

Title	スペクトル変形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
Author(s)	佐藤, 博幸
Citation	
Issue Date	2003-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/1662">http://hdl.handle.net/10119/1662</a>
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

修 士 論 文

変形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  
情報科学研究科情報処理学専攻

佐藤 博幸

2003年3月

修 士 論 文

変形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

指導教官 赤木正人 教授

審査委員主査 赤木正人 教授  
審査委員 小谷一孔 助教授  
審査委員 下平博 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

110054 佐藤 博幸

提出年月: 2003 年 2 月

## 概要

本稿では、音声生成と知覚の間にある相互関係を解明することを目的として、変形聴覚フィードバックによる実験を行った。発話音声のスペクトルに摂動を与え、知覚側へフィードバックした際に音声の生成側で起こる反応およびその反応までに要する時間の調査を行った。その結果、実験において知覚側で起こった変化を補正しようとする補償動作を音声生成側で確認することができた。

# 目次

第1章	序論	1
1.1	研究の背景	1
1.2	本研究の目的	2
第2章	実験系	3
2.1	変形聴覚フィードバック	3
2.2	フィードバック音声への摂動の与え方	3
2.3	フィードバックに要する時間の制限	4
2.4	実験系の概要	5
2.5	実験機器について	6
第3章	予備実験	8
3.1	目的	8
3.2	方法	8
3.3	実験結果	10
3.4	考察	11
第4章	刺激の準備	12
4.1	刺激の検討	12
4.1.1	実験に用いる発話音声	12
4.1.2	フィルタの種類	12
4.1.3	フィルタの除去幅	12
4.2	刺激の作成方法	13
4.3	フィルタの構成	18
第5章	本実験	20
5.1	目的	20
5.2	方法	20
5.3	分析方法	22
5.4	連続母音/aiai.../による実験結果	23
5.4.1	結果	23

5.4.2 考察 . . . . .	27
<b>第 6 章 結論</b>	<b>29</b>
6.1 本論文のまとめ . . . . .	29
6.2 今後の課題 . . . . .	30
<b>付 録 A</b>	<b>33</b>
A.1 本実験で用いたフィルタ変更のタイミング . . . . .	33

# 目次

2.1	TAFによる実験の概要図	3
2.2	スペクトル変化が母音同定におよぼす影響 [9]	4
2.3	実験系の概要	5
3.1	時間遅延を挿入した実験の結果:ヘッドホンを付けた状態(時間遅延 0ms)と遅延時間を挿入したときの発話のしやすさを被験者に評価してもらった。Y軸は発話のしやすさを表しており、この値が大きいほどヘッドホン装着時(遅延時間 0ms)に比べて発話しづらいことを意味する。	10
4.1	バタワースフィルタの極の位置 ( $N=2$ )	16
4.2	2次のIIRフィルタのブロック図	18
4.3	本研究で構成したフィルタのブロック図	18
5.1	分析方法	22
5.2	除去幅を徐々に広げた場合のスペクトル包絡例:除去幅を変化させてもスペクトル包絡の基本的な形状に変化はあまり見られない	24
5.3	除去幅を徐々に狭めた場合のスペクトル包絡例:除去幅が 500Hz の場合に除去した F2 の両側で大きなパワーの変化が起こっている。そして除去幅を小さくしていくと、F2 の両側で起こるパワーの変化も小さくなり、除去幅が 200Hz の場合にはスペクトル形状が通常発話の場合とかなり似た形状になっている。	28

# 表 目 次

4.1	設計したフィルタの基本仕様 . . . . .	13
5.1	通常発話時の F1 と F2 . . . . .	20
5.2	除去幅を広げていった場合の突出度 . . . . .	23
5.3	除去幅を狭めていった場合の突出度 . . . . .	25
5.4	ランダムに除去幅を変化させた場合の突出度 . . . . .	26
A.1	フィルタの変化パターン 1 . . . . .	33
A.2	フィルタの変化パターン 2 . . . . .	34
A.3	フィルタの変化パターン 3 . . . . .	34



# 第1章 序論

## 1.1 研究の背景

話をする時、話者は自分自身が発した音声を、聴覚器官を通して脳にフィードバックし、意図通りの発話になるように、絶えずコントロールを行っている。このフィードバックは聴覚フィードバックと呼ばれ、音声生成において重要な役割をになっている。聴覚障害者の多くは、自分自身の発話が聴覚器官から脳へ正確にフィードバックされないため、発話が困難であったり、発話が正確でなかったりする。

河原らは変形聴覚フィードバック (TAF)[2] と呼ばれる測定手法を提案し、この測定手法により、基本周波数制御における音声生成と音声知覚の間に実時間相互作用が存在することを示した。そして、その相互関係は基本周波数変動への補償レスポンスであることを明らかにした。また、B.S.Leeら [1] により、およそ 200ms の時間遅延をフィードバック音声に挿入すると、発話にどもりのような現象が起こり、発話が困難になる傾向があることが報告されている。河原らが着目したのは、基本周波数と時間遅延のみであり、スペクトルなど他の知覚情報の変化に対して、音声の生成側がどのように反応するかは、あまり報告されていない。その要因として次のようなことが考えられる。

- フィードバックに要することのできる時間が限られている。
- 計算機の処理能力が十分ではなかった。
- 実時間で音声の再合成を行うことが困難である。

現在では計算機やハードウェアの能力が向上したことにより、フィルタリングなど、単純な処理であれば実時間でフィードバック音声に変化を与えることができる可能性がある。

## 1.2 本研究の目的

本研究では、発話音声において重要な物理量であるスペクトルに着目する。そして、聴覚器官へフィードバックされる音声のスペクトルを制御し、知覚情報の1つであるスペクトルの変化に対して音声生成側で何らかの反応があるか、またその反応までの時間はどの程度であるかを調査することを目的とする。

## 第2章 実験系

### 2.1 変形聴覚フィードバック

変形聴覚フィードバック (TAF: Transformed Auditory Feedback)[4] は聴覚がどのような形で正常な状態の発話に関与しているかを機能レベルで明らかにすることを狙い、河原らによって提案された測定手法である。

発話器官と聴覚器官の間に擬似的なフィードバック経路を構成し、そのフィードバック経路において音声のパラメータに微少な摂動を与えてその発話に対する影響を解析する。TAF による実験の概念図を図 2.1 に示す。

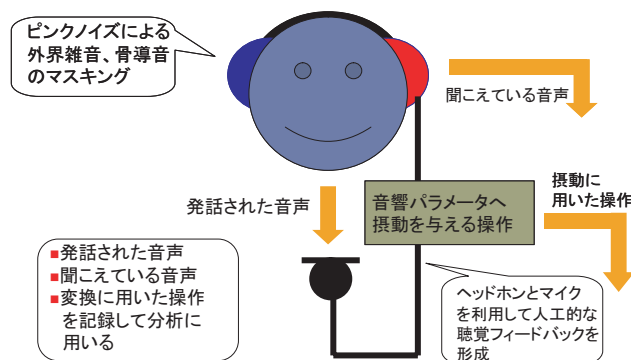


図 2.1: TAF による実験の概要図

### 2.2 フィードバック音声への摂動の与え方

フィードバック音声のスペクトルへ摂動を与える手段として、フィルタ処理、フォルマント位置の移動、ある帯域とある帯域におけるフォルマントの位置を入れ替えるといった手段が考えられる。図 2.2 は、いくつかのスペクトル変動が母音同定に及ぼす影響をまとめたものである。この図によれば、スペクトルの傾斜において、低域と高域周波数部の相対振幅を変化させるように中域周波数値を軸としてスペクトルを回転させているが、このようなスペクトル変化による影響はわずかである。スペクトルの谷の深さを変えるようなスペクトル変化も母音同定に及ぼす影響は比較的少ない。それに対してスペクトルのピークの相対位置の移動は多くの場合、母音知覚に決定的な影響を及ぼす。

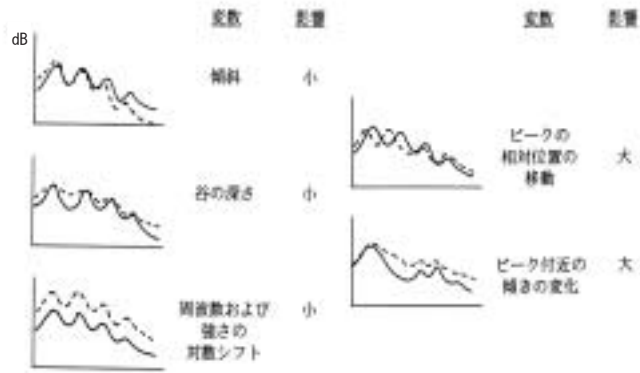


図 2.2: スペクトル変化が母音同定におよぼす影響 [9]

このことから、本研究では、処理に要する時間が短いこと、フォルマントを除去することで、それまでフォルマントの陰に隠れて強調されなかった帯域のパワーが強調され、擬似的にフォルマントが移動したような状況を容易に作り出すことが可能であることから、フィルタ処理をフィードバック音声へ摂動を与える手段として選択する。

## 2.3 フィードバックに要する時間の制限

発話に影響を及ぼす時間遅延に個人差はあるが、約 80ms の時間遅延で、発話に時間遅延による影響が現れたケースが報告されている。このことから、フィードバック音声の操作とフィードバックに利用できる時間は 80ms 以内であることがわかる。しかし、80ms という時間は、最低限の時間であり、実際には時間遅延の影響を取り除くために、もっと短い時間ですべての処理を終えなければならない。本研究では、どの程度の時間遅延がフィードバック音声に加わった場合に発話に影響が現れるかを調査するために、フィードバックに時間遅延を加えた実験を第 3 章で行う。

## 2.4 実験系の概要

図 2.3 に本研究で構築した実験系の概要を示す。コンデンサマイクロホンとヘッドホンをを用いて、音声生成と知覚の間に擬似的な聴覚フィードバック経路を構築する。その際、骨導音や外界からの雑音をマスキングするために 65 から 70[dB] のピンクノイズをフィードバック音声に付加している。

被験者が発した音声をコンデンサマイクロホンにより集音し、マイクアンプを通して計算機内の AD 変換ボードへ入力する。計算機内で信号処理を経た後に計算機内の DA 変換ボードから出力する。その出力とピンクノイズをミキサーにより混合した音声をアンプを通して、密閉型のヘッドホンにより被験者の両耳へフィードバックする。この際、分析用のデータとして発話音声、フィードバック音声、フィルタの変更点を示す信号を記録用の計算機内に記録する。

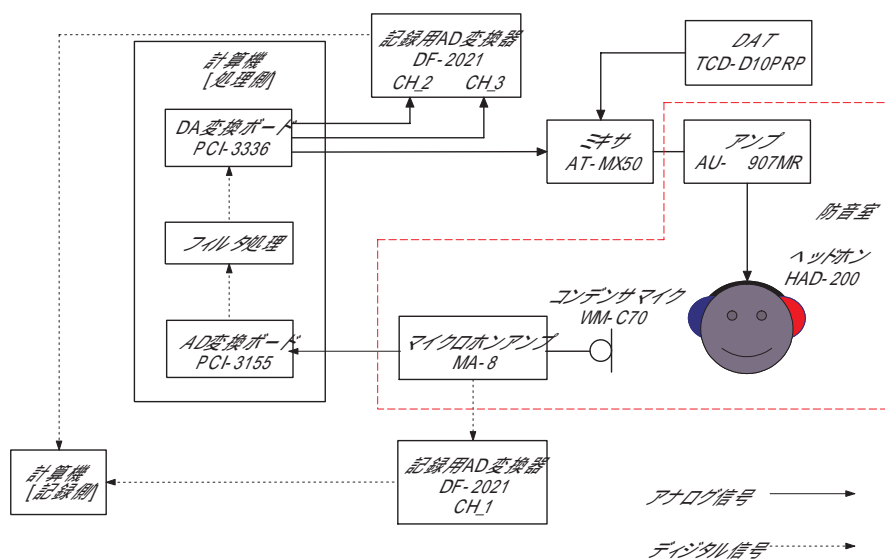


図 2.3: 実験系の概要

## 2.5 実験機器について

- PCI-3155(Interface)  
PCIバスに準拠した高性能、高速、FIFO搭載16ビットAD変換ボード。変換時間は、チャンネル固定時で $10\mu\text{s}$ 、シングルエンド16チャンネル、差動8チャンネル入力まで可能。
- PCI-3336(Interface)  
PCIバスに準拠した16ビット高精度DA変換ボード。セトリングタイムは、チャンネルあたり $5\mu\text{s}$ 、出力チャンネル数は4チャンネルまで可能。
- DF-2021(Pavec)  
データ記録側で用いた8チャンネル16ビットAD変換器。計算機とのインタフェースとしてSCSIを採用し、高性能、高速なAD変換が可能である。
- WM-C70(RAMSA)  
コンデンサ型マイクロホン。最大入力音圧148dB/SPL、暗騒音などを効果的に押さえることのできるローパスフィルタを内蔵する。
- MA-8(TASCAM)  
8チャンネルマイクロホンアンプ。
- AU- $\alpha$ 907MR(SANSUI)  
インテグレートッドアンプ。
- AT-MX50  
4チャンネル入力2チャンネル出力のアナログミキサ。
- 計算機 [処理側](自作 AT 互換機)  
CPUに数値演算に向くAthlon1900+、メモリ1024MBを搭載している。またオペレーティングシステムとして、VineLinux2.5をインストールした上にハードウェアリアルタイム機能を付加する拡張機能としてRTLinux3.0をインストールしてある。
- 計算機 [記録側](DELL)  
CPUにPentium3 700MHz、メモリ512MB、SCSIインタフェースを搭載している。AD変換器との間でデータの記録を行っている。
- ヘッドホン HDA-200(SENNHEISER)  
125Hzから16kHzの間で周波数特性がフラットな密閉型の耳掛けヘッドホン。ヘッドホンのみで、外界からのノイズをかなり防ぐことができる。

- 騒音計 B&K2231(B&K)  
IEC651 に規定される Sound Level Meter で、使用周波数は 1Hz ~ 70kHz。
- DAT TCD-D10PRO2(SONY)  
ポータブルタイプのデジタルオーディオレコーダー。DC 駆動のため、他の機器からノイズの影響を受けない。
- 防音室  
刺激音の音圧レベルを測定、聴取実験に暗騒音約 35 [dB(SPL)] の防音室を使用した。

## 第3章 予備実験

### 3.1 目的

フィードバック音声へ時間遅延を挿入した実験を行うことにより、構築した実験系の正常動作の確認、またフィードバック音声のヘッドホン聴取による影響と時間遅延の影響がどの程度で発生するかを調査する。

### 3.2 方法

#### 被験者

特別なボイストレーニングを受けていない、正常聴力を有する大学院生 4 名。

#### 準備した刺激

刺激であるフィードバック音声へ挿入する時間遅延として、0ms,10ms,20ms,30ms,60ms,70ms,80ms,90ms,120ms,150ms,180ms,190ms,200ms,210m, の 14 種類を準備した。

#### 発話音声

単母音/a/と短文「昔昔、あるところにおじいさんとおばあさんがいました。おじいさんは山へ洗濯に、おばあさんは川へ柴刈にいきました」の 2 種類とした。なお、短文を用いた実験では用紙に書いてある文を読んでもらうようにした。



## 実験方法

- 被験者への教示

発話被験者には、発話時の声の大きさ、声の高さ、発話のスピードは被験者が発話しやすいもので構わないが、できるだけ一定の発話をしてもらうようお願いした。

- 記録したデータ

フィードバック音声に時間遅延を挿入した際の発話音声とフィードバック音声の信号を計算機内のハードディスクへ記録した。また、マイクから 20cm 離れた位置における発話音圧をマイクへの受話音圧として、騒音計により測定し、記録した。被験者には各刺激ごとに次の質問に回答してもらった。

- 質問

- － ヘッドホン聴取に違和感を感じるか。
- － 時間遅延なしの場合に比べて発話のしやすさはどうか。
- － 時間遅延なしの場合に比べてフィードバック音声に違和感があるか。

## 評価方法

実際に発話した音声を聴いて、遅延時間がどの程度の長さになったときに発話に変化が現れているかを確認した。また質問に対する回答から、遅延時間の影響とヘッドホン聴取の影響について評価を行った。

### 3.3 実験結果

フィードバック音声へ時間遅延を挿入した際の実験結果を図 3.1 に示す。なお、この実験におけるマイクへの平均受話音圧は 73[dB/SPL] であった。

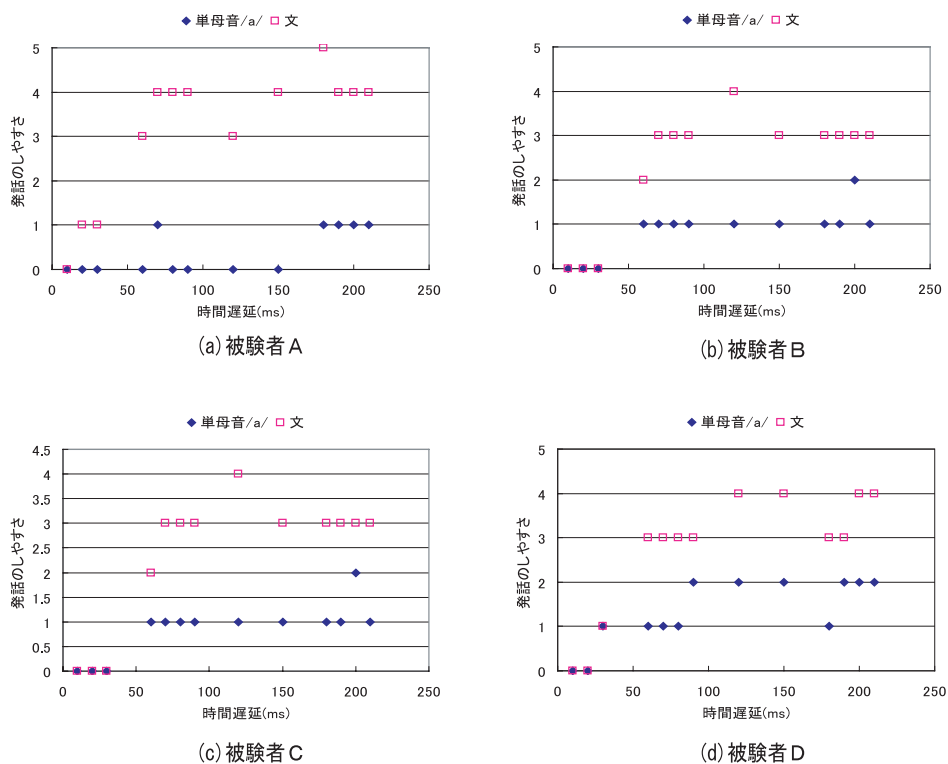


図 3.1: 時間遅延を挿入した実験の結果:ヘッドホンを付けた状態 (時間遅延 0ms) と遅延時間を挿入したときの発話のしやすさを被験者に評価してもらった。Y 軸は発話のしやすさを表しており、この値が大きいほどヘッドホン装着時 (遅延時間 0ms) に比べて発話しづらいことを意味する。

### 3.4 考察

4人の被験者全員が、実験前に通常時とヘッドホン装着時(遅延時間0ms)について発話とフィードバック音声の聴え方について比較を行ってもらった。その結果、発話に関しては通常時とヘッドホン装着時で違いを感じていなかった。しかし、フィードバックされる音声を「聴く」という点では通常の発話とヘッドホン装着(0ms)とで、被験者全員が違和感を感じていた。

図3.1の比較結果において、「発話する」という点では遅延時間が20msまでならば、遅延時間0msの場合との違いを感じていない。しかし、遅延時間を30ms以上に延ばしていくに従って、被験者による個人差はあるが、遅延時間0msの場合に比べ、発話に違和感あるいは支障をきたしていることがわかる。したがって、発話音声にフィルタ処理を施し、聴覚器官へフィードバックさせる実験を行うためには、処理による遅延を20ms以内としなければならない。

フィードバックされる音声を「聴く」という点に関しては、被験者全員が、遅延時間0msの場合でも通常の発話時に聴くフィードバック音声との違いを感じている傾向にあった。これは、骨導音がピンクノイズによりマスキングされているために、通常の発話時に聴くフィードバック音声との違いを感じていると予測できる。本研究では、ヘッドホン聴取による違和感をできるだけ軽減するため、本実験前に約10分間マイクとヘッドホンで構成された疑似的な聴覚フィードバック系を利用して会話を行った。

発話音声の違いについては単母音に比べ、文章の場合の方が遅延時間により、大きく発話に支障をきたしている。これは、発話音声は単母音の場合には、遅延時間の影響を受けるのが発話のはじめのみであり、それ以降は遅延時間0msと同じように音声はフィードバックされているような状況になるからだと考えることができる。

## 第4章 刺激の準備

### 4.1 刺激の検討

本研究では、TAFの実験系において、フィードバック音声のスペクトルに摂動を与える手法としてフィルタ処理を用いる。はじめに次のことについて検討を行った。

#### 4.1.1 実験に用いる発話音声

母音は第1、第2、第3フォルマント周波数という極めて単純な音響記述の集合で特徴づけることができる。本研究はフィードバック音声の変化が発話側に与える影響の調査を目的の1つとしている。発話音声を単母音とした場合には、口の形を固定した状態での発話となるため、フィードバック音声の変化による影響が音声生成側に現れずらい可能性がある。そこで、本研究で対象とする音声は連続母音/aiai.../とする。

#### 4.1.2 フィルタの種類

フィルタを用いて、フィードバック音声のスペクトルを変化させる際には、できるだけ正確に狙った帯域のスペクトルを除去できることが望ましい。そこで本研究では、通過領域のゲイン特性が平坦であるバターース型の高域通過フィルタと低域通過フィルタの2つを組み合わせてノッチフィルタを構成する。またIIR型のフィルタを選択することにより、低次数で鋭い遮断特性をもつフィルタを構成する。

#### 4.1.3 フィルタの除去幅

どの程度の除去幅で音声のスペクトルを除去した際に音声の音韻性、音質に変化が確認できるか調査するために、MATLAB上で種々の除去幅をもつノッチフィルタを作成した。そして、それらのフィルタを用い、ATR音声データベース男性話者MAUの発話音声5つの母音に対して、第1フォルマントと第2フォルマントの除去を行った。その結果、300Hz以上の除去幅をもつノッチフィルタでフォルマントの除去を行ったときに、音韻性あるいは音質に変化が感じられた。そこで、本研究では連続母音/aiai.../の/i/の第2フォルマントを200Hzから500Hzの幅で除去を行うノッチフィルタを100Hz刻みに構成し、

実験での刺激として用いる。

## 4.2 刺激の作成方法

本研究で用いるフィルタは、以下に示すような設計手順で設計を行った。

- 設計仕様の決定
- 次数の決定
- 基準ローパスフィルタの設計
- カットオフ周波数のプリワーピング
- 周波数変換
- 双1次  $s$ - $z$  変換

### 設計仕様の決定

本研究ではフィルタに通過域の特性が平坦であるバターース型の低域通過フィルタと高域通過フィルタを IIR フィルタで構成している。本研究で用いたフィルタのカットオフ周波数以外の仕様は表 4.1 の通りである。

表 4.1: 設計したフィルタの基本仕様

仕様項目	仕様値
阻止域エッジ周波数	カットオフ周波数 $\pm 100$ (Hz)
通過域リップル量	3(dB) 以内
阻止域減衰量	20(dB) 以上

## 次数の決定

仕様を実現するために必要なフィルタの次数は (4.1) 式により求めた。

$$N = \frac{\log k_1}{\log k}$$

本研究で、必要とする仕様を満たす低域通過フィルタと高域通過フィルタを構成するためには各々30次の次数を必要とした。

$$k_1 = \sqrt{\frac{10^{\frac{A_c}{10}} - 1}{10^{\frac{A_a}{10}} - 1}} \quad (4.1)$$

- 低域通過フィルタの場合

$$k = \frac{\tan \frac{\omega_c T}{2}}{\tan \frac{\omega_a T}{2}}$$

- 高域通過フィルタの場合

$$k = \frac{\tan \frac{\omega_a T}{2}}{\tan \frac{\omega_c T}{2}}$$

## 基準ローパスフィルタの設計

振幅特性  $|G_n(\omega)|^2$  の 2 乗特性  $|G_n(\omega)|^2$  が

$$|G_n(\omega)|^2 = \frac{1}{1 + \omega^{2N}} \quad (4.2)$$

となるようなフィルタが  $N$  次バターースフィルタである。  $\omega_c=1\text{rad/s}$  が遮断周波数である。

$\omega = p/j = -jp$  を (4.2) 式に代入すると

$$\begin{aligned} |G_N(p)|^2 &= G_N(s)G_N(-p) \\ &= \frac{1}{1 + (-jp)^{2N}} \\ &= \frac{1}{1 + (-p^2)^N} \end{aligned} \quad (4.3)$$

が得られる。(4.3) 式の極は  $N$  が奇数の場合、

$$p^{2N} = 1$$

を解くと

$$\begin{aligned} p_k &= e^{j\frac{k\pi}{N}} \\ &= \cos\left(\frac{k}{N}\pi\right) + j\sin\left(\frac{k}{N}\pi\right) \quad (k = 0, 1, 2, \dots, 2N - 1) \end{aligned} \quad (4.4)$$

となる。

また、 $N$  が偶数の場合には

$$p^{2N} = -1$$

を解くと

$$p_k = e^{j\frac{(2k+1)\pi}{2N}} \cos\left(\frac{2k+1}{2N}\pi\right) + j \sin\left(\frac{2k+1}{2N}\pi\right) \quad (k = 0, 1, 2, \dots, 2N - 1) \quad (4.5)$$

となる。

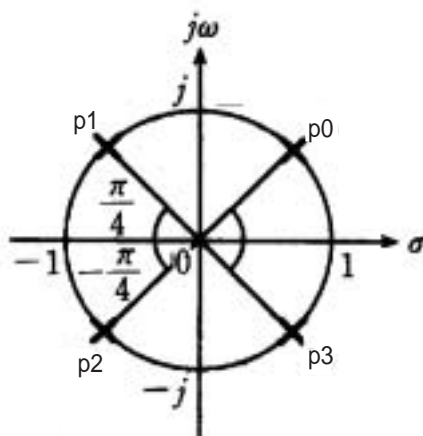


図 4.1: バタワースフィルタの極の位置 ( $N=2$ )

図 4.1 から安定な極は  $p_1 = e^{\frac{3}{4}\pi}$ ,  $s_1 = e^{\frac{3}{4}\pi}$  であるので  
2 次のバタワースフィルタの伝達関数は

$$G_2(p) = \frac{1}{(p - p_1)(p - p_2)} \frac{1}{p^2 - 2 \cos\left(\frac{3}{4}\pi\right)p + 1} \frac{1}{p^2 + 2 \cos\left(\frac{1}{4}\pi\right)p + 1} \frac{1}{p^2 + \sqrt{2}p + 1} \quad (4.6)$$

(4.6) 式が基準ローパスフィルタの伝達関数となる。



## カットオフ周波数のプリワーピング

本研究で用いた設計法では、 $z$  平面への変換に双 1 次  $s$ - $z$  変換を施している。この場合、アナログフィルタの振幅特性が歪んだ形でデジタルフィルタの振幅特性に写像されてしまい、好ましくない。この問題を回避するために (4.7) 式によってあらかじめアナログフィルタの設計の段階でカットオフ周波数を歪ませておいた。

$$\Omega_c = \frac{2}{T} \tan\left(\frac{\omega_c T}{2}\right) \quad (4.7)$$

$\omega_c$ : デジタルフィルタで要求されるカットオフ周波数

$\Omega_c$ : アナログフィルタの設計に使用するカットオフ周波数

## 周波数変換と双 1 次 $s$ - $z$ 変換

周波数変換と双 1 次  $s$ - $z$  変換に関しては (4.8) 式, (4.9) 式を用いた。

- 周波数変換基準ローパスフィルタの伝達関数

$$\frac{b}{p^2 + ap + b}$$

周波数変換後の伝達関数

$$\frac{H_1}{s^2 + A_1 s + B_1} \quad (A_1 = a\omega_c, B_1 = H_1 = b\omega^2) \quad (4.8)$$

- 双 1 次  $s$ - $z$  変換

$$\begin{aligned} H_2(s) &= \frac{Cs^2 + Ds + E}{s^2 + As + B} \\ H_2(z) &= G \frac{(\beta + D\alpha + E) + 2(E - \beta)z^{-1} + (\beta - D\alpha + E)z^{-2}}{1 + 2G(B - \alpha^2)z^{-1} + G(\alpha^2 - A\alpha + B)z^{-2}} \end{aligned} \quad (4.9)$$

ただし、

$$\alpha = \frac{2}{T}, \quad G = \frac{1}{\alpha^2 + A\alpha + B}, \quad \beta = C\alpha^2$$

### 4.3 フィルタの構成

本研究で用いたフィルタは前節で設計した2次のローパスフィルタとハイパスフィルタを縦続接続することにより必要な仕様を満たしている。図4.2と(4.10)式に2次のIIRフィルタのブロック図とその差分方程式を示す。

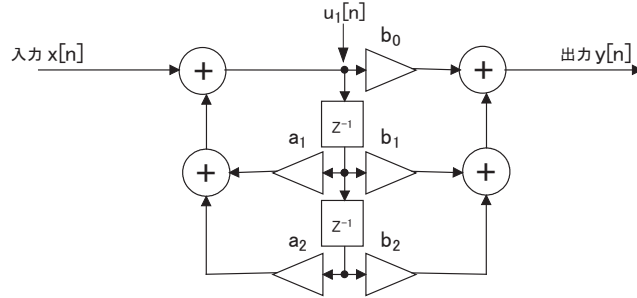


図 4.2: 2 次の IIR フィルタのブロック図

$$\begin{aligned} u_1[n] &= a_1 u_1[n - 1] + a_2 u_1[n - 2] + x[n] \\ y[n] &= b_0 u_1[n] + b_1 u_1[n - 1] + b_2 u_1[n - 2] \end{aligned} \quad (4.10)$$

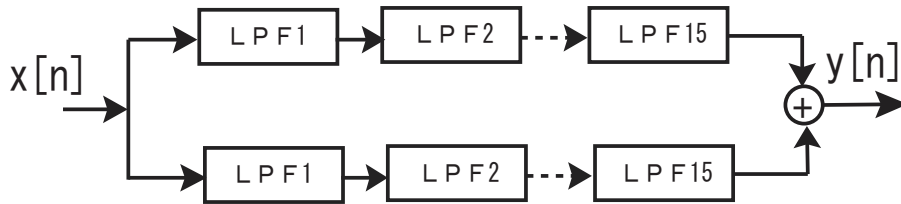


図 4.3: 本研究で構成したフィルタのブロック図

## 本実験で使用するフィルタによって生じる時間遅延

本研究で行うスペクトルに関して実験を行う際に生じる時間遅延は

- AD/DA に要する時間:数  $\mu s$
- フィルタが安定するまでの時間:7.5ms
- 計算に要する時間:125 $\mu s$

となり、計 8ms の時間遅延を生じる。

この 8ms という値は予備実験の際に得られた 20ms 以内にすべての処理を納めなければならないという条件を十分にクリアしている。

# 第5章 本実験

## 5.1 目的

フィードバック音声へフィルタ処理を施した実験を行うことにより、スペクトル制御における音声生成・知覚の相互関係について調査、考察を行う。

## 5.2 方法

### 被験者

特別なボイストレーニングを受けていない、正常聴力を有する大学院生3名。通常発話時のF1とF2は表5.1に示す通りであった。

表 5.1: 通常発話時のF1とF2

被験者	/a/		/i/	
	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)
A	680	1230	300	2300
B	850	1500	250	2000
C	900	1500	200	2400

### 発話音声

連続母音/aiaiai.../とした。

## 準備した刺激

フィードバック音声への刺激に用いるフィルタとして母音/i/の第2フォルマントを200Hzから500Hzの幅で除去を行うノッチフィルタ4種類を用意した。フィルタを切り替えるパターンとして、フィルタの除去幅を徐々に広げていく場合と徐々に狭めていく場合、そしてランダムに変化させる場合の3種類とした。またフィルタを掛けている時間幅は2000ms、3000ms、4000msとして、フィルタの切り替えの際には、間に通常発話音声(フィルタ操作なし)を1000ms挿入した。フィルタ切り替えのタイムチャート例を付録Aに示す。

## 実験方法

- 実験前の準備

被験者には、実験前にヘッドホン聴取による違和感をできるだけ取り除くためにしばらくマイクとヘッドホンからなる擬似的なフィードバック経路を介して、10分から15分程度の会話を行ってもらった。

- 被験者への教示

発話時の、声の大きさ、声の高さ、発話のスピードは被験者が発話しやすいもので構わないが、できるだけ一定の発話をしてもらうようお願いした。

- 記録したデータ

本研究の実験系はTAFの測定手法に準じている。TAFの測定手法では分析のために摂動に用いた信号(フィードバック音声)、その摂動に対する出力(発話音声)を記録する必要がある。本実験では、この2つの信号と合わせて、フィルタの切り替えを示す信号の3つを計算機内のハードディスクへ記録した。また、マイクから20cm離れた位置で発話音声をマイクへの受話音圧として、騒音計により測定した。

- 実験の試行回数

同一の実験を1人の被験者について、3セット行ってもらった。この際、被験者に同一の実験であることは伝えていない。

### 5.3 分析方法

本研究では、分析に高品質音声分析変換合成法 STRAIGHT(河原 2001)を用いた。STRAIGHTによる分析から得られたスペクトログラムから、ある時刻のスペクトル包絡を求め、フォルマント形状の変化を観測した。観測の際、着目した点として、除去を行った帯域のパワーが通常発話と比較して、フィルタを掛けた場合に変化が見られるか、除去を行った帯域周辺でどのような変化が起こるかという2点に着目した。息継ぎなどの理由から発話波形が定常状態に達していない時間のデータは分析の対象から除外した。分析方法の概要を図5.1に示す。

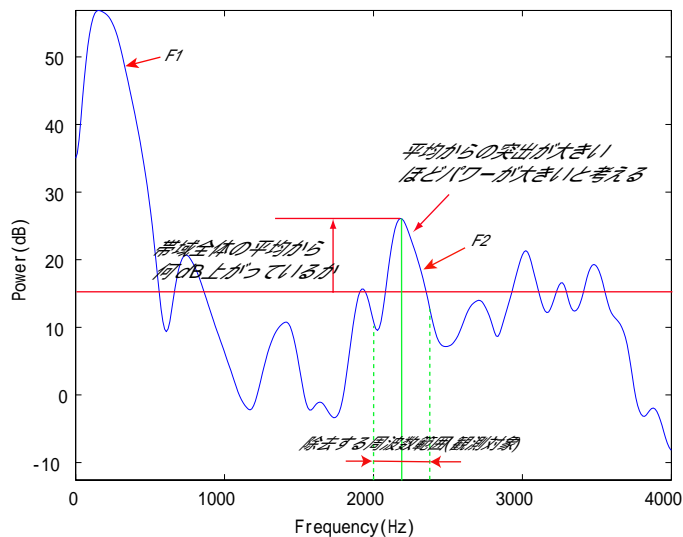


図 5.1: 分析方法

## 5.4 連続母音 /ai ai.../ による実験結果

### 5.4.1 結果

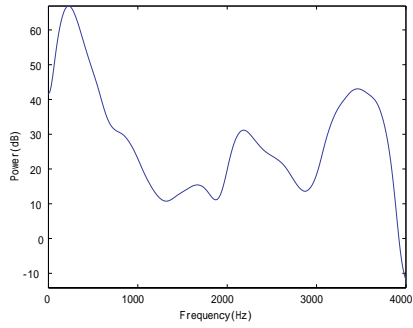
#### フィルタの除去幅を徐々に広げた場合

フィルタの除去幅を広げていった場合の結果を表 5.2 に示す。またそのときに起こるスペクトル包絡の例を図 5.2(被験者 A 実験 2) に示す。5.2 の結果から、フィルタの除去幅を広げていった場合にフィルタの除去幅が 200 から 300Hz の場合に通常発話と比較すると、フィルタ操作を行った場合には、除去したフォルマントのパワーが数 dB 大きくなる傾向にあることがわかる。そして、フィルタの除去幅を 400、500Hz と広げると、逆に通常発話よりも除去したフォルマントパワーが数 dB 小さくなる傾向にあることがわかる。しかし、フィルタ処理を行った際に生じるパワー変化の大きさがどの程度であるかを推測できる傾向は現れていない。

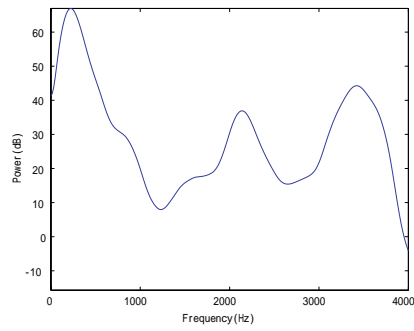
また、図 5.2 から、フィルタの除去幅を徐々に広げた場合には多少、パワーの増減があるが、スペクトル包絡の基本的な形状に目立った変化は起こっていないことがわかる。

表 5.2: 除去幅を広げていった場合の突出度

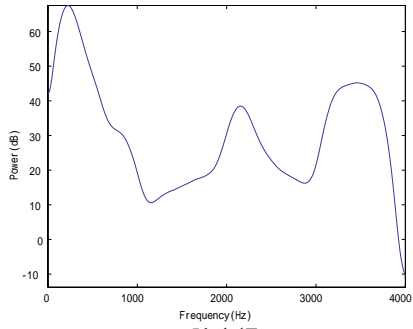
		通常発話時とフィルタ操作時の突出度の差 (db)			
除去幅の変化 (Hz)		200	300	400	500
被験者 A	実験 1	7.2	2.7	2.3	9.5
	実験 2	4.2	13	-6.1	-8.5
	実験 3	2.3	4.6	-3.1	-1.4
被験者 B	実験 1	0.2	3.2	-5	0
	実験 2	4.0	2.7	-3.4	12.9
	実験 3	4.7	2.9	-2.3	3.1
被験者 C	実験 1	-	-	-	-
	実験 2	3.5	1.4	-0.5	0.3
	実験 3	2.1	0.8	1.7	-0.2



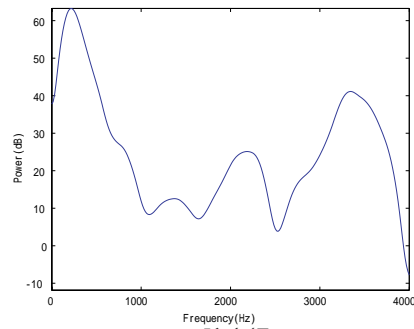
(a) 通常発話



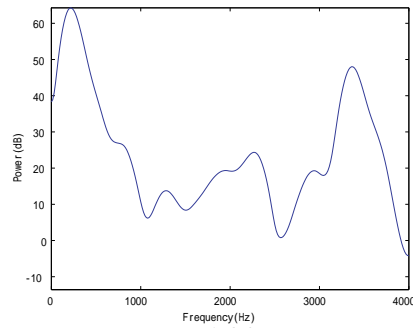
(b) 除去幅200Hz



(c) 除去幅300Hz



(d) 除去幅400Hz



(e) 除去幅500Hz

図 5.2: 除去幅を徐々に広げた場合のスペクトル包絡例:除去幅を変化させてもスペクトル包絡の基本的な形状に変化はあまり見られない



## フィルタの除去幅を徐々に狭めた場合

フィルタの除去幅を狭めていった場合の結果を表 5.3 に示す。またそのときに起こるスペクトル包絡の例を図 5.3(被験者 A 実験 6) に示す。フィルタの除去幅を徐々に狭めた場合にも、フィルタの除去幅を徐々に広げた場合と同様に除去した帯域でパワーの増減が見られるが、その増減の様子に規則性は見られない。しかし、図 5.3 に見られるようなフィルタの除去幅が大きい場合に除去した帯域周辺でスペクトル包絡の形状に大きなパワーの増減が起こり、除去幅を狭めていくと、その増減は小さくなり、通常発話音声にスペクトル包絡へ近づいていくという傾向がスペクトル包絡の変化に見られる。図 5.3 は変化の様子が顕著に現れた例である。

表 5.3: 除去幅を狭めていった場合の突出度

		通常発話時とフィルタ操作時の突出度の差 (db)			
除去幅の変化 (Hz)		500	400	300	200
被験者 A	実験 4	-7.9	-5.7	-0.2	-4.4
	実験 5	-4.2	6.3	7.3	4.6
	実験 6	6.7	7.6	6.2	5.9
被験者 B	実験 4	-4.4	1	1.6	8.1
	実験 5	-0.6	0.7	6.9	4.5
	実験 6	-3.6	-0.1	-2.2	2.4
被験者 C	実験 4	-9.7	-15	-9.7	-10.7
	実験 5	-4.4	-4.7	-2.5	-3.7
	実験 6	-5.6	-5.1	-4.4	-3.8

## フィルタの除去幅をランダムに変化させた場合

フィルタの除去幅をランダムに変化させた場合の結果を表 5.4 に示す。フィルタの除去幅をランダムに変化させた場合には、除去した帯域でパワーの増減は見られるが、その増減の様子に規則性は見られない。またスペクトル包絡の形状の変化にも何も傾向が現れなかった。

表 5.4: ランダムに除去幅を変化させた場合の突出度

		通常発話時とフィルタ操作時の突出度の差 (db)			
除去幅の変化 (Hz)		200	300	500	400
被験者 A	実験 7	5.5	5.7	7.3	1.5
	実験 8	4.2	-3.8	-8.2	0.6
	実験 9	2.4	-1.2	0.3	2.4
被験者 B	実験 7	3.8	4.1	4.5	6.4
	実験 8	-12.6	4.2	-6.4	-9.1
	実験 9	6.3	5.0	-0.3	3.1
被験者 C	実験 7	-4.1	-6.4	-3.4	-2.8
	実験 8	-0.7	-0.7	0.8	-12.9
	実験 9	-4.0	6.1	1.3	1.5

## 5.4.2 考察

母音の音韻性を保つためには、フォルマント周波数と各フォルマント同士のパワー比が重要であることがわかっている。表 5.2 から表 5.4 に示す除去した帯域のパワーの増減は、音韻性を保つために除去された帯域を強調してフォルマントを維持しようとする、あるいはフォルマント周波数のパワーを増減することによりフォルマント同士のパワー比を保とうとする音声生成と知覚の間に存在する補償動作の 1 つであると考えられる。

除去幅を徐々に広げていった実験で除去幅が 400、500Hz の場合に除去した帯域のパワーが減少する傾向があったが、この際のフォルマント幅が 200 から 300Hz 程度であることから補償動作とフォルマント幅の間に何らかの関係が存在することを示しているのではないかと考えられる。

また、フィルタの除去幅を徐々に狭めていった実験結果で見られたスペクトル包絡の形状変化は、除去された帯域周辺のパワーを上げることで、フォルマント周波数の位置を探し出そうとしている、あるいは除去された帯域周辺のパワーを上げることで、除去された帯域のパワーを上げようとしているのではないかと考えられる。

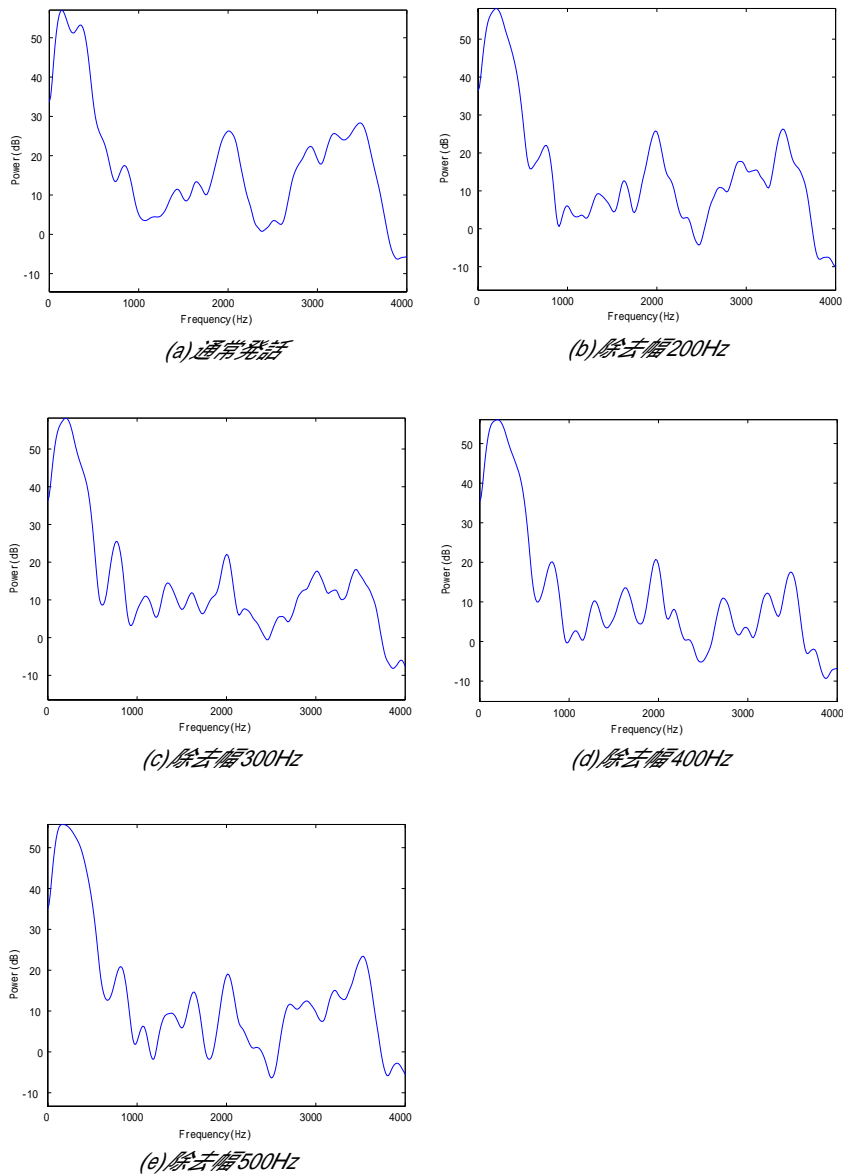


図 5.3: 除去幅を徐々に狭めた場合のスペクトル包絡例:除去幅が 500Hz の場合に除去した F2 の両側で大きなパワーの変化が起こっている。そして除去幅を小さくしていくと、F2 の両側で起こるパワーの変化も小さくなり、除去幅が 200Hz の場合にはスペクトル形状が通常発話の場合とかなり似た形状になっている。

# 第6章 結論

## 6.1 本論文のまとめ

本研究では、発話音声のスペクトルを制御した音声を聴覚器官へフィードバックし、そのフィードバック音声に対する音声生成側の反応を発話音声の分析から音声生成・知覚の間にある相互関係を明らかにすることを目的とした。

測定手法 TAF に準じた実験系を構築して、単母音/a/と連続母音/aiai/の第2フォルマントを除去のターゲットとして、実験を行った結果、音声生成と知覚の間に次のような動作の存在が推測された。

### 本研究から推測された補償動作

- 除去されたフォルマントを強調しようとする動作フィルタよりによって除去された帯域のパワーを数 dB 上げることで除去された帯域を強調しようとする動作。
- フォルマント形状を維持しようとする動作フィルタによって除去された帯域周辺のパワー上げることで、フォルマント形状を維持しようとする、あるいは正確なフォルマント位置を探し出そうとしている動作。

本研究は、音声生成と知覚のメカニズムの解明し、音声生成と音声知覚のモデルを構築しようとする研究分野に貢献している。また、本研究は、フィードバック音声の変化に対して、生成系が反応することを利用することで、発話障害者のリハビリ支援を目的とした分野へ応用される可能性がある。

### 本実験での問題点

- ヘッドホン聴取による影響  
ピンクノイズにより骨導音のマスクングを行っているために、フィルタ操作を行っていない場合でもヘッドホンへフィードバックされる音声に日常の発話との違和感を感じる。この違和感を取り除くためにイコライザなどを用いてヘッドホンからフィードバックされる音声を調整する必要がある。

- フィードバック音声に摂動を与える手法

本研究ではスペクトルに摂動を与える手段として、除去幅固定のノッチフィルタを用いている。被験者ごとにフィルタを作成してフォルマントの除去を行ったが、発話ごとにフォルマント周波数が 100Hz 前後の範囲で揺れているために常に同じ条件でフォルマントを除去できていない可能性がある。この問題を回避するために能動的にフォルマント位置を推定し、フォルマントに操作を加えることのできるシステムへ実験系を変更する必要がある。

## 6.2 今後の課題

- 筋電計やカメラを使って、実際に口の回りの動きなどを測定する。発話時の口唇周辺の動きを筋電計やカメラで測定、観測することにより、フィードバック音声に変化した際の音声生成側の反応を調査することにより、音声生成と知覚側の間にあるメカニズムをより明らかにする。

## 参考文献

- [1] B.S.Lee, Effect of Delayed speech feedback, J,Acoust.Soc.Am.22, 824-826 ,1950.
- [2] Hideki Kawahara, Julia Catalina Williams, Effects of Auditory Feedback on Voice Pitch Trajectories:Characteristic Respons to Pitch Perturbations, ATR HIP Res.Labs, Vol.2 of 2, pp145-160, 1996.
- [3] 河原英紀, 音声知覚・生成相互作用の伝達特性について, 情報処理学会研究報告, NL-84-13, pp95-102, 1991.
- [4] 河原英紀, 加藤比呂子, J.C. ウィリアムズ, 聴覚による発声の制御モデルとシミュレーション, 秀英出版, 1964.
- [5] 佐藤裕, 森浩一, 福島康弘 聴覚による F0 制御の時間特性と吃音検査への応用の可能性, 日本音響学会聴覚研究会資料, Vol.32, No.2, pp181-186, 2002.
- [6] 中山一郎, 岡田稔枝, 中川みかほ, 山田真司, 発話者が知覚する音声の音色-その客観的評価-, 日本音響学会講演論文集 1 of 2,pp349-350, 1996.
- [7] 中山一郎, 芝麻子, 三浦裕美子, 発話者が知覚する音声の音色-録音/再生音との比較, 聴覚研究会資料 H-95-86, 1995.
- [8] 森友一朗, 薬師輝久, 馬場秀忠, RTLinux リアルタイム処理プログラミングハンドブック, 秀和システム, 2000.
- [9] レイ・D・ケント / チャールズ・リード, 音声の音響分析, 海文堂出版株式会社, 1996.
- [10] 尾知博, デジタル・フィルタ設計入門, CQ 出版株式会社, 1992.
- [11] 河原秀紀, 高品質音声分析変換合成法 STRAIGHT, ATR 人間情報通信研究所, 和歌山大学, 2001.

# 謝辞

本研究を進めるにあたり、日頃から数多くの貴重な御助言、御指導頂きました、北陸先端科学技術大学院大学 情報科学研究科 赤木 正人教授、党 建武助教授、鷓木 祐史助手、並びに本学教官の皆様に深く感謝致します。また御多忙の中、御助言や御討論頂き、また、聴取実験に御協力頂いた赤木・党研究室の皆様に感謝致します。

最後に、研究を進めるにあたり日頃からあたたかく見守って下さった家族や友人に深く感謝致します。



# 付録A

## A.1 本実験で用いたフィルタ変更のタイミング

ここではフィードバック音声に対して施した各フィルタ操作を次のような略記で示している。

- 無し:フィルタ処理を行わない
- No1:/i/の F2 を中心周波数から 200Hz の幅でスペクトルを除去
- No2:/i/の F2 を中心周波数から 300Hz の幅でスペクトルを除去
- No3:/i/の F2 を中心周波数から 400Hz の幅でスペクトルを除去
- No4:/i/の F2 を中心周波数から 500Hz の幅でスペクトルを除去

表 A.1: フィルタの変化パターン 1

フィルタ操作	フィルタをかけている時間幅 (ms)		
	実験 1	実験 2	実験 3
無し	5000	5000	5000
No1	2000	3000	4000
無し	1000	1000	1000
No2	3000	4000	2000
無し	1000	1000	1000
No3	4000	2000	3000
無し	1000	1000	1000
No4	3000	4000	2000

表 A.2: フィルタの変化パターン 2

フィルタ操作	フィルタをかけている時間幅 (ms)		
	実験 4	実験 5	実験 6
無し	5000	5000	5000
No4	3000	4000	2000
無し	1000	1000	1000
No3	4000	2000	3000
無し	1000	1000	1000
No2	3000	4000	2000
無し	1000	1000	1000
No1	2000	3000	4000

表 A.3: フィルタの変化パターン 3

フィルタ操作	フィルタをかけている時間幅 (ms)		
	実験 7	実験 8	実験 9
無し	5000	5000	5000
No1	2000	3000	4000
無し	1000	1000	1000
No2	4000	2000	3000
無し	1000	1000	1000
No4	3000	4000	2000
無し	1000	1000	1000
No3	2000	3000	4000