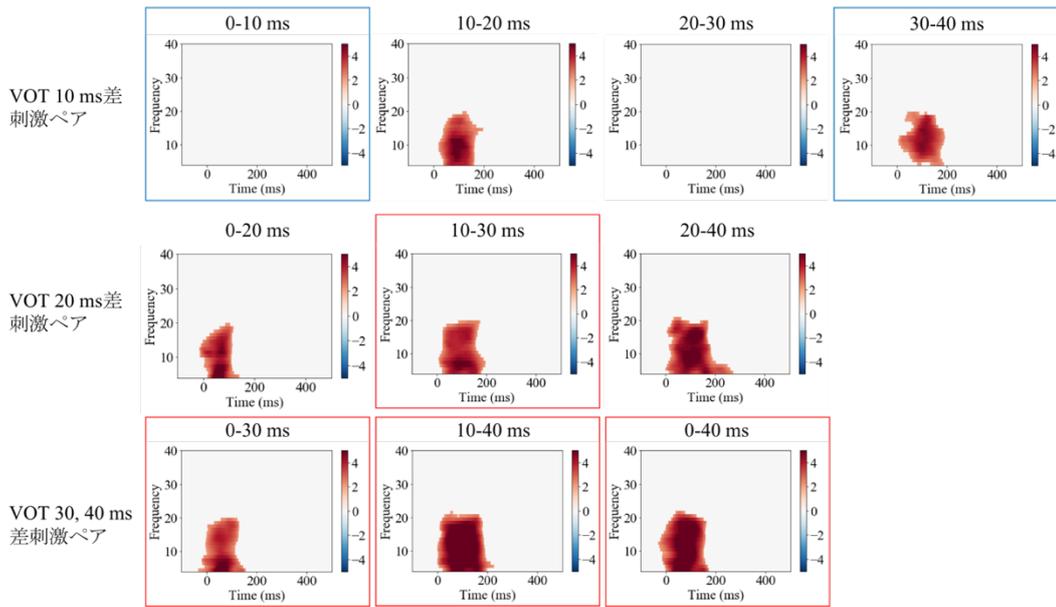


図 28. OP4 の ROI において各刺激ペアで有意 ( $p < 0.001$ ) な差が認められた潜時・周波数帯。青枠は知覚に違いがない刺激ペア。赤枠は知覚に違いがある刺激ペア。

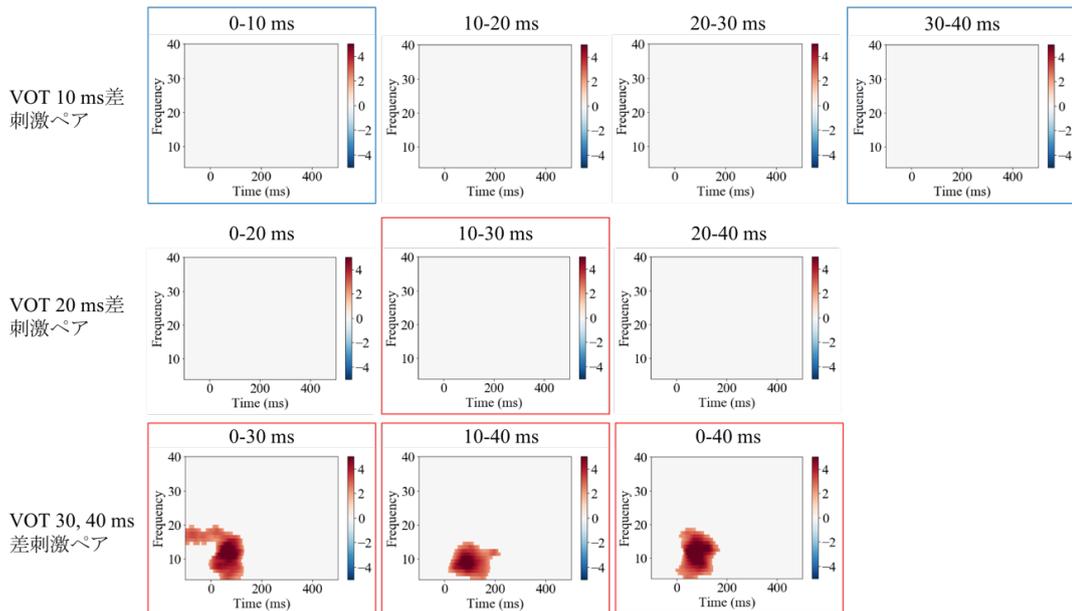


図 29. PMC の ROI において各刺激ペアで有意 ( $p < 0.001$ ) な差が認められた潜時・周波数帯。青枠は知覚に違いがない刺激ペア。赤枠は知覚に違いがある刺激ペア。

さらに、聴覚皮質や音声生成に関与する脳領域と OP4 の間の機能的結合を調べたところ (図 33) , 聴覚皮質の A1 や PBelt についてはどの刺激を聴いている際にも潜時

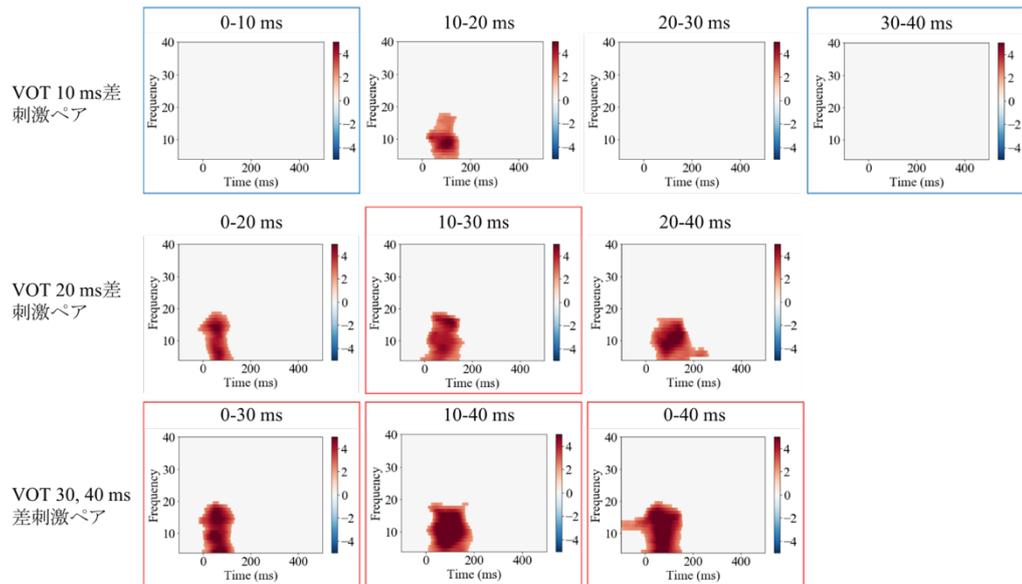


図 30. SC の ROI において各刺激ペアで有意 ( $p < 0.001$ ) な差が認められた潜時・周波数帯. 青枠は知覚に違いがない刺激ペア. 赤枠は知覚に違いがある刺激ペア.

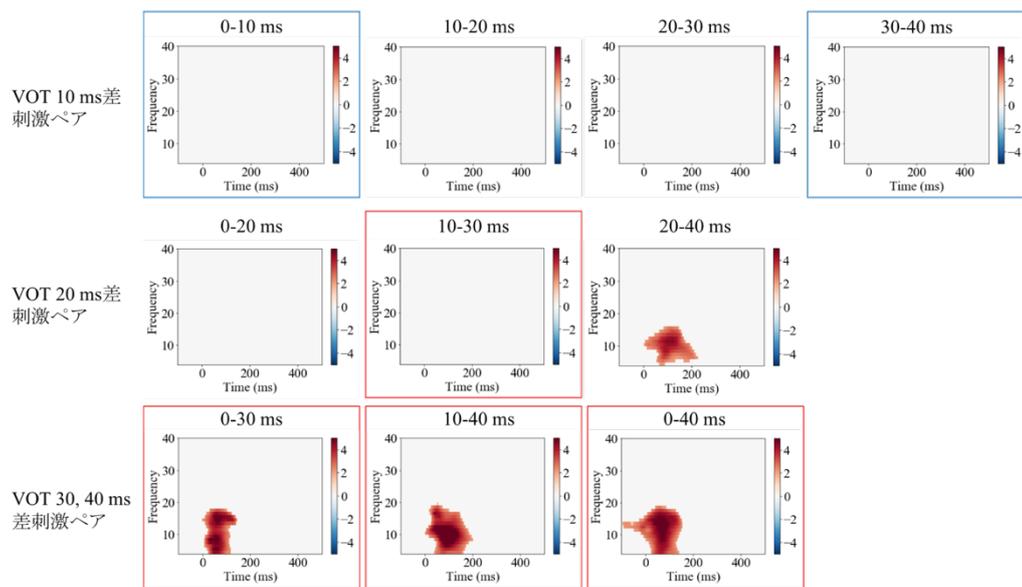


図 31. FOP1 の ROI において各刺激ペアで有意 ( $p < 0.001$ ) な差が認められた潜時・周波数帯. 青枠は知覚に違いがない刺激ペア. 赤枠は知覚に違いがある刺激ペア.

100 ms 付近で PSI が負の値を取ることが分かり、聴覚皮質から OP4 へ情報が伝達されていることを示唆する結果が得られた。音声生成に関与する脳領域については、SC ではどの刺激条件でも潜時 100 ms と 200 ms 付近で PSI が正の値を取ることが分かり、

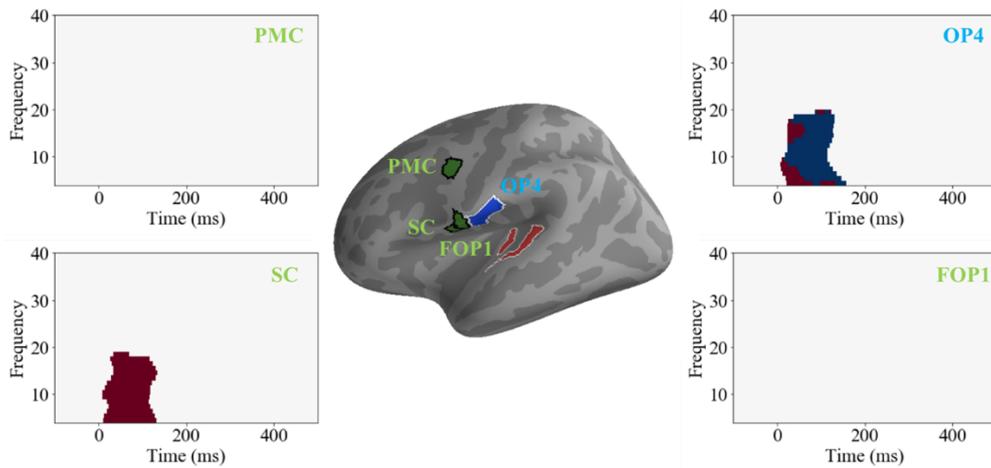


図 32. OP4, PMC, SC, FOP1 の ROI において知覚に違いがある全ての刺激ペアで有意差が確かめられる且つ知覚に違いがない刺激ペアで有意差が確かめられなかった潜時・周波数帯 (赤) . 知覚に違いがある全ての刺激ペアで有意差が確かめられる且つ知覚に違いがない刺激でも有意差が確かめられた潜時・周波数帯 (青)

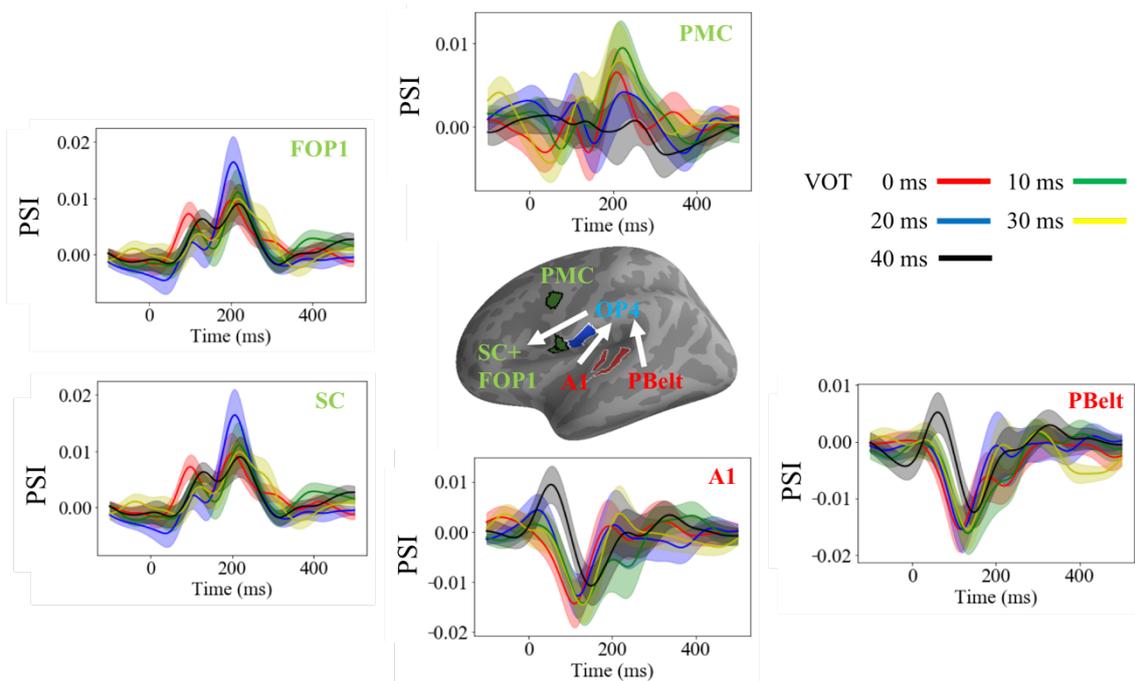


図 33. A1, PBelt, PMC, SC, FOP4 と OP4 の間の機能的結合 (PSI) の時間変化の参加者平均. 透明部は標準誤差を示す.

OP4 から SC へ情報が伝達されていることを示唆する結果が得られた。一方, PMC に関しては, VOT 20 ms の条件では潜時 200 ms 付近で PSI が正の値を取る傾向が見られた

が、その他の条件では同様の傾向が見られず、OP4 との機能的結合は確かめられなかった。

## 5.5 実験 2 考察

調音の時間的性質 (VOT) に特徴付けられる有声性を知覚する際の聴覚情報及び調音運動情報の利用に係る脳内メカニズムを検討するために、fMRI で有声性知覚時と生成時の両方で活動が生じることが特定された脳領域の活動を時間解像度に優れた MEG を用いて調べた。聴覚情報の利用については、A1 や PBelt といった聴覚皮質において有声性の知覚と相関する神経活動 (潜時 100 ms 以前の  $\beta$  帯域の神経振動) が見られることが分かった。英語母語話者を対象とした研究では、VOT に対する聴覚応答と有声性の知覚の関連が示唆されていたが、英語母語話者とは有声性の知覚 (VOT 境界) が異なる言語圏の話者 (スペイン語母語話者, フランス語母語話者) を対象とした EEG 研究では両者の関連性は見出されていなかった (Elangovan & Stuart, 2011; Hoonhorst et al., 2009)。日本語母語話者を対象とした本研究では、EEG よりも空間解像度が高い MEG を用いて聴覚皮質活動の信号源解析と神経振動の分析を行なうことで VOT に対する聴覚応答と有声性知覚の関連性を見出した。ただし、有声性の知覚には違いをもたらさない 10 ms や 20 ms といったわずかな VOT の違いであっても、 $\beta$  帯域よりも低い周波数帯域や潜時 100 ms 以降における聴覚皮質の神経活動に違いをもたらすことを考慮すると、聴覚皮質では有声性の知覚に依存した神経活動だけでなく刺激依存性の神経活動も生じていると考えられる。

調音運動情報の利用に関わる脳内メカニズムについて検討を行うために、音声生成に関与する脳領域の PMC や SC, FOP1, 聴覚皮質・運動皮質の両方と相互接続があるとされる OP4 の神経振動を A1 や PBelt と同様の方法で調べた。その結果、PMC, FOP1, PO4 における神経振動には有声性の知覚と相関するような傾向は見られなかったが、

SCでは潜時 100 ms 付近で有声性の知覚と相関する神経振動が見られることが分かった。さらに、聴覚皮質や音声の生成に関与する脳領域と OP4 の間の機能的結合についても検討した。その結果、潜時 100 ms 付近で A1 や PBelt から OP4 へ情報が伝達されることを示唆する結果が得られるとともに、音声生成に関与する脳領域については、潜時 100 ms 及び 200 ms 付近で OP4 から SC へと情報が伝達されることが示唆される結果が得られた。有声性の知覚判断に要する時間が刺激呈示開始後 400-500 ms 程度かかることを考慮すると (Pisoni & Tash, 1974) , 潜時 100-300 ms で生じる OP4 を介した聴覚皮質と SC の間の脳内ネットワーク活動は有声性の知覚に関与すると考えられるだろう。また, A1 や PBelt の神経振動では知覚に違いがない刺激ペアで有意差が認められる潜時・周波数帯が見出されたのに対して SC では有意差が認められる神経振動が見られなかったという結果を考慮すると, SC で生じている神経振動は聴覚皮質で生じている神経振動よりもカテゴリ知覚との相関性が高いと考えられる。

調音位置の違いに特徴付けられる/d/と/g/の音韻対のカテゴリ知覚を扱った Alho et al. (2014, 2016) の MEG 研究では、運動前野の活動とカテゴリ知覚の明確性の間に関連性が見られるとともに、聴覚皮質と運動前野との間で機能的結合が報告されている。一方、有声性の知覚を扱った本研究では、運動前野と知覚結果の関連性は見られず、聴覚皮質と運動皮質と相互接続が存在するとされる OP4 と運動前野の間の機能的結合も明確ではなかった。本研究の結果と先行研究の結果をまとめた上で考察をすると、音声知覚時の調音運動情報の利用に関わる脳内メカニズムは知覚する音韻弁別素性の種類によって異なり、調音位置素性の知覚には舌や軟口蓋などの調音器官の制御を行う運動前野・運動野、有声性の知覚には声帯の時間的制御を行う subcentral area という様に処理を担当する脳領域も異なると考えられる。調音の時間的制御に係る情報の処理については、その他にも廣谷ら (2016) によって発話リズムの自然性の知覚に発話のリズム制御に関与するとされる補足運動野の活動が関与することが示されてい

る。本研究と先行研究の知見を合わせて考察すると、調音運動情報に基づく音声知覚処理は、VOTなどの数ミリから数十ミリ単位の時間スケール、発話リズムなどの数百ミリ秒単位の時間スケールといった様々なスケールの時間情報を利用した音声知覚に関与すると考えられる。

本研究では、MEGを用いた実験で聴覚情報と調音運動情報に基づく音声知覚処理に係るとされている脳領域の両方で有声性のカテゴリ知覚と相関する脳活動が見られ、調音運動情報の利用に係る脳領域の活動の方がよりカテゴリ知覚との相関性が高いという結果が得られた。この結果は調音運動情報に基づく音声知覚処理が聴覚情報に基づく音声知覚処理を補完することでカテゴリ知覚の明確性を高めていることを示唆するものだと考えられる。しかしながら、脳機能計測実験では特定の脳領域における神経活動と知覚・認知処理の間の相関関係については議論が出来るものの、両者の因果関係については明らかに出来ないという課題点がある。そのため、今後はTMSを用いて聴覚皮質や *subcentral area* といった脳領域の活動を抑制あるいは活性化させることで知覚結果にどのような影響が生じるかを調べるための実験を行う必要があるだろう。また、英語母語話者を対象とした研究では2つの音事象の同時性知覚に係る聴覚処理の脳内メカニズムを検討することで有声性知覚にその聴覚処理特性が利用されていることを示唆する結果が示されていた (Simos et al., 1998a,b; Steinschneider et al., 2004)。しかしながら、本研究では日本語母語話者の有声性知覚に逆向性マスクングに関わる聴覚処理特性が密接に関与する可能性については検証出来ていない。そのため、今後は逆向性マスクングの脳内メカニズムを調べて、有声性知覚時の脳内メカニズムとの共通点や相違点について検討を行う必要があるだろう。



## 第6章 本論文のまとめ

本論文での主な検討課題は、音声知覚のメカニズムに関する包括的な検討を行うために、日本語母語話者の有声性知覚過程において聴覚情報及び調音運動情報がどのように利用されているのかを2つの行動実験で調べるとともにそれらの脳内メカニズムを特定することであった。研究1では、日本語母語話者の有声性カテゴリー知覚の不変的な特徴を見出す聴覚処理特性について検討を行った結果、逆向性マスクングに係る聴覚処理の非線形的な性質が日本語母語話者の有声性カテゴリー知覚に密接に関与しており、閉鎖子音における周期成分（母音部）から逆向性のマスクングを受ける雑音成分（子音部）が検出されるか否かが、無声音が知覚されるか有声音が知覚されるかの不変特徴になる可能性が示唆された。この結果は、非線形的な聴覚処理特性によって作り出される知覚的不連続点（natural psychophysical boundary）が音韻カテゴリー知覚の不変的特徴を生み出すという聴覚情報処理の役割を重要視する仮説（Diehl et al., 2007; Holt et al., 2004; Kuhl & Miller et al., 1978; Stevens, 1989）を支持するものだと考えられる。

研究2では、CAF課題を用いて有声性の生成におけるVOTを変化させることで有声性のカテゴリー知覚にどのような影響が生じるかを調べることで、調音運動情報を利用した音声知覚処理が有声性の知覚にどのように関与するかについて検討を行った。その結果、CAF課題で有声音と無声音のVOTの長さの違いが大きくなるのに伴って有声性の知覚が明確になること、そして、両者の変化の間に有意な正の相関があることが分かった。この結果は、有声性をはじめとして音韻の生成と知覚の相関を調べた先行研究（Bailey & Haggard, 1980; Perkell et al., 2004）の知見を補強するものであり、調音運動情報に基づいた音声知覚処理が音韻知覚の明確性に関与することを示唆するものだと考えられる。

研究3では、有声性知覚時の聴覚情報及び調音運動情報の利用に係る脳内メカニズムを検討したところ、聴覚皮質と音声生成に関与する脳領域の間でOP4を介した機能的結合が見られるとともに両方の脳領域において有声性のカテゴリー知覚と相関するような神経活動が見られた。音声生成に関与する脳領域については、主に声帯の時間的な制御の役割を担い、有声性を特徴付けるVOTの制御にも密接に関与すると考えられるsubcentral areaにおいて聴覚皮質との機能的結合と有声性の知覚と相関する神経活動が見られた。調音位置の違いに特徴づけられる音韻対のカテゴリー知覚には運動野など調音位置の制御の役割を担う脳領域の活動が関与することが先行研究では示されているのに対して（Pulvermüller et al., 2006; Chevillet et al., 2013; Möttönen & Watkins, 2009, 2012）、調音の時間的制御の違いに特徴づけられる有声性のカテゴリー知覚には調音の時間的制御の役割を担う脳領域の活動が関与するという研究3で得られた知見は、音声知覚時にその音声信号を生成した調音運動に関する情報を脳内で推定している（Barnaud et al., 2016; Fowler, 1986; Laurent et al., 2017; Liberman et al., 1967; Liberman & Mattingly, 1985; Schwartz et al., 2012）、あるいは、入力音声と脳内で生成した音声の照合をしている（Poeppe et al., 2008; Stevens & Halle, 1967）という音声知覚における音声生成系の役割を重要視する仮説を先行研究の知見と合わせてより強固にする証拠と言えるだろう。

以上の研究結果をまとめて章の冒頭で述べた検討課題について議論を行う。日本語母語話者の有声性知覚における聴覚情報及び調音運動情報の利用については、まず研究3の結果から両方の情報が日本語母語話者の有声性知覚に利用されていることがわかった。さらに、研究1,2の行動実験の結果から、言語圏によって具体的な処理方略が異なる可能性があるものの聴覚情報が有声性知覚の主要な手がかりになること、調音運動情報の利用が有声性カテゴリー知覚の明確性に関与することが示唆された。これらの結果をまとめると、英語母語話者を対象とした研究で示唆されているように調音

運動情報に基づく音声知覚処理が聴覚情報を基にした音声知覚処理を補完することで頑健な音声知覚システムが実現されるというメカニズムが日本語母語話者にも備わっていると考えられる。

最後に、本論文も含め数多くの先行研究では音韻知覚を題材として音声知覚メカニズムの検討が行われているが、今後は日常的な音声コミュニケーション場面で用いられる連続音声を知覚する際に聴覚情報および調音運動情報に基づいた音声知覚処理の役割について検討を進めるべきであろう。近年の脳機能計測研究では連続音声知覚時の聴覚情報および調音運動情報に利用に関する脳内メカニズムの検討が進められており、連続音声を持つ 10 Hz 以下の振幅変調（音節が出現する頻度と概ね一致する）に同期した神経振動が聴覚皮質で生じることやその神経振動の同期性が雑音下での音声知覚成績と関連することが示されている（Giraud & Poeppel, 2012; Howard & Poeppel, 2010; Peele et al., 2012; Doelling et al., 2014）。また、音声の生成に関与する脳領域でも聴覚皮質と同様の同期活動が生じていることや音声生成に関与する脳領域から聴覚皮質へのトップダウン的な処理を表す機能的結合が見られることが明らかにされたことから、音声生成系が入力信号によって誘起される聴覚情報を事前に予測する役割を担っている可能性も示唆されている（Onojima et al., 2017; Park et al., 2015; Rimmele et al., 2018）。これらの先行知見から調音運動情報に基づく音声知覚処理は日常的な音声コミュニケーション場面、特に雑音環境下での連続音声知覚では静寂下で音韻知覚課題を行う場合よりも多様な役割を担う可能性が考えられる。



## 謝辞

本論文の執筆にあたり、修士課程からの5年間ご指導頂きました森周司先生に深く感謝を申し上げます。研究室助教の廣瀬信之先生には、日頃の研究室活動を手厚くサポートしていただくとともに学術雑誌論文を執筆する際には多くのご助言をいただきました。元研究室特任助教で東北学院大学の伊藤一仁先生は研究に関する相談をいつでも聞いてくださり的確なアドバイスをいただきました。その他にも森研究室のスタッフや学生の皆様の助けがあってこの博士論文を執筆出来たと思っております。心より感謝申し上げます。

日本電信電話株式会社コミュニケーション科学基礎研究所の廣谷定男先生には本論文の副査およびアドバイザー委員を引き受けて頂いたのに加え、学会等でも本論文の研究課題に関するアドバイスを数多く頂きました。また、博士課程1年次には研究所でのインターンシップに招いていただき、日本トップクラスの研究環境を体験させていただいたことは大きな刺激になりました。九州大学大学院システム情報科学研究院の冨浦洋一先生、芸術工学研究院のジェラード・レメイン先生には本論文の副査、システム生命科学研究院の伊良皆啓次先生にはアドバイザー委員を引き受けて頂いて、本論文の研究課題に対して貴重なコメントをいただきました。

九州大学医学研究院の光藤崇子先生には、九州大学病院や帆秋病院で実施した脳機能計測実験を手伝っていただきました。また、本論文に関わる研究や他の研究についても数多く議論をさせて頂きました。廣永成人先生、谷口奈美先生には研究3を進める上で数多くの助言をいただくとともに実験や計測データの解析環境を整備するのを手伝って頂きました。金沢大学の木村岳裕先生には高知工科大学にてMRI構造画像の撮像を手伝っていただきました。上記の方々にサポートしていただいたことで研究3を円滑に進めることが出来たと思っております。心から感謝申し上げます。

九州大学病院精神科の鬼塚俊明先生、平野羊嗣先生には帆秋病院でMRI装置を利用する機会を提供していただくとともに来年度から統合失調症の言語性幻聴や聴覚情報処理に関する研究に携わる機会をいただきました。卒業後は、一研究者として精神医学の発展に寄与出来るよう頑張りたいと思います。帆秋病院の帆秋伸彦先生にはfMRI計測実験のためにMRI装置を貸して頂いたのに加え、実験参加者の募集についてもお力を貸して頂きました。また、九州大学病院精神科の中村一太先生や帆秋病院の放射線技師後藤旬平さんには帆秋病院で実施した実験を手伝っていただきました。

本論文の内容とは別の研究で共同研究を進めさせていただいている九州工業大学の水町光徳先生には本論文の研究内容についても貴重なコメントをいただきました。また、お忙しい中時間を割いて卒業後の進路の相談にも乗ってくださいました。人間環境学府の成儒彬さんとは韓国語母語話者の音声知覚研究に関する共同研究をさせていただき、一緒に実験を行って論文も投稿しました。同年代の研究者と議論をしながら研究を進める機会は私にとって非常に刺激的なことでした。

最後に、博士課程に進学をして研究を続けている私を優しく見守ってくれた両親とそばで支え続けてくれた婚約者の大野あゆみさんに心より感謝を申し上げます。



## 文献

1. Ackermann, H., Gräber, S., Hertrich, I., & Daum, I. (1997). Categorical speech perception in cerebellar disorders. *Brain and Language*, *60*(2), 323-331.
2. Agnew, Z. K., McGettigan, C., & Scott, S. K. (2011). Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. *Journal of Cognitive Neuroscience*, *23*(12), 4038-4047.
3. Alho, J., Lin, F. H., Sato, M., Tiitinen, H., Sams, M., & Jääskeläinen, I. P. (2014). Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. *Frontiers in Psychology*, *5*(394), 1-10.
4. Alho, J., Green, B. M., May, P. J., Sams, M., Tiitinen, H., Rauschecker, J. P., & Jääskeläinen, I. P. (2016). Early-latency categorical speech sound representations in the left inferior frontal gyrus. *NeuroImage*, *129*, 214-223.
5. Bailey, P. J., & Haggard, M. P. (1980). Perception-production relations in the voicing contrast for initial stops in 3-year-olds. *Phonetica*, *37*(5-6), 377-396.
6. Barnaud, M. L., Bessière, P., Diard, J., & Schwartz, J. L. (2018). Reanalyzing neurocognitive data on the role of the motor system in speech perception within COSMO, a Bayesian perceptuo-motor model of speech communication. *Brain and Language*, *187*, 19-32.
7. Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex*, *13*(1), 85-95.
8. Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: an fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, *17*(9), 1353-1366.

9. Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, 66(4), 1001-1017.
10. Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *The Journal of the Acoustical Society of America*, 67(2), 648-662.
11. Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13(11), 1428-1432.
12. Chao, S. C., Ochoa, D., & Daliri, A. (2019). Production variability and categorical perception of vowels are strongly linked. *Frontiers in Human Neuroscience*, 13(96), 1-9.
13. Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, 18(12), 2844-2854.
14. Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *Journal of Neuroscience*, 33(12), 5208-5215.
15. Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26(1), 55-67.
16. D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381-385.
17. Diehl, R. L. (2007). Acoustic and auditory phonetics: the adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 965-978.

18. Diehl, R. L., Lindblom, B., & Creeger, C. P. (2003, August). Increasing realism of auditory representations yields further insights into vowel phonetics. In *Proceedings of the 15th International Congress of Phonetic Sciences* (Vol. 2, pp. 1381-1384).
19. Denes, P. B., & Pinson, E. N. (1993). *The speech chain* (2nd ed.). New York: W. H. Freeman and Co.
20. Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761-768.
21. Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nature Communications*, *7*, 12241.
22. Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*(1), 99-109.
23. Elangovan, S., & Stuart, A. (2008). Natural boundaries in gap detection are related to categorical perception of stop consonants. *Ear and Hearing*, *29*(5), 761-774.
24. Elangovan, S., & Stuart, A. (2011). A cross-linguistic examination of cortical auditory evoked potentials for a categorical voicing contrast. *Neuroscience Letters*, *490*(2), 140-144.
25. Elliott, L. L. (1971). Backward and forward masking. *Audiology*, *10*(2), 65-76.
26. Evans, S., & Davis, M. H. (2015). Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. *Cerebral Cortex*, *25*(12), 4772-4788.
27. Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, *43*(4), 155-171.
28. Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception

- and production by Dutch speakers of English. *Speech Communication*, 6(3), 185-202.
29. Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1), 3-28.
  30. Fowler, C. A., Shankweiler, D., & Studdert-Kennedy, M. (2016). "Perception of the speech code" revisited: Speech is alphabetic after all. *Psychological Review*, 123(2), 125-150.
  31. Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... & Smith, S. M. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536, 171-178.
  32. Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511-517.
  33. Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 267.
  34. Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, 86, 446-460.
  35. Greenberg, S., & Christiansen, T. U. (2019). The perceptual flow of phonetic information. *Attention, Perception, & Psychophysics*, 81(4), 1-13.
  36. Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., ... & Parkkonen, L. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, 65(15), 349-363.
  37. Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1), 35-42.
  38. Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M.

- R., ... & Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, 7(3), 213-223.
39. Hamalainen, M. S., & Sarvas, J. (1989). Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Transactions on Biomedical Engineering*, 36(2), 165-171.
40. Hay, J., & Garcia-Sierra, A. (2005). Effects of linguistic experience on perception and learnability of non-speech categories. In *ISCA Workshop on Plasticity in Speech Perception*, 223-226.
41. Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.
42. Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.
43. Holt, L. L., Lotto, A. J., & Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *The Journal of the Acoustical Society of America*, 116(3), 1763-1773.
44. Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213-1216.
45. Houde, J. F., & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech I. *Journal of Speech, Language, and Hearing Research*, 45(2), 295-310.
46. Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, 104(5), 2500-2511.
47. Ivry, R. (1993). Cerebellar Involvement in the Explicit Representation of Temporal Information. *Annals of the New York Academy of Sciences*, 682(1), 214-230.

48. Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F 0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108(3), 1246-1251.
49. Jones, J. A., & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Current Biology*, 15(19), 1768-1772.
50. Kawahara, H. (1994). Interactions between speech production and perception under auditory feedback perturbations on fundamental frequencies. *Journal of the Acoustical Society of Japan*, 15(3), 201-202.
51. Kleinschmidt, D. F., & Jaeger, T. F. (2016). Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning?. *Psychonomic Bulletin & Review*, 23(3), 678-691.
52. Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69-72.
53. Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America*, 63(3), 905-917.
54. Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *Journal of Neuroscience*, 34(31), 10339-10346.
55. Laurent, R., Barnaud, M. L., Schwartz, J. L., Bessière, P., & Diard, J. (2017). The complementary roles of auditory and motor information evaluated in a Bayesian perceptuo-motor model of speech perception. *Psychological Review*, 124(5), 572-602.
56. Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36.

57. Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431-461.
58. Lisker, L. (1975). Is it VOT or a first-formant transition detector?. *The Journal of the Acoustical Society of America*, 57(6), 1547-1551.
59. Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384-422.
60. Lisker, L., & Abramson, A. S. (1970). The Voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Science* (Vol. 563, pp. 563-557). Prague, Czech Republic: Academia Prague.
61. Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. (1977). On pushing the voice-onset-time (VOT) boundary about. *Language and Speech*, 20(3), 209-216.
62. Massaro, D. W. (1973). A comparison of forward versus backward recognition masking. *Journal of Experimental Psychology*, 100(2), 434-436.
63. Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177-190.
64. McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
65. Medina, V., Hoonhorst, I., Bogliotti, C., & Serniclaes, W. (2010). Development of voicing perception in French: Comparing adults, adolescents, and children. *Journal of Phonetics*, 38(4), 493-503.
66. Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise–buzz sequences with varying noise-lead times: An

- example of categorical perception. *The Journal of the Acoustical Society of America*, 60(2), 410-417.
67. Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *The Journal of the Acoustical Society of America*, 73(6), 2124-2133.
68. Mitsuya, T., MacDonald, E. N., & Munhall, K. G. (2014). Temporal control and compensation for perturbed voicing feedback. *The Journal of the Acoustical Society of America*, 135(5), 2986-2994.
69. Molins, A., Stufflebeam, S. M., Brown, E. N., & Hämäläinen, M. S. (2008). Quantification of the benefit from integrating MEG and EEG data in minimum  $\ell_2$ -norm estimation. *NeuroImage*, 42(3), 1069-1077.
70. Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, 29(31), 9819-9825.
71. Möttönen, R., & Watkins, K. E. (2012). Using TMS to study the role of the articulatory motor system in speech perception. *Aphasiology*, 26(9), 1103-1118.
72. Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, 20(7), 895-903.
73. Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, 113(5), 2850-2860.
74. Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, 115(10), 2292-2307.

75. Nolte, G., Ziehe, A., Nikulin, V. V., Schlögl, A., Krämer, N., Brismar, T., & Müller, K. R. (2008). Robustly estimating the flow direction of information in complex physical systems. *Physical Review Letters*, *100*(23), 234101.
76. Ogasawara, N. (2011). Acoustic analysis of voice-onset time in Taiwan Mandarin and Japanese. *Concentric: Studies in Linguistics*, *37*(2), 155-178.
77. Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868-9872.
78. Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, *89*(13), 5951-5955.
79. Onojima, T., Kitajo, K., & Mizuhara, H. (2017). Ongoing slow oscillatory phase modulates speech intelligibility in cooperation with motor cortical activity. *PloS One*, *12*(8), e0183146.
80. Patri, J. F., Perrier, P., Schwartz, J. L., & Diard, J. (2018). What drives the perceptual change resulting from speech motor adaptation? Evaluation of hypotheses in a Bayesian modeling framework. *PLoS Computational Biology*, *14*(1), 1-38.
81. Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*(12), 1649-1653.
82. Parker, E. M. (1988). Auditory constraints on the perception of voice-onset time: The influence of lower tone frequency on judgments of tone-onset simultaneity. *The Journal of the Acoustical Society of America*, *83*(4), 1597-1607.

83. Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, 8(57), 1-14.
84. Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in Neuroscience*, 8(253), 1-13.
85. Peelle, J. E., Gross, J., & Davis, M. H. (2012). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6), 1378-1387.
86. Perrachione, T. K., & Ghosh, S. S. (2013). Optimized design and analysis of sparse-sampling fMRI experiments. *Frontiers in Neuroscience*, 7(55), 1-18.
87. Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, 116(4), 2338-2344.
88. Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *The Journal of the Acoustical Society of America*, 61(5), 1352-1361.
89. Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2), 285-290.
90. Poeppel, D., & Monahan, P. J. (2008). Speech perception: Cognitive foundations and cortical implementation. *Current Directions in Psychological Science*, 17(2), 80-85.
91. Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, 103(20), 7865-7870.

92. Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, 271(1-2), 16-25.
93. Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718-724.
94. Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, 22(2), 173-189.
95. Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, 22(10), 870-882.
96. Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, 35(3), 439-443.
97. Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), 131-141.
98. Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, 31(4), 307-314.
99. Schaefer, R. S., Morcom, A. M., Roberts, N., & Overy, K. (2014). Moving to music: effects of heard and imagined musical cues on movement-related brain activity. *Frontiers in Human Neuroscience*, 8(774), 1-11.
100. Schuerman, W. L., Nagarajan, S., McQueen, J. M., & Houde, J. (2017). Sensorimotor adaptation affects perceptual compensation for coarticulation. *The Journal of the Acoustical Society of America*, 141(4), 2693-2704.
101. Schuerman, W. L., Meyer, A. S., & McQueen, J. M. (2017). Mapping the speech code: cortical responses linking the perception and production of vowels. *Frontiers in Human Neuroscience*, 11(161), 1-16.

102. Schwartz, J. L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5), 336-354.
103. Scott, S. K., & Johnsruide, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26(2), 100-107.
104. Sepulcre, J. (2013). An OP4 functional stream in the language-related neuroarchitecture. *Cerebral Cortex*, 25(3), 658-666.
105. Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, 106(2), 1078-1083.
106. Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, 125(2), 1103-1113.
107. Shimizu, K. (1977). Voicing Features in the Perception and Production of Stop Consonants by Japanese Speakers. *Studia Phonologica*, 11, 25-34.
108. Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., & Papanicolaou, A. C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cognitive Brain Research*, 7(2), 215-219.
109. Simos, P. G., Breier, J. I., Zouridakis, G., & Papanicolaou, A. C. (1998). MEG correlates of categorical-like temporal cue perception in humans. *Neuroreport*, 9(11), 2475-2479.
110. Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., & Howard III, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials

- recorded directly from human auditory cortex. *Journal of Neurophysiology*, 82(5), 2346-2357.
111. Steinschneider, M., Fishman, Y. I., Volkov, I. O., & Howard III, M. A. (2004). Physiological evidence for a temporal processing mechanism underlying voice-onset time (VOT) encoding. *The Journal of the Acoustical Society of America*, 115(5), 2466-2466.
112. Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.
113. Stevens, K. N., Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In W. Walther-Dunn (ed.), *Models for the perception of speech and visual form* (pp. 88-102). Massachusetts, The MIT Press.
114. Summerfield, Q. (1982). Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops. *The Journal of the Acoustical Society of America*, 72(1), 51-61.
115. Szenkovits, G., Peelle, J. E., Norris, D., & Davis, M. H. (2012). Individual differences in premotor and motor recruitment during speech perception. *Neuropsychologia*, 50(7), 1380-1392.
116. Tamura, S., Ito, K., Hirose, N., & Mori, S. (2018). Psychophysical boundary for categorization of voiced–voiceless stop consonants in native Japanese speakers. *Journal of Speech, Language, and Hearing Research*, 61(3), 789-796.
117. Tamura, S., Ito, K., Hirose, N., & Mori, S. (2019). Precision of voicing perceptual identification is altered in association with voice-onset time production changes. *Experimental Brain Research*, 237(9), 1-8.
118. Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine & Biology*, 51(7),

- 1759.
119. Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., & Garnsey, S. M. (2018). The time-course of cortical responses to speech revealed by fast optical imaging. *Brain and Language, 184*, 32-42.
120. Toyoda, G., Brown, E. C., Matsuzaki, N., Kojima, K., Nishida, M., & Asano, E. (2014). Electrocorticographic correlates of overt articulation of 44 English phonemes: intracranial recording in children with focal epilepsy. *Clinical Neurophysiology, 125*(6), 1129-1137.
121. Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics, 21*(4), 289-297.
122. Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*(7), 701-702.
123. Van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., ... & Desain, P. (2009). The brain-computer interface cycle. *Journal of Neural Engineering, 6*(4), 041001.
124. Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America, 122*(4), 2306-2319.
125. 笥一彦. (1995). 音声知覚のモデル. *行動計量学, 22*(1), 30-39.
126. 廣谷定男. (2017). 「聞くと話す」の脳科学. *日本音響学会誌, 73*(8), 509-516.
127. 廣谷定男, Jasmin, K., Evans, S., Krisnan, S., Lima, C., Ostarek, M., Boebinger, D., & Scott, S. K. (2016). 非母語発話リズム音声聴取時における脳機能計測. *日本音響学会春季研究発表会講演論文集, 433-434*.