

Title	変形聴覚フィードバックの摂動量と補正動作の関係について
Author(s)	田中, 貴文
Citation	
Issue Date	2006-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1987
Rights	
Description	Supervisor: 党 建武, 情報科学研究科, 修士

修 士 論 文

変形聴覚フィードバックの摂動量と
補正動作の関係について

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

田中 貴文

2006 年 3 月

修 士 論 文

変形聴覚フィードバックの摂動量と 補正動作の関係について

指導教官 党 建武 教授

審査委員主査 党 建武 教授

審査委員 徳田 功 助教授

審査委員 鷗木 祐史 助教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

410078 田中 貴文

提出年月: 2006 年 2 月

概 要

音声知覚・音声生成過程がどのように関わっているのかを説明する研究が、これまで数多く行われてきた。しかし十分な実験に欠くため、人間本来がもつこのメカニズムについては未だに多くの解明されていないことが残っている。聴覚系と発話系の相互作用の存在を示すものとして、聴覚フィードバックがある。聴覚フィードバックとは発話音声を聴覚系にフィードバックしながら発話器官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。聴覚フィードバックが音声生成過程においてどのような役割を果たしているのかについて検討するため、実時間の変形聴覚フィードバック実験を行い摂動量と補正動作の関連について調べた。その結果補正反応は摂動量が多い場合、起きにくいことがわかった。また、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ことが示唆された。さらに、F1における摂動量と補正反応が生じた割合とF2における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る。

目次

第1章	序論	1
1.1	背景	1
1.2	目的	2
1.3	論文の構成	2
第2章	変形聴覚フィードバック実験システムの構成	3
2.1	聴覚フィードバック	3
2.2	変形聴覚フィードバック	4
2.3	実験要件	5
2.4	実験装置	6
2.5	フォルマント変形処理	7
第3章	これまでの実験における問題点について	9
3.1	音質の改良	9
3.1.1	フォルマント変形フィルタ	9
3.2	変形聴覚フィードバック実験	10
3.2.1	実験目的	10
3.2.2	実験手順	10
3.2.3	分析方法	11
3.2.4	実験結果	12
3.3	考察	12
第4章	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
4.1	変形聴覚フィードバック実験	16
4.1.1	実験目的	16
4.1.2	実験手順	16
4.1.3	実験結果	17
4.2	考察	28
第5章	まとめ	29
5.1	結論	29
5.2	今後の課題	29

第1章 序論

1.1 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”についての話者内部（脳内）での音声生成と音声知覚の相互交信を説明するため，Liverman らは 1960 年代より音声知覚の運動理論（Mortor 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，発話時の音声知覚・生成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聴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 [10]．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定量的な説明には不十分である．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換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 [11] [12] [13]，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，フォルマントの影響については明らかにされなかった．これに対して，佐藤 [14] と斉藤 [15] は短時間のフォルマント変形に対する発話器官の補正動作について，日本語母音に関して音声のスペクトルグラムを用いて分析を行った．その結果，スペクトルの変動は見られたが，それが補正動作か否かは確認されなかった．そこで，松岡 [16] は口唇を調音点とした中国語母音を音声資料として用いて，さらにフィードバック音声の音質の改善を行い，音声のスペクトルや，筋電信号，画像情報での分析を行ったところ，補正動作が確認された．さらに，複数の調音点の補正が相互に絡み合っている可能性が示唆された．

これらの先行研究における知見より音声のスペクトル形状の変形により，補正動作の存

在が示されている．しかし，フィードバック音声の摂動量や音素の違いに対する補正動作の生じる割合などについての定量的な観測は行われていない．

1.2 目的

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，フォルマントの変形において，摂動量とその補正動作の関係については十分に調べられているとはいえない．

本研究では，日本語母音に関して，発話音声のフォルマントを変形したフィードバックを行い，短時間の変形聴覚フィードバック実験において，音声スペクトルを観測する．その結果，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考察し，聴覚系と発話系の相互作用を定量化して，聴覚系と発話系のメカニズムを解明することが本研究の目的である．

1.3 論文の構成

本論文は5章により構成される．第1章は研究の背景等の本論文に関する導入部にあたる．第2章は一般的な変形聴覚フィードバック実験システムの構成に関する記述である．第3章はこれまでの先行研究における問題点の記述及び解決方法，また解決したシステムを用いた実験による結果の分析方法，分析結果，そして考察を述べる．第4章は摂動量と補正反応に関係における変形聴覚フィードバック実験である．それに関する記述及び実験による結果の分析方法，分析結果，そして考察について述べる．第5章はまとめとして本論文の結論と今回の実験の問題点や今後行われるべき内容について述べる．

第2章 変形聴覚フィードバック実験システムの構成

本研究におけるメインテーマである聴覚フィードバックについて、また、その実験システムについて述べる。

2.1 聴覚フィードバック

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。もの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も明瞭な発話が維持されることと一致する。また、言語獲得後に聴力を失っ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 [6] [7] [8]。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 [9] 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 [10]、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するうえで必要であることを示している。

2.2 変形聴覚フィードバック

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[10]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はたらくことが示された。このような河原らの一連の研究により[11][12][12]、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、発話音声のフォ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[21][22]。その結果、話者はフォ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。現在のところ、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。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フォルマント周波数に同様な補正的な変化が見られたと報告している[22]。この結果は、聴覚フィードバック条件下の方が補正的な効果が大きいため、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、フォ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正動作が存在する可能性を示した[14]。続いて斉藤も第1,第2フォルマントを変化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[15]。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可能性が示唆された。そこで、松岡[16]は口唇を調音点とした中国語母音を音声資料として用いて、さらにフィードバック音声の音質の改善を行い、音声のスペクトルや、筋電信号、画像情報での分析を行ったところ、補正動作が確認できた。

これらの先行研究における知見より音声のスペクトル形状の変形により、補正動作の存在が示されている。しかし、フィードバック音声の摂動量や音素の違いに対する補正動作の大きさや生じる割合などについての定量的な観測は行われていない。

2.3 実験要件

佐藤・斉藤らは、実時間によるフォ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と報告している[14][15]。

- 要件1 実時間による音声の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 変換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[17]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。たとえば、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、フォ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[10]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[14]。

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人性を保持したまま実時間でフォルマントを変換することである。Shimon らが指摘している [18] 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] [19] , 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わかるように [18] , 人はフィ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 [20] 。そのため、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正動作であった場合、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。このとき使用する摂動、あるいは音韻、音節によっては、補正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となり、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。さらに検知感度などについても考察する必要がある。たとえばフォ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いことから [17] , 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。

2.4 実験装置

本研究では、上記の実験要件を踏まえ、先行研究の用いたシステム [14] [16] を基本にしてシステムの構築を行った。実験システム概要を図 2.1 に示す。実験は防音室内（暗騒音 約 35dB(SPL)）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。また、骨導音や自然側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）とマイクロホン（WM-C70）を身に着けた状態で発話を行う。

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.

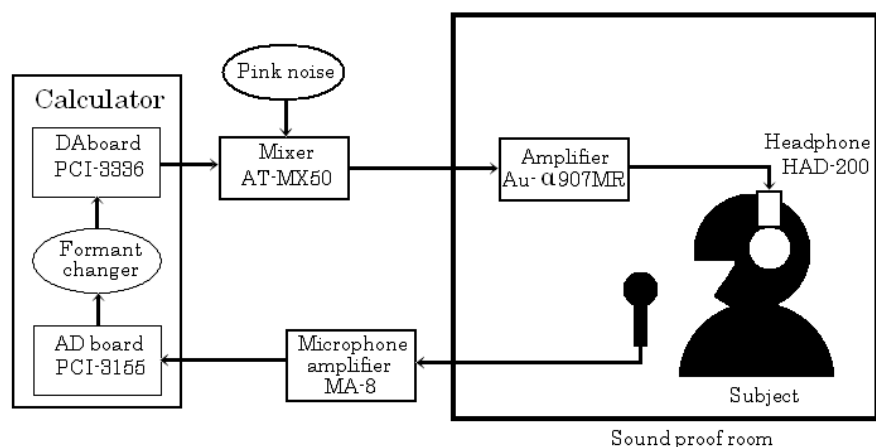


図 2.1: システム構成図

2.5 フォルマント変形処理

フォルマント変形処理には実時間による処理が必要であるという要請と高品質で自然性が高い音声が必要であるという要請を満たす必要がある. 一般にこの2つを同時に満たすことは困難である. ボコーダのような分析・合成系による方法では処理にかかる負荷が大きく実現は難しい. また群遅延等の問題で自然性が損なわれる危険性を含んでいる. そのためフォルマント変形処理は処理が簡易なフィルタ処理のみで実現する. フィルタ処理はFFTを用いた重複加算法による短時間合成を用いた畳み込みにより実現する. 図2.2はフォルマント変換処理の概要を表している.

入力音声 $x(n)$ は窓関数 $w(n)$ により切り出され, フーリエ変換により周波数表現 $X(k)$ を得る. 図中の右側のパスではフォルマント分析が行われ, 分析で得られた情報を基にフォルマントフィルタ $H(K)$ が生成される.

そして $X(k)$ に対してフィルタ処理がなされ, フィルタ出力 $Y(k)$ を逆フーリエ変換することでフィードバック音声 $y(n)$ が得られる.

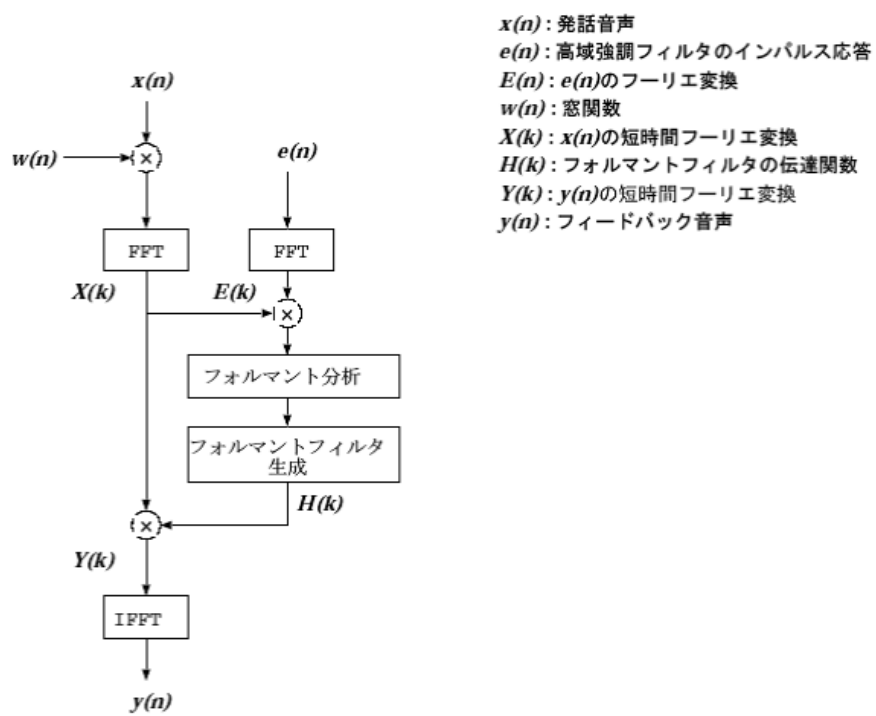


図 2.2: システム構成図

第3章 これまでの実験における問題点について

先行研究における問題点について，実験システムの改良を試みる．また，改良した実験システムを用いて変形聴覚フィードバック実験を行う．

3.1 音質の改良

日本語母音に関する変形聴覚フィードバック実験 [15] において補正動作が確認出来なかった理由として考えられることはフィードバック音声の音質が 2.3 における要件 3 を満たしていないために，被験者が補正動作を行うことができなかった可能性がある．そこで，フィードバック音声の改良を行った．改良を行ったのは以下の点である．

1. フォルマント変形フィルタを改善した．
2. サンプリング周波数を 8kHz から 16kHz へ改善した．
3. ソフトウェアによる時間軸をハードウェアによる時間軸に変更した．
(旧プログラムはソフトウェアクロックを使用していたため時間軸が歪み，音質が劣化していた．)
4. アーシングによる電氣的ノイズを低減させた．

3.1.1 フォルマント変形フィルタ

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

$$G(w) = A \exp \left(-\frac{(w - f)^2}{2B^2} \right) \quad (3.1)$$

f, 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．

$$|H(w)| = 10^{G(w)/20} \quad (3.2)$$

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

$$H(w)H(w)^{-1} = 1 \quad (3.3)$$

となる性質を持っている．この場合，除去する対象のフォルマントと追加したフォ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，除去する対象のフォルマントと追加したフォルマントの形状は必ずしも同一ではない．つまり，式 3.3 におけるフォ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，追加するフォルマントのバンド幅を調節した．その結果，フォルマントの削除，追加を確実なものにした．

3.2 変形聴覚フィードバック実験

3.2.1 実験目的

フィードバック音声の音質改善を行った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．

3.2.2 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フォルマント (F1) と第2フォルマント (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 を求めている．被験者の音響パラメータを表 3.1 に示す．

表 3.1: 被験者の音響パラメータ

	/a/		/e/	
被験者	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フォ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フォルマント変形なしでフィードバックし，その後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“摂動

なし”の音声を表示する．2種類の異なる振動開始時間のパターンと振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 3.1 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)	振動なし(2sec)	振動あり(2sec)	振動なし (1sec)	終了信号 (1sec)
開始信号 (1sec)	振動なし(2.5sec)	振動あり(2sec)	振動なし (0.5sec)	終了信号 (1sec)
開始信号 (1sec)	振動なし (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 3.1: トライアルごとの構成

3.2.3 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，フォ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．振動量はそれぞれ変形した音声の $F1$ と $F2$ と振動なし区間のそれとの差とし，ベクトル $(\Delta F1, \Delta F2)$ で表す．振動に対するフォルマントの移動量は振動あり区間の平均 $F1$ と $F2$ と同トライアルの振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振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，振動なし区間の音声は振動を加えていないので $(\Delta F1 - \Delta F2)$ 平面の原点となる．

ここで振動に対する応答ベクトル r を以下の式で定義する．

$$r = (\Delta f1, \Delta f2) - (\Delta F1, \Delta F2) \quad (3.4)$$

そして，

$$\|(\Delta f1, \Delta f2)\| < \|(\Delta F1, \Delta F2)\| \quad (3.5)$$

を満たすならば，補正反応があると判断する．

3.2.4 実験結果

フィードバック音声の音質改善を行った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 3.2 に示す。また、式 3.5 を満たしたことより補正反応があると判断した補正動作のみを図 3.3 に、補正反応がないと判断した補正動作のみを図 3.4 に示す。被験者 ML に関しても同様に図 3.5, 3.6, 3.7 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。

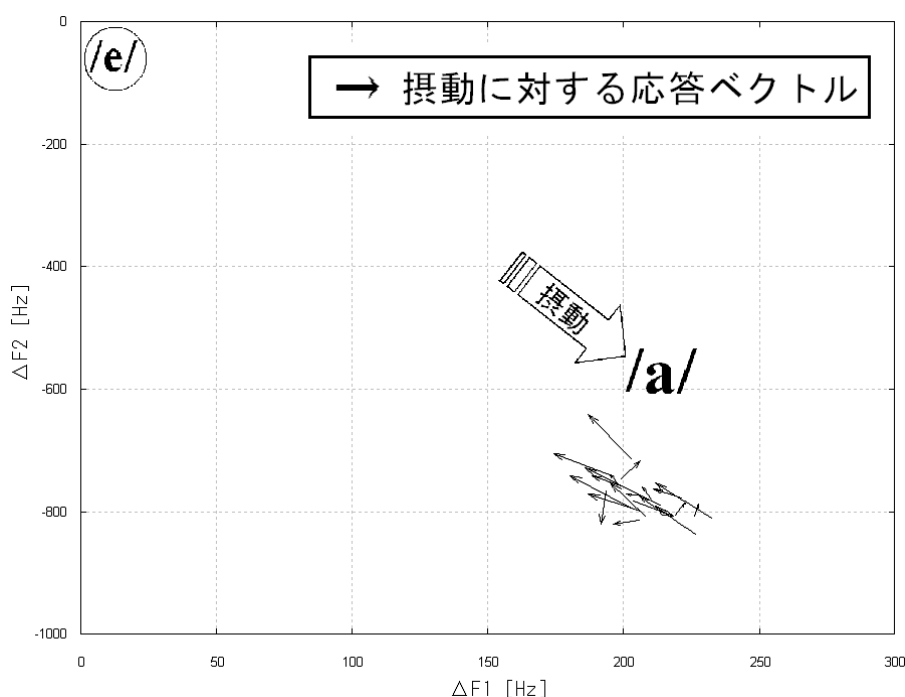


図 3.2: 被験者 SY の摂動量と補正動作 r の関係

3.3 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。これはフィードバック音声の音質が 2.3 における要件 3 を満たしたため、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出来たと考えられる。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の音質改善により補正反応が確認できたが、フィードバック音声の音

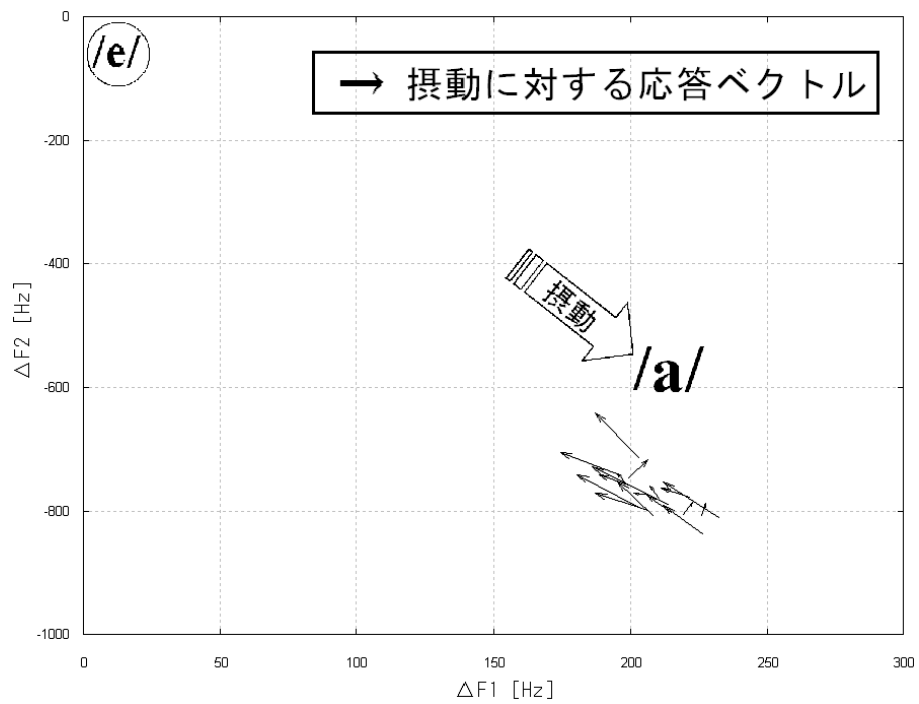


図 3.3: 被験者 SY の摂動量と補正動作 r の関係（補正反応がある場合）

質以外にも補正反応に影響があることが示唆される。

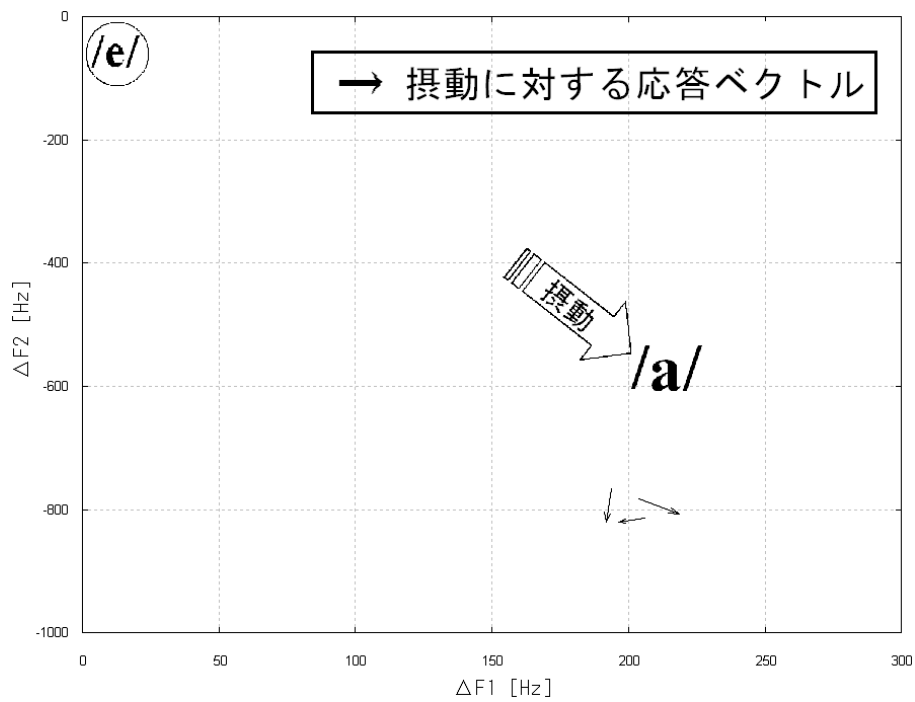


図 3.4: 被験者 SY の摂動量と補正動作 r の関係（補正反応がない場合）

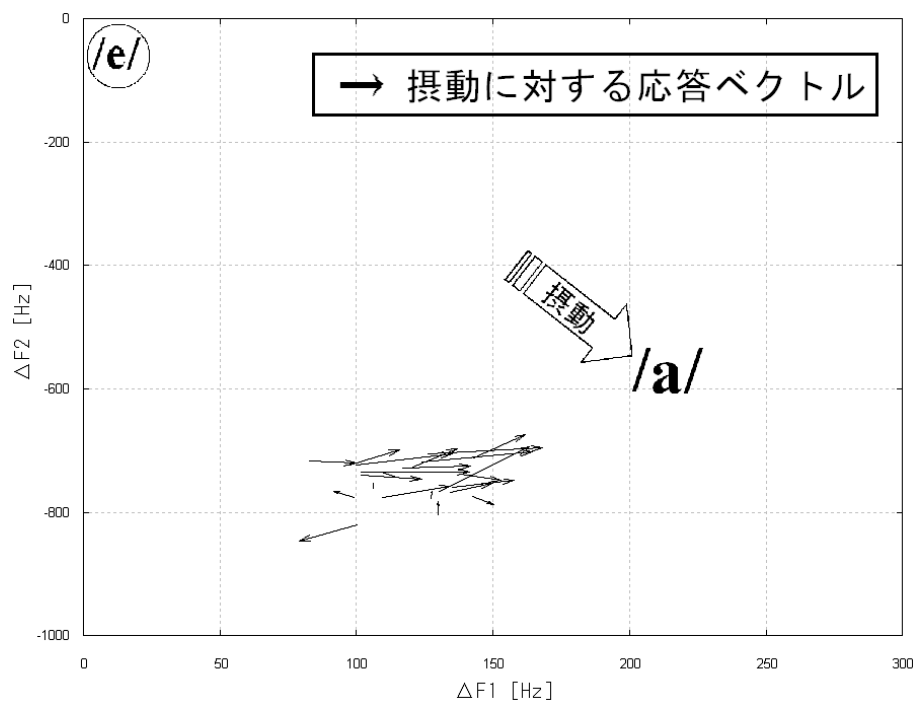


図 3.5: 被験者 ML の摂動量と補正動作 r の関係

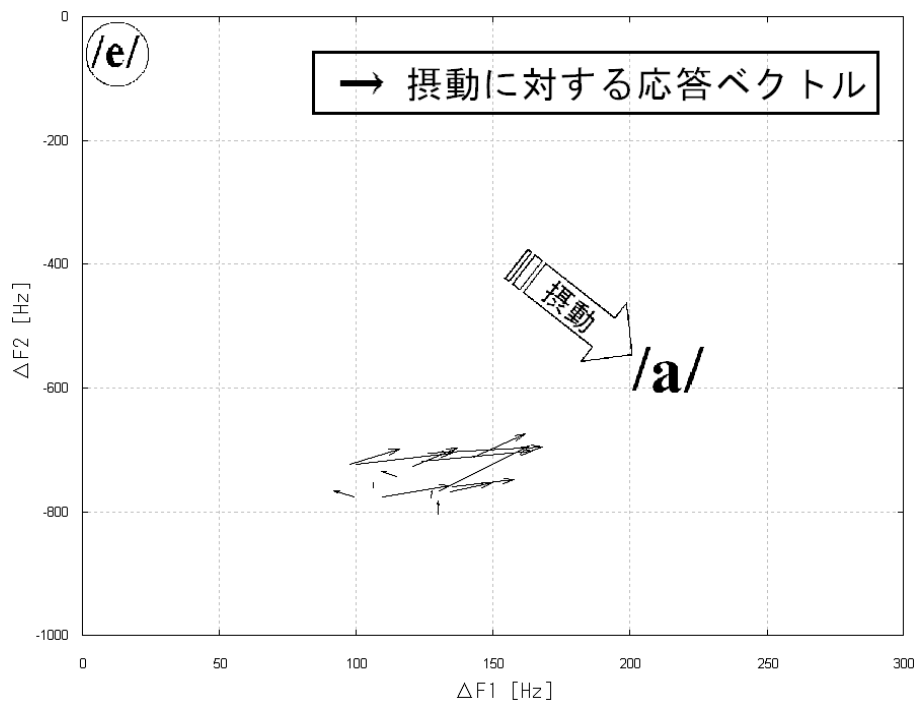


図 3.6: 被験者 ML の摂動量と補正動作 r の関係（補正反応がある場合）

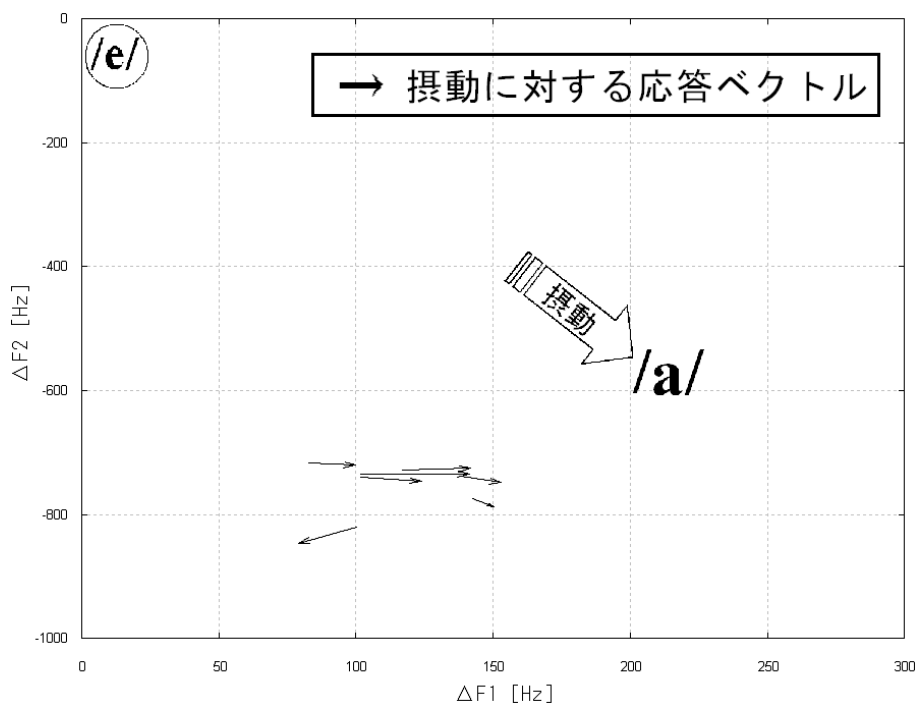


図 3.7: 被験者 ML の摂動量と補正動作 r の関係（補正反応がない場合）

第4章 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．

4.1 変形聴覚フィードバック実験

4.1.1 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．3.2 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．松岡は音声資料に中国母音の /i/ および /ü/ を用いた結果，補正反応を確認することが出来た [16] ． /i/ から /ü/ への調音動作は口唇の突き出しと丸めのみであり，声道形状はほぼ同じである．これは，2.3 の要件 4 を十分に満たしていると考えられる．したがって被験者は比較的簡単に補正を行うことで出来る．一方， /e/ から /a/ への調音動作は舌，顎による複数発話器官の相互作用による複数調音である．そのため，2.3 の要件 4 を満たすことが難しい．したがって，補正反応が被験者によって差異が生じた可能性が考えられる．

そこで，2.3 の要件 4 を満たすことが出来るように，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フォルマント変形量（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．

4.1.2 実験手順

実験の手順は 3.2.2 を基に行った．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 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 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 4.1 に示す．

表 4.1: /e/ → /a/ のフォ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

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（フォルマント変形）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．

4.1.3 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 4.1 ～図 4.9 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 4.2 ～図 4.10 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 4.2 に示す．フォルマント別では，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 4.11 ～図 4.19 と図 4.12 ～図 4.20 にそれぞれ示す．

表 4.2: 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

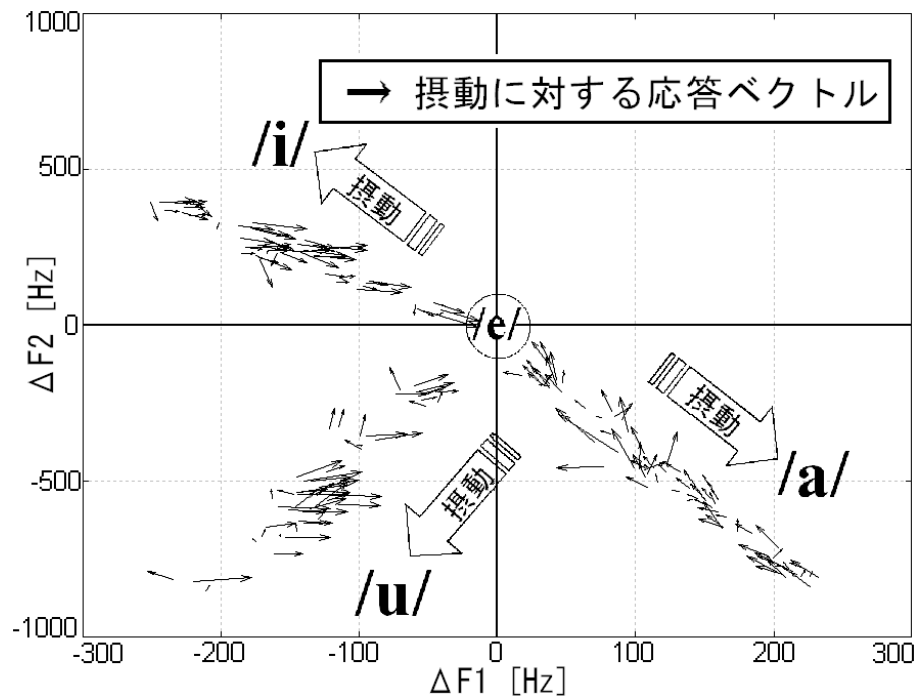


図 4.1: 被験者 SY における摂動量と補正動作 r の関係 (補正反応ありの場合)

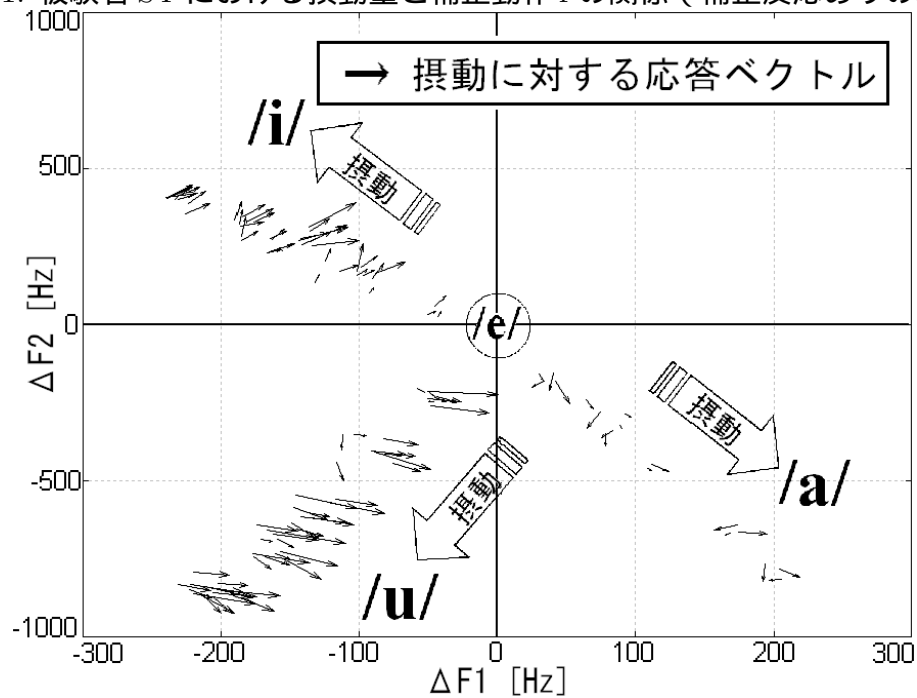


図 4.2: 被験者 SY における摂動量と補正動作 r の関係 (補正反応なしの場合)

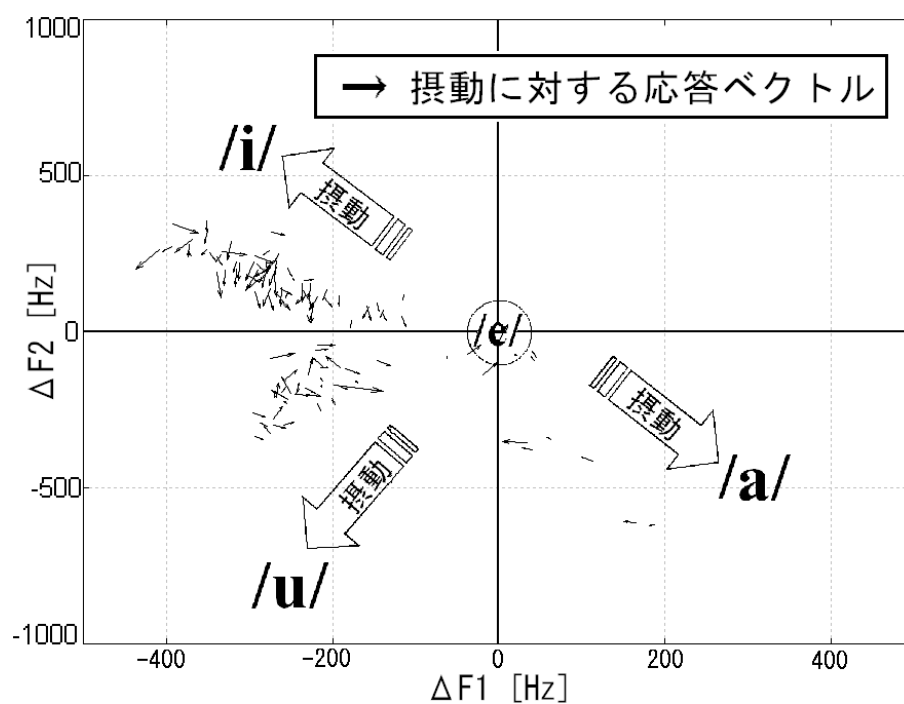


図 4.3: 被験者 HT における摂動量と補正動作 r の関係 (補正反応ありの場合)

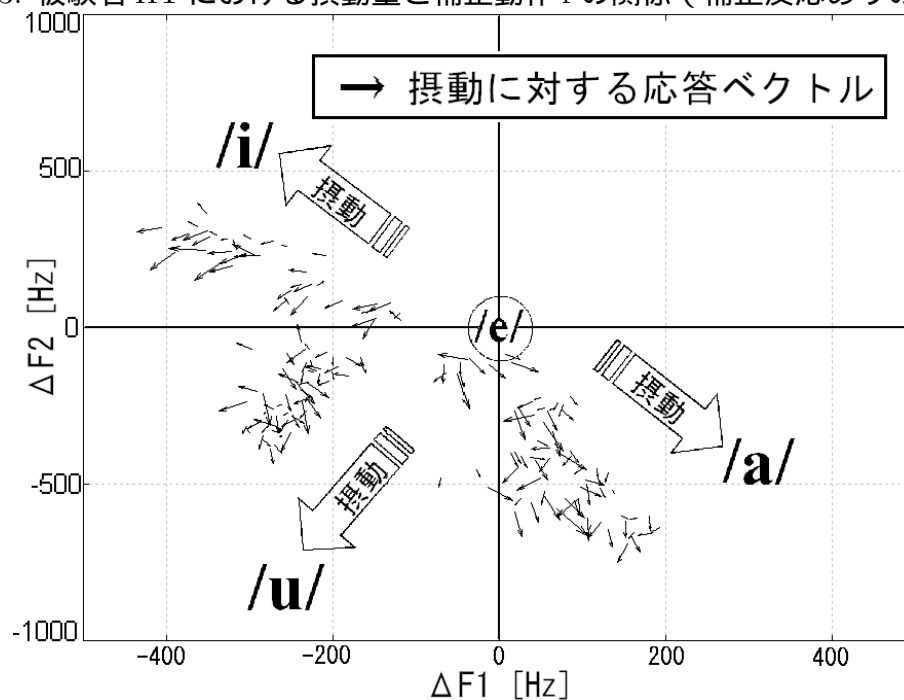


図 4.4: 被験者 HT における摂動量と補正動作 r の関係 (補正反応なしの場合)

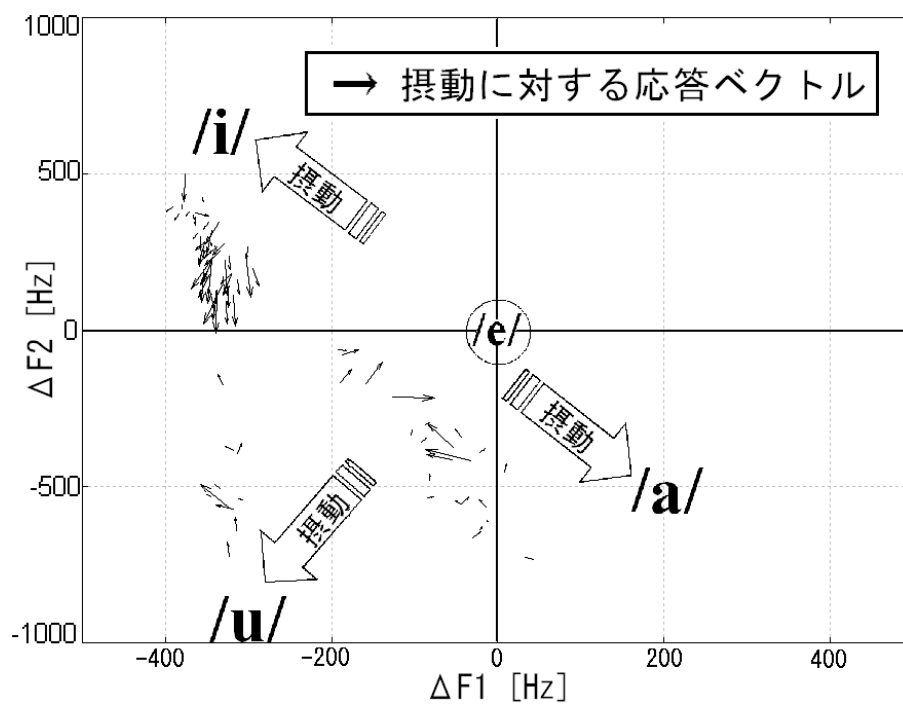


図 4.5: 被験者 MK における摂動量と補正動作 r の関係 (補正反応ありの場合)

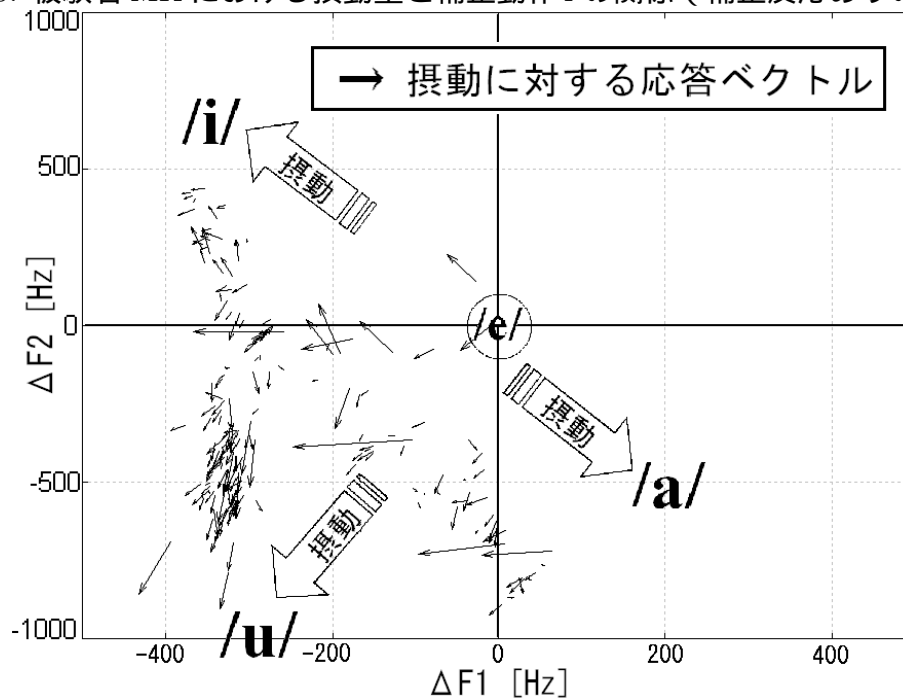


図 4.6: 被験者 MK における摂動量と補正動作 r の関係 (補正反応なしの場合)

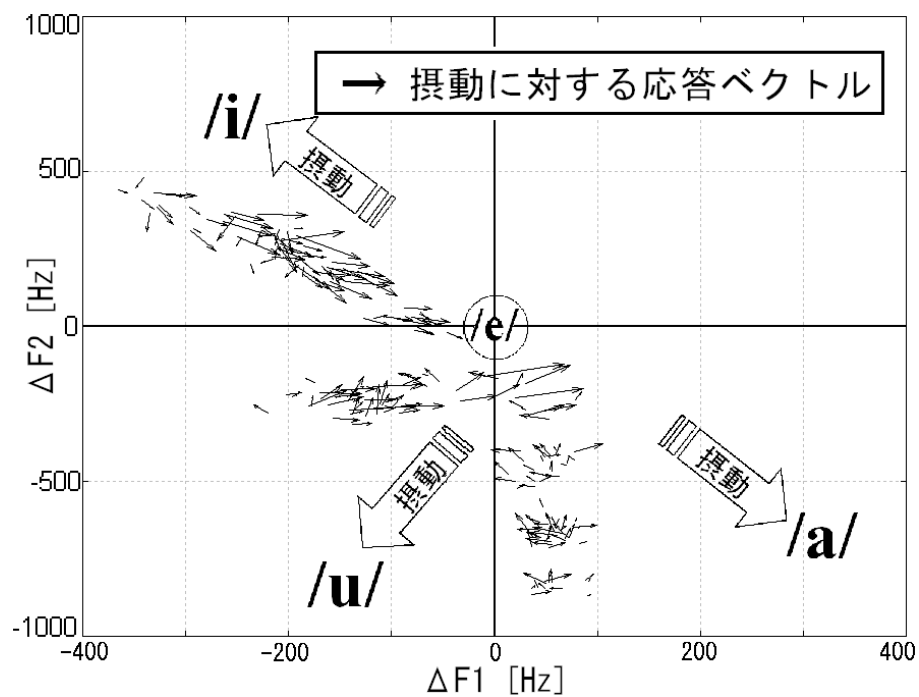


図 4.7: 被験者 ML における摂動量と補正動作 r の関係 (補正反応ありの場合)

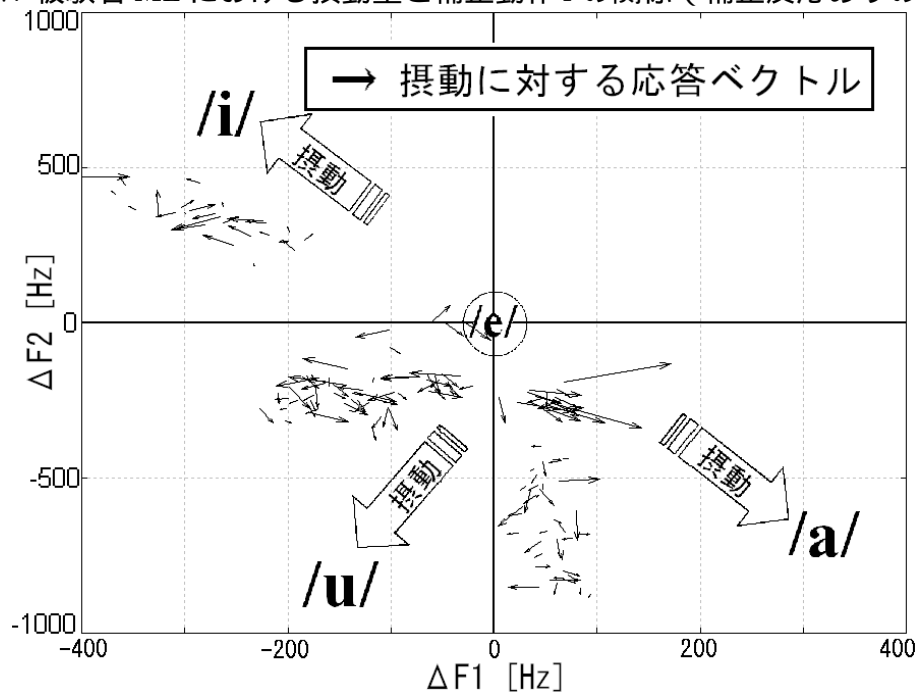


図 4.8: 被験者 ML における摂動量と補正動作 r の関係 (補正反応なしの場合)

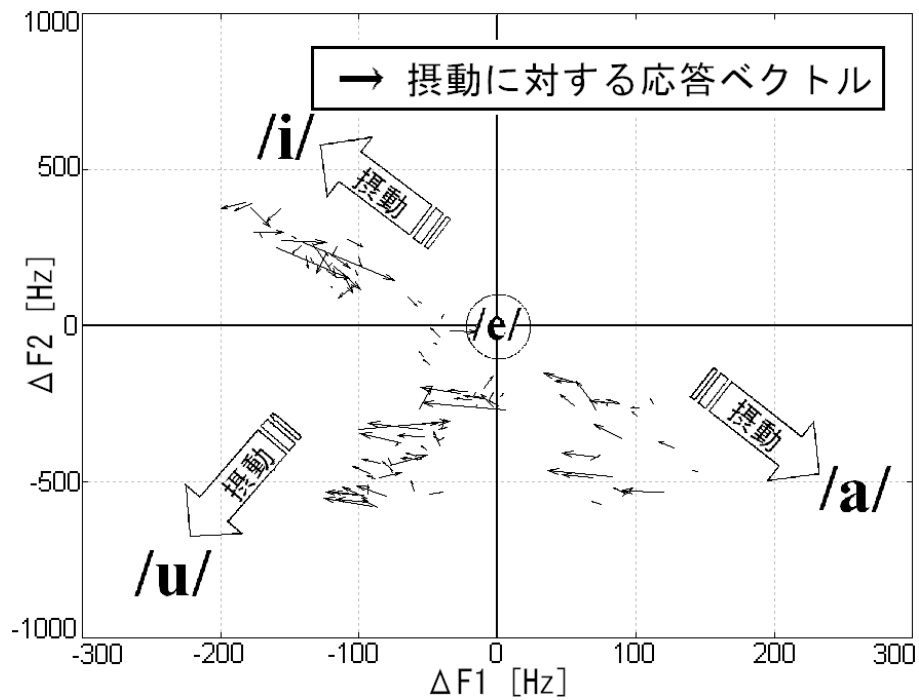


図 4.9: 被験者 MT における摂動量と補正動作 r の関係 (補正反応ありの場合)

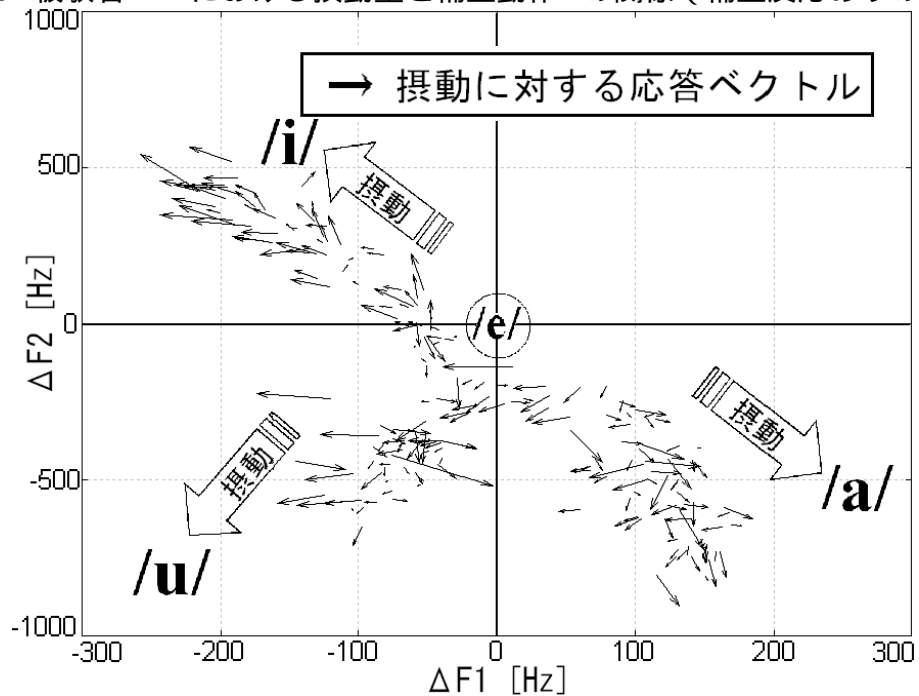


図 4.10: 被験者 MT における摂動量と補正動作 r の関係 (補正反応なしの場合)

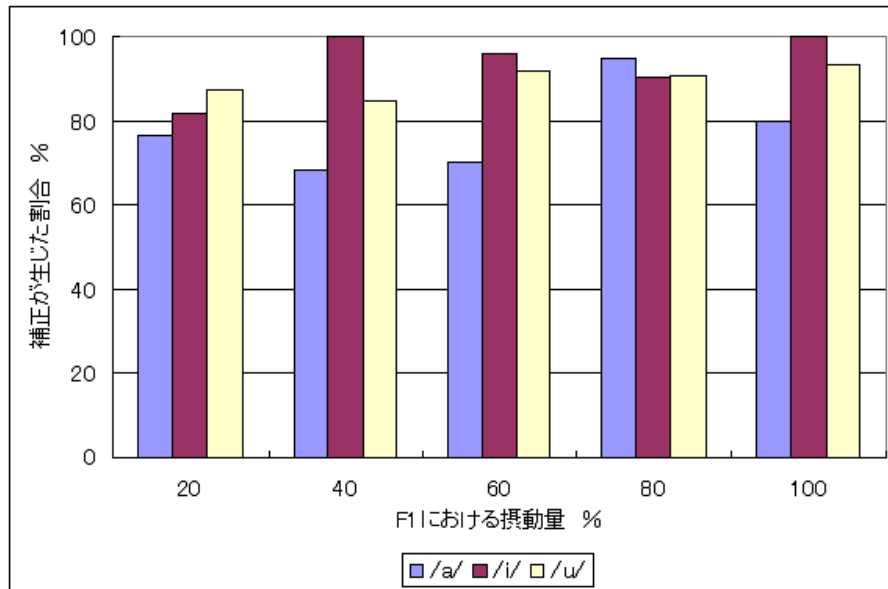


図 4.11: 被験者 SY において F1 における摂動量と補正反応が生じた割合の関係

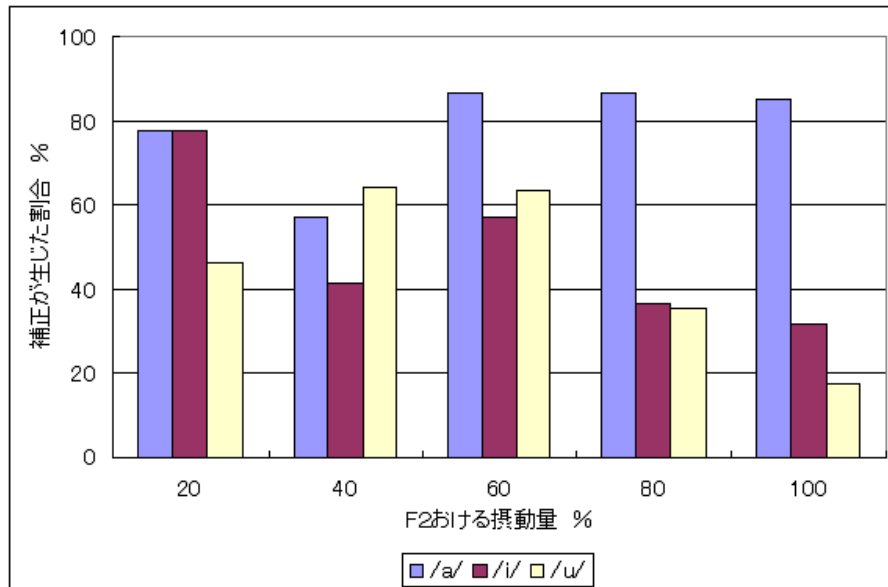


図 4.12: 被験者 SY において F2 における摂動量と補正反応が生じた割合の関係

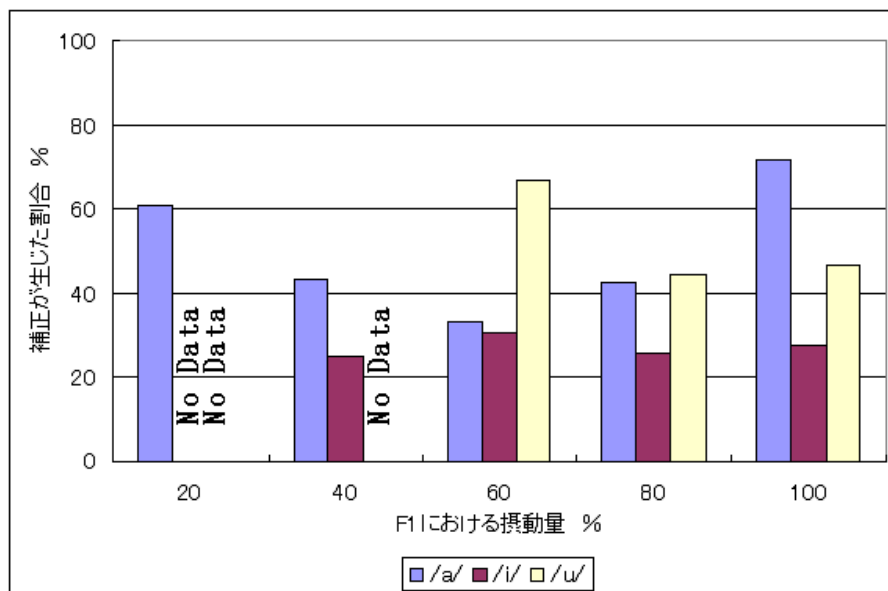


図 4.13: 被験者 HT において F1 における摂動量と補正反応が生じた割合の関係

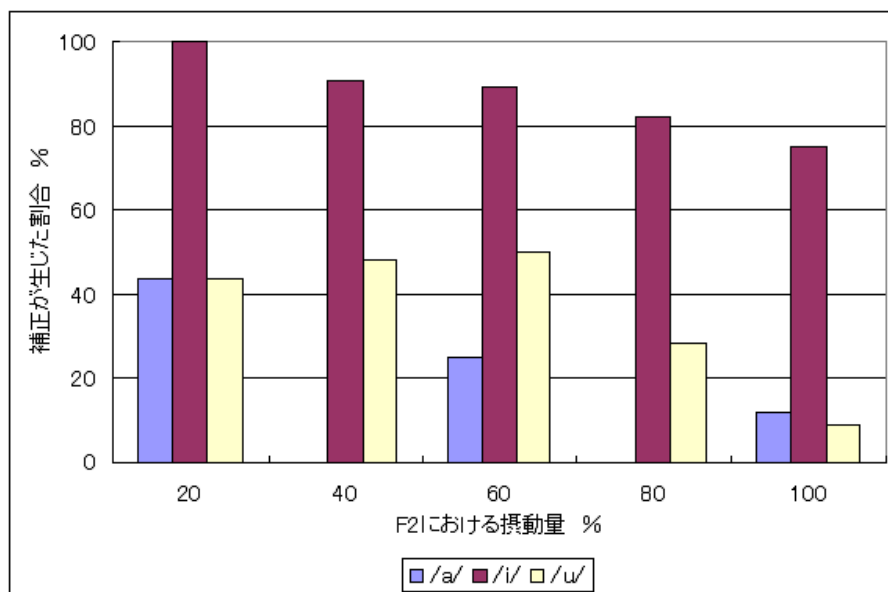


図 4.14: 被験者 HT において F2 における摂動量と補正反応が生じた割合の関係

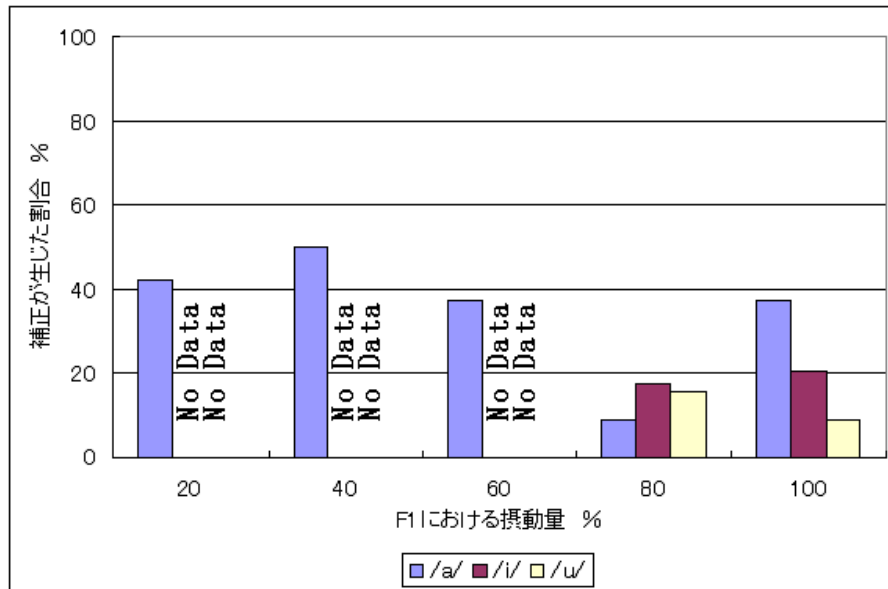


図 4.15: 被験者 MK において F1 における摂動量と補正反応が生じた割合の関係

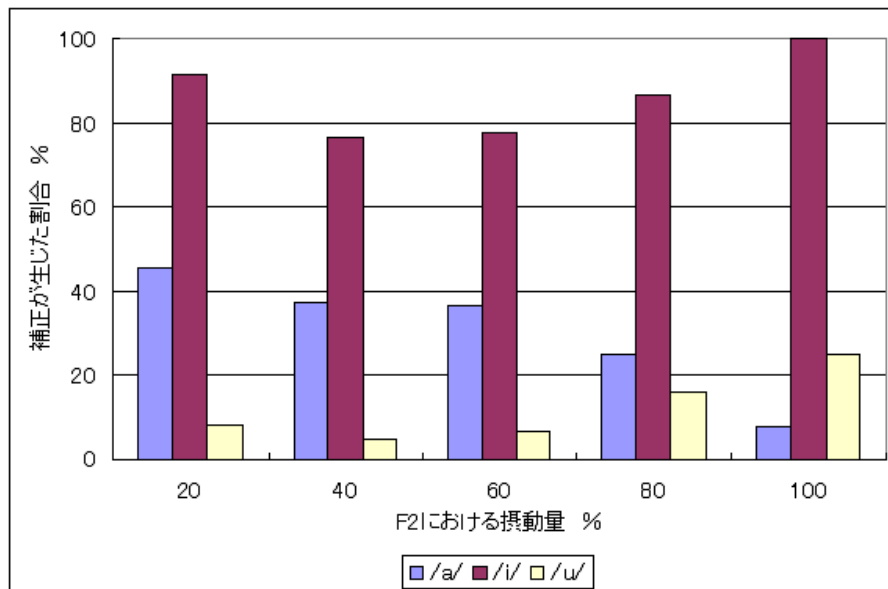


図 4.16: 被験者 MK において F2 における摂動量と補正反応が生じた割合の関係

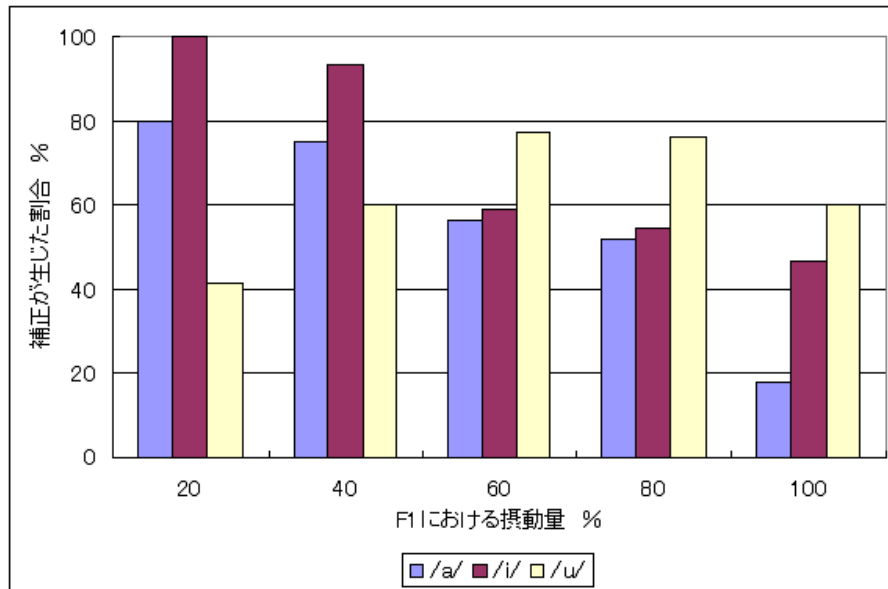


図 4.17: 被験者 ML において F1 における摂動量と補正反応が生じた割合の関係

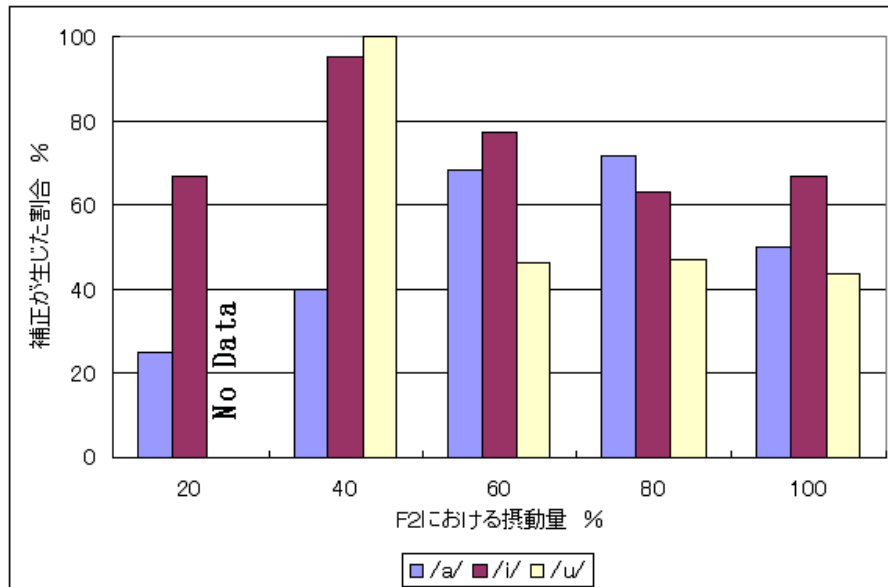


図 4.18: 被験者 ML において F2 における摂動量と補正反応が生じた割合の関係

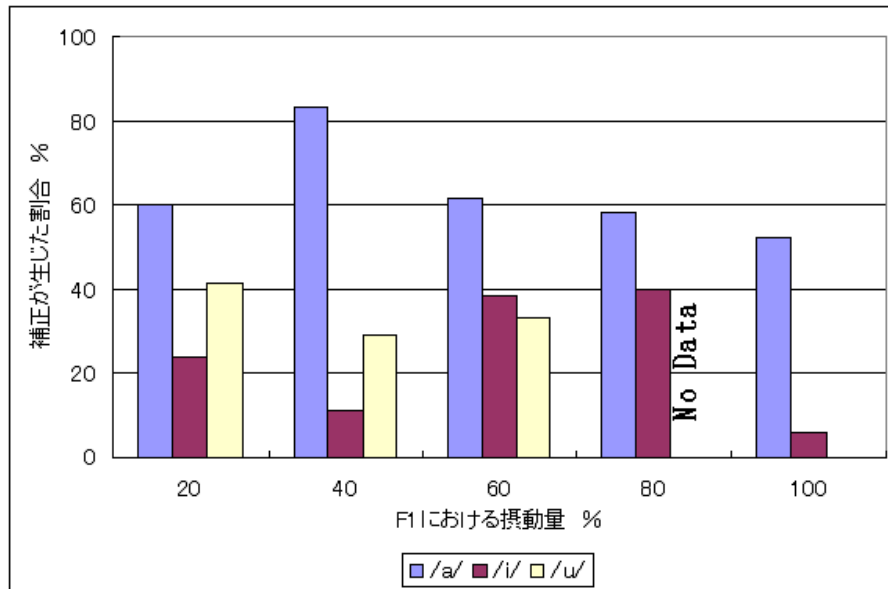


図 4.19: 被験者 MT において F1 における摂動量と補正反応が生じた割合の関係

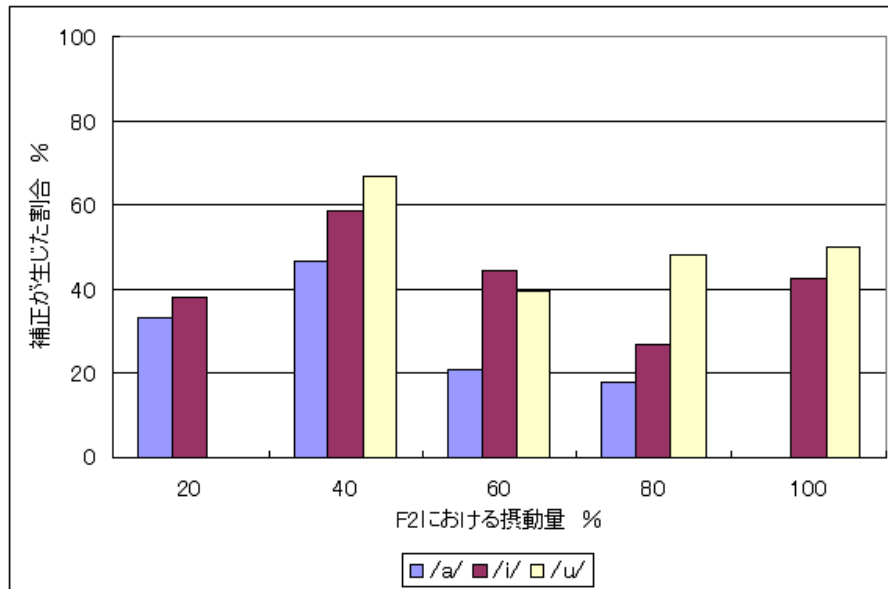


図 4.20: 被験者 MT において F2 における摂動量と補正反応が生じた割合の関係

4.2 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、摂動量が多い場合、補正反応が起きにくいことがわかる。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、フォ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、大きな摂動量を与えた場合には、音韻の変化が知覚されるため、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。フォルマント別では、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っ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。

第5章 まとめ

5.1 結論

本研究ではフィードバック音声の音質の改良を行ったシステムを用いて変形聴覚フィードバック実験を行った。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た。その結果、発話を強調する方向への補正動作が音響分析により確認されたが、補正反応が起きた割合に差異が生じた。そこで、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、被験者により差異があるが、全体的に補正反応は摂動量が多い場合、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る。

5.2 今後の課題

F1 と F2 とは補正動作には、それぞれ別々の発話器官が調音運動をしている可能性が考えられる。したがって、発話時の音声知覚・生成の相互作用についての検討には下顎や口唇、舌などの調音器官に関して筋電 (Electromyogram) 計測や EMA (Electro-magnetic Articulography) などを用いた定量的な分析が必要であると考えられる。

謝辞

本研究を進めるにあたり，日頃から多くの貴重な御助言，ご指導をいただきました，北陸先端科学技術大学院大学 情報科学研究科 党 建武教授，赤木 正人教授，Lu Xugang 助手，並びに党研究室，赤木研究室の皆様へ深く感謝いたします．また御多忙の中，御助言，御討論を頂き，また，実験にご協力していただいた皆様へ感謝致します．最後に，研究を進めるにあたり日頃からあたたくご指導いただきました私の周りの全ての関係者の方々にもう一度深く感謝するとともに，何年も続くこの研究の一年に関わることができましたことに深く感謝の意を表します．

参考文献

- [1] Denes, P. and Pinson, E. ,”The Speech Chain, 2nd Ed.”, New York: W.H.Freeman and Con, 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”, Annals 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] Levitt, H., Stromberg, H., Smith, C. and Gold, T. ,”The structure of segmental errors in the speech of deaf children”, Dec; 13(6), pp.419-441, 1980.
- [7] Osberger, M. and McGarr, N. ,”Speech production characteristics of the hearing impaired”, Speech and Language: Advances in Basic Research and Practice, Vol.8, pp.221-283, 1982.
- [8] Smith, CR. ,”Interjected sounds in deaf children’s speech”, Jun:8(2), pp.123-128, 1975.
- [9] Cowie, R. and Douglas-Cowie, E. ,”Speech production in profound postlingual deafness.” , In M. Lutman, and M. Haggard(eds.), Hearing science and hearing disorders. London : Academic Press, pp.183-230, 1983.
- [10] Lee, B.S. ,”Effect of Delayed speech feedback”, Journal of the Acoustical Society of America, 22, pp.824-826, 1950.
- [11] 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.

- [12] Kawahara, H. ,”Interactions between speech production and perception under auditory feedback perturbations on fundamental frequencies”, J.Acoust. Soc. Jap, Vol.15, pp.201-202, 1994.
- [13] 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.
- [14] 佐藤. ,”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”, 北陸先端大学修士論文, 2003.
- [15] 斉藤. ,”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”, 北陸先端大学修士論文, 2004.
- [16] 松岡. ,”聴覚系と発話系の相互作用に関する研究”, 北陸先端大学修士論文, 2005.
- [17] 沢田., 筧. ,”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, 2003.
- [18] Shimon Sapir, Elizabeth Derosier, Andrea M.Simonson and Amy Wohlert, ”Effects of frequency modulated tones and vowel formants on perioral muscle activity during isometric lip rounding”, Journal of Voice and Hearing, Vol.4, No.2, pp.152-158, 1990.
- [19] Lane, H. and Webster, J. ,”Speech deterioration in postlingually deafened adults”, Journal of the Acoustical Society of America, 89, pp-859-866, 1991.
- [20] 甘利., 外山., ”脳科学大辞典”, 朝倉書店, 2000.
- [21] Houde, J., Jordan, M. ,”Sensorimotor adaptation in speech production”, Science, 279, pp.1213-1216, 1998.
- [22] Houde, J., Jordan, M. ,”Sensorimotor adaptation of speech I: Compensation and Adaptation”, Journal of Speech Language, and Hearing Research, 45, pp.295-310, 2002.