
音声コミュニケーションに おける知覚と生成の 相互作用に関する研究

16300053

平成16年度～平成18年度科学研究費補助金
(基盤研究(B)) 研究成果報告書

平成19年6月

研究代表者 赤木正人
北陸先端科学技術大学院大学 情報科学研究科 教授

<はしがき>

本研究は、音声生成と音声知覚の密接な関係を示す一例として「聴覚フィードバック」を取り上げ、知覚・生成の相互作用の解明を図ることを目的としている。実験では、聴覚フィードバックに摂動を与え、摂動が音声生成系にどのように影響を与えるかについて測定した。具体的には、音声資料として、中国語母音対[i]-[y]と日本語母音対[e]-[a], [e]-[i], [e]-[u]を採用し、発話者が母音対の最初の母音を継続的に発話している途中で、摂動として2番目の母音を聴覚にフィードバックさせた。そして、発話音声のホルマント周波数、発話に関わる筋肉の筋電（EMG）および舌運動（EMA）の変化を測定した。測定の結果、

1. 被験者の摂動への反応は摂動の方向と反対方向であり、発話において摂動に対する補正が行われていることが確認できた。
2. ホルマント周波数分析から、第1, 第2ホルマントにおいて明確な補償動作が観測された。これは、スペクトルに関する聴覚フィードバックにおいて短時間での補正反応を捕らえた初めての結果である。
3. EMG および EMA の分析結果から、摂動に対してこれを補償するような筋肉および舌の動きが観測された。
4. 補償動作は、摂動開始から約 150 ms で始まり、290 ms で最大値に到達した。

これらの観測結果から、聴覚フィードバックは発話時に頻繁に起きていること、そして、発話者はリアルタイムで自身の発話をモニタリングしており、摂動によって生じた「誤り」を補償しようとしていることが示唆された。引き続き、詳細なデータを収集中である。

<Abstract>

This study employed an auditory feedback paradigm with perturbed fed-back speech to investigate interaction between speech perception and production by measuring simultaneous fluctuations of speech production organs using the electromyographic (EMG) signals, articulatory movements, as well as spectral analyses, where the articulatory data were obtained by the electromagnetic articulographic (EMA) system. Chinese vowels pair [i]-[y] and Japanese vowels pairs [e]-[a], [e]-[i] and [e]-[u] were chosen as the experimental objects. When the speaker is maintaining the first vowel, the feedback sound is randomly changed from the first vowel to the second one in each pair by manipulating the first three formants. Spectral analysis showed that a clear compensation was seen in the first and second formants of the vowels. Analyses of EMG and EMA signals also showed muscle reactivation and tongue movements to compensate for the perturbations. Latency of the compensating response is about 150 ms to start and about 290 ms for maximum compensation from the onset of the perturbation. According to the measurements, it seems that in most cases the speaker attempts to compensate for the "error" caused by the auditory perturbation by a real-time monitoring, and the auditory feedback takes place simultaneously often during speech production.

<研究組織>

研究代表者	赤木正人	北陸先端科学技術大学院大学	情報科学研究科	教授
研究分担者	党 建武	北陸先端科学技術大学院大学	情報科学研究科	教授
	鵜木祐史	北陸先端科学技術大学院大学	情報科学研究科	助教授
	Lu Xugang	北陸先端科学技術大学院大学	情報科学研究科	助手

<交付決定額（配分額）>

年度	直接経費	間接経費	総額
平成 16 年度	8,300,000	0	8,300,000
平成 17 年度	3,000,000	0	3,000,000
平成 18 年度	1,400,000	0	1,400,000
計	12,700,000	0	12,700,000

<研究発表>

- [1] 松岡, Lu, 党, 赤木(2004). “調音情報を考慮した聴覚系と発話系の相互作用に関する検討”, 音響学会聴覚研究会資料、H-2004-103.
- [2] Matsuoka, R., Lu, X., Dang, J., and Akagi, M. (2004). “Investigation of interaction between speech perception and speech production,” Proc. KIT Int. Sympto. Brain and Language 2004, 27-28.
- [3] 松岡, Lu, 野田, 党, 赤木(2005). ”聴覚系と発話系の相互作用に関する検討”, 平成 17 年春季音響学会講演論文、3-P-33.
- [4] Dang, J., Akagi, M., and Honda, K. (2006). "Communication between speech production and perception within the brain - Observation and simulation," J. Comp. Sci. & Tech., 21, 1, 95-105.
- [5] 田中, Lu, 党, 赤木(2006). ”変形聴覚フィードバックにおける摂動量と補正動作の関係について”, 平成 18 年春季音響学会講演論文、2-3-11
- [6] 内山田, Lu, 党, 赤木(2006). "変形聴覚フィードバックによる発話系の補償動作について", 平成 18 年秋季音響学会講演論文、3-P-9.
- [7] Akagi, M., Dang, J., Lu, X., and Uchiyamada, T. (2006). "Investigation of interaction between speech perception and production using auditory feedback," J. Acoust. Soc. Am., 120, 5, Pt. 2, 3253.
- [8] 内山田, Lu, 党, 赤木(2007). ”変形聴覚フィードバックに対する発話系の反応の計測”, 平成 19 年春季音響学会講演論文、2-Q-18

成果報告

<目次>

I. 序論	1
1. 1 研究背景	3
1. 2 研究目的・手法	3
1. 3 音声知覚・生成過程に関する研究	3
1. 4 計測システムの構成	8
1. 5 測定対象	10
II. 実験1：音響計測	11
2. 1 予備実験	12
2. 2 聴覚フィードバック実験における摂動量と補正反応の関係	16
III. 実験2：筋電計測	28
3. 1 生理学的観測を取り入れた実験方法	29
3. 2 予備実験	32
3. 3 本実験	36
3. 4 考察	39
3. 5 実験2－筋電計測－のまとめ	41
IV. 実験3：舌運動計測	42
4. 1 調音器官の補償動作の測定	43
4. 2 考察	53
4. 3 まとめ	53
V. 結言	54
5. 1 本研究のまとめ	55
5. 2 今後の課題	56
参考文献	58
成果資料	60

I . 序論

<要旨>

音声知覚・生成は、音声によるマン・マンコミュニケーションの根幹を成すものである。また、マン・マシンコミュニケーションにおいても、ヒトの音声生成・知覚機構を基礎として、これを工学的に実現した音声合成・認識が重要な役割を果たそうとしている。本来、音声知覚・生成は、音声コミュニケーションにおいて表裏一体を成すものであり、コミュニケーションを円滑に保つためには双方が一体となって働く必要がある。また、音声知覚・生成をつかさどる脳内の場所は同じ言語野にあり、密接な関係があるものと想像される。そのため、音声知覚・生成を総合的に研究する意義はますます増大している。ところが、今までの研究の流れを見ると、音声生成は主に発声機構の生理学的見地から、音声知覚は心理学的見地からそれぞれ独立に研究されてきたため、個別の機能解明はできても、相互的な機能解明に至らない可能性が非常に高い。本研究では、知覚・生成を個別に研究対象とするだけでなく、これらの相互作用も含め、音声コミュニケーションの中での音声知覚・生成の総合的な解明を目指す。

このため、本研究では、音声生成と音声知覚の密接な関係を示す一例として「聴覚フィードバック」を取り上げ、知覚・生成の相互作用の解明を図ることを目的とする。「聴覚フィードバック」とは、話者が発した声を自分の耳で聞き取りながら発話機構を制御するための音声によるフィードバックのことであり、これを用いて話者は正確な発話を行っている。「聴覚フィードバック」が発話に重要な役割を演じているのならば、フィードバックされる音声を人為的にリアルタイムで変形し耳に呈示した場合、その変形によって音声生成に何らかの影響があるはずである。影響は、音声知覚機構、音声生成タスクプランニング機構、音声生成機構（調音器官）の順に伝播する。本研究では、リアルタイムで変形されたフィードバック音声の変形量と生理的指標の関係および生成された音声との関係を調べることで、音声知覚とタスクプランニングも含めた音声生成、およびこれらの相互作用を議論する。

このために、以下の項目について研究を行う。

- (1) 発話者の音声をリアルタイムで分析、特徴変形、再合成が行える装置の構築、
- (2) この装置によって変形された音声を発話者にフィードバックするシステムの構築、
- (3) 発話者の音声の特徴のみならず、調音器官の動き、筋活動、呼気流などの生理的指標を同時に計測できるシステムの構築、および
- (4) 変形聴覚フィードバックによる音声生成への影響および音声生成と知覚との関連について、申請者が構築した知覚・生成モデルを用いての考察、さらには知覚・生成相互機構の新たなモデル化とその応用可能性の検討

これらの検討の結果、実験装置を作成し、次の測定結果を得た。

1. 被験者の摂動への反応は摂動の方向と反対方向であり、発話において摂動に対する補正が行われていることが確認できた。
2. ホルメント周波数分析から、第1、第2ホルメントにおいて明確な補償動作が観測された。これは、スペクトルに関する聴覚フィードバックにおいて短時間での補正反応を捕らえた初めての結果である。
3. EMG および EMA の分析結果から、摂動に対してこれを補償するような筋肉および舌の動きが観測された。
4. 補償動作は、摂動開始から約 150 ms で始まり、290 ms で最大値に到達した。

しかし、まだモデル化までには至っていない。今後の検討課題である。

1. 1 研究背景

音声生成と知覚は表裏一体の関係にあり、多様な音韻を生成するためには多様な音韻を聞き分ける能力が不可欠である。このため人間は、音声生成過程とその逆過程の音声知覚の繰り返しにより母国語を含む言語音声知覚機構の発達を遂げた。このような発達には音声の生成と知覚とを密接に結ぶ情報交換の通路（ことばの鎖）が大きく関与している[1]。このような”ことばの鎖”について話者内部（脳内）での音声生成と音声知覚との相互交信を説明するため、Lieberman らは 1960 年代より音声知覚の運動理論（Motor theory of speech perception）を提唱してきた[2][3]。しかし、話者内部（脳内）において、音声生成・知覚がどのように関連し存在しているのかについてはまだ明らかにされていない。

これまでの先行研究では、発話時の音声生成・知覚の相互作用を明らかにするために、聴覚フィードバックに着目した研究が数多く行われてきた。その中で Lombard は雑音環境下では通常の会話より発話音声が大きくなり、基本周波数も高くなる現象を観測した[4][5]。その現象は Lombard 効果として知られている。また、発話音声を遅延して話者に聞かせる遅延聴覚フィードバック実験（DAF）では、吃音や発話速度が遅くなる等の現象が生じる[6]。これは、聴覚フィードバックの効果を表す顕著な例である。しかし、上記の報告は、定性的な性質を述べるにとどまっており、これらの現象を起こすメカニズムについての説明は不十分である。また、DAF のような発話過程を破壊するような実験では、発話過程自体が破綻してしまうため、発話への影響に対する定量的な分析は困難である。

そこで河原らは、非破壊的な実験パラダイムとして変形聴覚フィードバック（TAF）を提案した。河原らの一連の研究により、聴覚系が発話系の基本周波数制御に重要な役割を果たすことが明らかとなったが、ホルマントの影響については明らかにされなかった[7][8][9]。このため、佐藤と斉藤が、日本母音に関して短時間のホルマント変形に対する発話器官の補償動作について、音声のスペクトルグラムを用いて検証を行った。その結果、スペクトルの変動は見られたが、それが補償動作か否かは確認できなかった。

これらの先行研究における知見より、基本周波数やホルマント周波数の変化を含めたスペクトル形状の変形により、補償動作が存在する可能性が示されている。しかし、これらの研究は、音声スペクトルを分析するのみにとどまっており、調音器官の動作については観測が行われてはいない。このため、音声スペクトルにははっきりとあらわれないが、調音器官がホルマント形状の変形を補うような方向で補償をおこなっている可能性が考えられる。

1. 2 研究目的・手法

以上のような背景から考えられる可能性を明らかにするため、本研究では、フィードバック時に存在可能である瞬時的な調音器官の補償動作や応答特性を、音声信号以外の測定手法を加えて明らかにしていく。具体的には音声信号に加え、筋電信号（EMG）、および、磁気センサシステム（EMA）を用いて調音運動からの観測も同時に行う。このように、本研究では、音声知覚の面からだけでなく調音運動のレベルでも考察できるような音声以外の計測手法を導入する。このような手法から、各々の調音器官の調音運動における補償動作を定量的に計測する。

また、発声・発話動作における聴覚フィードバックの影響に関して、基本周波数においては様々な知見が得られているが、ホルマントの変形において報告された例は少ない。そこで、本研究では、発話音声のホルマントを変形したフィードバック音声を用い、短時間の変形聴覚フィードバック実験における発話器官の逐次的な応答特性を観測する。

以上のような手法を用いて、各発話期間の調音活動を定量的に測定することにより、聴覚フィードバックにおける補償動作の存在を明らかにし、聴覚系と発話系の相互作用を定量化することが本研究の目的である。

1. 3 音声知覚・生成過程に関する研究

音声知覚と生成過程は、人と人との音声コミュニケーションにおいて重要な機能を果たすものであり、コミュニケーションを円滑に保つためには両方の過程が一体となって働く必要がある。音声認識・合成は、人間の音声知覚・生成機構の工学的な実現で、人と機械とのコミュニケーションのインターフェイスとして期待されている。

近年、音声認識・合成は、コンピュータの発展と共に大きな進歩を遂げ、理想的な環境で応用できるレベルに近づいた。音声認識に関して言えば、現在用いられている音声認識方法は、理想的な環境ではほとんど行き着くところまで行き、次のステップとして、どのような環境においても高性能な音声認識を行う手法の開発が課題となってきている。ところが、現在の手法では、あらゆる環境で高性能な音声認識を行うためには、あらゆる環境を想定しあらゆる実例を用意することが必要である。例えば、現在の音声認識システムは、雑音環境では認識率が著しく低下し、ほとんど無能に近い。それに対して、人間の音声認識能力は雑音環境にほとんど左右されない。

音声は本来人間により認知され人間により生成されるものである。“音（音声）を聴くとはどういうことであるのか”という原点にもう一度立ち帰り、音声知覚と音声生成という人間本来のメカニズムを解明することは、音声認識・合成の発展に大きく貢献すると考えられる。

1. 3. 1 ことばの鎖 ～Speech Chain～

音声は本来人間により生成され人間により認知されるものである。人間は、音声生成過程とその逆過程の音声知覚の繰り返しにより母国語を含む言語音声を学習している。このような学習により音声の生成と知覚とを密接に結ぶ情報交換の通路（ことばの鎖）ができたのである[1]。“音（音声）を聴くとはどういうことであるのか”を解明するためには、“ことばの鎖”の中での知覚機構を考える必要があり、また同様に、“音声を生成するとはどういうことであるのか”を解明するためには、生成機構を考える必要がある。

“ことばの鎖 ～Speech Chain～”とは、図 1.1 にあるように、言語学的段階から生理学的段階を経て音響学的段階へと移行し、また逆の過程をたどる、人間同士がコミュニケーションを円滑に行うためのサイクルのことをいう。つまり、人間が相互にコミュニケーションを行う場合、言葉を発して相手に自分の考え、感情などを伝えようとする一方で、相手が伝えてきた考え、感情などの情報を受け取り、理解して、そして、適切な応答を行うサイクルのことである。自分自身の中でこのサイクルが上手くまわることにより、コミュニケーションが円滑に保たれている。

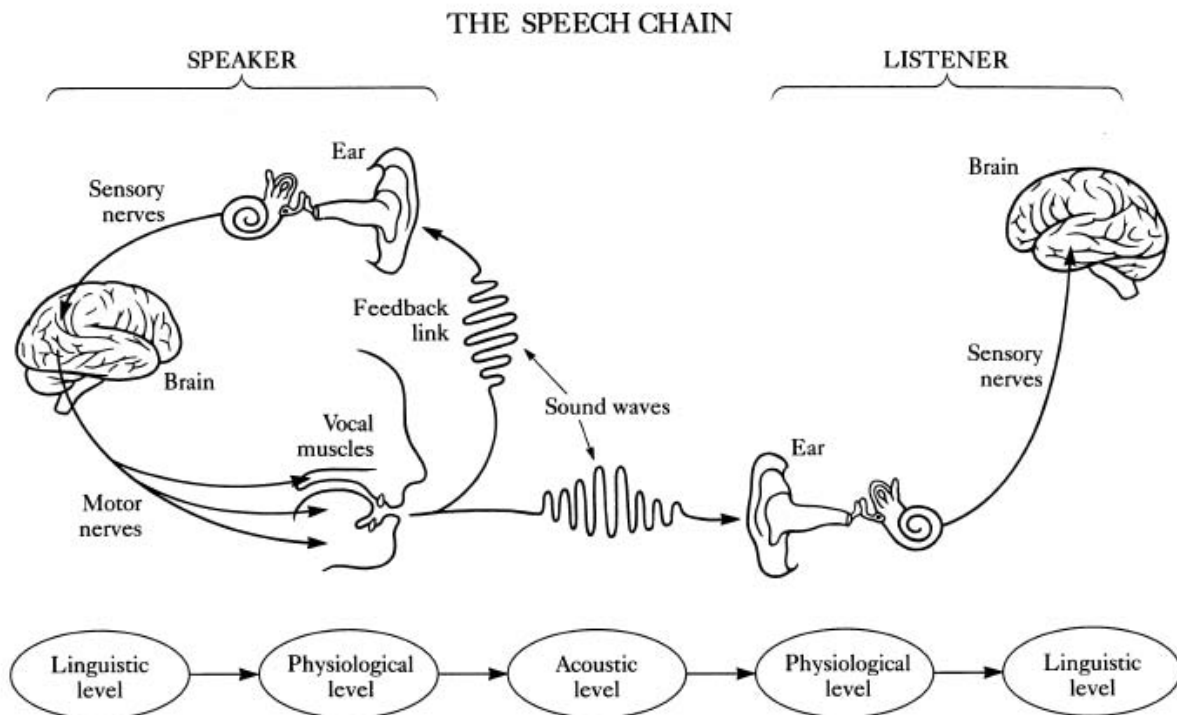


図 1.1 ことばの鎖～Speech Chain～（文献[10]より引用）

会話音声の場合、発話速度の変化や感情などの影響と、発話器官の生理学的な制約により生成された

音声は変形され、最終的には音声波形と音素系列とのはっきりした対応関係は見られなくなる。しかし、人間には音声の生成と知覚との情報交換通路、つまり、“ことばの鎖”が存在しているため、音声と音素列の対応関係を正しく見つけ出すことが可能である。ところが、“ことばの鎖”についての話者の内部（脳内）での存在実態はまだ明らかにされてはいない。

人間が音声を発する時にまず行うことは、話したい内容を表現する語句を選択し、それを文法的に正しい言語形式に変換することである。次に、脳から発話器官（舌や口唇など）に運動神経指令が出され、発話器官の筋肉が動いて、空気振動としての音声を作られる。生じた音声信号は、空気中や伝送路（電話回線など）を伝わって相手の耳に届き、今度は聞き手の聴覚器官（耳）によって神経信号として相手の脳に伝えられる。そしてその内容が理解されて初めて音声によるコミュニケーションが成立する。

音声を話す際には、話者自身の耳にもその信号がフィードバックとして伝わる。このフィードバックは非常に重要であり、話し手は絶えず発話器官を制御することができる。このような、話者が発した声を自分の耳で聞き取りながら発話機構を制御するための音声によるフィードバックのことを“聴覚フィードバック”と呼ぶ。自分の音声が入るフィードバックされた時、聴覚の神経、脳の神経を経過して自分の発声・発話器官の筋肉をコントロールするという力が人間には備わっている。聴覚障害とは、この“ことばの鎖～Speech Chain～”の輪が切断されてしまうことである。聞くことを通して、自分の音声をコントロールし、発話方法を習得する“ことばの鎖”が切断されれば、これらの活動は困難になる。

このように我々健常者に備わっている音声聴覚システム（ことばの鎖）のメカニズムを解明することは、上記に述べた工学的な場面だけでなく、医学的な分野、例えば障害を有する人の聴覚の仕組みを解明する分野などにも多くの知見をもたらすだろう。

1. 3. 2 聴覚フィードバック

“ことばの鎖”の実態を調べるため、これまで音声知覚と生成の相互作用に関して様々な実験を通して仮説が提案されてきた。

聴覚系と発話系の相互作用の存在を示すものの一つとして、“聴覚フィードバック”がある。“聴覚フィードバック”は発話音声を聴覚系にフィードバックしながら発話動作の制御を行うもので、この機能により正常な発話が可能となっていると考えられている。もし聴覚フィードバックが発話過程で常に働いているなら、フィードバック音声を人為的に変形し耳に呈示した場合、その変形によって発話系に何らかの影響があるはずである。このような前提をもとに、古くから発話時における“聴覚フィードバック”の役割に関する実験が数多く行われてきた。

Lombard らは、初期の研究において、雑音によって発話音声をマスクした環境下でも発話が可能であることを示している[4][5]。このことは、言語習得後に聴覚を失った後天性難聴者においては明瞭な発話が維持されること[11][12]と一致する。また、言語獲得前に聴力を失った子供においては徹底した訓練を施しても正確な発話を身に着けることが困難であることが報告されている[13][14][15]。これらの知見は聴覚フィードバックが言語獲得過程においては極めて重要な役割を果たすが、それ以降はあまり関与しないことを示す根拠となっている。しかし、Lane らは雑音環境下においても発話は可能であるが、発話者の発話音量を雑音の音量に応じて増大させている（Lombard 効果）ことも示している[5]。また言語習得後においても聴力を失うと、/s/のような無声摩擦子音の生成や基本周波数の制御が急速に悪化すること[11][16]や、発話音声に数十～数百 ms の遅延を挿入する（遅延聴覚フィードバック：DAF）ことで発話動作を著しく崩壊させることから[6]、聴覚フィードバックが音声生成に必ずしも決定的に関与するとは考えにくい、いくつかの知見は少なくとも正確に発話をするうえで必要であることを示している。

1. 3. 3 変形聴覚フィードバック実験

これまで、上記で述べた“聴覚フィードバック”に関する実験を通じて、音声知覚と生成過程が相互にどのように関連しているのかを調べる研究が、様々な手法を用いて行われてきた。聴覚からの入力と音声生成過程及び言語習得過程での発話動作に関連があるならば、聴覚からの入力に変化が起こった場合、発話動作にも変化が現れることが予想される。変形聴覚フィードバックとは、聴覚への入力に対して変化（摂動）を与えることで発話動作に現れる変化を測定する実験方法である。聴覚と発話動作との関連を調べる方法として変形聴覚フィードバックによる実験が用いられてきた。Lombard ら[4][5]による雑音で発話音声を遮断する実験や Lee[6]による発話音声に遅延時間を挿入する実験（DAF）などは変形

聴覚フィードバックの一種である。

しかし、これらの実験による報告は、聴覚フィードバックの定性的な性質を述べるにとどまっており、これらの現象を起こすメカニズムについての説明は不十分である。特に、DAFのような発話過程を破壊するような実験では、発話過程自体が破綻してしまうため、発話への影響に対する定量的な分析は困難である。

そこで、河原らは、音響パラメータを実時間で変換することにより非破壊的で定量的な分析を可能にする変形聴覚フィードバック (Transformed Auditory Feedback : TAF) を提案した。河原らは発声過程から遅延聴覚フィードバックに伴う発話への影響を避けて、正常な発話に近い状態で聴覚と発話との関係を探ろうとし、実時間で基本周波数に微小な摂動を与えた。この実験により発声された音声の基本周波数の変化による影響として摂動を与えてから約 150ms 程度の遅れを伴い、変化を打ち消す方向への応答が働くことが示された。このような河原らの一連の研究により [7][8][9]、基本周波数制御における聴覚フィードバックの役割を示す重要な知見が得られた。

一方、スペクトルの変動に対しての聴覚フィードバック実験も、様々な手法を用いて行われている。Houde と Jordan は、発話音声のホルマント周波数を変化させた変形聴覚フィードバックの環境下で被験者に長時間発話させる実験を行った [17][18]。その結果、話者はホルマントの変化に対して、発話動作を調整、適応化させていることが示された。現在の所、聴覚フィードバックと発話動作との関わりは、リアルタイムのフィードバック機能というより、むしろ発話動作の結果として起こる音響現象をモニタし、発話者が意図した音響現象と異なる場合に発話動作を調整 (適応化) する機能、あるいは発話動作の運動計画を行う際に用いられる発話動作と音響現象との因果関係を学習によって獲得する機能にあるとされている [16][19]。さらに、Houde らは、発話動作が適応化した後、マスキングノイズで聴覚フィードバックを遮断した状態においても発話音声のホルマント周波数に同様な補償的な変化が見られたと報告している [18]。この結果は、聴覚フィードバック条件下の方が補償効果が大きいこと、変形聴覚フィードバックに対する学習効果以外にも、瞬時的な発話動作の調整がなされているという可能性を示唆している。しかし、本来、このような長時間学習の実験には、人間に備わっている音声知覚・生成メカニズム解明に焦点はあてられてはいなかった。

そこで、本学の音情報処理学講座では、ホルマントに変化を加える短時間の変形聴覚フィードバック実験を行ってきた。佐藤はノッチフィルタを用いて発話中に母音/i/の F2 付近である 2kHz 周辺の周波数成分を除去し、発話者の聴覚にフィードバックする実験を行った。その結果、除去された帯域、及びその周辺で通常発話より数 dB 上昇していることを確認した。佐藤はこの結果より、スペクトルの変形による補償動作が存在する可能性を示した [20]。続いて斎藤も第 1、第 2 フォルマントを変形させたフィードバック音を用いて聴覚フィードバックの実験を行った [21]。しかし個人により分散が激しいため、定性的な知見は得られなかったが、この結果からフィードバック音声が発話動作に何らかの影響を与えているものと解釈できる。

これらの知見より、基本周波数やホルマント周波数の変化を含めたスペクトル形状の変形により、補償動作が存在する可能性が示されているが、上記の研究は、音声スペクトルを分析するのみにとどまっており、調音器官の動作については観測が行われてはいない。このため、音声スペクトルにははっきりとあらわれないが、調音器官がホルマント形状の変形を補うような方向で補償をおこなっている可能性が考えられる。

1. 3. 4 音声知覚の運動理論 ～Motor Theory～

実験により様々な仮説が提案される一方で、“ことばの鎖”について話者内部 (脳内) での音声生成と知覚の相互交信を説明するような、音声の知覚機構に関する多くのモデル化が提案されてきた。

中でも、1960 年代より Liberman らは、音声知覚の運動理論 (Motor theory of speech perception) を提唱してきた [2][3]。この Motor Theory では、言葉を聞きとる時、音声の音響的特徴の詳細を解析することなく、言葉を作り出した調音運動ないし運動企画を“直接”知覚するのだと仮定する。発話には他の無声シグナルとは違う特別な知覚特性があるという説である。つまり、音声生成と知覚の間には強いリンクが存在し、音声の知覚過程は、調音動作の内部表象、あるいは知識を参照しながら実現すると強調している。

サルを使った神経科学的研究では、ある運動の企画と実行を司る神経群が他の固体が行う同種の運動を知覚する場合にも活動することが報告されており、ミラーニューロンと名付けられている [22][23][24]。

ある特定の行為をするときに活動するミラーニューロンは、他のサルや人間が同じことをしている場面を観察するだけで活動すると報告されている[23][24]。他者の行為を自分自身の内的な運動表象に対応づけることで、自分の行為と他者の行為を結びつける役目を果たすミラーニューロンの存在により、運動の企画と知覚が共通の神経回路網によって遂行されている可能性が指摘されている。

このように、人間において、音声生成と知覚の仕組みが平行して発達し、機能的に結びついているという可能性は、ある程度 Motor Theory の裏づけとなっている。しかし、ミラーニューロンは猿の非言語的な手の運動に対して発見されたため、音声知覚に特化して言えるのかは明らかではない。

一方、Action Theory では、音声に限らず一般に信号を形成する原因となった運動や事象を“直接”検知することが知覚なのだとしており、音声知覚では調音運動が“直接”検知されると主張する[25]。“直接”とは推測や仮説検証などの思考過程を経ないという意味である。しかし、どのような機構によって“直接”検知が可能なのか、まだ十分に解明されていない。

運動の“直接”知覚に重点をおくこれらの視点に対して音響と聴覚の密接な関係に焦点を当てる仮説も多いが、音声の知覚機構に関する知見は、脳科学などの発展に伴って今後更に具体化、精緻化していくものと期待される。

1. 3. 5 脳内での音声知覚・生成過程の検討

1995年に Savariaux らは、発話器官は音響信号により運動制御を行っているとして述べた[26]。そして2002年、菅田らは、顎に突然起こった摂動時に発話器官がどのように反応するのかを調べたところ、通常の発話を行おうとする方向への強い補正反応を起こすことを発見した[27]。このように、調音器官に摂動を与え、発話がどのような活動を示すのかを観測を行うことで Motor Theory の検証が行われている。しかし、発話と知覚が脳内でどのように相互作用を起こしているのかは明らかにされていない。

音声生成は、“ことばの鎖”の中ではフィードフォワードの役目であり、運動計画、筋肉活動、調音、発声などの複数のプロセスにより構成されている。また、音声生成と知覚は、運動計画という脳内での高レベルと、音響伝達という低いレベルで互いに結びついている。fMRI や MEG などの画像による手法に加え、筋電図は高レベルでの発話生成の働きを示す方法の一つである。本多は、舌筋の筋電信号を用い、人間の脳内での音声生成と知覚の相互依存について調べた結果、運動野と言語野の両方に、同様の母音のパターンが観測され、簡単な位相マッピングが存在しているという推論に達した[28]。図 1.2 は、脳の運動野と言語野で見られる、母音の調音物質と聴覚物質の位相マッピングを示している。矢印で示された聴覚と調音のつながりは、“ことばの鎖”のループとして閉じており、意図する調音運動や聴覚知覚を直接マッピングすることでより効果的な相互交信をしていると考えられている。

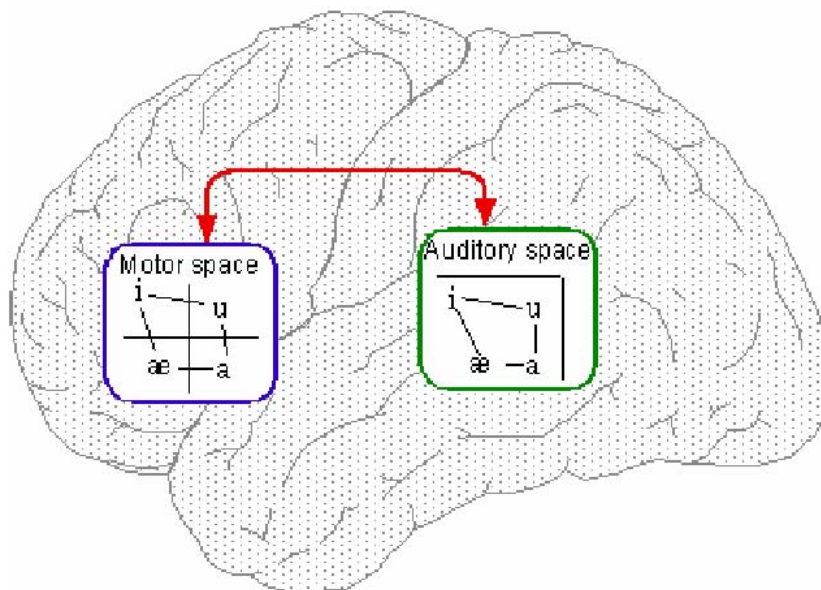


図 1.2 脳のマッピング

脳内で音声生成と知覚過程がどのように関わっているかについて検討を行う時、まだ何の訓練もして

いない乳幼児などの、フィードバックメインのコントロールを理解する必要がある[3][29]. **Motor Theory** においては、音響信号が調音動作を感知するとしている[3]. しかし、他の研究では、発話スキルを身に付けた後にフィードフォワードを繰り返し使うので、フィードバックメインのコントロールは最終的にはなくなってしまい、と力説されている[29]. 仮に運動説が正しいならば、フィードバック音声に変換を加えた時、音声生成と知覚のプロセスの間で、瞬時的な相互作用が行われていると仮定できるであろう.

本研究ではこの仮説をもとに、変形聴覚フィードバック (TAF) を用い、人間本来に備わっている音声知覚・生成過程のメカニズムについての検討を試みる.

1. 4 計測システムの構成

1. 4. 1 実験要件

実験を行うにあたり、実時間によるホルマント変形聴覚フィードバック実験を行う上で、以下の要件が満たされる必要がある.

要件 1 実時間による音声パラメータ変換を行う

要件 2 遅延を最小限にする.

要件 3 発話者の音響物理量をできるだけ多く残した自然性の高い変換を行う.

要件 4 変化に対して被験者が修正可能な摂動を与える.

要件 5 被験者が変化を知覚できる摂動を与える.

要件の 1 つ目はフィードバック音声の変換は実時間処理が行われる必要があるというものである. その理由はフィードバック音声に基本周波数や振幅のような時間変化する音響物理量が保持されていることが重要であるためである. 沢田らの報告[30]によれば、基本周波数や振幅包絡の時間情報もフィードバック音声として重要な音響物理量であることから、これらの情報の損失はフィードバック音声として不適切となる可能性がある. 例えば、フィードバック音声として予め被験者の音声を録音したものや合成したものを実験に用いたとすると、発話時の時々刻々と変化する基本周波数や振幅包絡の時間情報等がフィードバック音声に反映されない、という問題が生じる. そのため基本周波数の時間変化や振幅包絡の時間情報等を保持したまま、ホルマントのみが変換されなければならない.

2 つ目の要件は、フィードバック音声の遅延は発話動作の破壊[6]など実験に望ましくない影響を与える可能性があるため、遅延を最小限に抑えることも重要である. 佐藤は実験により、遅延が 30ms 以下であれば、発話に影響しないと報告している[20].

3 つ目の要件は、フィードバック音声は被験者の音声にできるだけ近い音声を利用すること、つまり、出来る限り音声に話者の個性を保持したまま実時間でホルマントを変換することが望ましい. Shimon らが指摘している[31]ように、聴覚フィードバックの実験で用いるフィードバック音声の自然性は重要である. 自然側音により発話が妨害されないことや[5][12], 多くの情報が失われたフィードバック音声が発話に影響しないことから分かるように[31], 人はフォードフォワード制御により自分の声とそれ以外の音声について弁別する能力が高い[32]. そのため、音響物理量あるいは自然性が損なわれることで、音声の話者の個性を失ってしまい、他人の音声として認知されるという問題が生じる. これは、聴覚フィードバック実験としては不適切な条件となり得る.

4 つ目の要件は、摂動として与える変化が被験者により修正可能であることである. もし摂動に対する応答が補償動作であった場合、被験者の発話動作は発話の変化に対して元に戻そうと働くことになる. この時使用する摂動、あるいは音韻、音節によっては、補償の方向が発話機構などの物理的な制約や言語習得時に獲得したモデル内に存在しないことにより、応答が十分に現れない可能性がある. また、発話動作が破綻するような摂動も応答の測定にはふさわしくない. このため、これらの問題を避けるような摂動、および対象とする音声資料 (音韻、音節) の選択が重要となってくる.

5 つ目の要件は、摂動が被験者によって知覚可能なことである. 摂動に対する応答が、反射運動のような自動レベルでの応答であるのか、意識レベルの応答であるのかは実験結果により判別できるが、それ以前に被験者に摂動が知覚されなければ、当然発話動作での応答は確認できない. つまり、観測すべき応答が摂動に対する随意運動のような意識レベルによる応答であるか、摂動がなくても反応が起こる自律レベルによる応答であるのかを区別するためには、被験者が摂動を検知する必要がある. さらに検知感度

などについても考察する必要がある。例えばホルマントの変化に対する検知感度は基本周波数のものほど高くなく[30]，基本周波数に比べて系の制御が安定していることから，観測対象とする調音器官の調音運動を十分考慮した摂動を与える必要がある。

1. 4. 2 測定システム

1. 4. 1の議論をもとに，以下の計測システムを構築した。筋電計測の場合を図 1.3，図 1.4 に示す。このシステムでは，バンド抑圧およびバンド強調フィルタによって，被験者が発話した音声のホルマント周波数を自由にリアルタイムで変形し，被験者に呈示することが可能である。さらに，変形に同期して，発話音声，フィードバック音声，および，各種の生理指標を同時計測することが可能である。

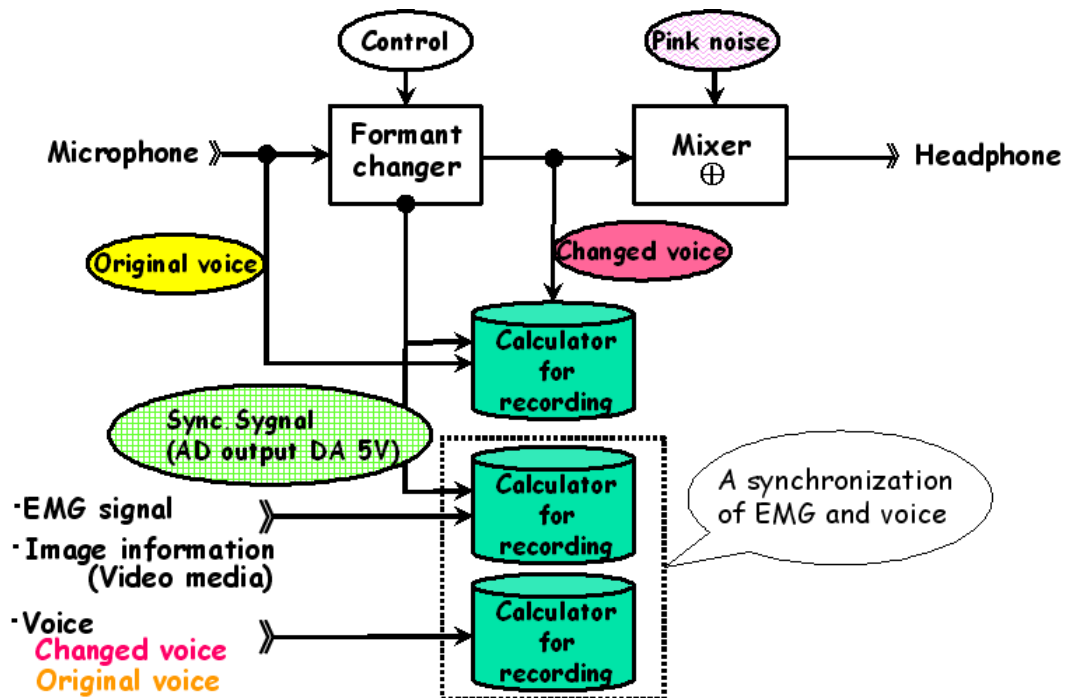


図 1.3 システム設計図（筋電計測の場合）

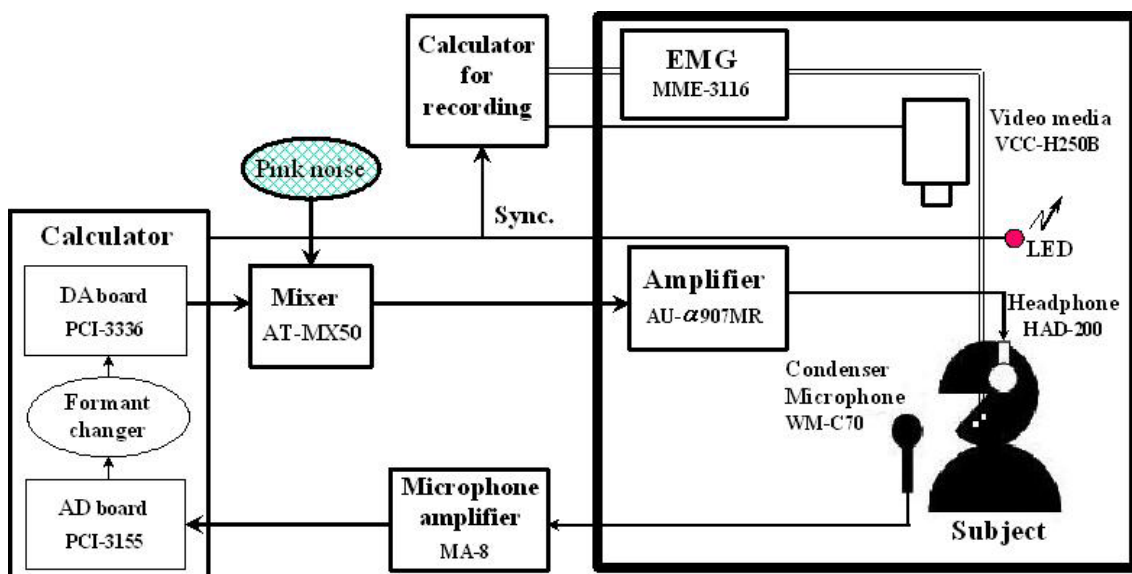


図 1.4 システム構成図（筋電計測の場合）

1. 5 測定対象

本研究では、以下のものを聴覚フィードバック実験の測定対象とし、1. 4節までで議論した内容について検討することとする。すなわち、スペクトルを始めとする音声特徴のみならず、何らかの生理指標を測定することにより、音声スペクトルにははっきりとあらわれないが、調音器官がホルマント形状の変形を補うような方向で補償をおこなっている証拠を見つけ出し、短時間聴覚フィードバックの存在を明らかにする。

音性特徴分析（聴覚フィードバックの摂動に対応したホルマント周波数の変化を測定する）
筋電計測（聴覚フィードバックの摂動に対応した発話にかかわる筋肉の反応を測定する）
舌運動計測（聴覚フィードバックの摂動に対応した舌運動の変化を計測する）

II. 実験 1 : 音響計測

2. 1 予備実験

2. 1. 1 ホルマント変形フィルタの作成

先行研究において、フィードバック音声の音質が悪いと言う課題があげられている。そこで、次のようなフィードバック音声の改良を行い、変形聴覚フィードバック予備実験において補正反応が現れるかどうかの確認を行った。

2.1.1.1 ホルマント変形フィルタ

ホルマント変形は、入力音声に関してホルマント分析により得られたパラメータに基づき、ホルマント形状を Gauss 関数により近似を行う。

$$G(f) = A \exp\left(-\frac{(f-f_0)^2}{2B^2}\right) \quad (2.1)$$

f_0 , A , B はそれぞれフィルタの中心周波数(Hz), 利得(dB), 帯域幅(Hz) を表す。伝達関数の振幅特性は次のようになる。

$$|H(f)| = 10^{G(f)/20} \quad (2.2)$$

ここで、先行研究の用いたホルマントフィルタは1つのホルマントに対して2つのフィルタが対応している。すなわち発話音声のホルマント除去を行うホルマント逆フィルタ $H(f)^{-1}$ とホルマントの追加を行うホルマントフィルタ $H(f)$ である。これらは互いに

$$H(f)H(f)^{-1} = 1 \quad (2.3)$$

となる性質を持っている。この場合、除去する対象のホルマントと追加したホルマントの形が同一である必要がある。しかし、実際の音声において、除去する対象のホルマントと追加したホルマントの形状は必ずしも同一ではない。つまり、式 2.3 におけるホルマントフィルタを用いると、実際の音声とは差異が生じてしまう。そこで、Distance algorithm を用い、それぞれの中心周波数の距離により削除、追加するホルマントのバンド幅を調節した。その結果、ホルマントの削除、追加を確実なものにした。

2. 1. 2 変形聴覚フィードバック予備実験

2.1.2.1 実験目的

フィードバック音声の音質改善を行ったホルマント変形聴覚フィードバック音声を呈示したときの発話音声を測定し、どのような傾向が見られるのか確認する。そして、本システムの有効性を計る。

2.1.2.2 実験手順

被験者に持続母音/e/を発話するように指示し、発話途中で被験者の母音/e/の第1ホルマント(F1)と第2ホルマント(F2)を/a/の方向へ変動してフィードバックした。それを実現するために、被験者の音声/e/及び/a/を事前に分析してF1とF2を求めておいた。被験者の音響パラメータを表2.1に示す。

表 2.1 被験者の音響パラメータ

被験者	/a/		/e/	
	F1 Hz	F2 Hz	F1 Hz	F2 Hz
SY	705	1109	497	1874
ML	657	1123	593	1834

変形された音声ヘッドホンを通じて被験者に提示される。摂動ありの場合、摂動を与えるまで被験者の音声をホルマント変形なしでフィードバックし、その後2秒間でホルマントの変形による“摂動あり”の音声をフィードバックする。その後再び“摂動なし”の音声を呈示する。2種類の異なる摂動開始時間のパターンと摂動なしのパターンのうち一つをランダムに被験者の発話途中に提示した。ト

ライアルの構成を図 2.1 に示す. 1 回発話を 1 トライアルとし, 1 セットは 3 トライアルからなり, 10 セット連続して実験を行った.

開始信号 (1sec)	摂動なし(2sec)	摂動あり(2sec)	摂動なし (1sec)	終了信号 (1sec)
開始信号 (1sec)	摂動なし(2.5sec)	摂動あり(2sec)	摂動なし (0.5sec)	終了信号 (1sec)
開始信号 (1sec)	摂動なし (5 sec)			終了信号 (1sec)
1 トライアル (7 sec)	休憩 (3 sec)	1 トライアル (7 sec)	休憩 (3 sec)	1 トライアル (7 sec)

図 2.1 トライアルごとの構成

2.1.2.3 分析方法

サンプリング周波数 16kHz で収録した音声に関して, ホルマント周波数は音響解析ソフトウェア WaveSurfer を用いてフレーム長 49msec, フレームシフト 10msec で算出した. 摂動量はそれぞれ変形した音声の F1 と F2 と摂動なし区間のそれとの差とし, ベクトル $(\Delta F1, \Delta F2)$ で表す. 摂動に対するホルマントの移動量は摂動あり区間の平均 F1 と F2 と同トライアルの摂動なし区間のそれとの差とし, ベクトル $(\Delta f1, \Delta f2)$ で表す. 被験者の発話音声とフィードバック音声から摂動量と補正量をそれぞれ計算して変形した音声の $(\Delta F1 - \Delta F2)$ 平面を表示する. ここで, 摂動なし区間の音声は摂動を加えていないので $(\Delta F1 - \Delta F2)$ 平面の原点となる. ここで摂動に対する応答ベクトル r を以下の式で定義する.

$$r = (\Delta f1, \Delta f2) - (\Delta F1, \Delta F2) \quad (2.4)$$

そして,

$$\|(\Delta f1, \Delta f2)\| < \|(\Delta F1, \Delta F2)\| \quad (2.5)$$

を満たすならば, 補正反応があると判断する.

2.1.2.4 実験結果

フィードバック音声の音質改善を行ったホルマント変形聴覚フィードバック音声を呈示したときの, 発話音声をの音響分析の結果を報告する. 被験者 SY のフィードバック音声を/e/から/a/の方向へ変化させた場合の摂動量と補正動作 r の関係を図 2.2 に示す. また, 式 2.5 を満たしたことから補正反応があると判断した補正動作のみを図 2.3 に, 補正反応がないと判断した補正動作のみを図 2.4 に示す. 被験者 ML に関しても同様に図 2.5, 2.6, 2.7 に示す. 補正反応があると判断されるトライアルの割合はそれぞれ 85 %, 68 %となった.

2. 1. 3 考察

実験の結果から, フィードバック音声の音質改善を行ったことにより, 与えた摂動に対して補正反応が起きることが確認できた. 被験者は与えられた摂動に対して補正反応を起こすことが示されたと考えられる. 被験者 SY の補正反応があると判断されるトライアルの割合は 85%であったが, 被験者 ML のそれは 68 %で, 被験者 SY と比較すると少ない. これより, フィードバック音声により補正反応が確認できたが, なぜあまり高率にならないのかを考える必要がある.

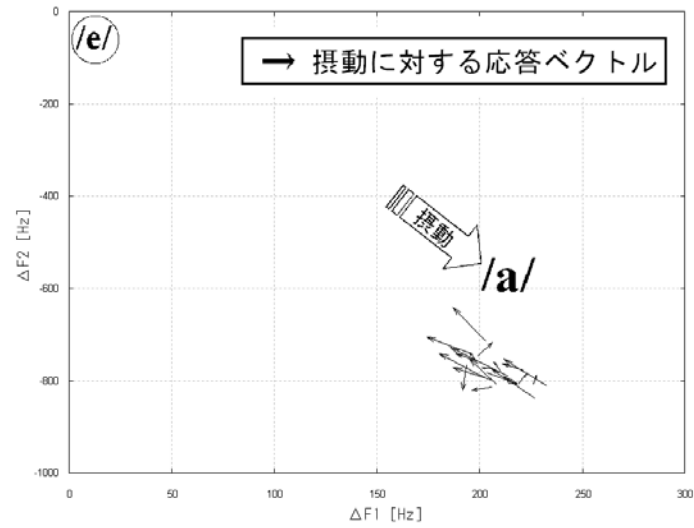


図 2.2: 被験者 SY の摂動量と補正動作 r の関係

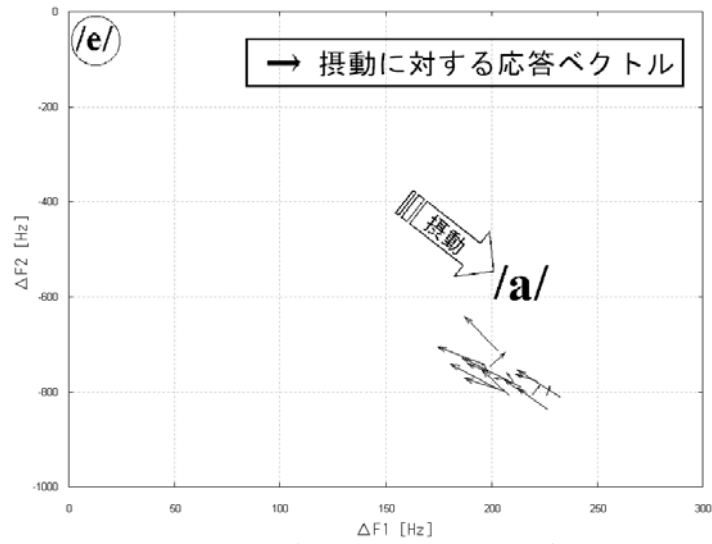


図 2.3: 被験者 SY の摂動量と補正動作 r の関係 (補正反応がある場合)

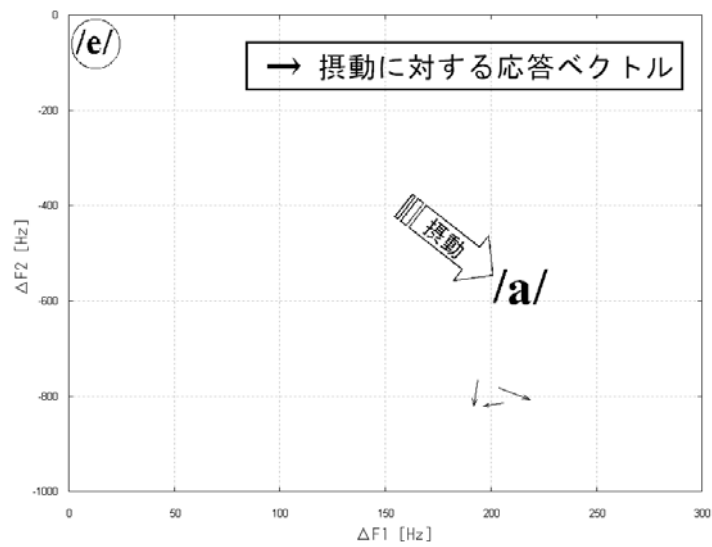


図 2.4: 被験者 SY の摂動量と補正動作 r の関係 (補正反応がない場合)

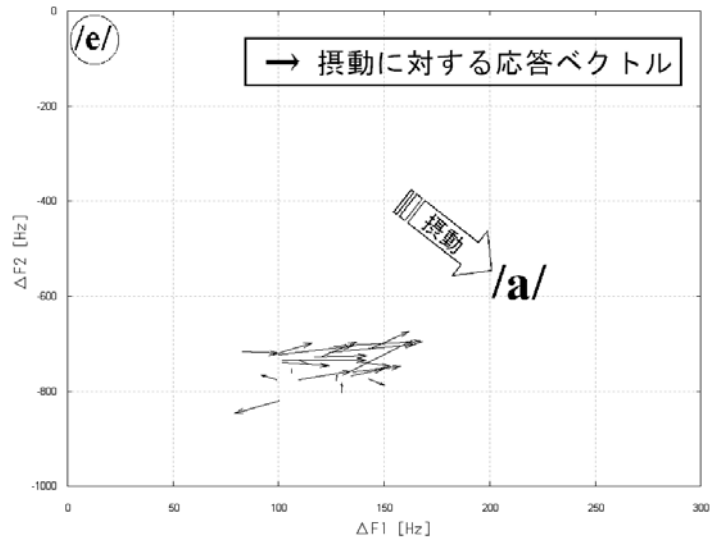


図 2.5: 被験者 ML の摂動量と補正動作 r の関係

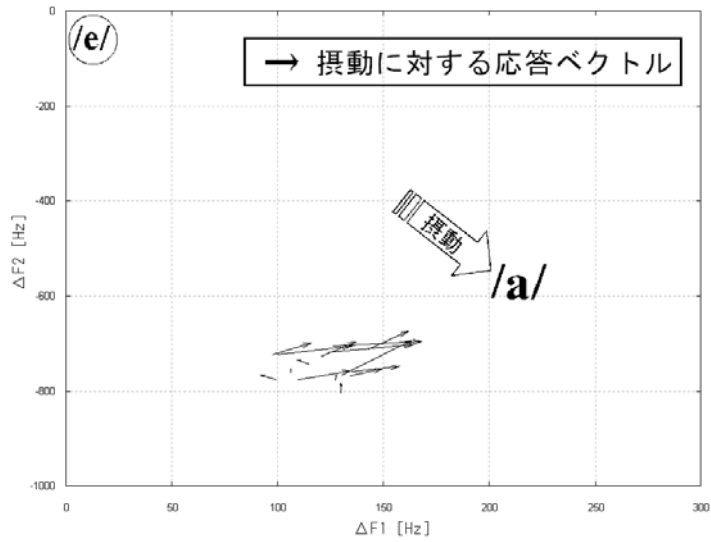


図 2.6: 被験者 ML の摂動量と補正動作 r の関係 (補正反応がある場合)

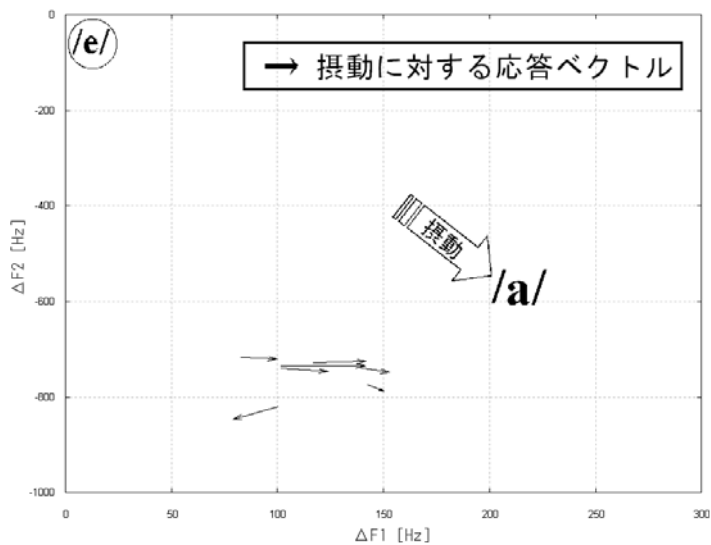


図 2.7: 被験者 ML の摂動量と補正動作 r の関係 (補正反応がない場合)

2. 2 聴覚フィードバック実験における摂動量と補正反応の関係

フィードバック音声の音質改良により、日本語母音において補正反応が確認できた。この章では変形聴覚フィードバックにおける補正反応についてさらに深く追求する。

2. 2. 1 変形聴覚フィードバック実験

2.2.1.1 実験目的

本実験では、さらに補正反応が起きる可能性を追求する。2. 1 節において、補正反応が起きる割合に差異があったことより、補正反応が起きる可能性はフィードバック音声の音質以外にもあることが示唆された。そこで、変形聴覚フィードバック実験において摂動として与える変形フィードバック音声のホルマント変形量（摂動量）を変化して、摂動量と補正動作の関連について調べ、補正反応が起きる条件について考察する、また、/e/から/a/のみならずその他の日本母音についても同様に考察する。

2.2.1.2 実験手順

実験の手順は2. 1 節の実験と同様である。被験者に持続母音/e/を発話するように指示し、発話途中で被験者の母音/e/の第1ホルマント(F1)と第2ホルマント(F2)を/a/, /i/, /u/の方向へ20,40,60,70,80,100%の割合でそれぞれ変動してフィードバックした。摂動量が100%である場合フィードバック音声は完全に/a/, /i/, /u/となる。それを実現するために、被験者の音声/a/, /i/, /u/を事前に分析してF1とF2を求めておいた。/e/から/a/の変形を一例として表2.2に示す1回発話を1トライアルとし、1セットは3トライアルからなる。各摂動量（ホルマント変形）に関して10セット連続して実験を行った。音声収録はサンプリング周波数16kHzで行った。

2.2.1.3 実験結果

被験者SY, HT, MK, ML, MTにおいて、与えた摂動量に対する応答ベクトル r の関係における、補正反応ありと判断されるトライアルのみを表示したものを図2.8, 図2.10, 図2.12, 図2.14, 図2.16と補正反応なしと判断されるトライアルのみを表示したものを図2.9, 図2.11, 図2.13, 図2.15, 図2.17にそれぞれ示す。また、摂動量として与える母音に対して補正が生じた割合を表2.3に示す。ホルマント別では、F1及びF2において最大の摂動量を100%とした場合の摂動量と補正反応ありと判断されるトライアルの割合の関係を図2.18, 図2.20, 図2.22, 図2.24, 図2.26と図2.19, 図2.21, 図2.23, 図2.25, 図2.27にそれぞれ示す。

表 2.2: /e/→/a/のホルマント変形量の一例

-	変形量 %	F1 Hz	F2 Hz
/e/	0	483	1908
↓	20	512	1782
↓	40	541	1656
↓	60	570	1530
↓	70	584	1467
↓	80	599	1403
/a/	100	628	1278

表 2.3: 摂動量として与えた母音に対して補正が生じた割合

補正が生じた割合 %			
-	/a/	/i/	/u/
SY	81.3	57.1	51.5
HT	15.1	62.5	38.4
MK	28.8	54.9	10.0
ML	59.3	71.8	44.3
MT	23.8	34.3	44.9

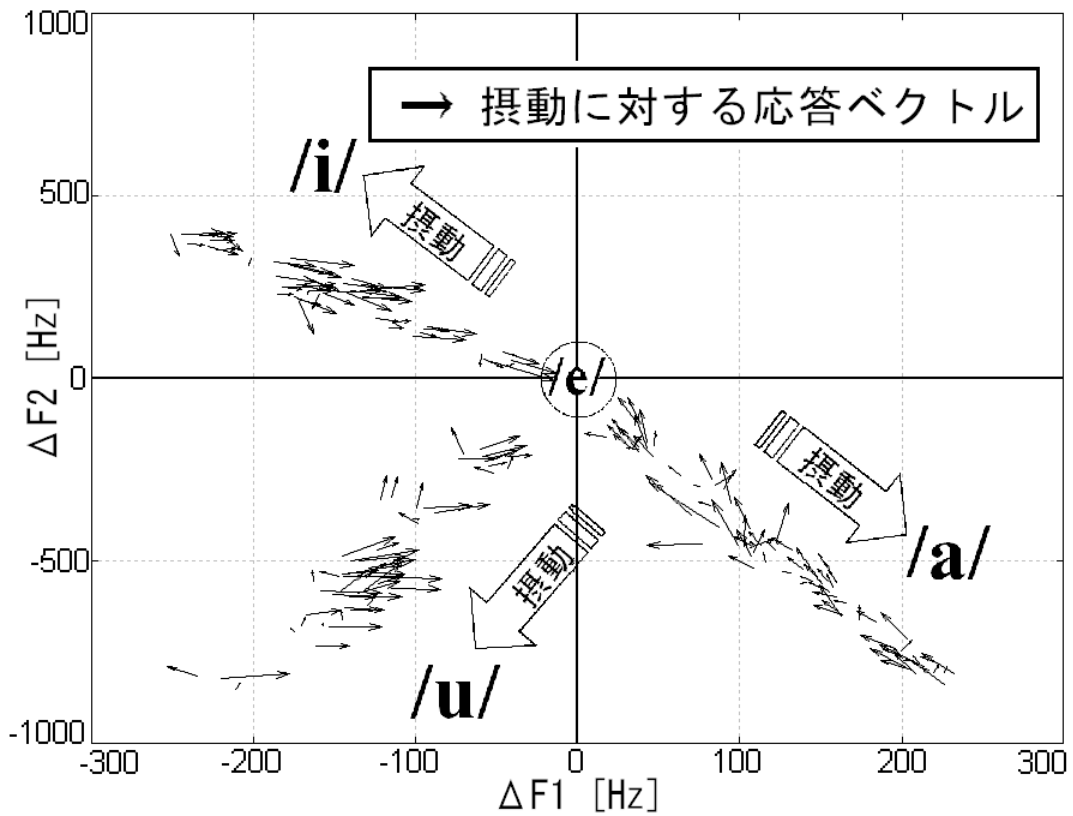


図 2.8: 被験者 SY における摂動量と補正動作 r の関係 (補正反応ありの場合)

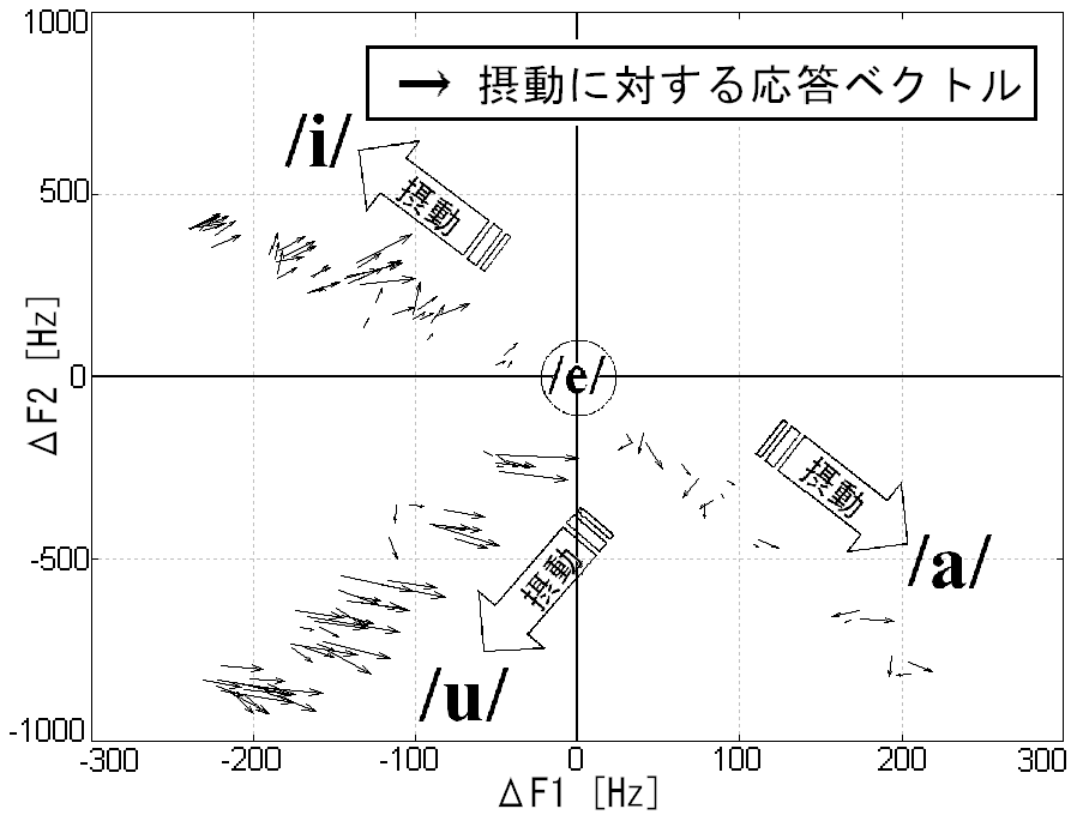


図 2.9: 被験者 SY における摂動量と補正動作 r の関係 (補正反応なしの場合)

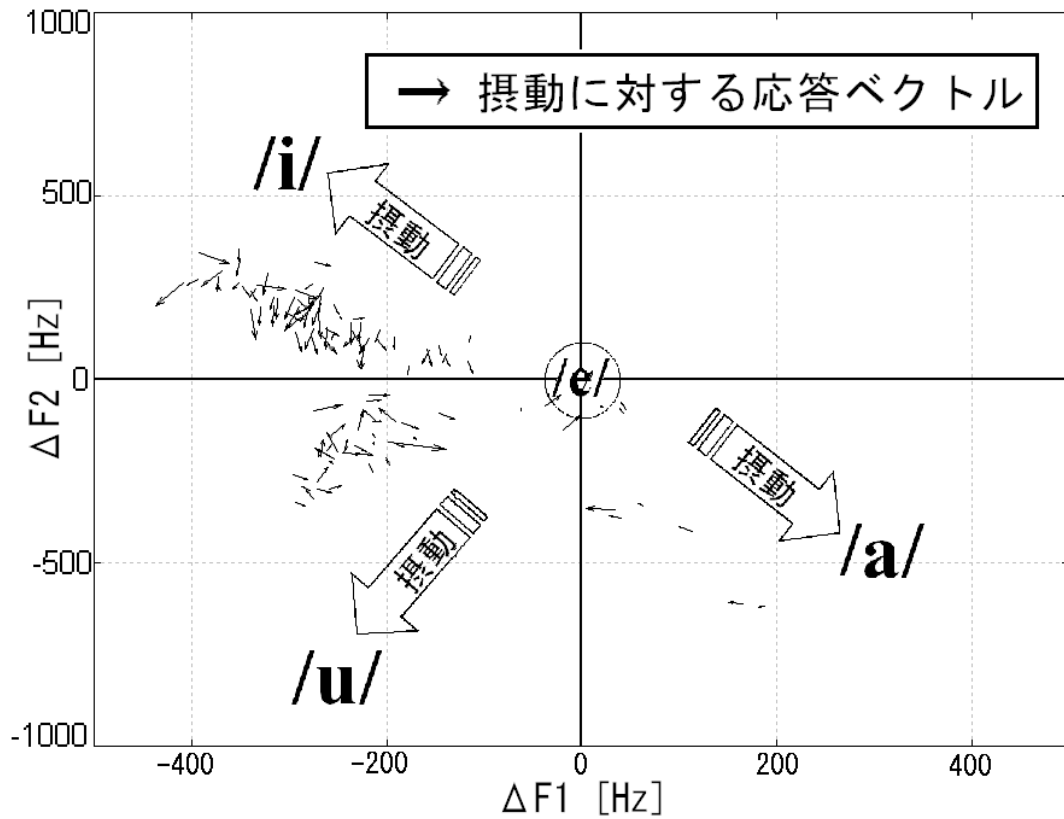


図 2.10: 被験者 HT における摂動量と補正動作 r の関係 (補正反応ありの場合)

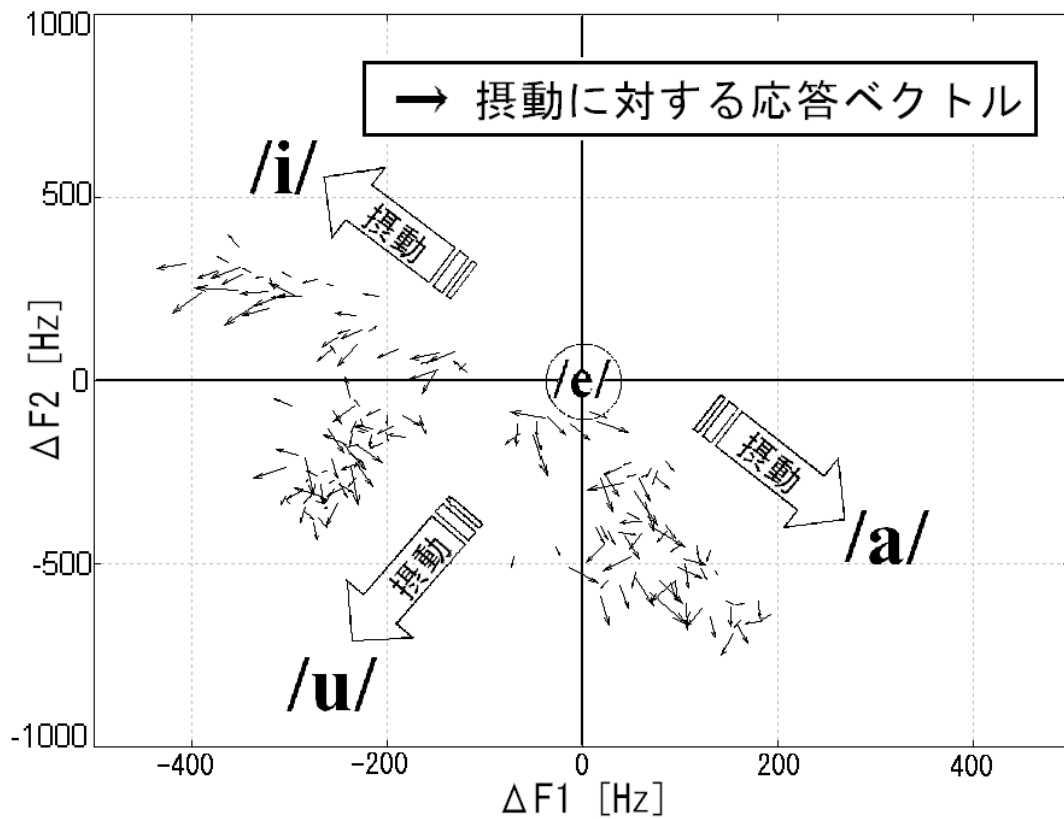


図 2.11: 被験者 HT における摂動量と補正動作 r の関係 (補正反応なしの場合)

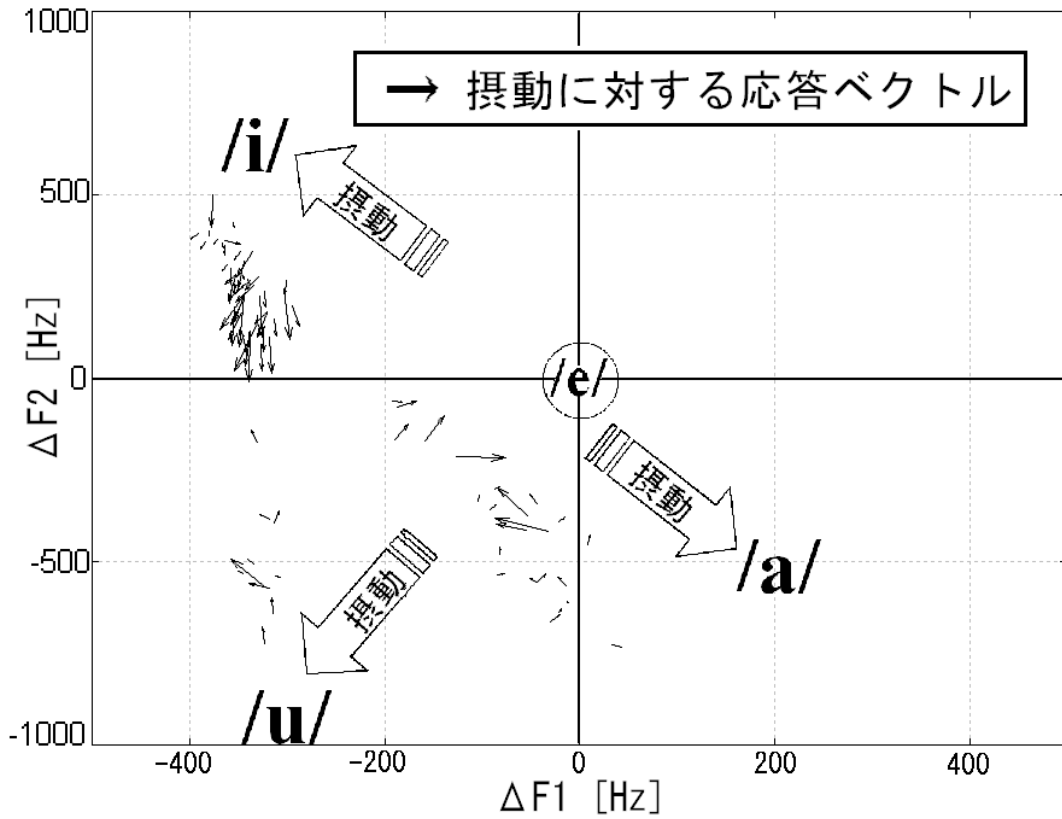


図 2.12: 被験者 MK における振動量と補正動作 r の関係 (補正反応ありの場合)

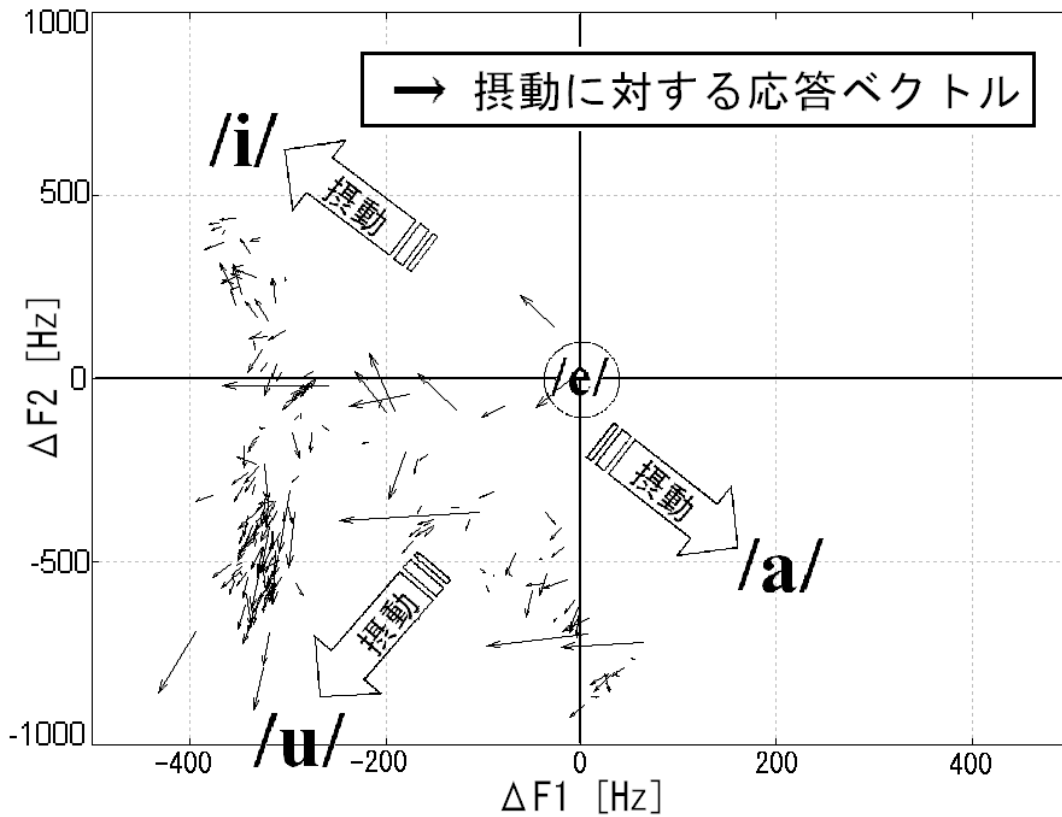


図 2.13: 被験者 MK における振動量と補正動作 r の関係 (補正反応なしの場合)

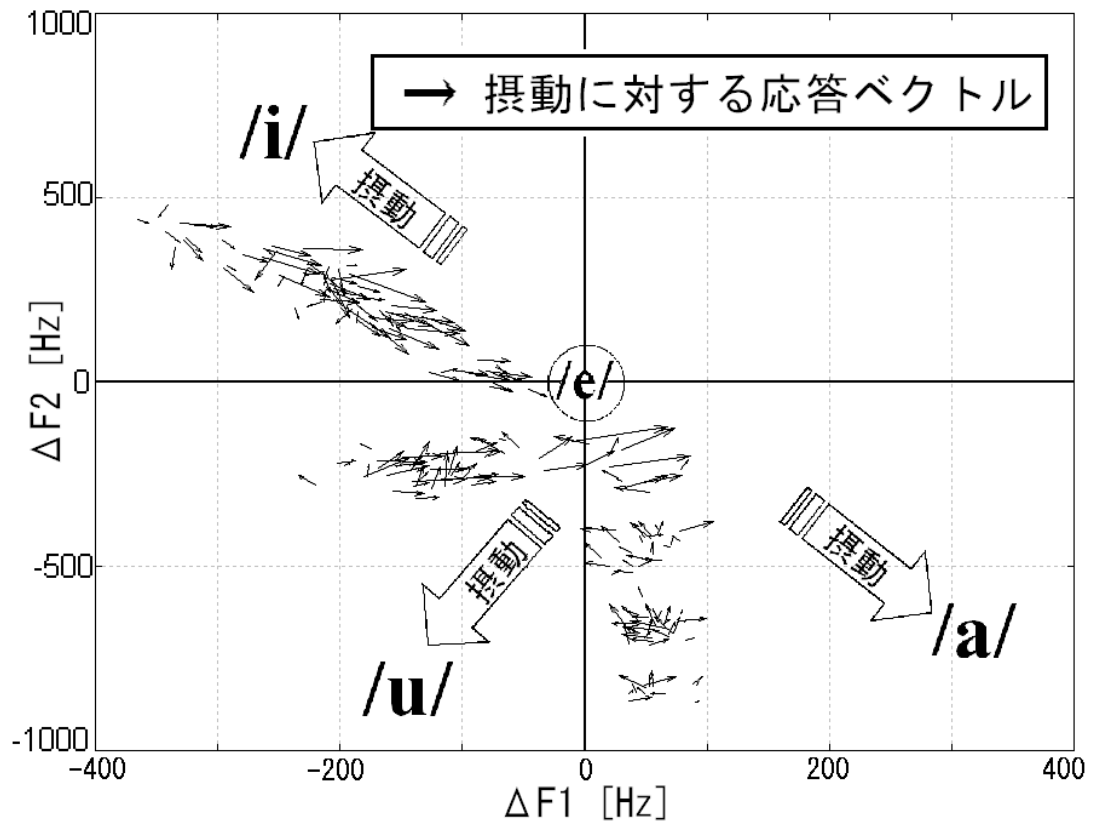


図 2.14: 被験者 ML における振動量と補正動作 r の関係 (補正反応ありの場合)

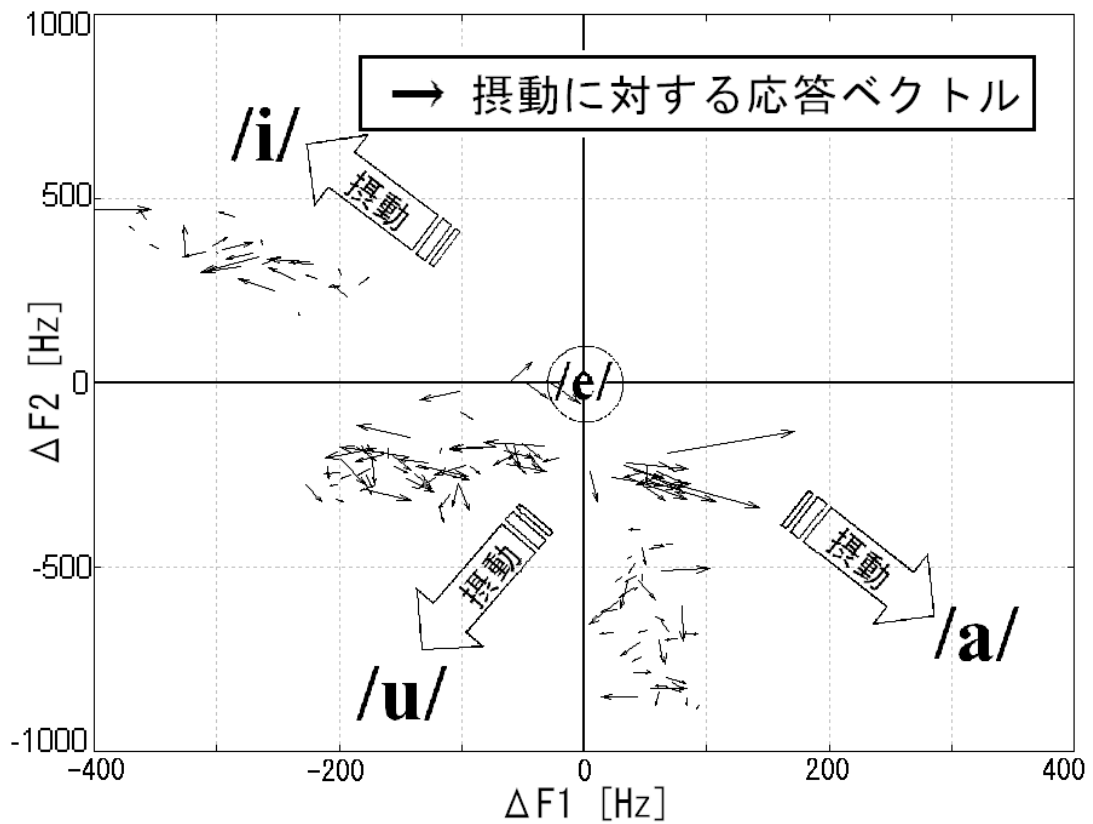


図 2.15: 被験者 ML における振動量と補正動作 r の関係 (補正反応なしの場合)

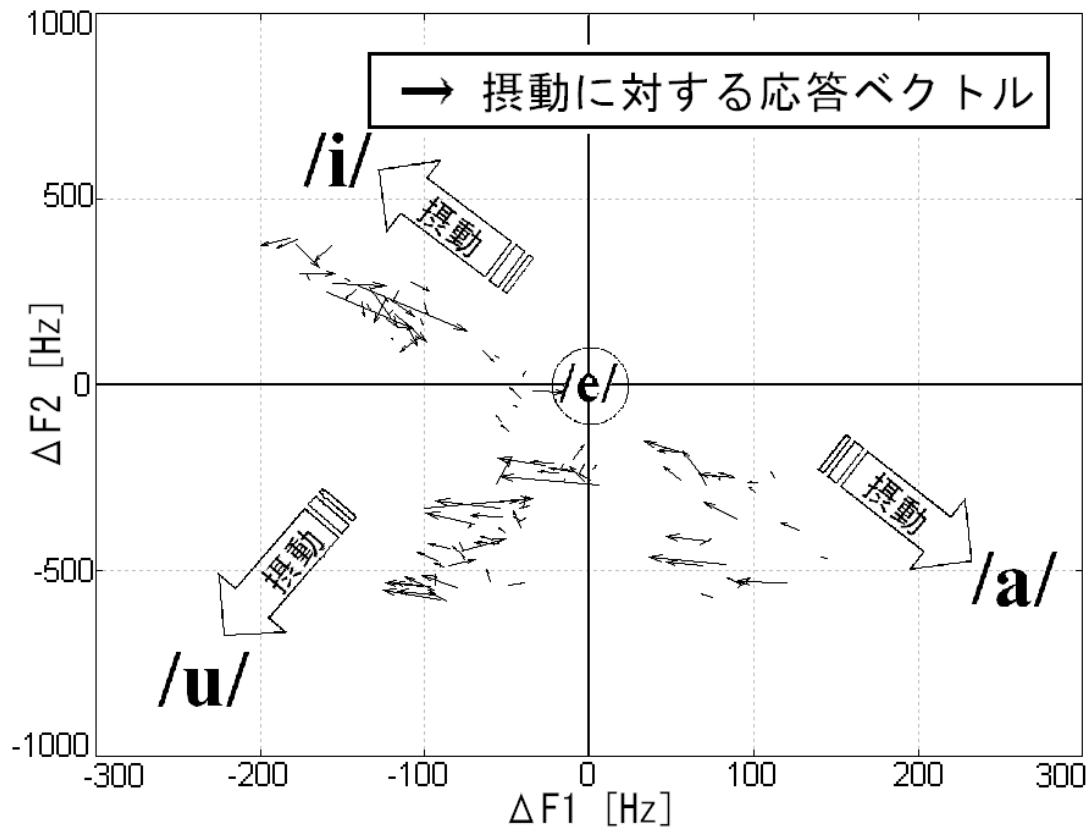


図 2.16: 被験者 MT における摂動量と補正動作 r の関係 (補正反応ありの場合)

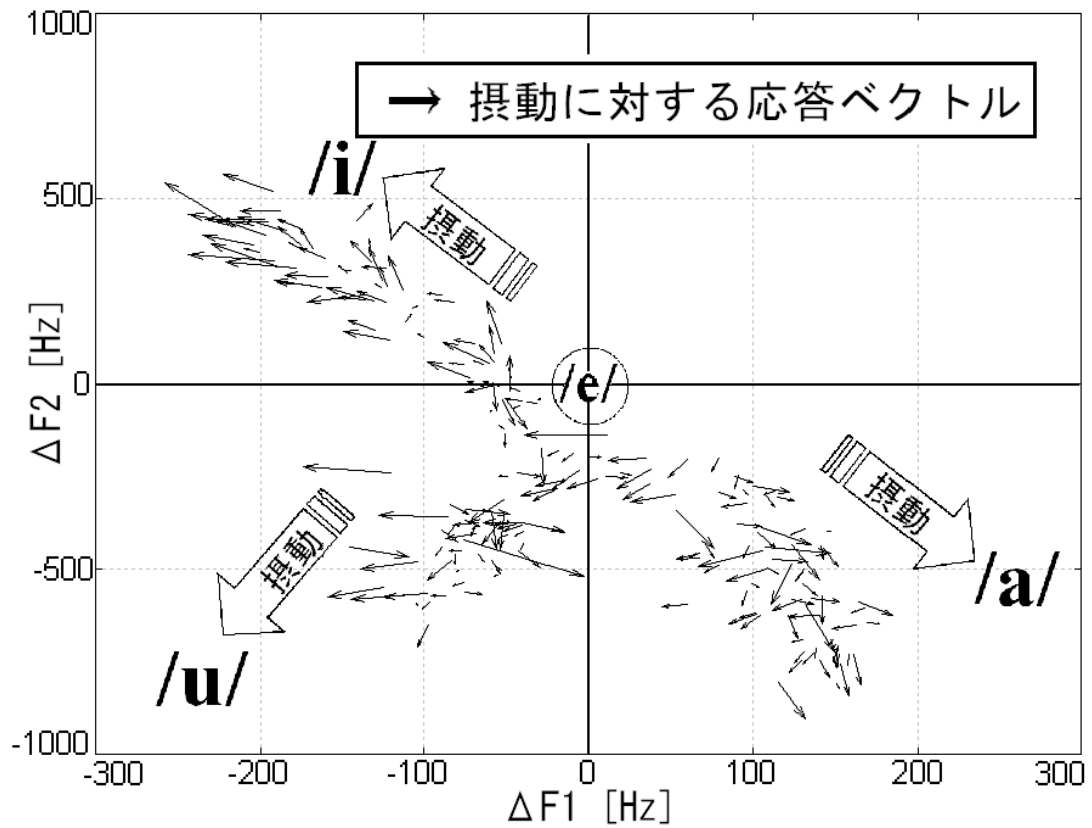


図 2.17: 被験者 MT における摂動量と補正動作 r の関係 (補正反応なしの場合)

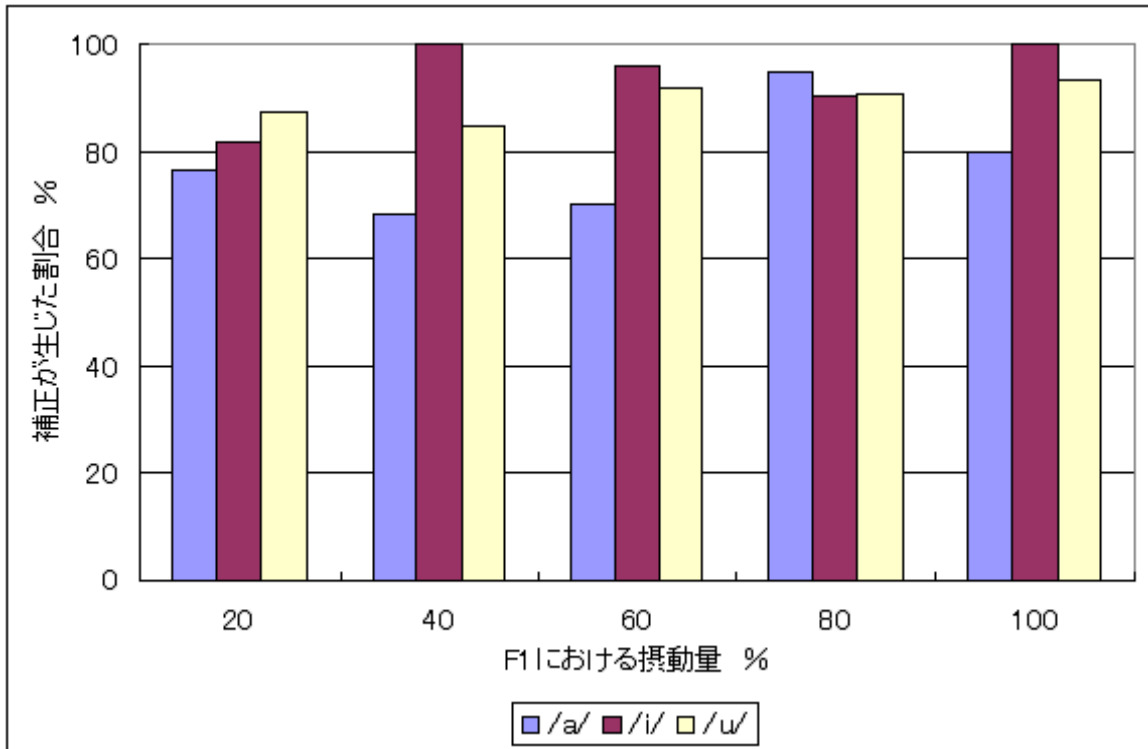


図 2.18: 被験者 SY において F1 における摂動量と補正反応が生じた割合の関係

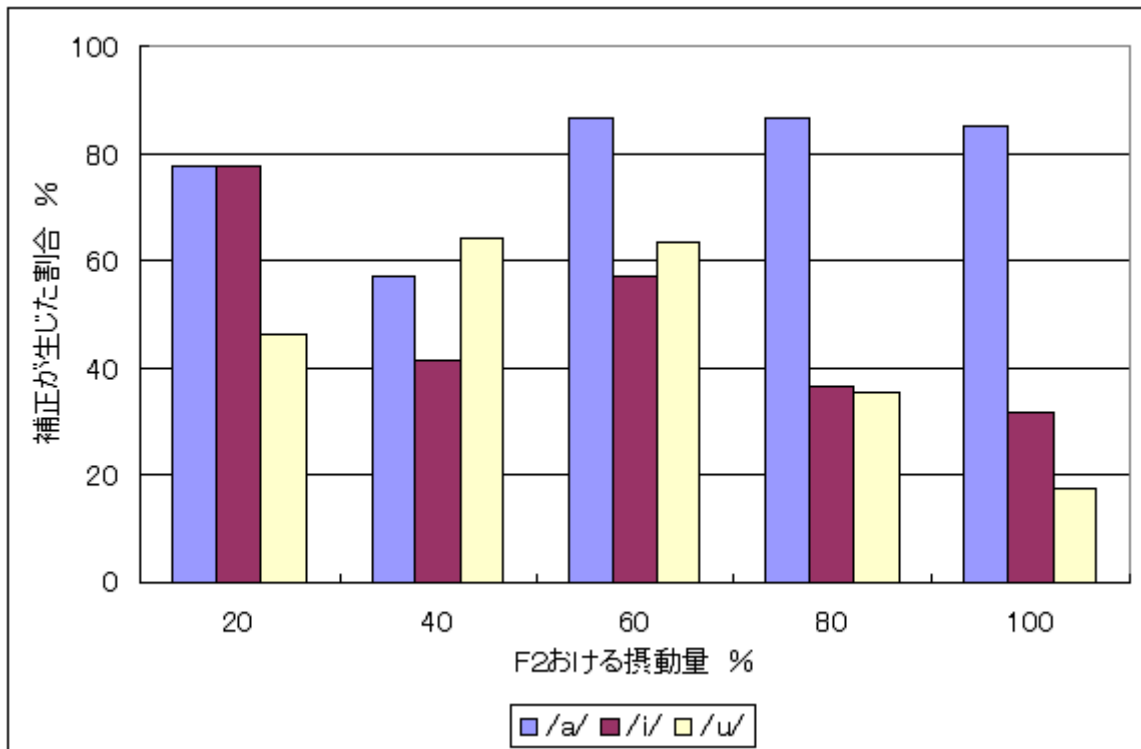


図 2.19: 被験者 SY において F2 における摂動量と補正反応が生じた割合の関係

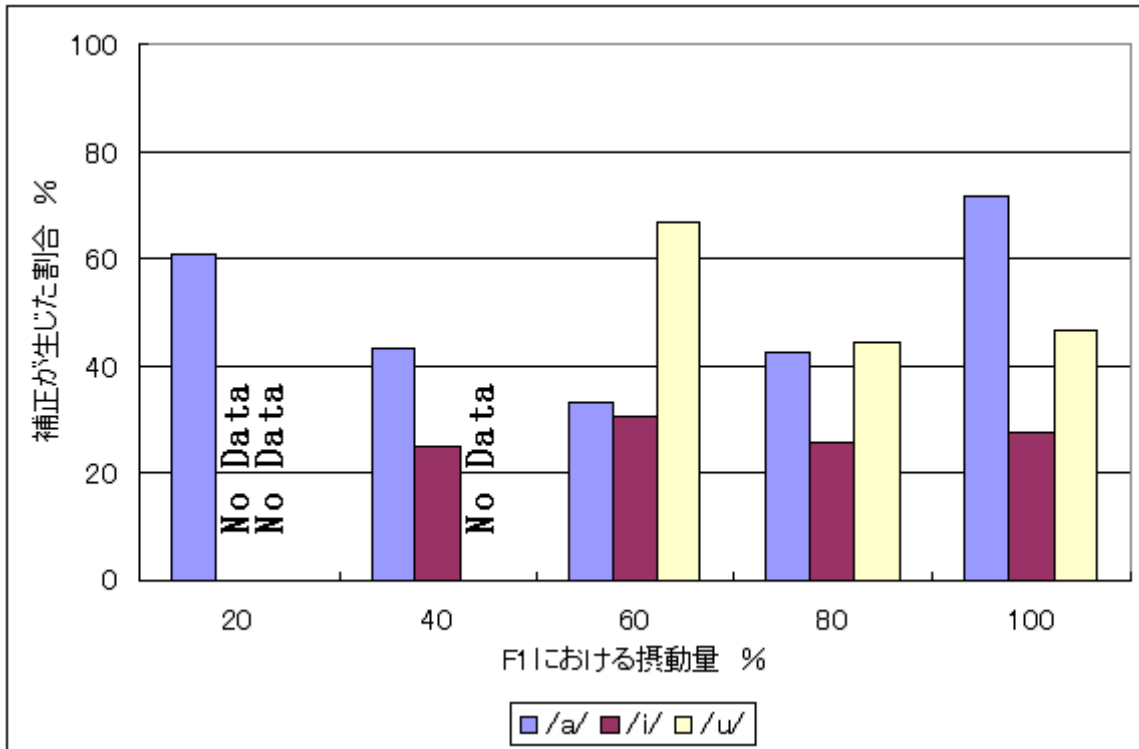


図 2.20: 被験者 HT において F1 における摂動量と補正反応が生じた割合の関係

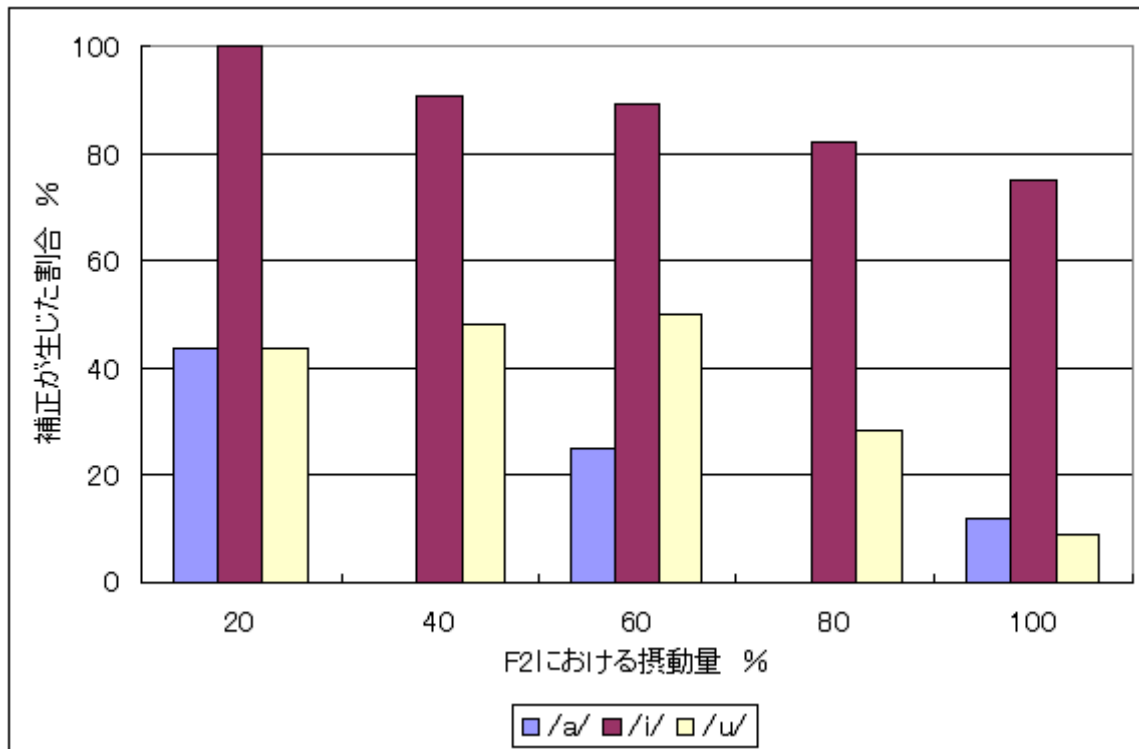


図 2.21: 被験者 HT において F2 における摂動量と補正反応が生じた割合の関係

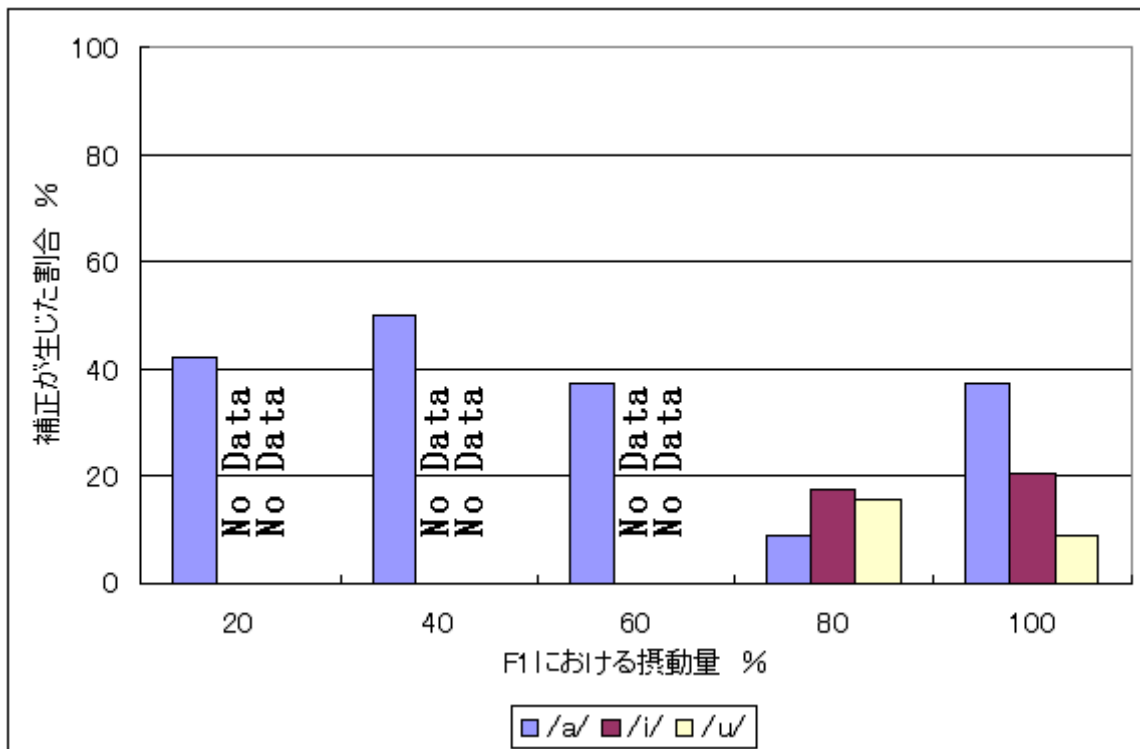


図 2.22: 被験者 MK において F1 における摂動量と補正反応が生じた割合の関係

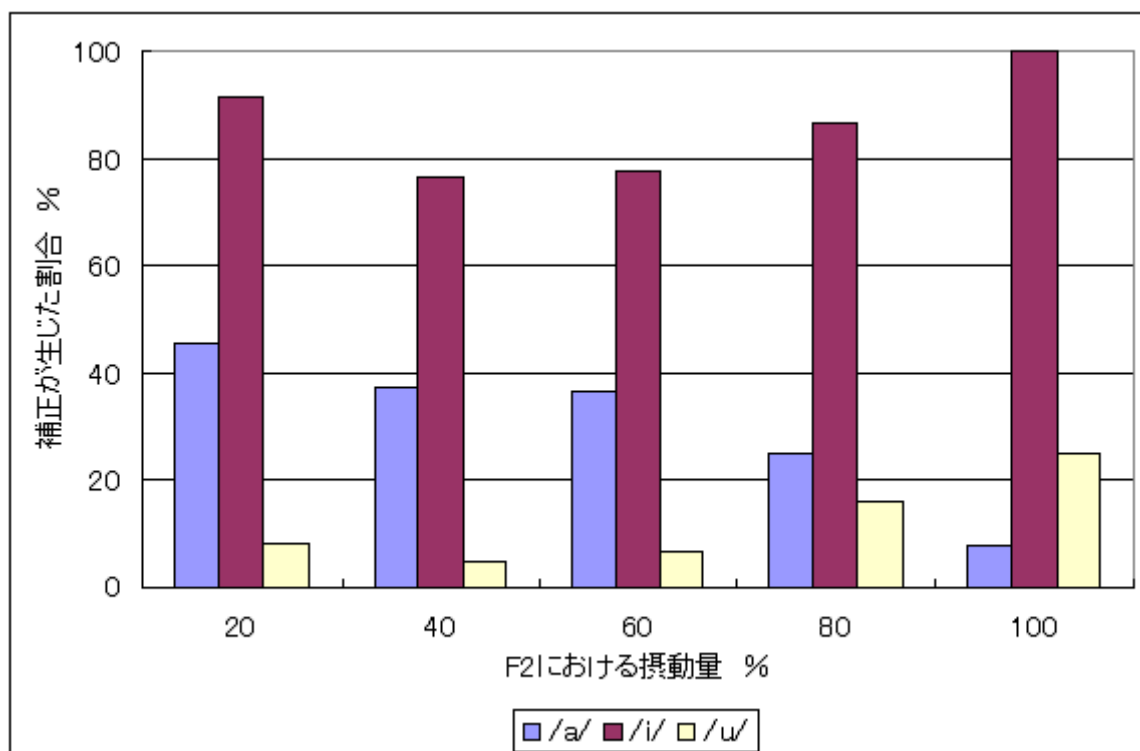


図 2.23: 被験者 MK において F2 における摂動量と補正反応が生じた割合の関係

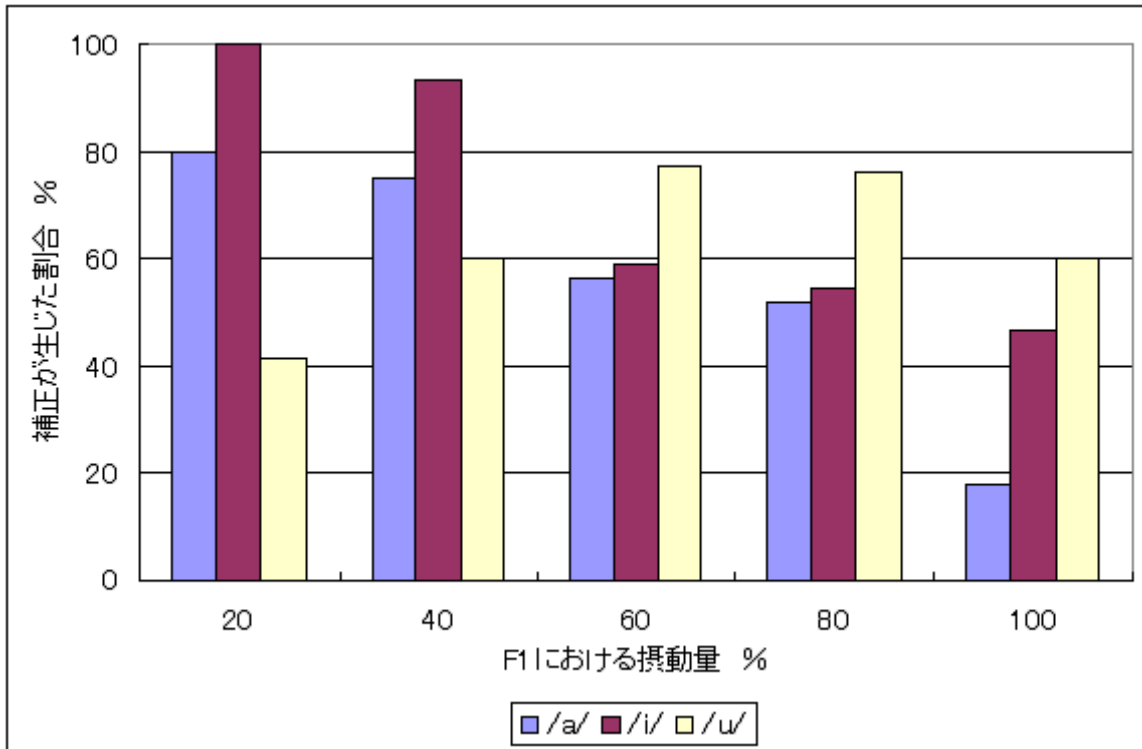


図 2.24: 被験者 ML において F1 における摂動量と補正反応が生じた割合の関係

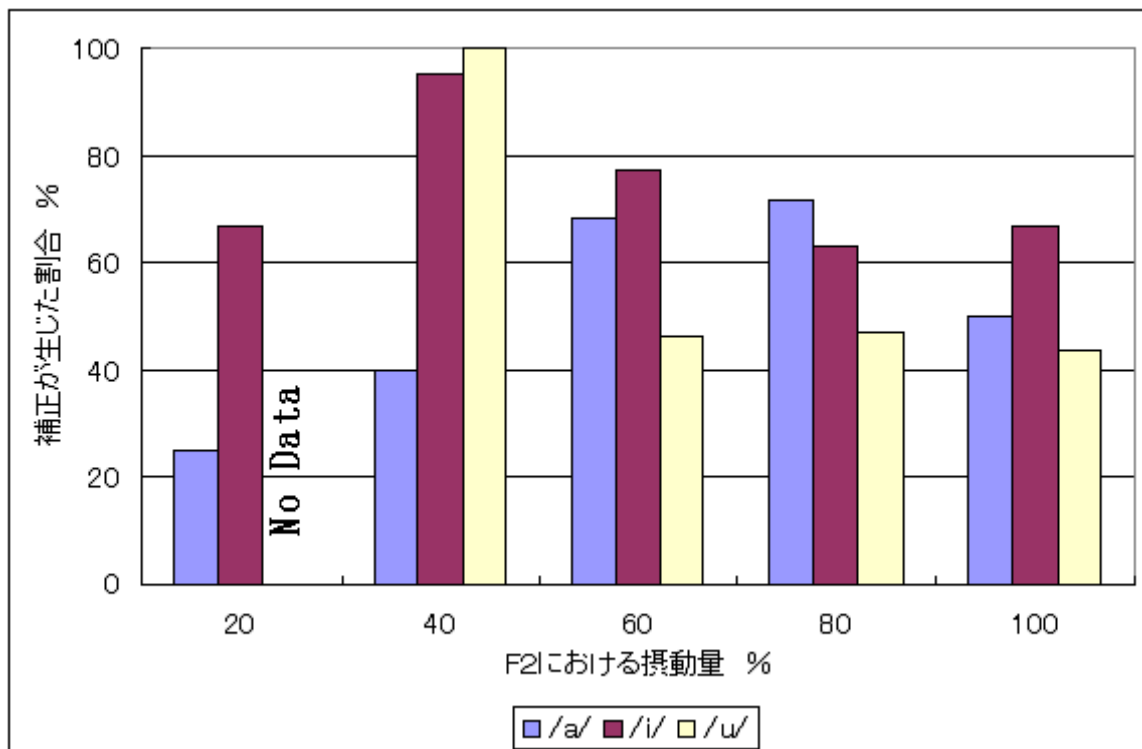


図 2.25: 被験者 ML において F2 における摂動量と補正反応が生じた割合の関係

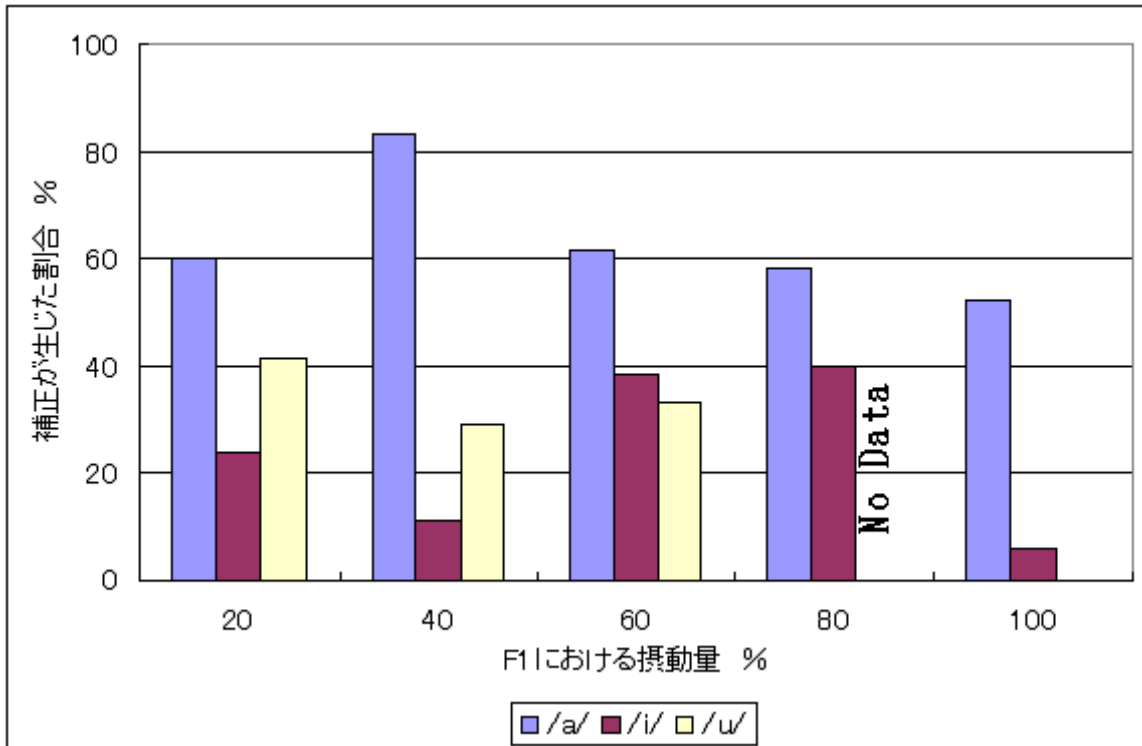


図 2.26: 被験者 MT において F1 における摂動量と補正反応が生じた割合の関係

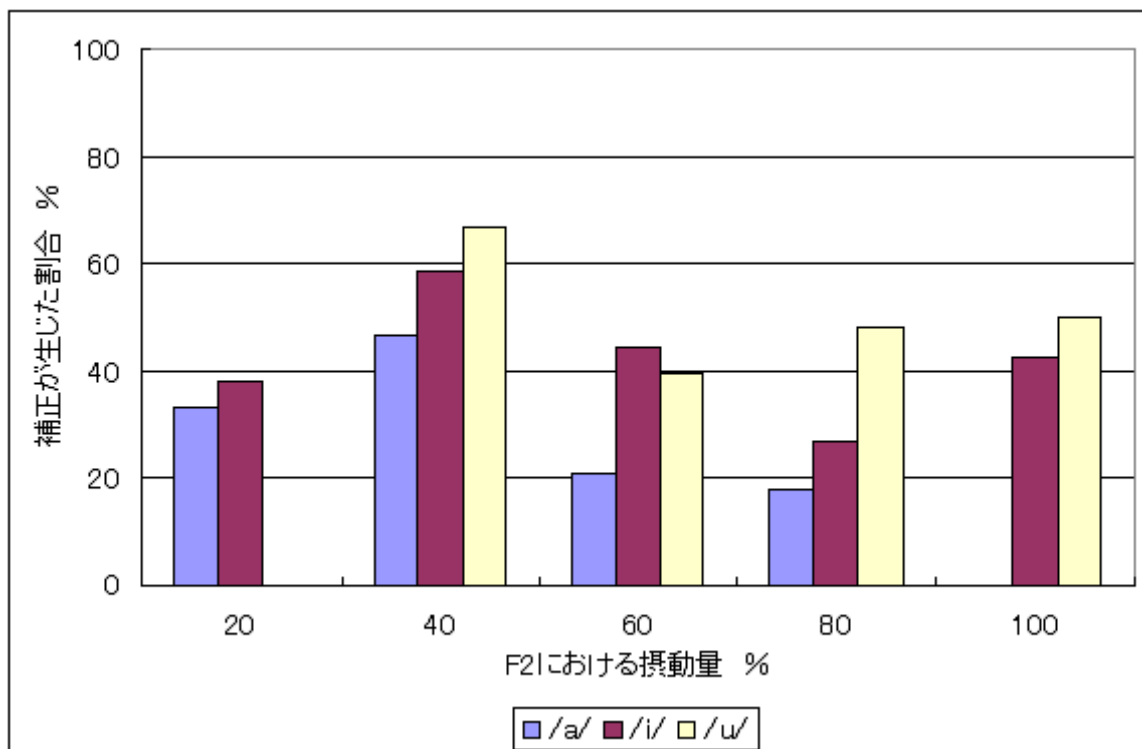


図 4.20: 被験者 MT において F2 における摂動量と補正反応が生じた割合の関係

2. 2. 2 考察

補正反応ありと判断されるトライアルの割合は、フィードバック音声を/a/, /i/及び/u/の方向へ変化させた場合でそれぞれ異なる傾向を示した。このことにより補正反応は摂動量として与える母音に依存することがわかった。

また、摂動量がある閾値を越えると、増加するにつれて補正が生じたトライアルの割合が減少する傾向が多く見られた。これより、摂動量により音韻の曖昧さが大きくなる場合補正反応がおきやすいが、それ以上になった場合被験者が意図している音声との差異が大きくなり、補正反応が起きにくくなることがわかる。また、補正反応の生じた割合が摂動量の60%をピークとした山形となる傾向も比較的多くみられた。これは小さな摂動量を与えた場合では、ホルマントが変化しても音韻の変化が知覚されないため補正反応が生じないのに対し、与えた摂動量が大きくなるにつれて音韻の変化が知覚されやすくなるため、補正反応が生じるということである。つまり、摂動量による音韻の変化を知覚したことに対して補正動作を行っているのではないかと考えられる。一方、摂動量が大きすぎる場合、被験者はフィードバックされた音声が自分の発話でないと認知される可能性があるため、かえって補正反応は少なくなる。

ホルマント別では、F1における補正反応の生じた割合はF2のそれよりも大きくなっている傾向を示した。これより、F1とF2との補正動作には、それぞれ別々の調音器官がはたらいている可能性が考えられる。F1は主に下顎や口唇が動きやすいので比較的補正が起こりやすい。一方F2は舌の前後運動に深く関連するため、生理学的な拘束により補正しにくくなると考えられる。

III. 実験 2 : 筋電計測

3. 1 生理学的観測を取り入れた実験方法
 3. 1. 1 実験システム

聴覚フィードバックによる構音に関する筋肉の筋電を計測するために、次に示す計測システムを構築した。実験システムの構成図を図 3.1 に示す。この実験システムは、主に話者の音声収録、音声スペクトルの変形、音声フィードバック、および計測部分からなる。本システムでは、計測部分には、音声スペクトル録音機能以外、筋電信号 (EMG) 計測器とビデオメディアを追加した。ビデオメディアは主に調音動作 (口の開閉など) の観測に用いる。

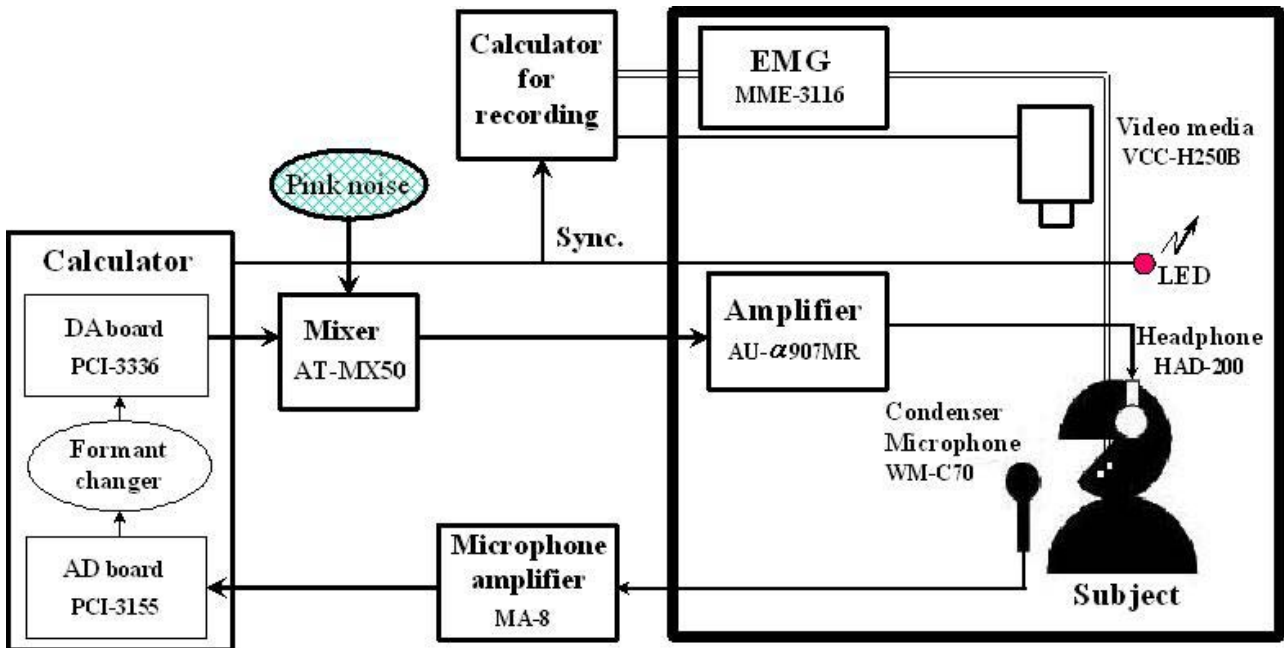


図 3.1 システム構成図

3. 1. 2 音声資料の選択

日本語の 5 母音の発声時には、口唇形状に比較的顕著に特徴が現れるが、調音点としては舌の活動がより重要となってくる。口を閉じた状態で発声ができるのも、日本語母音は舌の調音運動が中心となっているからである。しかし、中国母音の /ü/ の発声は、口唇を前に突き出す調音動作を伴わなければ成立しない。また、/ü/ の調音特徴は、口唇の突き出しと丸め以外、/i/ の声道形状とほぼ同じである。そこで、本研究では、筋電信号の計測とビデオメディアの観測を実施しやすいよう、音声資料として中国語母音の /i/ と /ü/ を選択した。中国語の /ü/ の国際的な音声記号は [y] である。以下、/i/ を [i]、/ü/ を [y] と表記する。

また、両者はフォルマントにおいても類似な構造を持っている。図 3.2 は [i] の running spectra で、図 3.3 は [y] の running spectra である。両者の第 1 フォルマント (F1) はほぼ同じで、第 2 フォルマント (F2) は [y] の方が 15% 程度低くなっている。

図 3.2、図 3.3 に示したように、[i] と [y] の音響的特徴は、主に第三フォルマント (F3) が異なることである。このようなフォルマント構造は、変形聴覚フィードバックにおいて、両母音間のフォルマントの変形が比較的容易である。

片方の母音を持続発声しながら、話者の音声をもう一方の母音に変形し、聴覚へのフィードバック音声に入れる。もし調音動作に補償があるならば、まず口唇の動きを特徴付ける口裂周囲の筋電位に現れることが期待できるだろう。これらの特徴を考慮したうえで、本研究では中国語母音の [i] と [y] を音声資料とした。

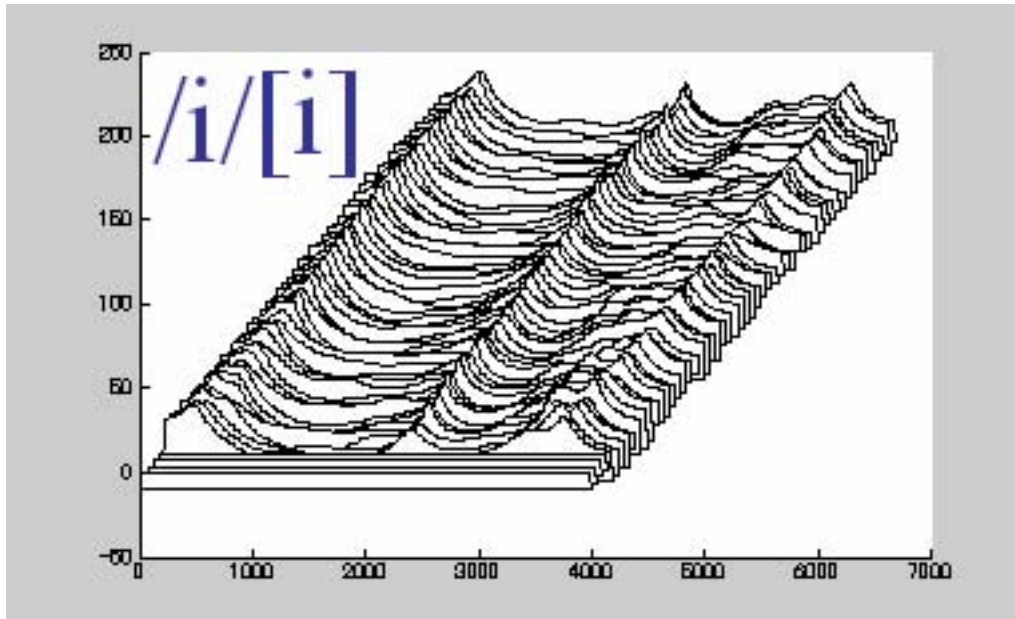


図 3.2 [i]の running spectra

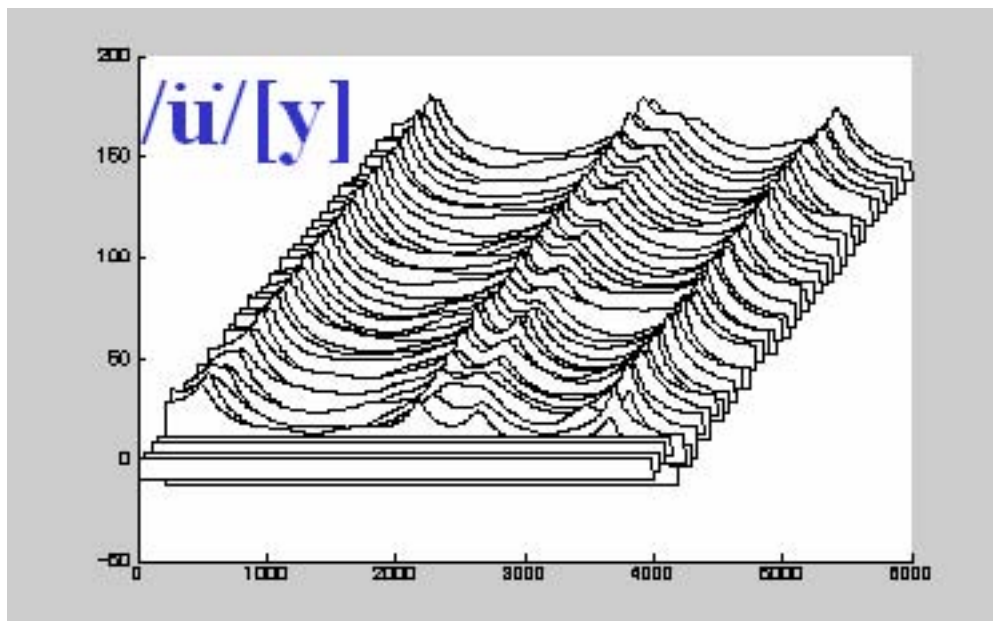


図 3.3 [y]の running spectra

3. 1. 3 筋電計測について

3.1.3.1 表面筋電

筋電図の測定では、針電極と表面電極との2種類の電極を用いて実施することができる。針電極は、配置位置が皮膚表面より深い筋に対して有効であり、位置的に正確に計測することができる。計測する際、専門的な知識を要することと、被験者の負担などの面から考えると本研究にとって最も適切な方法とは思えない。一方、表面電極は、配置位置が浅く、比較的大きな筋の測定に対して効率的である。そのゆえ、本研究は表面電極を用いて筋電信号の計測を行うことにした。

3.1.3.2 口裂周辺の筋

人は、舌・顎・唇などの複数の調音器官を巧みに操ることにより音声生成を実現している。特に、口唇は舌と並んで、柔軟な変形や微妙なせげめを表現し、声道の共鳴により音韻音色を制御している。多

くの調音筋のなかでは、顔面筋（口裂周囲の筋）は最も計測しやすい。口裂を囲んで口唇を形成している筋を口裂周囲の筋という。顔面筋の中で、発話動作に最も深くかかわっている筋肉は、口裂周囲の筋である。口唇形状は口裂周囲の筋肉（口筋）の緊張状態と顎の開閉状態とによって決定される。そこで本実験では、口唇の応答特性、補償動作を明らかにできるような口裂周囲の筋の測定を行う。

口裂周囲の筋を図 3.4 に示す。口裂周囲の筋のほとんどは、口角に集まり、口唇に付着している。その中で、頬の引き上げに関与している筋肉は、頬筋・笑筋であり、また下唇の下げに関与している筋肉は下唇下制筋である。これらの筋肉は[i]の発話に関わっていると言われている[33]。人によって異なるが、[i]の発話時には、頬を引き上げる運動と下唇を下げる運動が大きく関与しているためである。一方、[y]の発話には、口唇の突き出しやすぼめ、丸めなどの動作を促す口輪筋とオトガイ筋が最も積極的に関与している。上記で紹介した活動筋と発話動作との対応関係を表 3.1 に示す。

表 3.1 発話動作と活動筋の対応表

発話動作	主に働く筋肉	
[i]を発話した時	頬筋/笑筋	下唇下制筋
[y]を発話した時	口輪筋	オトガイ筋

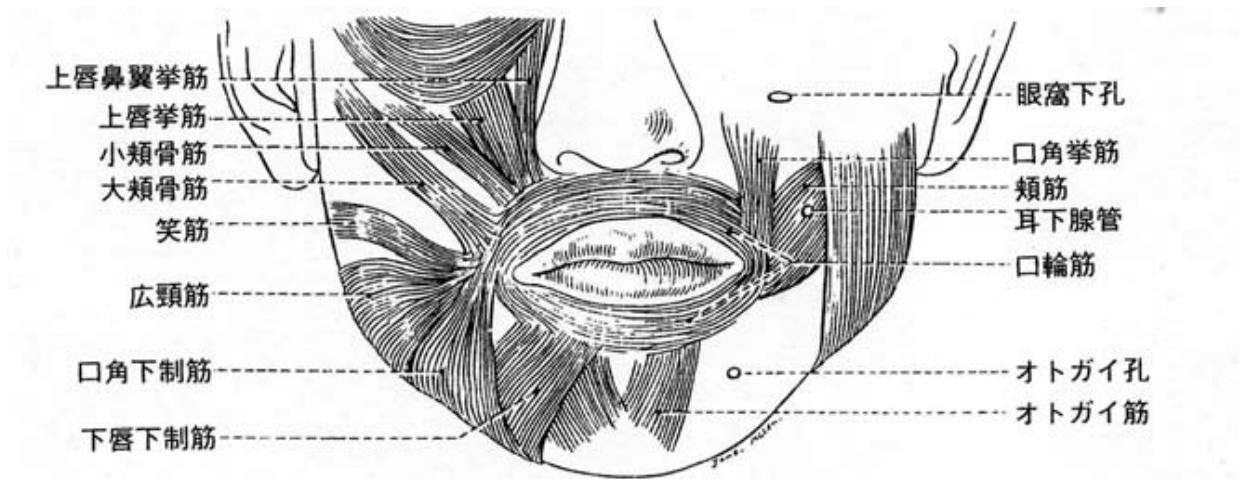


図 3.4 口裂周囲の筋（文献[34]より引用）

3. 1. 4 実験プロトコル

本研究では、上記の音声資料を用いて、正常な発話・聴覚能力を有する中国人男性話者を被験者として実験を実施する。

音声フォルマント変形の準備として、予め被験者の音声に対してフォルマントを計測しておく。変形聴覚フィードバックを実施する際、F1 はそのままに維持し、[i]から[y]へ F2 と F3 を切り替えてフォルマントの変形を実現する。

詳しい実験プロトコルについては、予備実験と本実験で異なるため、各章に実験プロセスとして記載した。

また、各被験者には以下のことを承諾してもらう。

- ・ピンクノイズを聴いてもらうことにより、本人の声や周囲の音が聞こえなくなること。
- ・実験開始信号の合図の後に、[i]と 5 秒間連続して発話を行うこと。
- ・顔面筋の位置を印すために、ペンで印をつけること。
- ・アルコールとスクラブでの肌洗浄の際に、痛みが伴う場合があること。
- ・テープやアルコール、ペーストなどへのアレルギーがないこと。

3. 2 予備実験

3. 2. 1 目的

予備実験は、音響的特徴や調音的特徴を加味した上で選択を行った音声資料を用いて実験を行うことで、本実験の有効性を計ることが目的である。また、計測システムの構成や実験プロトコルの改善点を見つけ出し、本実験の改善点を見つけ出すことにある。

3. 2. 2 実験条件

3.2.2.1 被験者

人間本来に備わっている音声知覚・生成のメカニズムを解明することが本研究の目的であるため、[i]と[y]の音声・調音情報を脳内にもっている被験者が望ましい。

そこで、正常な発話・聴覚能力を有する 33 歳の中国人男性話者 1 名 (XL) を被験者として予備実験を実施した。

3.2.2.2 被験者の音響パラメータ

音声フォルマント変形の準備として、予め被験者 XL の音声資料に対するフォルマントを計測した。その結果を表 3.2 に示す。この音声資料に対して、F1 はそのままにし、[i]から[y]へ F2 と F3 を切り替えてフォルマントの変形を実現する。

表 3.2 被験者 XL の[i]と[y]のフォルマントの値

母音	[i]			
フォルマント	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
被験者 XL	343	2115	2988	3426
母音	[y]			
フォルマント	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
被験者 XL	321	1791	2235	3251

3.2.2.3 計測装置

図 3.1 に記載された計測装置を用いて、4 チャンネルの筋電信号、画像情報、2 チャンネルの音声信号を同時に収録した。計測装置の概要について説明を行う。

実験は防音室内で行う。被験者により発話された音声は、リアルタイム OS である RT-Linux 上のプログラムにより実時間で変換が行われ、被験者にフィードバックされる。骨伝導や機器などの周囲の雑音をマスクするため、フィードバック音声には 60dB 程度のピンクノイズを付加する。

発話音声、及び変換されたフィードバック音声はそれぞれ記録用の計算機に記録され、分析に用いる。被験者は、ヘッドホン (HDA-200) を身に着けた状態で、マイクロホン (WM-C70) へ発話を行う。発話された音声はマイクロホン、マイクロホンアンプ (MA-8) を経て、一方は計算機内の AD 変換ボード (PCI-3155) に入力され、もう一方は記録用の計算機に接続された AD 変換器 (DF-2021) を経て記録用の計算機に記録される。計算機内の AD 変換ボードに入力された音声は実時間処理によりフォルマントの変換処理が行われ、DA 変換ボード (PCI-3336) を通じて外部に出力される。

DA 変換ボードから出力された音声は一方はミキサ (AT-MX50) によりピンクノイズを付加され、防音室内のアンプ (AU- α 907MR) を経て被験者にフィードバックされる。もう一方は、AD 変換器を経て記録用の計算機に記録される。

筋電信号の測定装置は MME-3116 であり、サンプリング周波数 6kHz、感度 250 μ V/div、フィルタの遮断周波数は高域 5kHz、低域 10kHz で最大 8 チャンネルの信号を測定することができる。

今回の予備実験では、中国語母音[i]と[y]の音声資料を用いるため、主に頬筋・笑筋、下唇下制御筋、口輪筋とオトガイ筋の 4 つの調音筋を観測した (筋の収縮と調音動作との対応関係は表 3.1 を参照)。表面電極とフェイスマーカー配置位置を図 3.5., 図 3.6 に示す。口唇の動きを観測するため、ビデオメディア

(VCC-H8000) により 60FPS で画像情報も筋電信号と同期させて収録する。一部の表面筋電極はモーショントラッキングのマーカとしても利用している。

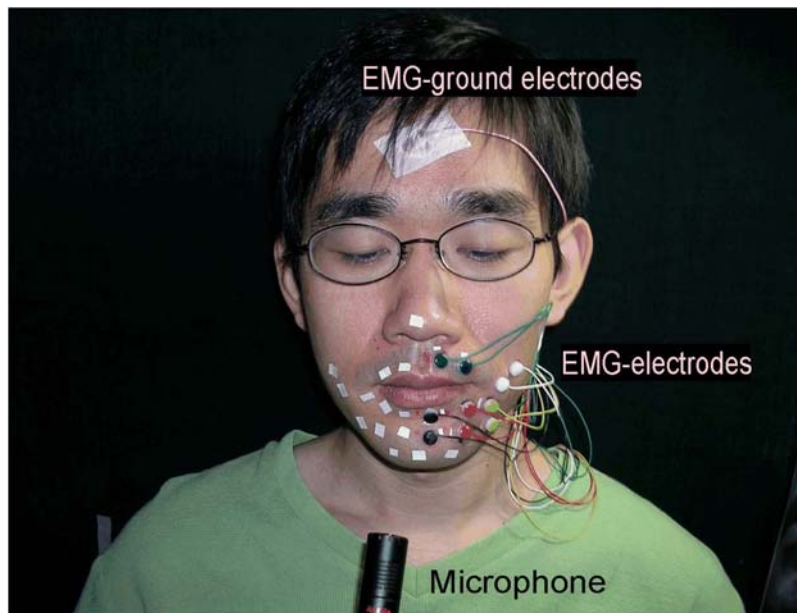


図 3.5 被験者 XL の表面筋電とフェイスマーカ配置位置 (正面)

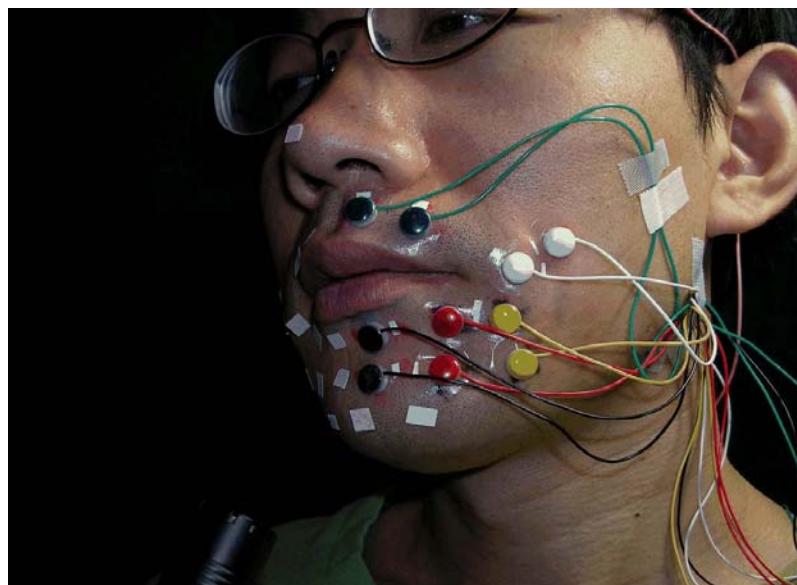


図 3.6 被験者 XL の表面筋電の配置位置 (側面)

3. 2. 3 実験プロセス

実験のプロセスについて説明をする。被験者の発話は母音[i]であり、フィードバック音声は被験者の[i]のフォルマントを[y]に変換したものを聞いた。それを摂動として被験者の発話途中ランダムに与える。発話期間を 5 秒に設定し、発話開始合図と終了合図(各 1 秒)はヘッドホンを通じて被験者に通知する。最初の 2.2 秒間は“摂動なし”で被験者の音声をそのままフィードバックする。2.2~4.4 秒の 2.2 秒間はフォルマントの変換による“摂動あり”の音声フィードバックされる。最後の 0.6 秒はまた“摂動なし”の音声を呈示する。摂動ありの場合の予備実験の構成略図を図 3.7 に示す。摂動がない場合は、摂動区間が“摂動なし”となる。このような 1 回発話は 1 トライアルとし、1 セットは 3 トライアルからなる。各トライアル間は 10 秒間の休憩を挟み、10 セット (30 トライアル) 連続して実験を行った。10 セット中の 3 割は摂動なしのトライアル、7 割は摂動ありのトライアルをランダムに行った。



図 3.7 予備実験の構成略図

3. 2. 4 実験結果

変形聴覚フィードバックを用いた予備実験における、筋電信号、音響分析、画像分析の結果を報告する。

3.2.4.1 筋電信号の分析結果

[i]の発話時に、笑筋の収縮により口を横に引き上げ、下唇下制筋の活動により下唇を外下方に引く。中国母音[i]の発話では、特に下唇下制筋は活発に活動することを確認した。上記の実験プロトコルに従って、[i]の持続発話時にフィードバック音を[i]から[y]に切り替えた後、もし聴覚フィードバックが常に存在するとしたら、筋電信号は発話動作が[i]を強調する方向への動くはずである。聴覚フィードバックに変形（摂動）がありとなしの場合に収録した下唇下制筋の筋電信号の一例を図 3.8 に示す。この図の左図は摂動なしの場合の筋電信号で、右図は摂動ありの筋電信号である。上段は筋電の波形信号で、下段は筋電信号のエネルギーである。聴覚に摂動のない場合、筋電信号は発話の最初に電位が高く、その後、活動は徐々に減っていく。それに対して、聴覚に摂動ありの場合、発話の最初に筋の活動が大きく、その活動が一旦減少し、フィードバック音が[i]から[y]に切り替えられた後、下唇下制筋は[i]を強調するためもう一回活発的に活動する傾向を示した。この現象は、発話時に聴覚フィードバックが働いていることを示唆している。全体のデータを分析したところ、摂動ありのトライアルの 71%にこのような活動が見られた。

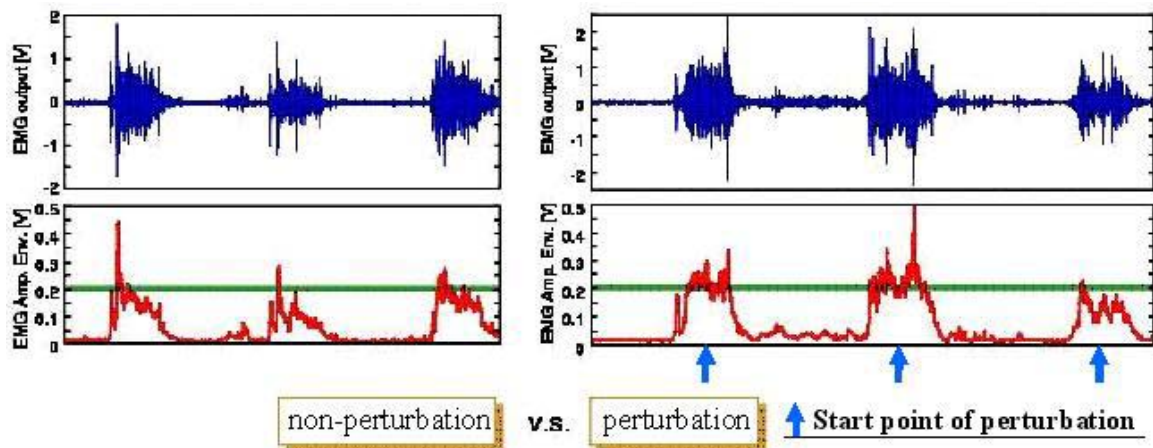


図 3.8 下唇下制筋の筋電の波形信号とエネルギー (XL)。左図：聴覚フィードバックに変形なし。右図：聴覚フィードバックに変形あり

3.2.4.2 音声スペクトルの分析結果

従来、先行研究で分析対象とされていた音声スペクトルに対しても、摂動ありと摂動なしの場合に分けて比較分析した。スペクトル分析では、窓長 49ms の hamming 窓を用い、LPC により 10ms ごとにフォルマントを求めた。時系列にフォルマント平均値を示したものを図 3.9 に示す。この図から聴覚フィードバックに摂動あり（実線）の場合と摂動なし（点線）の場合、フォルマント軌跡には差が見られた。摂動のある場合、つまり [i]の発話時、途中にフィードバック音声を [y]に切り替えた場合、F2, F3 の周波数が

上がる現象が確認された。この動きは[i]を強調することを示している。

さらに図3.9の摂動ありの値から摂動なし値を引いたフォルマントの差を図3.10に示す。その結果、摂動時に、特にF2に、他のフォルマントに比べて周波数の上昇を確認することができた。

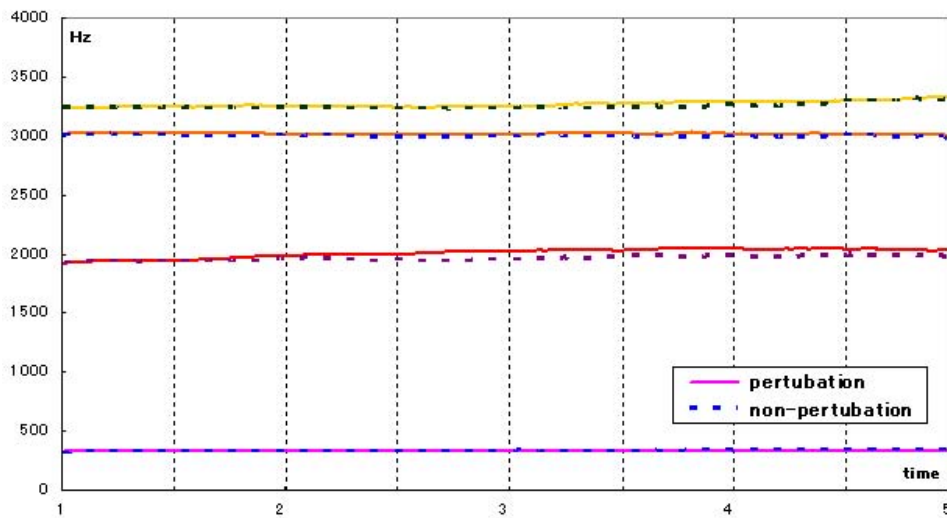


図 3.9 聴覚フィードバックの摂動ありと摂動なしの場合の各フォルマントの軌跡 (XL)。実線：摂動あり，破線：摂動なし

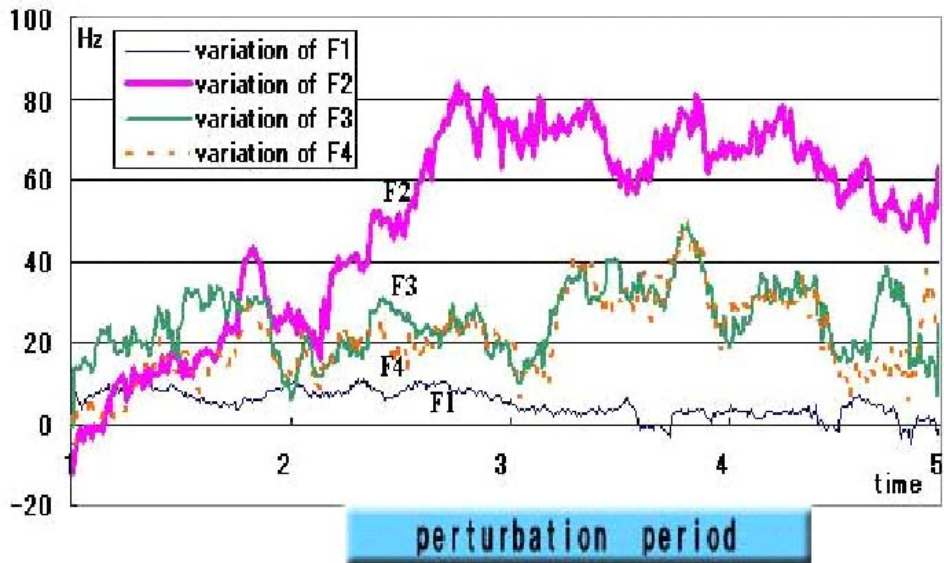


図 3.10 聴覚フィードバックの摂動ありの値から摂動なしの値を引いた各フォルマントの差 (XL)

3.2.4.3 画像の分析結果

上記の筋電信号、音声スペクトルの結果より、被験者は、下唇下制筋を働かせることにより、フィードバック音声の摂動に対して発話に補正動作があることが分かった。下唇下制筋は、下唇を下げる筋肉であることから、被験者は口を開く調音運動により発話を補正しているのであろうか。それとも下唇下制筋に意識を向けて力を入れているだけなのだろうか。

この疑問点を解決するため、口唇間の距離を調音運動の指標として口の開きを考察した。その結果、摂動時に微小な伸縮は見られるものの、はっきりとした傾向は確認できなかった。筋電信号に見られた摂動の影響が調音動作に見られなかったのは、画像の分解能が低いためか、または調音器官には慣性があるため口唇の動きにいたらなかったなどの可能性が示唆される。

3. 3 本実験

3. 3. 1 目的

フォルマント変形聴覚フィードバック音声を呈示したときの、発話音声、顔面筋電、画像情報を測定し、どのような傾向が見られるのか確認する。また、被験者を追加して測定を行うことで、より定性的な反応の観測を行うことも目的である。

3. 3. 2 実験条件

3.3.2.1 被験者

予備実験と同様に、正常な発話・聴覚能力を有する 29 歳の中国人男性話者 1 名 (PX) を被験者として予備実験を実施した。

3.3.2.2 被験者固有の音響パラメータ

音声フォルマント変形の準備として、予め被験者 PX の音声資料に対するフォルマントを計測した。その結果を表 3.3 に示す。両母音に対して、被験者 PX の F1 における差は 4% 以下、F2 の差は 10% 程度で、F3 の差は 20% 前後となっている。予備実験の被験者 XL の場合、F1 の差は 7% 以下、F2 の差は 15% 程度で、F3 の差は 25% 前後となっており被験者 XL と PX において F1 は 3%、F2 は 5%、F3 は 5% 程度、被験者 XL の方が大きく、この差は個人性と考えられる。

表 3.3 被験者 PX の [i] と [y] のフォルマントの値

母音	[i]			
フォルマント	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
被験者 XL	301	2180	2970	3668
母音	[y]			
フォルマント	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
被験者 XL	289	1965	2381	3340

この音声資料に対して、F1 はそのままにし、[i] から [y] へ F2 と F3 を切り替えてフォルマントの変形を実現する。

3.3.2.3 計測装置

本実験で用いた実験システムは、図 3.1 に示したものであり、この図に記載された計測装置を用いて、4 チャンネルの筋電信号と 2 チャンネルの音声信号画像情報、また 60FPS で画像情報を同期させて記録し、同期信号も収録した。計測装置の概要について説明を行う。

本実験は予備実験と同様、防音室内で行う。被験者により発話された音声は、リアルタイム OS である RT-Linux 上のプログラムにより実時間で変換が行われ、被験者にフィードバックされる。骨伝導や機器などの周囲の雑音をマスクするため、フィードバック音声には 60dB 程度のピンクノイズを付加する。

被験者は、ヘッドホン (HDA-200) を身に着けた状態で、マイクロホン (WM-C70) へ発話を行う。発話された音声はマイクロホン、マイクロホンアンプ (MA-8) を経て、AD 変換ボード (PCI-3155) に入力され、実時間処理によりフォルマントの変換処理が行われ、DA 変換ボード (PCI-3336) を通じて外部に出力される。

DA 変換ボードから出力された音声はミキサ (AT-MX50) によりピンクノイズを付加され、防音室内のアンプ (AU- α 907MR) を経て被験者にフィードバックされる。発話音声、及びフォルマント変形されたフィードバック音声は、AD 変換器を経て同計算機内に記録され、分析に用いる。

筋電信号の測定装置は MME-3116 であり、サンプリング周波数 6kHz、感度 250 μ V/div、フィルタの遮断周波数は高域 5kHz、低域 10kHz で最大 8 チャンネルの信号を測定することができる。

本実験においても、中国語母音 [i] と [y] を音声資料として用いるため、主に頬筋・笑筋、下唇下制御筋、

口輪筋とオトガイ筋の4つの調音筋を観測した(筋の収縮と調音動作との対応関係は表 3.1 を参照). 表面電極とフェイスマーカー配置位置を図 3.11, 図 3.12 に示す. 口唇の動きを観測するため, ビデオメディア (VCC-H8000) により 60FPS で画像情報も筋電信号と同期させて収録し, 同期信号も同時に記録する. また, 一部の表面筋電極の殻はモーショントラッキングのマーカーとしても利用している.

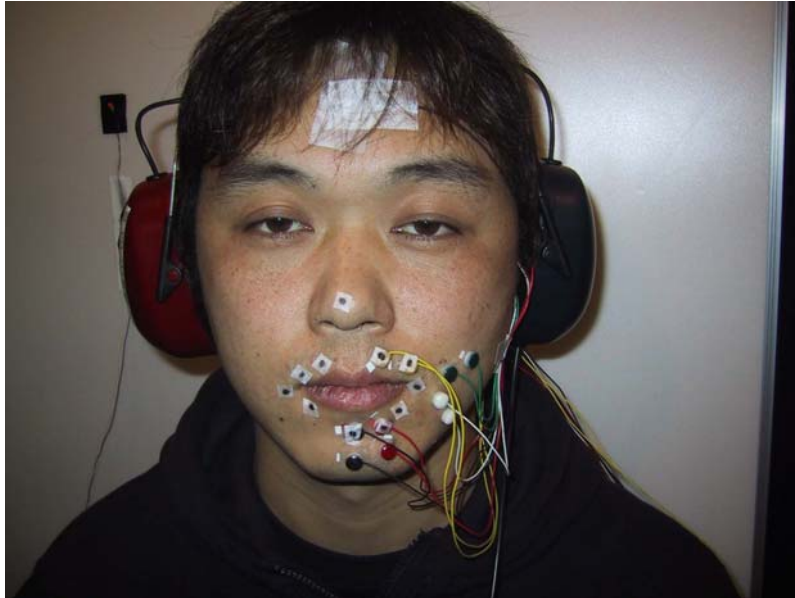


図 3.11 被験者 PX の表面筋電とフェイスマーカー配置位置 (正面)

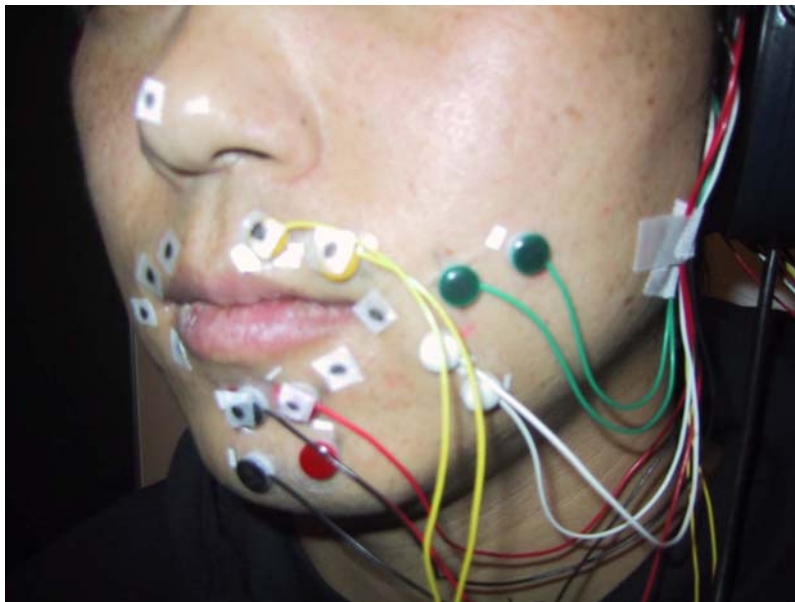


図 3.12 被験者 PX の表面筋電の配置位置 (側面)

3. 3. 3 実験プロセス

実験のプロセスは, 被験者に持続母音[i]を発話するよう指示し, 発話途中で被験者の音声[i]のフォルマントから[y]に変換してフィードバックを行った. それを摂動として被験者の発話途中ランダムに実行する. 発話開始信号 (各 1 秒間) から終了信号までの発話期間を 5 秒とし, 2 秒間の摂動を開始信号後の 2 秒目と 2.5 秒目の 2 パターンで与えた.

これらは全てヘッドホンを通じて被験者に通知され, 骨伝導を抑えるため 60dB のピンクノイズを常にフィードバック音声に付加した. 摂動ありの場合, 摂動を与えるまで被験者の音声をフォルマント変形なしでフィードバックし, その後の 2 秒間でフォルマントの変形による“摂動あり”の音声をフィー

ドバックする。その後再び“振動なし”の音声を呈示する。本実験の振動ありの場合（2パターン）の実験構成を図3.13に示す。

このような1回発話を1トライアルとし、1セットは3トライアルからなる。各トライアル間は5秒間の休憩を挿み、15セット（45トライアル）連続して実験を行った。45トライアル中の17トライアル（約4割）は振動なし、28トライアル（約6割）は振動ありのトライアルをランダムに実行した。

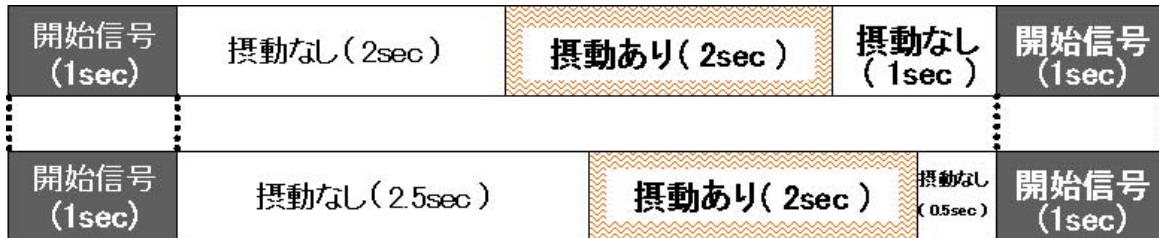


図 3.13 1trial の実験構成（振動ありの場合 2パターン）

3. 3. 4 実験結果

変形聴覚フィードバックを用いた本実験における、筋電信号、音響分析の結果を報告する。

3.3.4.1 筋電信号の分析結果

聴覚フィードバックに変形（振動）がありとなしの場合に収録した下唇下制筋の筋電信号の一例を図3.14に示す。3周期のように見られるのは3トライアルであり、左から順に、2秒目から振動ありの場合、2.5秒目から振動ありの場合、最後に振動なしの場合の筋電信号のエネルギーの時間波形である。薄い実線は開始・終了信号、網掛けの部分には振動区間を表している。

聴覚フィードバックに振動のない場合、筋電信号は発話の最初に活動エネルギーが高く、その後、筋活動は徐々に減っていく。それに対して、聴覚に振動ありの場合、発話の最初に筋の活動が大きく、その活動が一旦減少し、フィードバック音が[i]から[y]に切り替えられた後、話者が[i]を強調するため下唇下制筋はもう一回活発に活動する傾向が見られている。

話者 PX について、2秒目から振動ありの場合には78%、2.5秒目から振動ありの場合には57%、全体の振動ありのトライアルの67%にこのような活動が見られた。

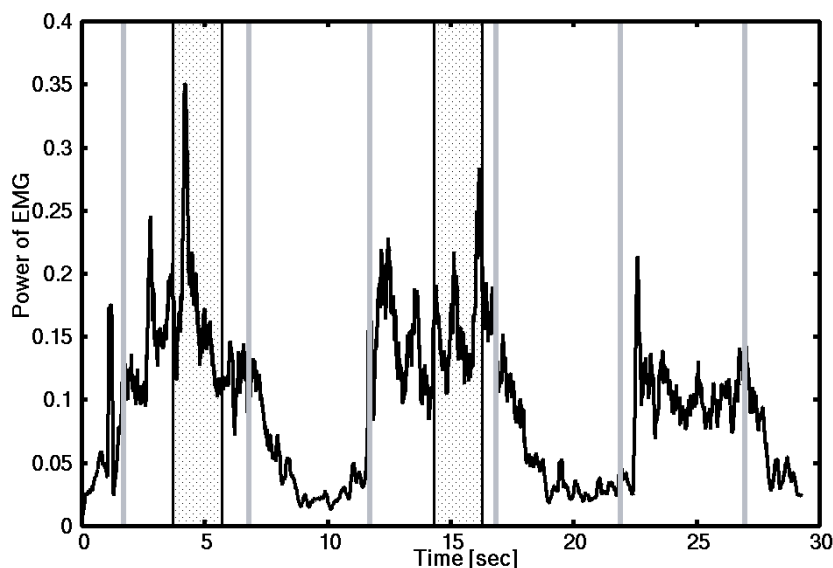


図 3.14 下唇下制筋の筋電信号エネルギー（PX）

左：2秒目から振動ありの場合
中央：2.5秒目から振動ありの場合。右：振動なしの場合

3.3.4.2 音声スペクトルの分析結果

音声スペクトルを分析し、摂動ありと摂動なしの場合で比較した。スペクトル分析では、窓長 49ms の hamming 窓を用い、LPC により 10ms ごとにフォルマントを求めた。得られたフォルマントを摂動ありとなしに分けてそれぞれ平均値を求め、摂動ありの値から摂動なしの値を引いたフォルマントの差を図 3.15 $\text{\ref{fig:2sec_perturbation}}$ 、図 3.16 $\text{\ref{fig:2.5sec_perturbation}}$ に示す。

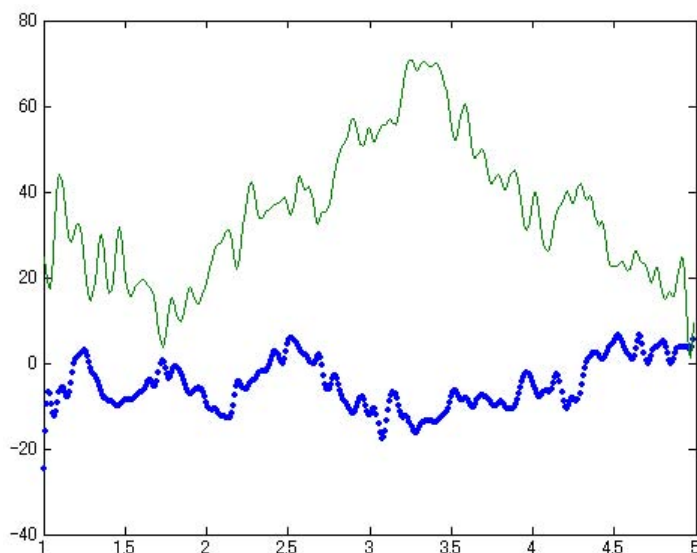


図 3.15 2 秒目から摂動開始のフォルマントの変化量 (PX)

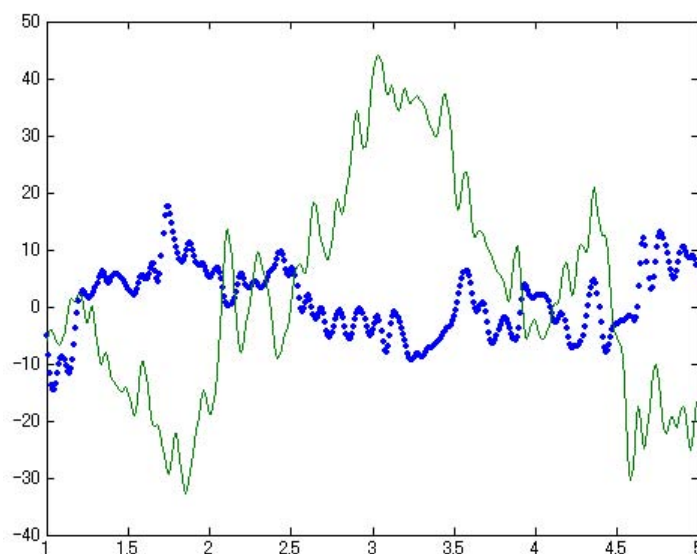


図 3.16 2.5 秒目から摂動開始のフォルマントの変化量 (PX)

3. 4 考察

3. 4. 1 筋電の分析結果から

本研究では、聴覚や発話に問題のない中国人被験者が、5 秒間母音[i]を発話し続けた。予備実験ではトライアルの 7 割に、本実験においてはライアルの約 6 割に、摂動を加えたフードバック音声を被験者の耳に返している。摂動は発話途中の 2 秒間に与えられ、被験者の発話音声[i]から[y]に変換されたものである。口唇の動きをつかさどる口裂周囲の顔面筋の中でも特に、口唇を下に引く動作を促す下唇下制筋が摂動時に活発に活動した。TAF が行われていない場合、筋電信号は始めに大きな値を示し、後になるにつれて小さくなっていく。言い換えると、下唇下制筋は激しい収縮の時に活性化し、時間が経つにつれ少しずつおさまってくる。TAF がおこなわれている場合、各トライアルの開始から設定された時間に

摂動が起きる。筋電の活性は始めは激しく、次第に弱まるが、TAF が加えられると再度活性する。フィードバック音声に変換により[i]と異なってきたとき、発話者は母音[i]を強調するために筋肉の活動を促そうとしたと考えられる。このような現象から、発話生成が知覚と関連づいて制御されていると考察できる。この相互作用は、被験者 XL については、摂動ありのトライアルの 71%に見られ、被験者 PX については、2 秒目から摂動ありの場合には 78%, 2.5 秒目から摂動ありの場合には 57%, 全体の摂動ありのトライアルの 67%に観測された。被験者 2 名を平均して、約 69%の確率で[i]をもう一度強調するような活動が観測された。この現象は、発話時において聴覚フィードバックが働いており、口唇により瞬時的な発話動作の調整がなされているという可能性を示唆している。

3. 4. 2 音響の分析結果から

被験者 XL の場合、摂動をトライアルの 2.2 秒目に 2 秒間加えた。図 3.17 に示したように、摂動の影響によって F2 には他のフォルマントに比べ周波数の上昇が見られた。摂動区間では、変形フィードバックで[i]から[y]への切り替えにより F2 が 200Hz 下がったのに対して話者は、F2 を約 50Hz 上昇させてその摂動を補償することを図った。同様な現象を被験者 PX の実験結果で確認できた。このことは、聴覚フィードバックの働きにより発話系は実時間的に反応していることを示していると考えられる。

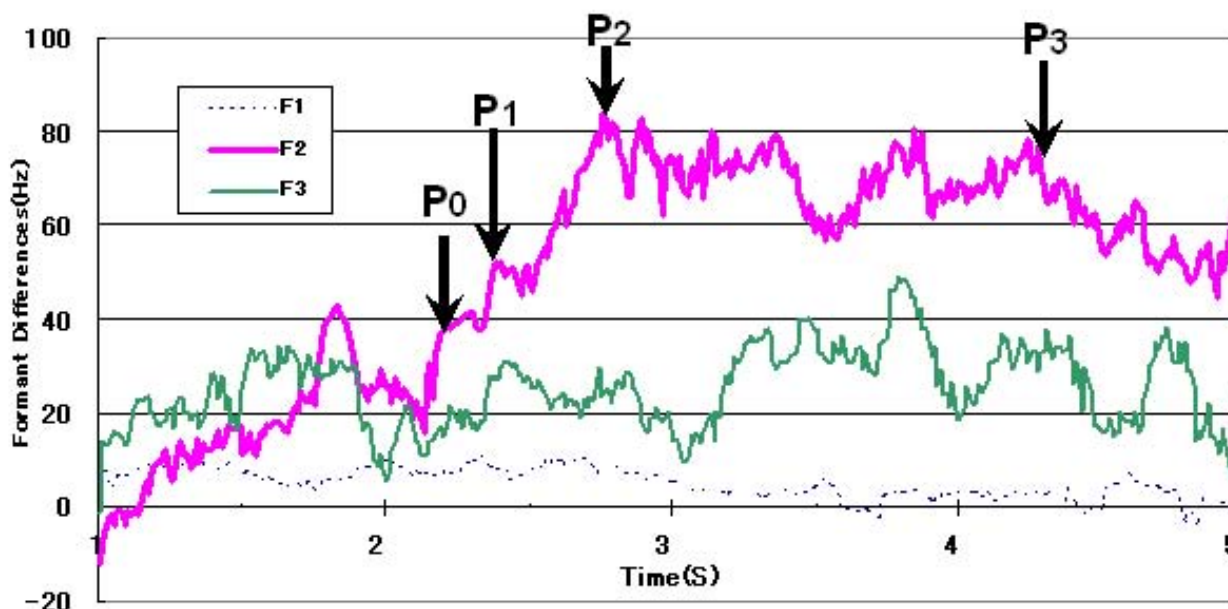


図 3.17 各フォルマントの変化量 (被験者 XL)
P0:摂動開始ポイント. P1:摂動時に起こる補正の開始ポイント.
P2:最大の補正反応ポイント. P3:摂動終了ポイント

また、被験者 XL に関して細かく分析を行った。図 3.17 は、被験者 XL の F1 から F3 まで摂動がありの場合から摂動なしの場合をひいたフォルマントの差である。図を見て分かるように、F2 に大きな違いが見られる。[i]から[y]への変換の際、フォルマントの周波数は F2 では 220Hz, そして F3 では 750Hz 下がっている。フィードバック音声に変換が加えられた場合には、F2 はおよそ 50Hz 上昇した。この値の増加は、フォルマントを上昇させる事でフォルマントの減少の補正を行おうとしていると説明できる。しかし、TAF により被験者 XL の[i]から[y]へ変化を起こすため、F3 を 2990Hz から 2240Hz へと減少させているが、F3 には特に目立った補正はない。

P0 とは 2.2 秒に摂動を行った箇所である。F2 の変化に注目すると、2.35 秒の P1 は、摂動時に起こる補正のスタートポイントであると考えられる。摂動を開始し、F2 に明らかな変化が見られるまでに、およそ 150ms かかっている。この反応時間は、聴覚が様々なプロセスを踏んでいると考えれば妥当な長さである。F2 で起こった最大の補正反応は、開始から、290ms の 2.74 秒目に観測された。この時間は発話調

音器官が本来費やす時間と言われている 330ms [35]とほぼ同じである。また, P3 で表しているように, 4.4 秒に摂動を終了すると, フォルマントは単調なものになった。このことは, 発話生成が音響信号によって制御されていることを暗示している。

3. 4. 3 画像情報の分析結果から

被験者 XL, PX の画像情報について摂動区間を観測した。摂動時, 唇がピクピクと動く補償的な口の開閉は確認できたものの, 摂動に対する随意的な反応であるのか, 人間にとって普遍的に起こる自律的な反応であるのかをはっきりと区別することが出来なかった。今後さらに分析する必要がある。また, 被験者の顔を固定していないため, ノイズが多くなり, はっきりとした傾向が出にくかった可能性も考えられる。

3. 5 実験2ー筋電計測ーのまとめ

本実験では, 聴覚フィードバックの有無を調べるため, 音響レベルの分析だけでなく, 筋電信号, 画像情報などの調音レベルでの分析も同時に行った。発話時に, 摂動を与えたフィードバック音声を被験者に聞かせたところ, 発話を強調する方向への補償動作が筋電信号により確認された。また, 音声スペクトルにも補償的な変動が見られ, 特に F2 に大きく現れていた。筋電信号と音響分析の結果から, 知覚系と発話系とに実時間的な相互作用が存在していることが分かった。聴覚フィードバックにより摂動に対して生成系は実時間レベルな補償を行っていると考えられる。しかし, 口の開きにははっきりとした応答特性は確認できなかった。

これらの結果から, 筋活動から調音運動として表れ出る過程で, 他の調音点での補正が相互に絡み合い, 打ち消されている可能性が考えられる。特にスペクトルの変動, 音声生成の観点からみて, 舌の補償動作が関与している可能性が高い。このため, 我々は EMA (Electro-magnetic Articulography) の分析を加え, 舌の応答特性を統合的に分析するため, 次の実験を行った。

IV. 実験 3 : 舌運動計測

4. 1 調音器官の補償動作の測定
 4. 1. 1 変形聴覚フィードバック実験
 4.1.1.1 実験目的

実験 1 から音響分析では日本母音/e/から/a/, /i/, /u/において補償が確認された。実験 2 では、調音器官の補償について口唇形状との関係のみを考察した。用いた音素は中国母音/i/から/u/であった。この実験から音響分析、口唇の動き共に補償動作が確認されているが、音響分析の結果から補償の起こる割合は 60~70%であり、調音器官の動きにより打ち消されている可能性が示されている。

そこで本実験では、日本母音について舌・口唇のような調音器官における補償動作について測定を行った。

4.1.1.2 実験システム

本研究では、上記の実験要件と実験アプローチを踏まえ、実験 1 で用いたシステムを基本にしてシステムの構築を行った。実験システムの概要図を図 4.1 [¥ref{fig:system.eps}](#)に示す。実験は防音室内で行う。被験者により発話された音声はリアルタイム OS である RT-Linux 上のプログラムにより実時間で変換が行われ、被験者にフィードバックされる。骨導音、自然側音等をマスクするため、フィードバック音声には 60dB 程度のピンクノイズを付加する。

発話音声、及び変換された聴覚フィードバック音声はそれぞれ記録用の計算機で記憶され、分析に用いる。被験者はヘッドホン (ATH-PRO700) を身に着けた状態で、発話された音声をマイクロホン (EGM-G5M)、マイクロホンアンプ (MA-8) を経て、一方は計算機内の AD 変換ボード (PCI-3155) に入力され、もう一方は記録用の計算機に接続された AD 変換機 (DF-2021) を経て記録用の計算機に記録される。計算機内の AD 変換ボードに入力された音声は実時間処理によりフォルマントの変換処理が行われ、DA 変換ボード (PCI-3336) を通じて外部に出力される。

DA 変換ボードから出力された音声は一方はミキサ (AT-MX50) によりピンクノイズが付加され、防音室内のアンプ (AT-α 907MR) を経て被験者にフィードバックされる。もう一方は、AD 変換機を経て記録用の計算機に記録される。

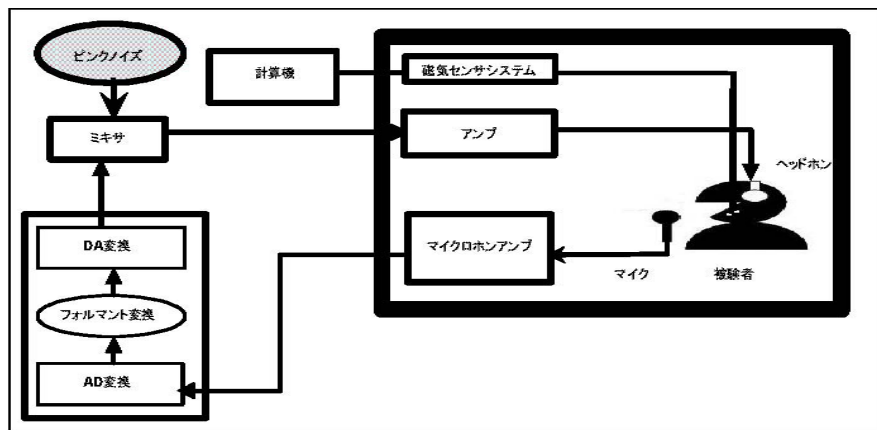


図 4.1 実験システム構成図

4.1.1.3 磁気センサシステム

ここで、本実験で舌・口唇のような調音器官を測定するために用いた磁気センサシステム (Electro-magnetic-Articulography : EMA) について説明する。磁気センサシステムとは、発話器官上の離散的なポイントに磁気センサを貼り付けてその運動を観測する装置である。今回用いた磁気センサシステムは Casterns 社製の AG500 である。このシステムは x,y,z の 3 次元的位置にセンサの前後と左右の傾斜を合わせた 5 次元の測定が可能なシステムである[36]。この磁気センサシステムを図 4.2 に示す。

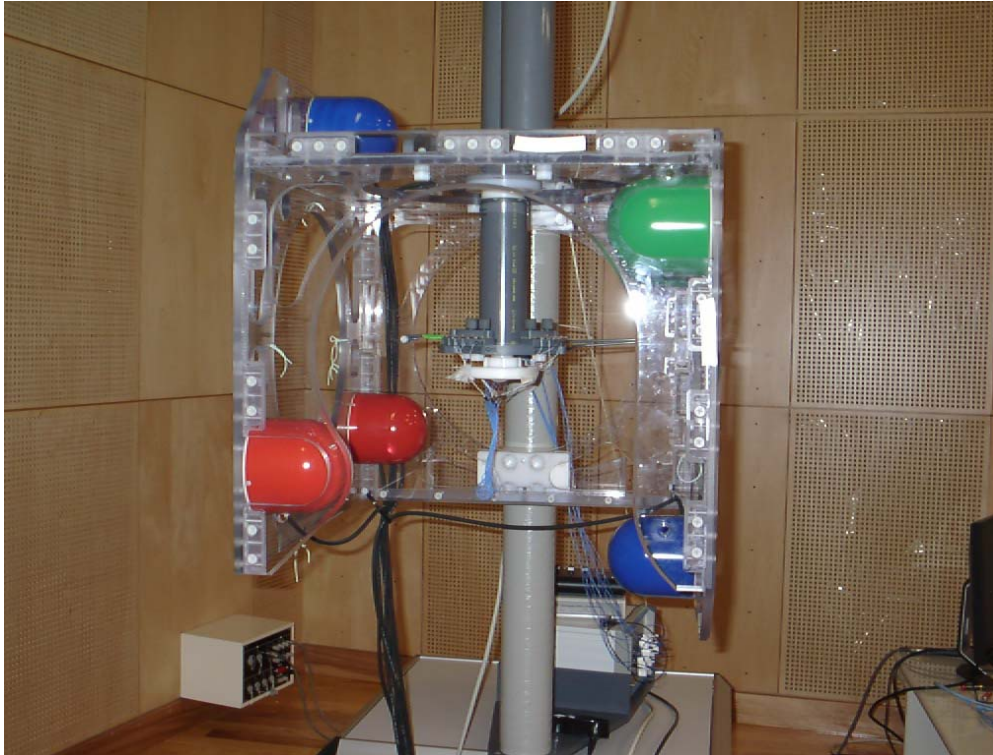


図 4.2 磁気センサシステム

4.1.1.4 実験手順

被験者の舌・口唇に磁気センサを装着し、日本語の中性母音である/e/を発話するように指示し、発話途中で被験者の母音/e/の第1フォルマント (F1) と第2フォルマント (F2) を/a/, /i/, /u/の方向へ0, 20, 40, 60, 70, 80, 100%の割合でそれぞれ変動してフィードバックした。摂動量が0%である場合は/e/, 100%である場合はフィードバック音声は完全に/a/, /i/, /u/となる。それを実現するために、被験者の発話音声/a/, /i/, /u/, /e/を事前に分析してF1とF2を求めた。/e/から/u/への変形を一例として表4.1に示す。前の実験と同様に1回の発話を1トライアルとし、1セットは3トライアルからなる。各摂動量(フォルマント変形)に関して10セット連続して測定し、音声収録はサンプリング周波数16kHzで行った。

表 4.1 /e/→/u/のホルマント変換量

	変形量	F1[Hz]	F2[Hz]
/e/	0%	533	2059
↓	20%	499	1961
↓	40%	466	1863
↓	60%	432	1766
↓	70%	416	1717
↓	80%	399	1668
/u/	100%	365	1570

本実では被験者に図4.3のように磁気センサを装着した。図4.3を正中断面から見ると、磁気センサの位置は図4.4のようになる。また、実験風景を図4.5に示す。磁気センサシステムによる調音運動はサンプリング周波数200Hzで行った。

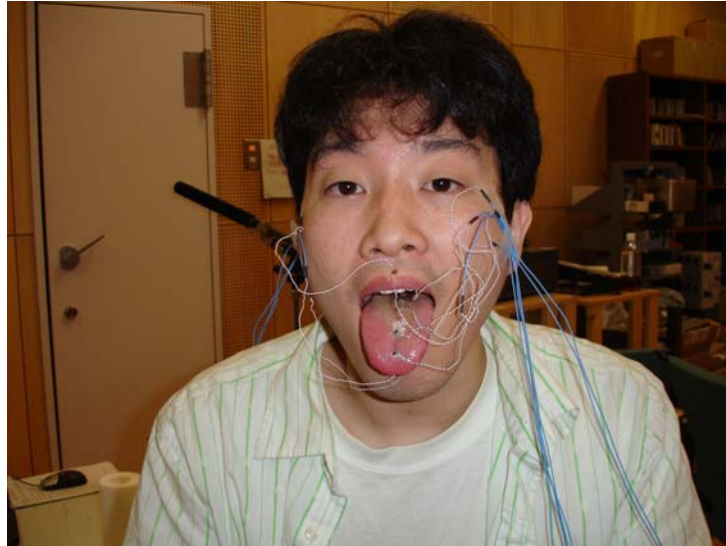


図 4.3 磁気センサ装着図

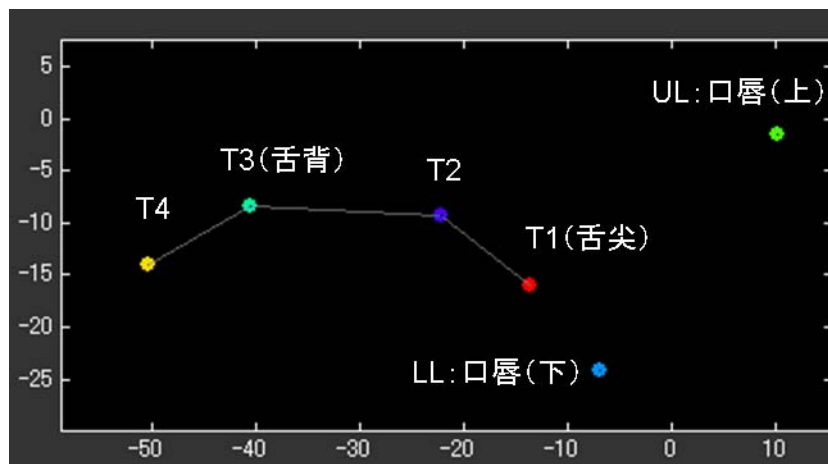


図 4.4 磁気センサの位置



図 4.5 実験風景

4.1.1.5 分析方法

本実験では、収録した音声による音響分析と磁気センサシステムで測定した調音運動について分析を行った。音響分析については、補償動作を示すベクトルとの相関が 0.3 以上のものを補償動作とする。また、調音運動の分析に関しては主成分分析を行った。しかし、/a/ の摂動を与えた場合の測定では、正常な測定結果である図 4.6 $\ref{fig:right.eps}$ に比べ、図 4.7 $\ref{fig:wrong.eps}$ のような磁気センサの一部が舌の動ける範囲とは考えられない動きを見せる正常とはいえない結果が得られなかったため、/a/ の測定結果を除く /i/, /u/ を摂動として与えた場合の測定結果を分析に用いた。なお、主成分分析については次に述べる。

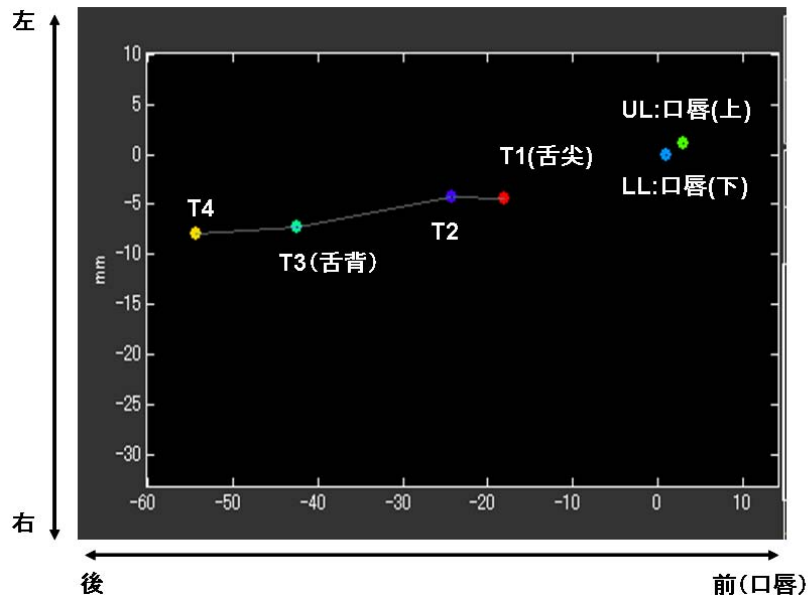


図 4.6 正常な測定結果

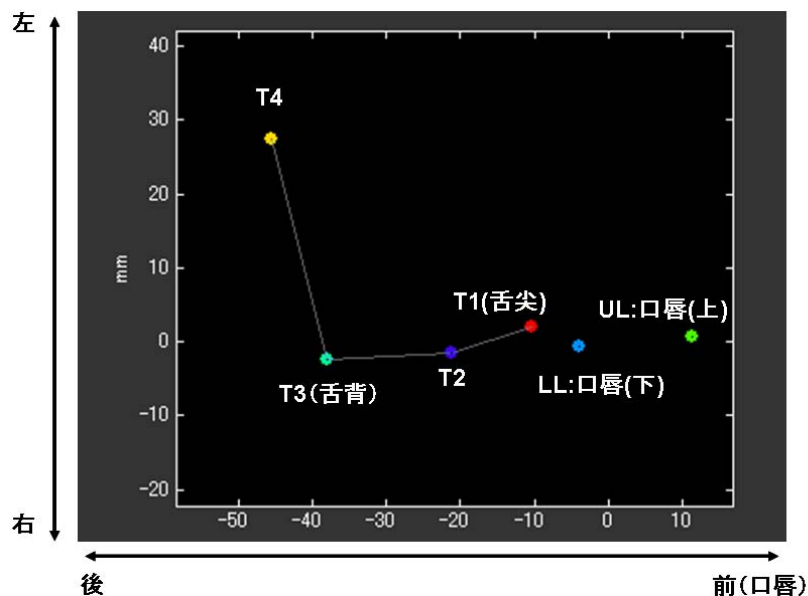


図 4.7 異常な測定結果

4.1.1.6 主成分分析

本実験の調音器官の動きの分析に用いた主成分分析について説明する。主成分分析 (PCA: Principal Component Analysis) とは、ある対象について観測された多数の項目 $(x_1, x_2, x_3, \dots, x_p)$ を個々に分析する

のではなく、それらの項目全体が織り成す意味合いを解釈するために使用する手法である。項目を個々に分析する方法は一変量解析と呼ばれ、個々の項目ごとに平均や分散を求め各項目の持つ意味を解釈するが、主成分分析は多数の項目が全体として持っている意味合いを解釈するために使用するもので多変量解析の一方法論である。

4.1.1.7 実験結果

まず、/e/から/i/へ摂動を与えた際の音響分析の結果を図 4.8 に示す。その中から、補償動作があると判断したものを図 4.9 に示す。補償動作の起こる確率は約 31.7%となっている。また、/e/から/u/へ摂動を与えた場合には、図 4.10 より補償動作が約 61.7%の確率で起こっている。/e/から/i/の場合では被験者が摂動につられてしまう場合が多く、/e/から/u/の場合では、実験 1 の結果と同等の割合（60~70%）で補償動作を確認できた。

また、磁気センサシステムの測定結果から、各磁気センサの x 方向（前後）、z 方向（上下）における主成分分析による各主成分の固有ベクトルを図 4.12、図 4.13、図 4.14 に示す。これらの T1~LL は図 4.4 の装着した磁気センサの位置に対応している。この 3 つの図より、第 1 主成分は舌後部の上下の動き、第 2 主成分は下唇の前後、舌尖の上下、舌背の上下の動き、第 3 主成分は舌前部の前後の動きの指標となっている。更に、図 4.15、図 4.16、図 4.17 に/e/から/i/へ、図 4.18、図 4.19、図 4.20 に/e/から/u/へ摂動を与えたときの各主成分の関係を示す。ここで、図 4.15~図 4.20 のベクトルは全トライアルデータに対して求めた各主成分の係数を用いて、各トライアルの摂動前の平均係数が始点、摂動中の平均係数が終点となっている。3 つの図において各主成分が示す正の方向は、第 1 主成分が舌後部が上方へ、第 2 主成分は下唇が前方、舌尖が上方、舌背が下方へ、第 3 主成分は舌前部が前方へ動いていることを示す。よって、摂動が/i/である場合、補償動作は第 1 主成分が負、第 2 主成分が正、第 3 主成分が負の方向へそれぞれ動いた時となり、摂動が/u/である場合、補償動作は第 1 主成分が負、第 2 主成分が正、第 3 主成分が正の方向へそれぞれ動いた時となる。また、補償動作は各主成分の補償動作の方向を示す単位ベクトルから相関が 0.3 以上になるものとした。その結果、/i/へ摂動を与えた場合には、図 4.15 では全トライアル中の約 14.7%、図 4.16 では約 20.2%、図 4.17 では約 20.2%となっている。/u/へ摂動を与えた場合には、図 4.18 は全トライアル中の約 46.4%、図 4.19 では約 58.8%、図 4.20 では約 54.6%の割合で現れた。また、この 3 つの場合のいずれかで補償が生じる割合は約 79.4%となった。それに対して、/i/の摂動に対する補正動作が生じる割合は約 44.0%となった。以上の結果をまとめたものを表 4.2 に示す。

表 4.2 音響分析と主成分分析の結果

摂動	/i/	/u/
音響分析で補償動作の起こる確率	31.7%	61.7%
第 1, 第 2 主成分で補償動作の起こる確率	14.7%	46.4%
第 1, 第 3 主成分で補償動作の起こる確率	20.2%	58.8%
第 2, 第 3 主成分で補償動作の起こる確率	20.2%	54.6%
3 つのいずれがで補償動作の起こる確率	44.0%	79.4%

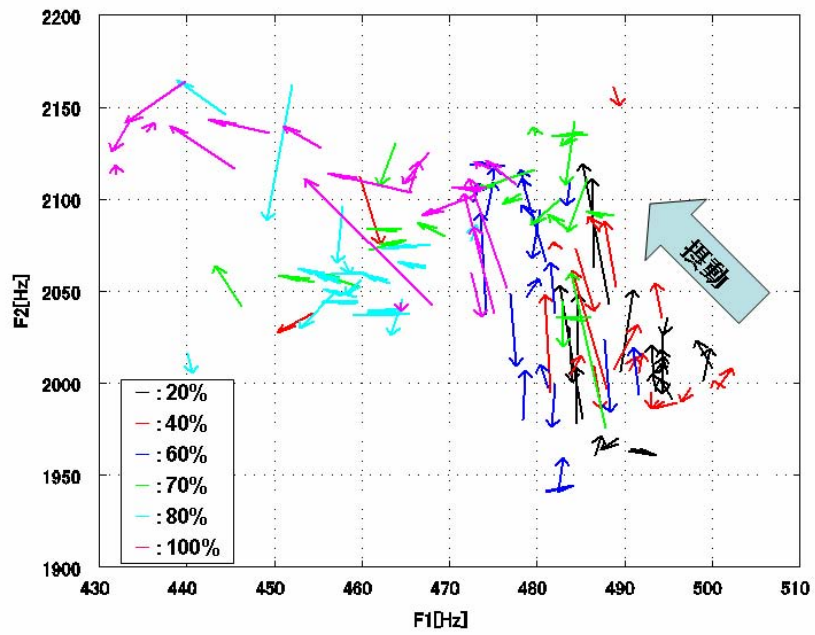


図 4.8 擾動を与えた場合の反応 (/e/→/i/)

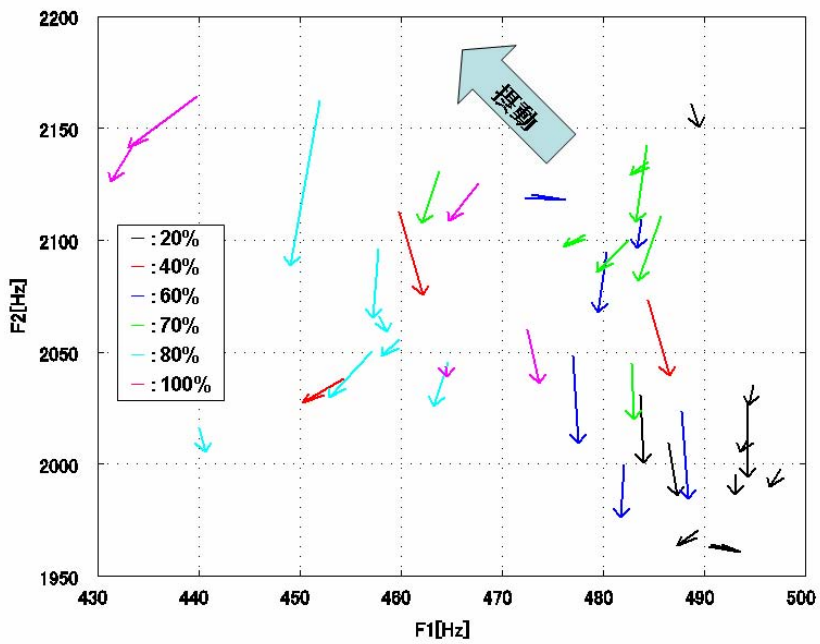


図 4.9 補償動作ありの場合 (/e/→/i/)

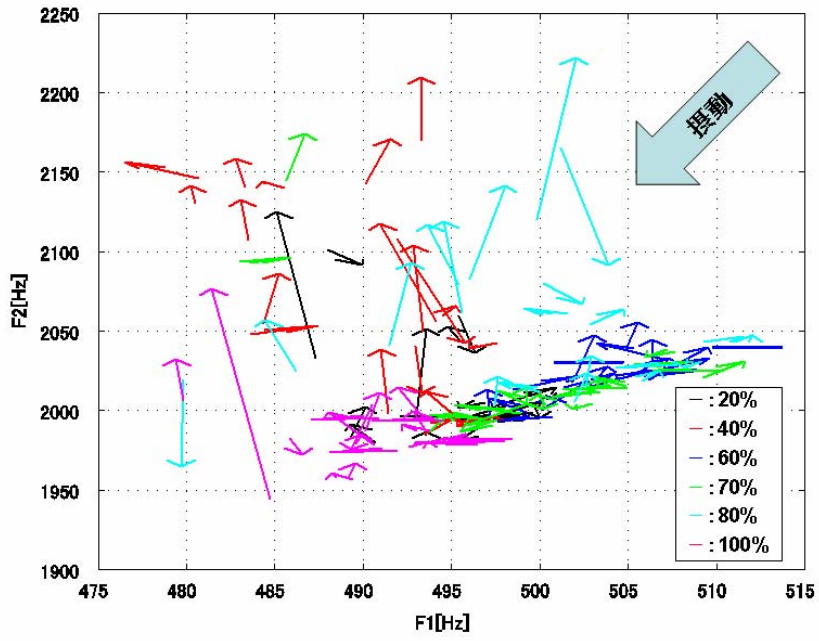


図 4.10 擾動を与えた場合の反応 (/e/→/u/)

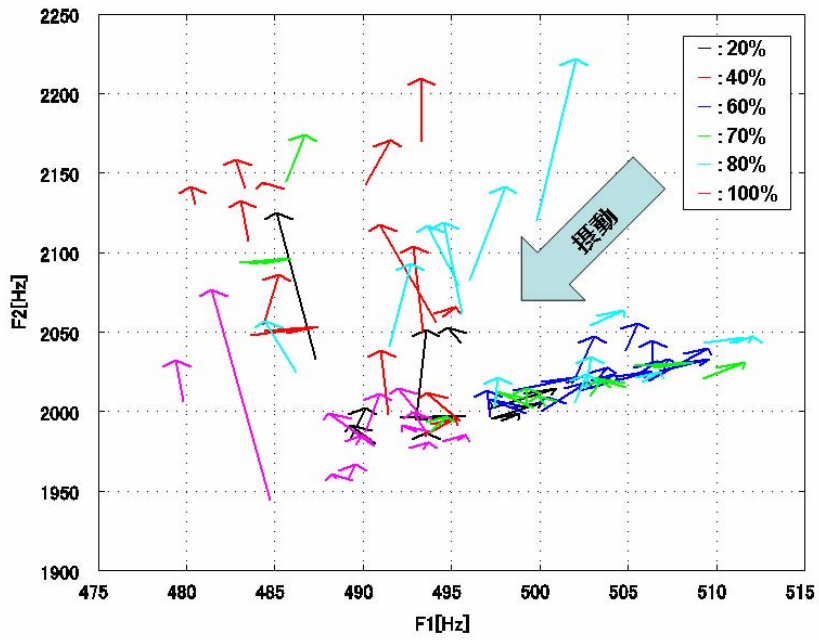


図 4.11 補償動作ありの場合 (/e/→/u/)

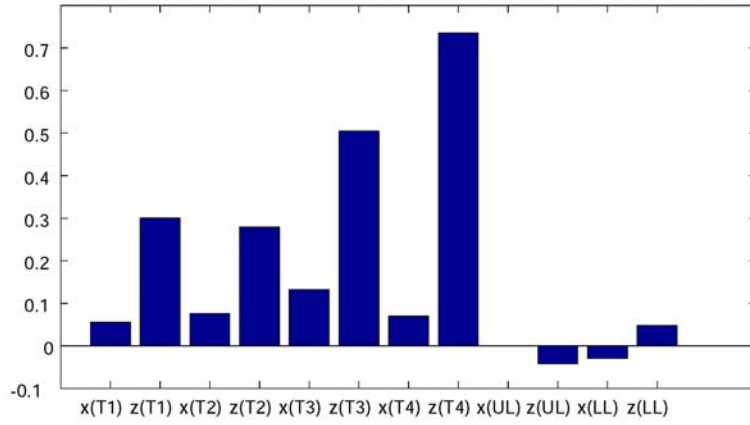


図 4.12 第 1 主成分の固有ベクトル

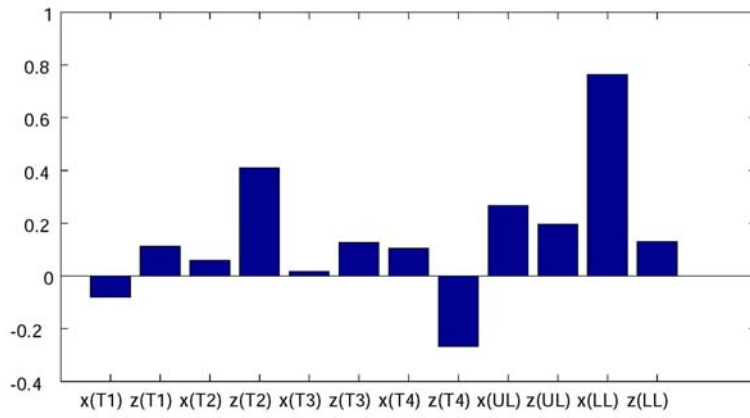


図 4.13 第 2 主成分の固有ベクトル

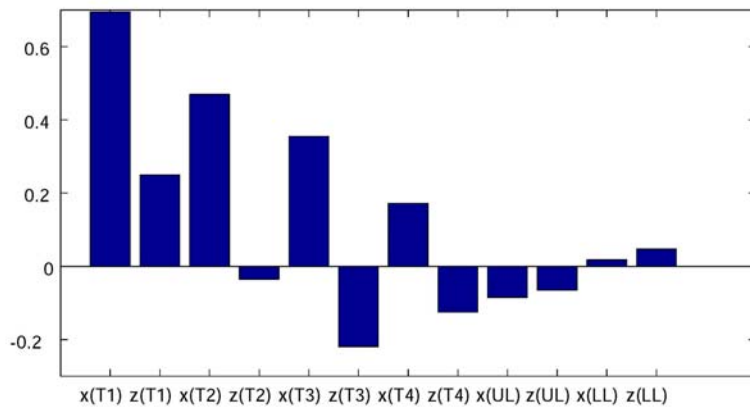


図 4.14 第 3 主成分の固有ベクトル

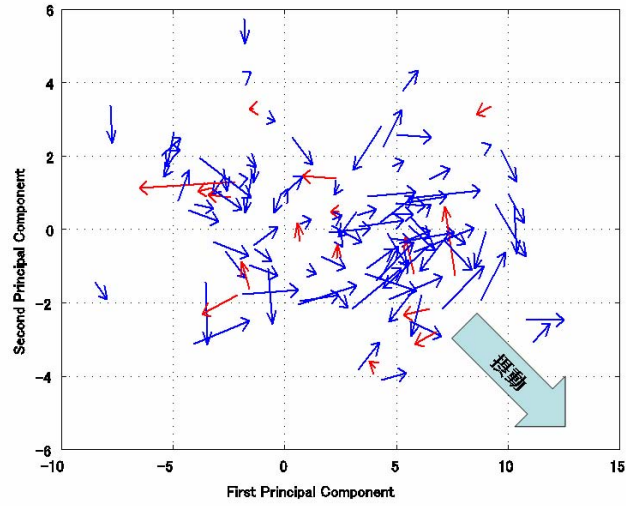


図 4.15 I/I の第 1, 第 2 主成分の関係 (赤: 補償あり, 青: 補償なし)

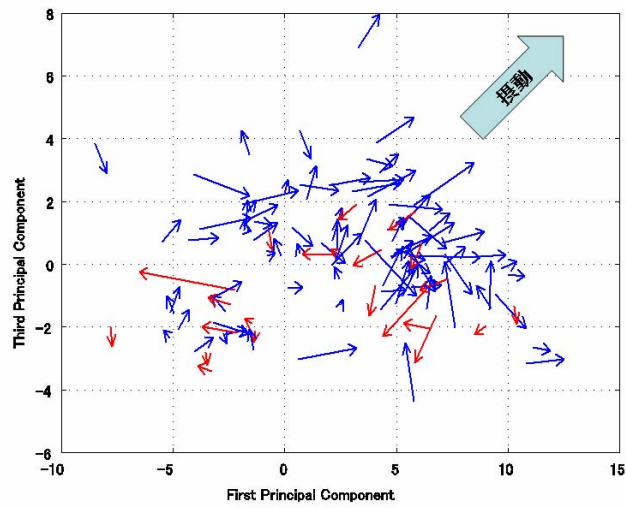


図 4.16 I/I の第 1, 第 3 主成分の関係 (赤: 補償あり, 青: 補償なし)

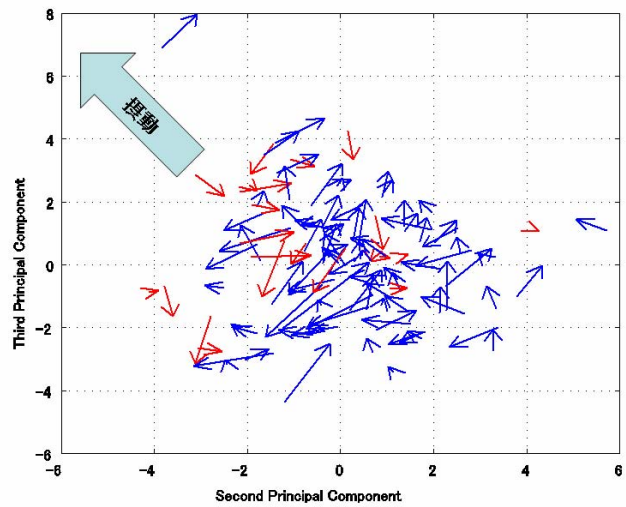


図 4.17 I/I の第 2, 第 3 主成分の関係 (赤: 補償あり, 青: 補償なし)

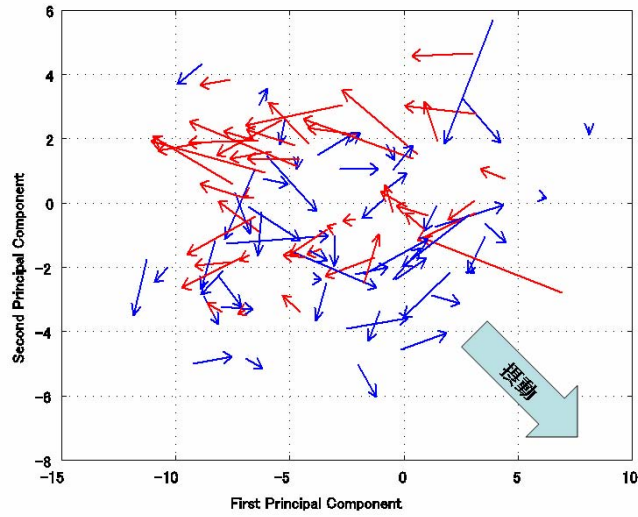


図 4.18 /u/の第1,第2主成分の関係 (赤:補償あり, 青:補償なし)

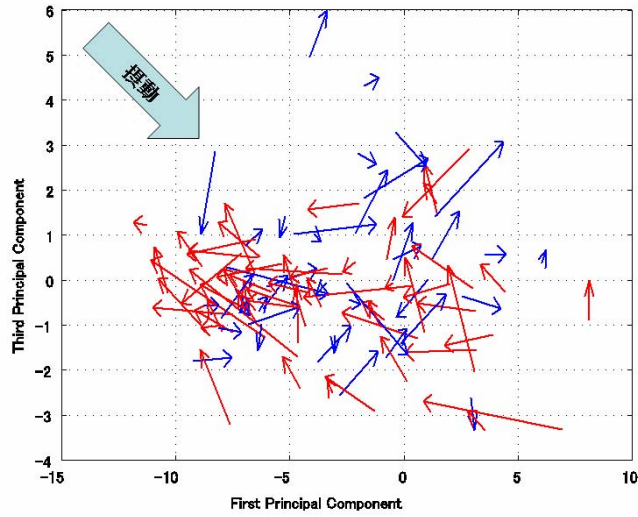


図 4.19 /u/の第1,第3主成分の関係 (赤:補償あり, 青:補償なし)

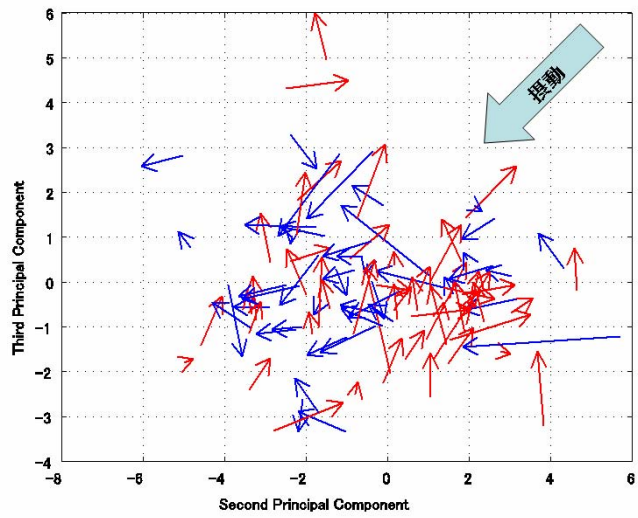


図 4.20 /u/の第2,第3主成分の関係 (赤:補償あり, 青:補償なし)

4. 2 考察

一般的に図 4.21 のように/e/は/i/に比べ舌の調音位置が後方かつ下方に移動する。また、/e/は/u/に比べ舌の調音位置が前方かつ下方に移動する。従って、/e/の発話時に/i/の方向へ摂動を与えた時に舌の後方かつ下方への動きが見られたとき、また同様に/u/の方向へ摂動を与えたときに舌の前方かつ下方への動きが見られたときに、それぞれ補償動作があったと判断した。

今回、主成分分析を用いて各摂動に対する舌全体の補償動作について調査を行った。図 4.18～4.20 より、各主成分の関係から、/u/へ摂動を与えた際の補償動作を確認することが出来た。この補償動作は第 1, 第 3 主成分での動きが大きく、第 2 主成分についての動きが小さいことから舌の上下前後の変化が大きく、口唇形状の変化が小さいことがわかる。よって/u/へ摂動を与えた場合には舌に補償動作が起こっていることが示唆される。また、舌の下方かつ前方、下方のみ、前方のみのいずれかの補償動作が起こっている割合は約 79.4%となり、先行研究で行われた音響分析の約 60～70%に比べ高い割合で生じた。これは音声を生成する際に、聴覚系が常にモニタリングをしていることを示唆している。/i/へ摂動を与えた場合に、/u/と比較して、補償動作の割合が低くなっているが、今回の実験結果では、補償動作と逆方向の反応が目立ち、被験者が摂動/i/に対してつられている可能性が考えられる。しかし、舌の下方かつ後方、下方のみ、後方のみいずれかの補償動作が起こっている割合は約 44.0%とチャンスレベルであり、/u/と同様に音響分析の結果に比べ高い割合となっている。今後の課題として被験者を増やし、一般的な傾向を見ていく必要がある。

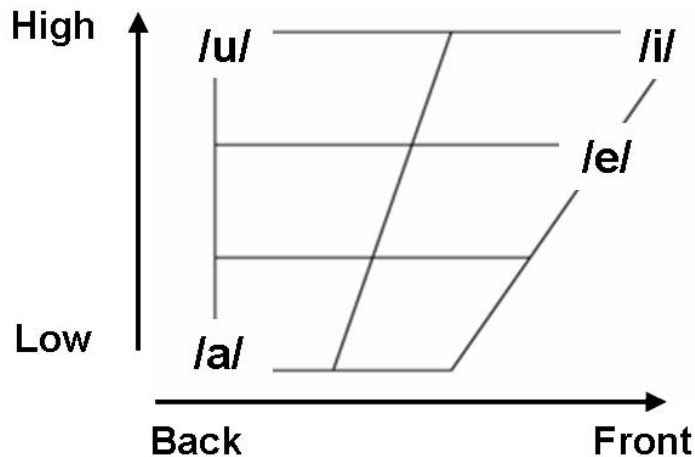


図 4.21 日本母音の調音位置

4. 3 まとめ

本研究では、磁気センサシステムを用いて、舌・口唇のような調音器官の反応を計測した。この位置情報から主成分分析を行った結果、摂動/i/に対しては、被験者が摂動につられている可能性が高い事が音響分析、磁気センサシステムの結果から確認された。しかし、調音器官の補償に関しては、舌の下方かつ後方、下方のみ、後方のみいずれかの補償動作が起こっている割合は約 44.0%とチャンスレベルであった。摂動/u/に対しては、図 4.15～図 4.17 各主成分の関係から、/u/へ摂動を与えた際の補償動作を確認することが出来た。この補償動作は第 1, 第 3 主成分での動きが大きく、第 2 主成分についての動きが小さいことから舌の上下前後の変化が大きく、口唇形状の変化が小さいことがわかる。よって/u/へ摂動を与えた場合には舌に補償動作が起こっていることが示唆できる。音響分析では 61.7%と先行研究と同程度 (60～70%)の割合で補償動作が確認されたが、磁気センサシステムからは、舌の下方かつ前方、下方のみ、前方のみのいずれかの補償動作が起こっている割合は約 79.4%となり、音響分析に比べ高い割合で補償動作を確認できた。

これは音声を生成する際に、聴覚系が常にモニタリングをしていることを示唆している。/i/へ摂動を与えた場合に、/u/と比較して、補償動作の割合が低くなっているが、今回の実験結果では、補償動作と逆方向の反応が目立ち、被験者が摂動/i/に対してつられている可能性が考えられる。

V. 結言

5. 1 本研究のまとめ

本研究では、音声生成と音声知覚の密接な関係を示す一例として「聴覚フィードバック」を取り上げ、知覚・生成の相互作用の解明を図ることを目的として、音響特徴のみならず生理指標をも計測することを試みた。

「聴覚フィードバック」とは、話者が発した声を自分の耳で聞き取りながら発話機構を制御するための音声によるフィードバックのことであり、これを用いて話者は正確な発話を行っている。「聴覚フィードバック」が発話に重要な役割を演じているのならば、フィードバックされる音声を人為的にリアルタイムで変形し耳に呈示した場合、その変形によって音声生成に何らかの影響があるはずである。影響は、音声知覚機構、音声生成タスクプランニング機構、音声生成機構（調音器官）の順に伝播する。

本研究では、リアルタイムで変形されたフィードバック音声の変形量と生理的指標の関係および生成された音声との関係を調べることにより、音声知覚とタスクプランニングも含めた音声生成、およびこれらの相互作用を議論した。

5. 1. 1 音響計測の結果

本研究では、持続母音/e/の発話途中に F1 と F2 に摂動を加えて、それぞれ/a/の方向へ変動してフィードバックし、生成される音声の特徴の変化を計測した。その結果、発話を強調する方向への補正動作が確認されたが、補正反応が起きた割合に差異が生じた。そこで次に、F1 と F2 をそれぞれ/a/,/i/,/u/の方向へ 20,40,60,70,80,100 %の割合でそれぞれ変動してフィードバックをして、変形聴覚フィードバックの摂動量と補正反応との関連を調べた。その結果、被験者により差異はあるが、全体的に摂動量による音韻の曖昧さが大きくなる場合補正反応は出やすいが、摂動量が大きすぎると補正が起きにくいことがわかった。また、補正反応の頻度は母音に依存している。また、F1 における摂動量と補正反応が生じた割合と F2 のそれとに差異が生じたことより、F1, F2 それぞれにおける補正には別々の調音器官がはたらく可能性が考えられた。

5. 1. 2 筋電計測の結果

本研究では、実時間における聴覚フィードバックの有無を調べるため、音響レベルの分析だけでなく、筋電信号、画像情報などの調音レベルでの分析も同時に行った。発話時に、摂動を与えたフィードバック音声を被験者に聞かせたところ、発話を強調する方向への補償動作が筋電信号により確認された。また、音声スペクトルにも補償的な変動が見られ、特に F2 に大きく現れていた。筋電信号と音響分析の結果から、知覚系と発話系とに実時間的な相互作用が存在していることが分かった。聴覚フィードバックにより摂動に対して生成系は実時間レベルな補償を行っていると考えられる。しかし、口の開きにははっきりとした応答特性は映像では確認できなかった。

これらの結果から、筋活動から調音運動として表れ出る過程で、他の調音点での補正が相互に絡み合い、打ち消されている可能性が考えられる。特にスペクトルの変動、音声生成の観点からみて、舌の補償動作が関与している可能性が高い。

5. 1. 3 舌運動計測の結果

本研究では、日本語の中性母音/e/の発話途中に F1 と F2 を変形した上で摂動として与え、舌運動へ与える影響の計測を行った。

舌の位置情報から主成分分析を行った結果、摂動/i/に対しては、被験者が摂動につられている可能性が高い事が音響分析、磁気センサシステムの結果から確認された。しかし、調音器官の補償に関しては、舌の下方かつ後方、下方のみ、後方のみ、後方のみ、いずれかの補償動作が起こっている割合は約 44.0% とチャンスレベルであった。摂動/u/に対しては、/u/へ摂動を与えた際の補償動作を確認することが出来た。この補償動作は第 1, 第 3 主成分での動きが大きく、第 2 主成分についての動きが小さいことから舌の上下前後の変化が大きく、口唇形状の変化が小さいことがわかる。よって/u/へ摂動を与えた場合には舌に補償動作が起こっていることが示唆できる。音響分析では 61.7% と先行研究と同程度(60~70 %)の割合で補償動作が確認されたが、磁気センサシステムからは、舌の下方かつ前方、下方のみ、前方のみ、いずれかの補償動作が起こっている割合は約 79.4%となり、音響分析に比べ高い割合で補償動作を確認

できた。

これは音声を生成する際に、聴覚系が常にモニタリングをしていることを示唆している。/i/へ摂動を与えた場合に、/u/と比較して補償動作の割合が低くなっているが、今回の実験結果では、補償動作と逆方向の反応が目立ち、被験者が摂動/i/に対してつられている可能性が考えられる。

5. 2 今後の課題

本研究は、発話生成と知覚間でどのような情報伝達が行われているのかについて説明するための TAF を用いた実験である。これまで行ってきた TAF 実験による結果により、人間が行う発話伝達は、例えば“運動野と言語野間の位相的なマッピング”などの効率的な方法で行われているということ仮定することができる。我々は、この変形聴覚フィードバック実験観察とともに、生理学的な調音モデルを用いたシミュレーションを用いて、脳内のこのような仮説について検証を行っている。このように今後は実験の結果だけでなく、様々なシミュレーションも行い、両方の結果を考察することによって、音声知覚と生成の相互作用について検討を行っていく必要があると考える。

5. 2. 1 筋電測定に関して

本研究では、予備実験と本実験の両実験において、摂動時に、被験者の下唇下制筋がもう一度活発に活動する現象を確認したが、この実験における筋電信号の定量的な分析手法を見つけることはできなかった。今後は筋電信号における定量的な分析方法の確立が必要であると考えられる。

実験を行う前に control 群として、以下の 5 つの条件での筋電測定を行った。

1. 何もしないもの（自律的な動きがないかを確認するため）
2. 普通の[i] の発話（フィードバック実験条件下でないもの）
3. 普通の[y] の発話（フィードバック実験条件下でないもの）
4. 最大に顔面筋を活動させた最大筋活動時の[i]
5. 最大に顔面筋を活動させた最大筋活動時の[y]

両母音において普通の発話での筋電位を測定したのは、フィードバック実験条件下で、摂動に対する反応がみられたとしても、それが被験者の通常発話の癖ではないのか確認するためである。最大の筋活動（Maximum voluntary contraction: MVC）を測定したのは、筋活動の積分値の比を求めるため、MVC を基準として定量的な分析が可能だと考えたからである。

しかし、運動器官に直接負荷を与える実験ではなく、刺激は聴覚からの信号であるため、この実験において筋の MVC データから比較検討はできない。そのため、一回一回の発話における筋電位について検討することにした。始めの発話時の筋電位は一般的には一番高く、その箇所筋電位と摂動区間における筋電位がどれだけの割合かによって、摂動による反応があるのかないのかを判定した。今回分析した手法は、始めの筋電位のピークの平均を 1 とし、摂動区間内の筋電位のピークの平均が 1 以上であれば、摂動ありと判断した。このような分析手法では、摂動における定性的な応答特性を確率により示すことは可能であるが、定量化することはできない。そこで今後は、このような SN 比の高い筋電信号を定量的に分析する手法が必要だと考える。

従来、変形聴覚フィードバック実験において、主に音声スペクトルに着目して分析が行われてきたが、十分な結果がでなかった。これは、フィードバック音の摂動に対する補償的な応答が、音声生成過程を経る際、調音器官内で打ち消されているため、音声信号として放出された時には、変化があらわれなからだと考えられてきた。このような考察から、音声のスペクトルに加え、筋電信号（EMG）や画像情報の観測も行ってきた。しかし、筋電信号はあくまで音声信号に付随した形で分析に用いられることが望ましいと考える。筋電信号は、自律レベルでの活動が大きいいため SN 比が高く、摂動に付随して発生した補償的な活動であるため定量的に評価することが難しいと思われるからである。今後、他の調音器官の観測も同時に行われ、女性の被験者による実験など、データの数が増えることで、より定量的な分析方法を確立することが課題である。

5. 2. 2 舌運動測定に関して

/i/へ摂動を与えた場合に、/u/と比較して、補償動作の割合が低くなっているが、今回の実験結果では、補償動作と逆方向の反応が目立ち、被験者が摂動/i/に対してつられている可能性が考えられる。しかし、舌の下方かつ後方、下方のみ、後方のみいずれかの補償動作が起こっている割合は約 44.0%とチャンスレベルであった。本実験から調音器官に関する分析方法を確立することができたので、今後の課題として被験者を増やし、一般的な傾向を見ていく必要がある。また、実時間の反応を計測する際に、被験者へ摂動を与える時間（本研究では 2 秒間）など、実験パラダイムについても更に検討していく必要があると考えられる。

<参考文献>

- [1] Denes,P. and Pinson,E. "The Speech Chain, 2nd Ed.", New York : W.H Freeman and Co. (1993)
- [2] Liberman,A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M., "Perception of the speech code", Psych.Rev.,74(6),pp.853-870. (1967)
- [3] Liberman,A.M. and Mattingly, I.G., "The motor theory of speech perception revised", Cognition,21,pp.1-36. (1985)
- [4] Lombard, E., "Le signe de l'elevation de la voix", Annuals Maladies Oreille, Larynx, Nez, Pharynx,37,pp.101-119.(1911)
- [5] Lane, H. and Tranel, B., "The Lombard sign and the role of hearing in speech", Journal of speech and Hearing Research, 14, pp.677-709. (1971)
- [6] Lee, B.S., "Effect of Delayed speech feedback", Journal of the Acoustical Society of America, 22, pp.824-826. (1950)
- [7] Kawahara, H., "Transformed auditory feedback: Effects of fundamental frequency perturbation", Journal of the Acoustical Society of America, Vol.94, No3, Pt.2, p.1883. (1993)
- [8] Kawahara, H., "Interactions between speech production and perception under auditory feedback perturbations on fundamental frequencies", J.Acoust. Soc. Jap, Vol.15, pp.201-202 (1993)
- [9] 河原., "音声知覚・生成相互作用の伝達特性について", 音響学会聴覚研究会資料, H-95-35, pp.223-226.(1995)
- [10] 切替, 藤村, 神山, 戸塚. (1966). "話しことばの科学 その物理学と生物学," 東京大学出版会
- [11] Cowie, R. & Douglas-Cowie, E. (1983). "Speech production in profound postlingual deafness." In M. Lutman, ¥& M. Haggard (eds.), Hearing science and hearing disorders. London : Academic Press, pp.183-230.
- [12] Lane, H. & Webster, J. (1991). "Speech deterioration in postlingually deafened adults," Journal of the Acoustical Society of America, 89, pp.859-866.
- [13] Levitt H, Stromberg H, Smith C, Gold T.(1980). "The structure of segmental errors in the speech of deaf children," 1980 Dec; 13(6) : pp.419-41.
- [14] Osberger, M., & McGarr, N. (1982). "Speech production characteristics of the hearing impaired," Speech and Language: Advances in Basic Research and Practice, Vol. 8, pp.221-283.
- [15] Smith CR. (1975). "Interjected sounds in deaf children's speech," 1975 Jun;8(2): pp.123-8.
- [16] Perkell, J. (1996). "Articulatory processes," In The handbook of phonetic science, Academic Press, pp.221-284.
- [17] Houde, J., & Jordan, M. (1998). "Sensorimotor adaptation in speech production," Science, 279, pp.1213-1216.
- [18] Houde, J., & Jordan, M. (2002). "Sensorimotor adaptation of speech I: Compensation and Adaptation," Journal of Speech Language, and Hearing Research, 45, pp.295-310.
- [19] Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., & O'Connell, M., et al. (1997). "Changes in sound pressure and fundamental frequency contours following changes in hearing statue," Journal of the Acoustical Society of America, 101, pp.2244-2253.
- [20] 佐藤. (2003). "スペクトル変型聴覚フィードバックによる音声生成・知覚の相互作用に関する研究," 北陸先端科学技術大学院大学.
- [21] 斎藤. (2004). "音声生成過程におけるフォルマント変換音声フィードバックの影響に関する研究,"

北陸先端科学技術大学院大学.

[22] Dipellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti G. (1992). "Understanding motor events - a neurophysiological study," *Exp Brain Res* 91, pp.176-180.

[23] Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G. (1996). "Action recognition in the premotor cortex," *Brain* 119, pp.593-609.

[24] Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L. (1996). "Premotor cortex and the recognition of motor actions," *Cognitive Brain Res* 3, pp.131-141.

[25] Fowler, C. A. (1996). "Listeners do hear sounds, not tongues," *JASA* 99, pp.1730-1741.

[26] Savariaux, C., Perrier, P., and Orliaguet, J.(1995). "Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production," *J. Acoust. Soc. Ame.*, 98(5), pp.2428-2442.

[27] Honda, M., Fujino, A., and Kaburagi, T.(2002). "Compensatory responses of articulators to unexpected perturbation of the palate shape," *J. Phonetics*, 30, pp.281-302.

[28] Honda, K. (1996). "Organization of tongue articulation for vowels," *J. Phonetics*, 24, pp.39-52.

[29] Callen, D.E., Kent, R.D., Guenther, F.H., & Vorperian, H.K. (2000) "An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system," *Jornal or Speech, Language, and Hearing Research*, 43, pp.721-736.

[30] 沢田, 寛. (2003). "聴覚フィードバックに利用される音声情報の物理的特徴," *日本音響学会聴覚研究会資料*, Vol.33, No.2, H-2003-21 pp.117-122.

[31] Shimon Sapir, Elizabeth Derosier, Andrea M. Simonson, and Amy Wohlert. (1990). "Effects of frequency modulated tones and vowel formants on perioral muscle activity during isometric lip rounding," *Jornal of Voice and Hearing*, Vol.4, No.2, pp.152-158.

[32] 甘利, 外山. (2000) "脳科学大辞典," 朝倉書店.

[33] Hollinshead, H. (1982). "Anatomy for Surgeons. Vol.1: The Head and Neck, 3rd Ed." New York: Hoeber Medical Division, Harper and Row Publishers.

[34] 新美, 西尾, JOEL C.KAHANE. (1999). "発話メカニズムの解剖と生理," インテルナ出版.

[35] Masaki, S., & Honda, K. (1994). "Estimation of temporal processing unit of speech motor programming for Japanese words based on the measurement of reaction time," *Proc. ICSLP*, 94, Yokohama Japan, pp.663-666.

[36] <http://www.articulograph.de/>

成果資料

- [1] 松岡, Lu, 党, 赤木(2004). “調音情報を考慮した聴覚系と発話系の相互作用に関する検討”, 音響学会聴覚研究会資料、H-2004-103.
- [2] Matsuoka, R., Lu, X., Dang, J., and Akagi, M. (2004). “Investigation of interaction between speech perception and speech production,” Proc. KIT Int. Sympo. Brain and Language 2004, 27-28.
- [3] 松岡, Lu, 野田, 党, 赤木(2005). ”聴覚系と発話系の相互作用に関する検討”, 平成 17 年春季音響学会講演論文、3-P-33.
- [4] Dang, J., Akagi, M., and Honda, K. (2006). "Communication between speech production and perception within the brain - Observation and simulation," J. Comp. Sci. & Tech., 21, 1, 95-105.
- [5] 田中, Lu, 党, 赤木(2006). ”変形聴覚フィードバックにおける摂動量と補正動作の関係について”, 平成 18 年春季音響学会講演論文、2-3-11
- [6] 内山田, Lu, 党, 赤木(2006). "変形聴覚フィードバックによる発話系の補償動作について", 平成 18 年秋季音響学会講演論文、3-P-9.
- [7] Akagi, M., Dang, J., Lu, X., and Uchiyamada, T. (2006). "Investigation of interaction between speech perception and production using auditory feedback," J. Acoust. Soc. Am., 120, 5, Pt. 2, 3253.
- [8] 内山田, Lu, 党, 赤木(2007). ”変形聴覚フィードバックに対する発話系の反応の計測”, 平成 19 年春季音響学会講演論文、2-Q-18