修 士 論 文

連続発話音声中に含まれる男声·女声知覚に寄与する音響特徴量に関する研究

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

柴田 武志

2008年3月

修 士 論 文

連続発話音声中に含まれる男声·女声知覚に寄与する音響特徴量に関する研究

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　鵜木祐史 准教授
審査委員　党建武教授

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

610043 柴田 武志

提出年月: 2008 年 2 月

概 要

本論文では、連続発話音声中に含まれる音響特徴量を静的な特徴、および動的な特徴に分類し、これらがどのような順序で男声 · 女声知覚に寄与しているか明らかにすることを目的とする。男声 · 女声音声から得られた各特徴量を表すパラメータ値について、各パラメータ値が男声 · 女声という分類で違いがあるかを確かめるために MDS で分析を行った。その結果、男声と女声の音声では静的特徴（平均基本周波数、スペクトル包絡、ゲイン）、および、動的特徴（スペクトル変化、音韻長、基本周波数の変化）に違いがあることがわかった。次に分析によって、違いが見られた各特徴量を表すパラメータ値が男声 · 女声知覚にどう影響を与えているかどうかを調べるために、実験 1 では男声 · 女声音声から得られた各特徴量を表すパラメータ値を平均した音声に、男声 · 女声の静的特徴（平均基本周波数、スペクトル包絡、ゲイン）、および、動的特徴（スペクトル変化、音韻長、基本周波数の変化）を付加した合成音声を用いて、男声 · 女声を判別する聴取実験を行った。実験 1 の結果、動的成分に比べて、静的成分の男声 · 女声知覚に対する寄与が高いことが明らかになった。男声 · 女声知覚について、静的な特徴である基本周波数とスペクトル包絡の影響が大きいということは、先行研究の結果を支持するものである。実験 2 では，実験 1 で影響の強い平均基本周波数とスペクトル包絡を固定し，実験 1 の結果から動的特徴の中で男声 · 女声知覚に影響を与えた “基本周波数の変化”，“語尾の変化”，“音韻長” といった特徴が知覚にどのような影響を与えているか調査した。実験 2 の結果から、語尾が動的成分の中で一番影響を与えていることが明らかになった。全体の傾向として，女声と判断された動的特徴を付加していくと，女声らしく知覚されるという結果が得られた. この結果は先行研究における女声らしさには話し方が影響を与えているという知見を支持するものであった。実験 1 と実験 2 の結果，男声 · 女声知覚には静的特徴である平均基本周波数とスペクトル包絡が大きな影響を与えており，次いで，動的特徴である基本周波数の変化と音韻長が影響を与えており，スペクトルの変化とゲインのダイナミックレンジはあまり影響を与えていないことが明らかになった.

目 次

図 目 次

表 目 次

第1章　序論

1.1　はじめに

人間は音声を聞くことで話者が男性か女性という性別の情報を得ることができる．これは人間が音声に含まれる男声 · 女声知覚に寄与する音響特徴量を知覚しているからである．これまで，男声 · 女声知覚は発声器官および調音器官の形状に起因する平均基本周波数やスペクトルである静的特徴が重要であるといわれてきたが，発声器官および調音器官の運動による基本周波数やスペクトルの時間変化パターンである動的特徴についてはあまり議論されていない [15][16]．本稿では，静的特徴である平均基本周波数やスペクトルと動的特徴である基本周波数の変化やスペクトルの変化が男声 · 女声知覚にどのような順序で寄与しているかを明らかにすることを目的とする．

1.2　研究背景

男声 · 女声の判別知覚は静的特徴が重要であるといわれている一方で動的特徴が男声 · 女声知覚に寄与しているといえるのだろうか? 動的特徴が寄与している例として個人性知覚における知見を述べる。個人性とは話者の特徴のことを示している。そして話者の特徴とは誰が話しているかという情報と「通る声」や「澄んだ声」などの声質に関する情報がある。このことから、個人性知覚における話者の特徴を性別の違いによる特徴の変化と捉えれば男声 · 女声の判別知覚は個人性知覚と近い関係であるといえる。今までに個人性知覚を明らかにするために多くの研究が行われている [1] [2][3][4][5][6][7][8][9][10]。その中で動的特徴が知覚に影響を与えている例として、家永と赤木が, 単語音声に含まれる基本周波数の時間変化パターンを対象に研究を行い, 基本周波数の時間変化パターンに個人性が多く含まれることを示している [11][12]。この例で個人性知覚では動的特徴が影響を与えていることから、男声 · 女声知覚に対しても動的特徴が寄与している可能性がある。

1.2.1　男声 · 女声知覚に関する研究

男声 · 女声知覚に関する先行研究では、基本周波数とフォルマント情報が男声 · 女声知覚に重要であることを示した [15][16]。さらに櫻庭らの報告によると、女性声を獲得するには、高くなく低くない声の高さと、女性らしい話し方が必要である。声の高さが重要

であることはわかっているが男性の基本周波数をあげただけでは女性らしく聞こえない。声の高さと同様に重要なものに女性らしい話し方がある。語尾を伸ばす、語尾をあげる、抑揚に富む、軟起声を使用する、鼻音化する、などがあげられる [17][18][19]。先行研究では、男声 · 女声知覚では静的な特徴である基本周波数とフォルマント情報が重要であり、さらに女声らしく話すためには話し方つまり動的な特徴が重要ではないかという知見が得られている。

1.2.2 声質変換に関する研究

男声 · 女声知覚に関係すると思われる特徴量だけを変形した音声を作成するために、声質変換手法を採用する。本研究では、ソースフィルタモデルに基づいて静的、動的特徴を扱う。静的特徴を声帯と声道の形状に起因する特徴、そして動的特徴を声帯と声道の動きに起因する特徴とする。そこで、ソースフィルタモデルを基にした声質変換手法の先行研究について説明する。これまで個人性の研究に声質変換を用いた例としては，藤野らは母音スペクトルと基本周波数に着目し、目標話者の母音のみの音声データでの声質変換を試みた [20]。里地らは、重み付け法を適用してスペクトル交換を行い、静的成分のスペクトル近似を試みた。その結果単独発話母音での、重み付け法の有効性がみられたが、単語音声ではスペクトルの十分な類似性がみられなかった [21]。Nguyen と赤木は文音声を用いて静的特徴である基本周波数とフォルマント情報を利用して STRAIGHT(Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum)[25][26][27][28] と TD(Temporal Decomposition)[33] および GMM(Gaussian mixture model)[23][24] を用いた声質変換を行い男声から女声にかえた。その結果女声から男声 100 %変化し、女声から男声は約 83 %という結果になった [22]。

ここで紹介した先行研究はさきほど説明した静的特徴を目標話者に向けて変形させた例であるが、静的特徴と動的特徴を考慮した声質変換モデルは考えられていない。最初に男声 · 女声に関係すると思われる静的 · 動的特徴量を得るために、静的特徴と動的特徴を考慮した声質変換モデルを構築する必要がある。

1.3 本研究の目的

本研究では人間が男声 · 女声を知覚するときに静的、動的といった音響特徴量がどのように寄与しているか明らかにするために、ソースフィルタモデルを基として動的特徴を考慮した声質変換モデルを提案する。そしてモデルを利用して分析、聴取実験を行うことで静的特徴と動的特徴が男声 · 女声知覚に与える影響を明らかにすることを目的とする。本研究によって、静的特徴と動的特徴の関係が明らかになることで、男声 · 女声知覚における新しい知見を得ることができる。今後静的特徴と動的特徴を考慮した声質変換にも応用できると考えられる。

1.4　研究方法

本研究では、連続発話音声中に含まれる男声・ 女声知覚に寄与している音響特徴量を静的な特徴、および動的な特徴に分類し、これらがどのような順序で寄与しているか明らかにするために、以下に示す 3 つの目標をたて問題を解決した。

[**声質変換モデルの構築**]

STRAIGHT と TD を基本コンセプトとした声質変換手法を用いることで、静的特徴と動的特徴を考慮した声質変換モデルを提案する。

[**声質変換モデルから得たパラメータ値の分析**]

MDS を用いて分析することで静的特徴と動的特徴を表すパラメータ値が男声と女声に違いがあるのかを明らかにする。

[**聴取実験**]

男声と女声で MDS による分析を行い、その結果違いがみられたパラメータ値を用いて刺激音を作成する。静的特徴 · 動的特徴を表すパラメータ値がどの程度男声 · 女声知覚に影響を与えているかを聴取実験により確かめる。

1.5　本論文の構成

本論文の構成を以下に示す。

第 1 章では、本論文の対象としている研究背景に関する研究分野の現状と問題点を指摘し、本論文の目的を明らかにする。

第 2 章では、声質変換手法の概要について述べる

第 3 章では、提案手法の説明、評価を行う

第 4 章では、多次元尺度構成法を用いた分析を行う。

第 5 章では、聴取実験について説明する

第 6 章では、全体の考察を行う。

第 7 章では、本研究で明らかになったことと今後の課題について説明する

以上の全体の構成を最後図 1.1 にまとめる。

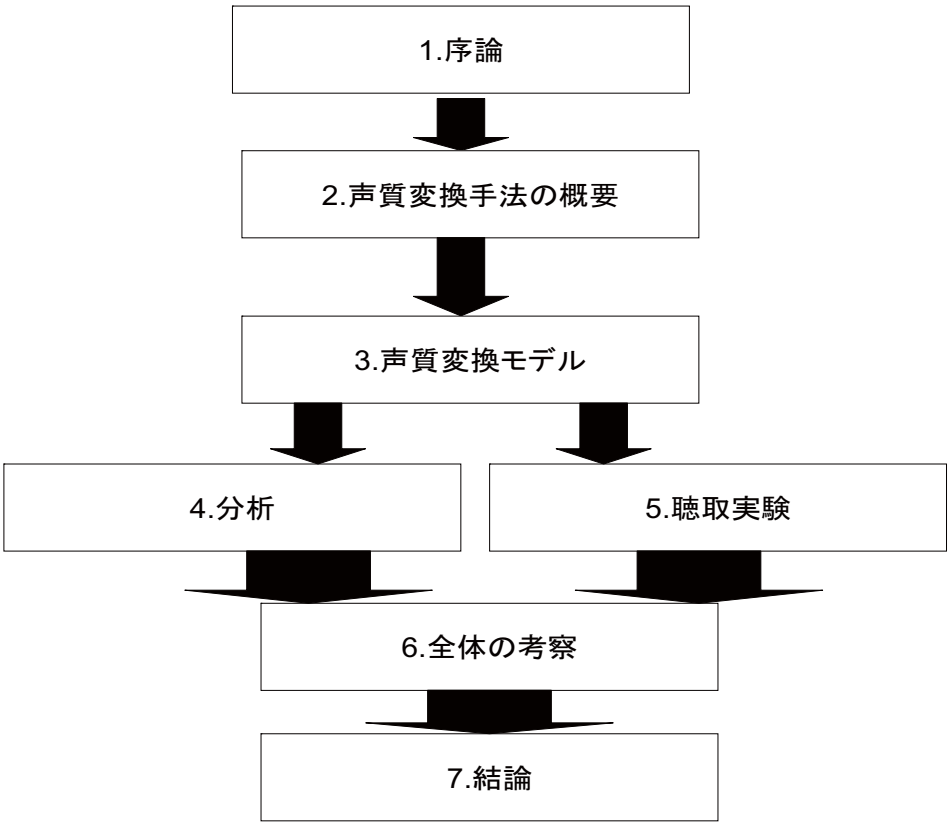

図 1.1: 全体の構成

第2章　声質変換手法の概要

2.1　目的

本研究の目的を達成するためには、静的、動的特徴を制御できる声質変換が行えることが必要である。そこで、2 章、3 章ではソースフィルタモデルに基づき静的、動的な特徴を分けることができる声質変換モデルを提案する。静的、動的な特徴を用いた声質変換を行うことができれば、それぞれのどの特徴が知覚に影響を与えているか、与えていないかを調べることができる。

まず、声質変換モデルがもつ条件として以下のように設定する。

条件 1：**聴取実験で使うことができる音質のよい合成音声**

条件 2：**動的特徴が扱えること**

条件 3：**男性と女性でパラメータの対応がとれること**

条件 4：**時間変化をパラメータとして扱えること**

STRAIGHT[25][26][27][28] と MRTD[30][31] を用いる。これらを用いて、条件を満たすようなモデルを作成することを目的とする。

2.2　声質変換モデルの流れ

声質変換モデルの流れを図 2.2 に提示する。図 2.2 で、まず音声を入力する。そして、その入力音声を最初のブロックである STRAIGHT 分析を行い 3 つのパラメータとして、基本周波数、スペクトル包絡、非周期成分を出力する。そして、スペクトル包絡は LSF(Line spectral frequency) というパラメータに変換する。そして、次に MRTD を用いてスペクトルパラメータ (LSF) をイベントターゲット (静的成分）とイベント関数 (動的成分) に時間分解する。そしてイベント関数は、男声、女声の動きに変形して、MRTD によって合成する。変形方法に関しては 3 章で説明する。得られた各パラメータを同じイベント関数で記述することで静的特徴と動的特徴をわけている。一方イベントターゲットは変形を行わず MRTD 合成する。その結果、男声から女声の静的、動的な特徴を変形することが可能になる。スペクトル包絡、基本周波数などのパラメータの変形後、最後に STRAIGHT を用いて合成する。その結果として合成音声が完成する。次の節からは、モデルの基礎となっている STRAIGHT と MRTD について説明する。

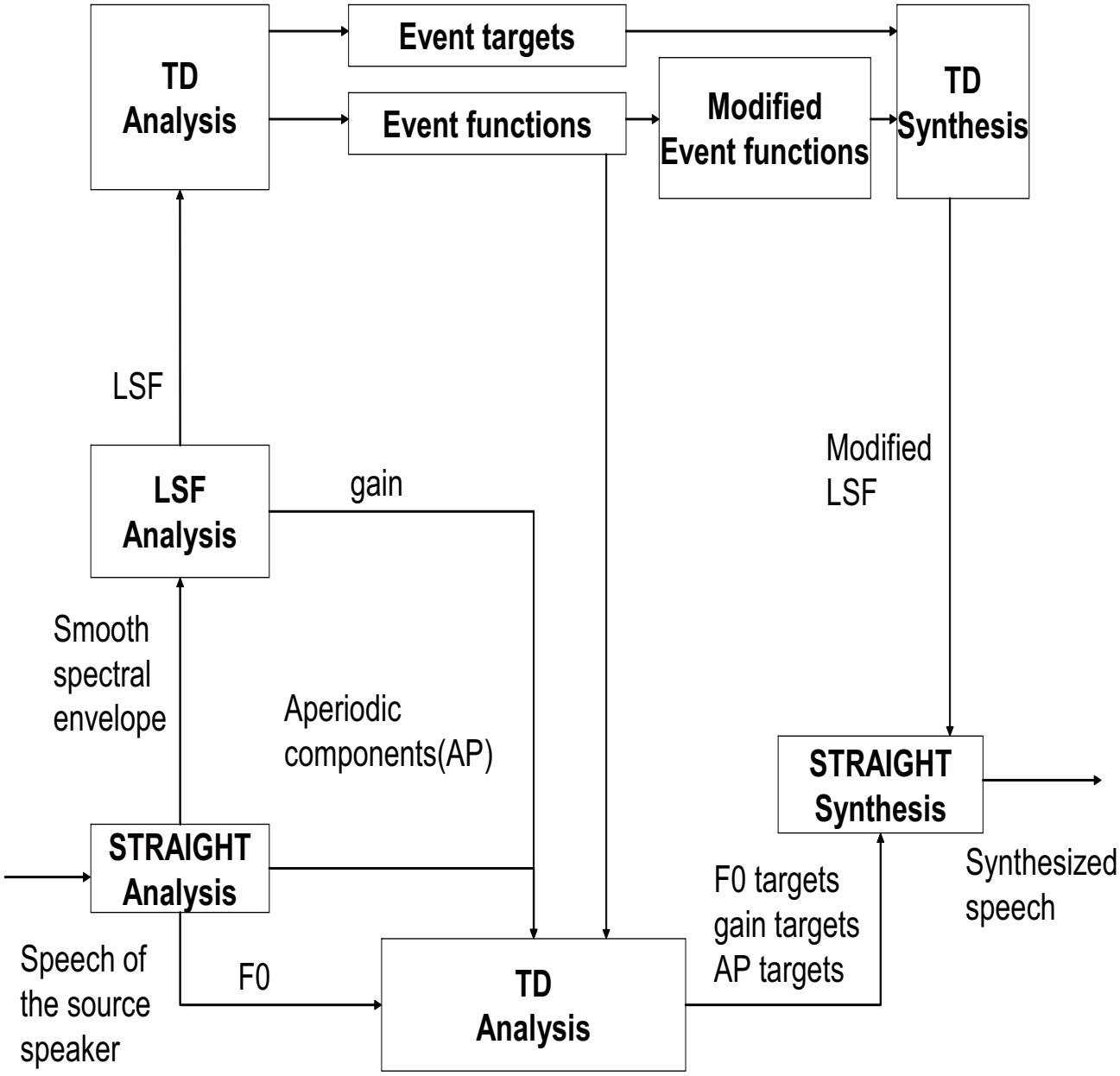

図 2.1: 声質変換モデルのブロックダイアグラム

2.3 音声分析合成系

2.3.1 STRAIGHT

STRAIGHT[25][26][27][28] は、STRAIGHT-core、SPIKES、TEMPO2 の 3 つの主要な部分から構成されている。STRAIGHT-core は、音声の励振の周期性による干渉の影響のない時間周波数表現を抽出する方法である。その中心的なアイディアは基本周期、基本周波数を節点とする区分的線形関数による補間と等価な時間周波数領域の平滑化を行なうことにある。SPIKES は、合成に用いる駆動音源の位相特性を操作することにより、VOCORDER 特有の buzzy な音色を軽減する方法である。ここでは、同一のパワースペクトルであっても群遅延を操作して時間的な微細構造を変えることで音色が変化することを利用している。TEMPO2 は、2 つのフィルタ出力の微分の特性を基に、音声の基本周波数を推定する方法である。特別なフィルタ設計と搬送対雑音比（C/N 比）の組み合わせにより、基本周波数の推定が正確なものになっている。本研究では STRAIGHT バージョン 40-006b を使用した。

2.3.2 Temporal Decompositon

TD は以下のように、イベントベクトルの線形結合によってスペクトルパラメータの時間変化を近似する [33]。

$$\hat{\mathbf{y}}(n) = \sum_{k=1}^{K} \mathbf{a_k} \phi_k(n),\ 1 \leq n \leq N, \tag{2.1}$$

ここで、 $\mathbf{a_k}$、$\phi_k(n)$ は、それぞれ k 番目イベントターゲット、イベント関数である ($K \ll N$) 。$\hat{\mathbf{y}}(n)$ は、n 番目スペクトルパラメータの近似値である。式 2.1 を行列表示すると以下のようになる。

$$\hat{\mathbf{Y}} = A\Phi \quad \hat{Y} \in R^{P \times N}\ ,\ A \in R^{P \times K}\ ,\ \Phi \in R^{k \times N}. \tag{2.2}$$

ここで、P、N そして K は、それぞれスペクトルパラメータの次数、音声区間におけるフレーム数、イベントの数である。発話中における各イベントは、時間と共に徐々に増加、減少していき、それらは隣同士重なり合う。よって、時間変化パターンを表すイベント関数には以下の特性が考えられる。

- 各イベント関数には始まりと終りの時間が存在する、すなわち時間間隔が存在する。
- 各イベント関数はその存在期間においては非負である。
- 各イベント関数は実際の発話における音声生成と同様、緩かな増加と減少で表せる。

2.3.3 MRTD

この節では、STRAIGHT によって得られたスペクトル包絡を LSF というスペクトルパラメータに変換し、イベントターゲット（静的な成分）とイベント関数（動的な成分）に分解する。

次のような理由から LSF[32] を用いる。

1. 線形補完性が優れている
2. TD に用いられているスペクトルパラメータの中で、より歪みが少なく再現性が良い

STRAIGHT から得られるスペクトルから LSF への変換方法は [9][21][20] を参照。

MRTD を用いる理由は

1. スペクトルパラメータを静的な特徴と動的な特徴に分解するため
2. 分解することによって、動的特徴をイベント関数という式で簡単に表現できる

である。

2.3.4 イベントターゲット

MRTD におけるイベントターゲットを求めるアルゴリズムは [29] を参照。

MRTD では最大スペクトル安定基準に基づく初期のイベント中心位置を決定する方法を Spectral Feature Transition Rate(SFTR) という。SFTR がどうやってイベント中心位置を決定しているかは [34] を参照。

Spectral Feature Transition Rate **による初期イベント位置決定の問題点**

ここでは、SFTR を用いた場合に、男声と女声で音質とイベントターゲット数がどのように変化するか調べるために SFTR 窓の幅を変化させた音声を作成し、音質と窓幅とイベントターゲット数の関係を調べた。

分析合成に用いる音声は ATR 音声データベース男性話者 mht と女性話者 ffs に関する情報を用いた。音声データは文音声の「はい、こちらでけっこうです」を用いた。実行条件は下記の通りである。

- sampling 周波数：8kHz
- 分析窓長：40ms
- 分析シフト幅：1ms
- FFT 長：1024
- LSF 次数：32

この実行条件の元で、合成音声をを作成した。

SFTR 窓を変化させた音声データ

作成した音声はSFTR 窓を 10、20、40、60、80、100、120(ms) の窓を用いて分析合成させた。各窓のときのイベント数は下記の通りである。

- SFTR 窓：10ms イベントの数（男性:276 女性:313)
- SFTR 窓：20ms イベントの数（男性:188 女性:208)
- SFTR 窓：40ms イベントの数（男性:126 女性:138)
- SFTR 窓：60ms イベントの数（男性:96 女性:112)
- SFTR 窓：80ms イベントの数（男性:83 女性:90)
- SFTR 窓：100ms イベントの数（男性:74 女性:83)
- SFTR 窓：120ms イベントの数（男性:71 女性:79)

以上の条件で音声を合成させた結果この文章では 40ms での音質が一番よいことがわかった。SFTR 窓を短くするとイベント数が増加し音質がよくなる。しかし窓長を短くしすぎても音質が悪くなる。逆に SFTR 窓を大きくするとイベントターゲットの数が少なくなり、子音の部分の音が劣化していて音質が悪くなることがわかった。男性と女性ではイベントターゲットの数がどの窓幅でも同じにならないことが明らかになった。そして、男性と女性で各イベントターゲットの数が違ってしまうと、同じ発話内容であっても男性の発話音声と女性の発話音声で対応関係がとれないために、分析が行うことができない。SFTR を用いるとイベントターゲット数を固定できないことから、本研究の最初の目的である声質変換モデルでは使えないことが明らかになり、SFTR 以外の手法を提案する必要がある。

2.4 まとめ

2 章では声質変換手法の概要について説明した。STRAIGHT を用いることで音声から基本周波数とスペクトル包絡と非周期成分を得ることができ、さらに、MRTD を用いることで、スペクトルパラメータを静的な特徴 (イベントターゲット) と動的な特徴 (イベント関数) を分解することが可能になった。しかし、問題点としてイベントターゲットを決定する場合に SFTR を用いると、イベントターゲットの数が発話者によって違うため、用いることができない。そして、イベント関数をどうやって制御するかという問題が未解決のままである。この問題を解決する方法は 3 章で説明する。

第3章　提案手法

3.1　目的

STRAIGHT と MRTD を用いることでスペクトルパラメータの静的、動的特徴に分けることができた。次に SFTR を用いたときに起こる問題点として、イベントターゲットの数が発話者によって違うため SFTR を用いることができないという問題を解決するために新たなイベント位置決定方法について述べる。次に、各特徴量を表すパラメータ値が男声 · 女声知覚にどう影響を与えているかどうかを調べるために、イベントターゲットの数を同じ発話内容であれば男声 · 女声で同じになること、さらに基本周波数、ゲイン、非周期成分を同じイベント関数で記述し、さらにイベント関数を制御する方法を提案する。

3.2　イベント位置の決定方法

同じ連続発話音声であれば発話者が男声 · 女声でも同じイベントターゲットの数にし、各特徴量を入れ替え、分析ができるようにするためイベント位置決定方法について考える。2 章で述べたが、イベント位置を決定する場合に考えなければならないこととして、

- イベントの数が多すぎても音質が悪くなる
- イベントの数が少なすぎても音質が悪くなる
- 母音より子音の劣化が激しい

といったことが、2 章の最後で SFTR を変化させたときにわかったことである。イベント数を多すぎず少なすぎないイベントターゲット数を考える必要がある。そこで音声の発話時間が記録されているラベル情報を用いることで効率のよいイベントターゲットの設定を行う。そこで、いくつかの条件でイベントターゲットの設定を行った。

1. 母音中心部をイベント位置とした音声
2. 母音と子音の中心をイベント位置とした音声
3. 母音と子音の開始時間と終わる時間と母音と子音の中心をイベントターゲットとした音声

4. 子音と母音の開始と最後と中間地点、最初と中間の中間点、中間と最後の中間点をイベントターゲットとした音声

以上の条件でイベントを設定し、合成音声を作成した。上の条件ほどイベントターゲットの数は少ない。これらの条件と、2 章の後半で説明した SFTR を用いたときにもっとも音質のよかった音声とを何人かに聞かせて、どの条件が一番音質がよいか調査した。その結果、4 番目の条件の場合に SFTR を用いたときより音質がよいとう結果を得た。この結果を元にさらに閉鎖子音の場合を付加したものが最終的なイベントターゲット決定方法である (図 3.2)。

1. 子音と母音の開始点と終了点に 1 イベント
2. 開始点と終了点の中間点に 1 イベント
3. 開始点と中間点、中間点と終了点の間に 1 イベント
4. 子音の ch や d など一度舌で閉じて発話されるような音は、ラベル情報があるものだけを開始点と終了点とその中間点とした。(これに対してはさらに中間点をとるということはしていない)

以上により、同じ発話内容なら話者が異なった場合でも、各パラメータの対応関係がとれる。

3.3 TD を用いた各パラメータの分解

これまで、分解したのはスペクトルパラメータ (LSF) であるが、基本周波数やゲイン非周期成分も同じイベント関数で TD 分解する必要がある。基本周波数やゲイン、非周期成分を同じイベント関数で記述する。STRAIGHT で合成する場合にはスペクトルだけでなくほかパラメータも用いるため、無視できない。さきほど求めたイベント位置で、ほかのパラメータのイベントターゲットを求め、LSF で求めたイベント関数を用いることで、各パラメータが一つのイベント関数でが記述できる。

3.3.1 基本周波数の分解

基本周波数パラメータは、基本周波数のイベントターゲットとスペクトルパラメータから抽出したイベント関数を用いて以下のように再構成できる。

$$\hat{p}(n) = \sum_{k=1}^{K} p_k \phi_k(n),\ 1 \leq n \leq N. \tag{3.1}$$

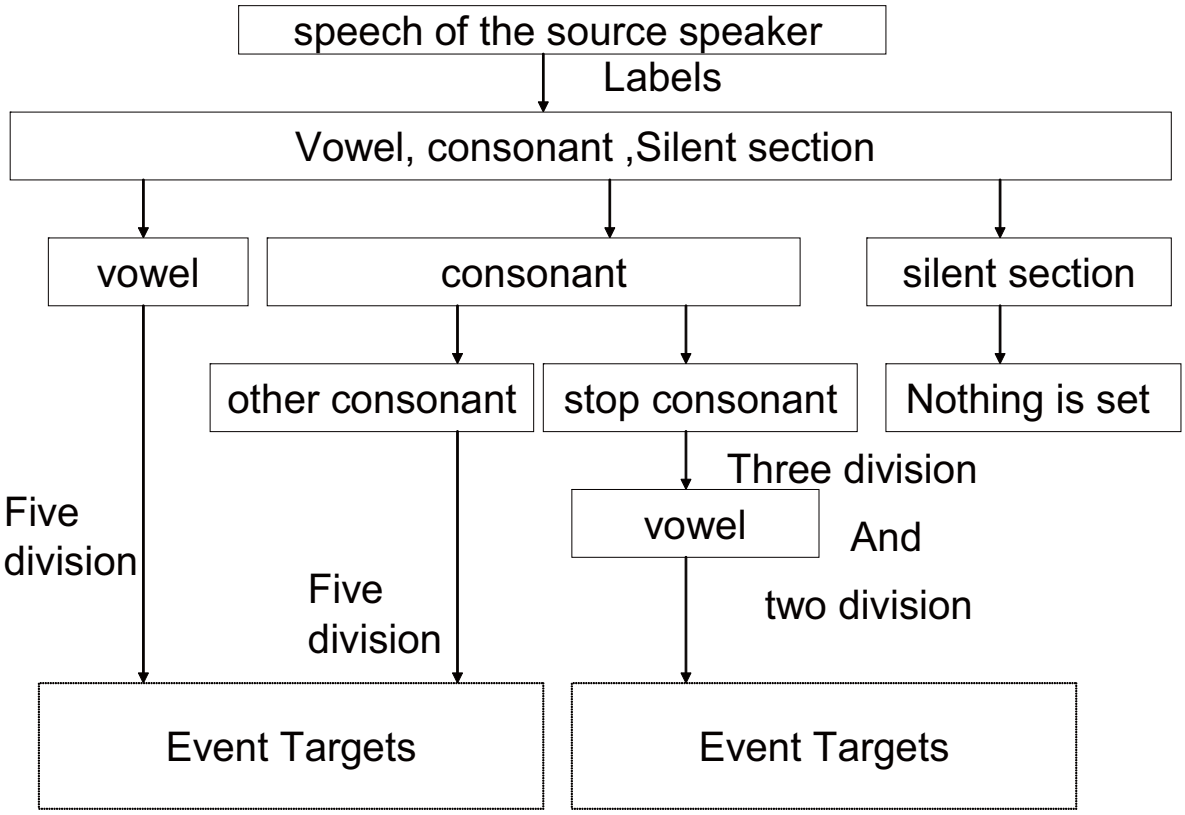

図 3.1: イベントターゲット決定方法

$\hat{p}(n)$ と p_k は、それぞれ n 番目のフレームに対して再構成した基本周波数パラメータと基本周波数ターゲットである。式 3.1 は行列表記すると次のように書かれる。

$$\hat{P} = A_p \Phi, \tag{3.2}$$

$\hat{P}$ と A_p は、それぞれ再構成した基本周波数ベクトルと基本周波数ターゲットベクトルである。A_p は元の基本周波数パラメータと再構成した基本周波数パラメータの二乗誤差を最小にするように決定される。A_p は以下のように表される。

$$A_p = \hat{P}\Phi^T(\Phi\Phi^T)^{-1}. \tag{3.3}$$

以上により、基本周波数ターゲットとスペクトルのイベント関数を用いて基本周波数パラメータを構成した。これにより、基本周波数ターゲット（静的特徴）、イベント関数（動的特徴）に分けることができた。

3.3.2 ゲインの分解

ゲインパラメータは、ゲインのイベントターゲットとスペクトルパラメータから抽出したイベント関数を用いて以下のように再構成できる。

$$\hat{g}(n) = \sum_{k=1}^{K} g_k \phi_k(n),\ 1 \leq n \leq N. \tag{3.4}$$

$\hat{g}(n)$ と g_k は、それぞれ n 番目のフレームに対して再構成したゲインパラメータとゲインターゲットである。式 3.4 は行列表記すると次のように書かれる。

$$\hat{G} = A_g \Phi, \tag{3.5}$$

$\hat{G}$ と A_g は、それぞれ再構成したゲインベクトルとゲインターゲットベクトルである。A_g は元のゲインパラメータと再構成したゲインパラメータの二乗誤差を最小にするように決定される。A_g は以下のように表される。

$$A_g = \hat{G}\Phi^T(\Phi\Phi^T)^{-1}. \tag{3.6}$$

以上により、ゲインターゲットとスペクトルのイベント関数を用いて基本周波数パラメータを構成した。これにより、ゲインターゲット（静的特徴）、イベント関数（動的特徴）に分けることができた。

3.3.3 非周期成分の分解

非周期成分パラメータは、非周期成分のイベントターゲットとスペクトルパラメータから抽出したイベント関数を用いて以下のように再構成できる。

$$\hat{I}(n) = \sum_{k=1}^{K} i_k \phi_k(n),\ 1 \leq n \leq N. \tag{3.7}$$

$$\hat{I} = A_i \Phi \quad \hat{I} \in R^{L \times N}\ ,\ A_i \in R^{P \times K}\ ,\ \Phi \in R^{k \times N} \tag{3.8}$$

ここで、L、N そして K は、それぞれ FFT 長の半分、音声区間におけるフレーム数、イベントの数である。

$$\hat{I} = \ A_i \Phi, \tag{3.9}$$

$\hat{I}$ と A_i は、それぞれ再構成した非周期成分ベクトルと非周期成分ターゲットベクトルである。 A_i は元の非周期成分パラメータと再構成した非周期成分パラメータの二乗誤差を最小にするように決定される。 A_i は以下のように表される。

$$A_i = \hat{I}\Phi^T(\Phi\Phi^T)^{-1}. \tag{3.10}$$

以上により、非周期成分ターゲットとスペクトルのイベント関数を用いて基本周波数パラメータを構成した。これにより、非周期成分ターゲット（静的特徴）、イベント関数（動的特徴）に分けることができた。

これで、基本周波数、ゲイン、非周期成分を同じイベント関数 Φ で表すことができた。その結果、イベントターゲット（静的成分）とイベント関数（動的成分）として取り扱うことが可能になった。

3.4 イベント関数の制御方法

MRTD で抽出されたイベント関数は分析が行えるが、制御ができない。そこで、イベント関数（動的成分）を制御するために、イベント関数をモデル化する。モデル化する前に、MRTD によって得られるイベント関数について整理する。

3.4.1 イベント関数

MRTD ではイベント関数に 2 つの制約が加えられる。1) 時間のどの瞬間においても、隣接する 2 つイベント関数だけで記述する。2) どの時刻においても隣接するイベント関

数の合計は 1 である。この制約を用いれば式 2.1 は次のようになる。また、図 3.2 に隣接するイベント関数の例を示す。$C(k) \leq n \leq C(k+1)$ に対して

$$\begin{aligned} \hat{\mathbf{y}}(n) &= a_k\phi_k(n) + a_{k+1}\phi_{k+1}(n) \\ &= a_k\phi_k(n) + a_{k+1}(1-\phi_k(n)), \end{aligned} \tag{3.11}$$

ここで、$C(k)$、$C(k+1)$ は、それぞれイベント k、$k+1$ の中心位置である。ただし、

$$\begin{aligned} \phi_k(C(k)) = 1\ ,\ \phi_k(C(k+1)) = 0 \\ 0 \leq \phi_k(n) \leq 1 \quad for\ C(k) \leq n \leq C(k+1), \end{aligned} \tag{3.12}$$

さらに制約として，イベント関数はイベントの生成と消滅を扱うため，単峰性の関数であるとする。最終的に $\phi_k(n)$ は、次のように決定される。

$$\phi_k(n) = \begin{cases} 1-\phi_{k-1}(n), & if\ C(k-1) < n < C(k) \\ 1, & if\ n = C(k) \\ min(\phi_k(n-1)), & \\ max(0, \phi_k(n)), & if\ C(k) < k < C(k+1) \\ 0, & otherwise, \end{cases} \tag{3.13}$$

ここで

$$\hat{\phi_k} = \frac{\langle (y(n) - a_{k+1}), (a_k - a_{k+1}) \rangle}{||a_k - a_{k+1}||^2}. \tag{3.14}$$

3.4.2 イベント関数のフィッティング

MRTD におけるイベント関数の定義より、イベント関数の縦軸は 1 から 0 までしか変化しないことから、切片は 1 で固定する。横軸の時間は、イベントターゲットから次のイベントターゲットまでの距離が横軸の長さになる。よって、横軸の長さはイベントターゲットと一個先のイベントターゲットとの距離で求めることができる。イベント関数は単峰性であり式 3.14 に従って変化していることからイベント関数は図 3.2 のような外形をしている。先ほど縦軸と横軸は示したので、次にイベント関数の変化を求めるために、以下の式を用いた非線形最小二乗法を用いたカーブフィッティングを行う。

$$Z = -\left(\frac{X}{c}\right)^M + e. \tag{3.15}$$

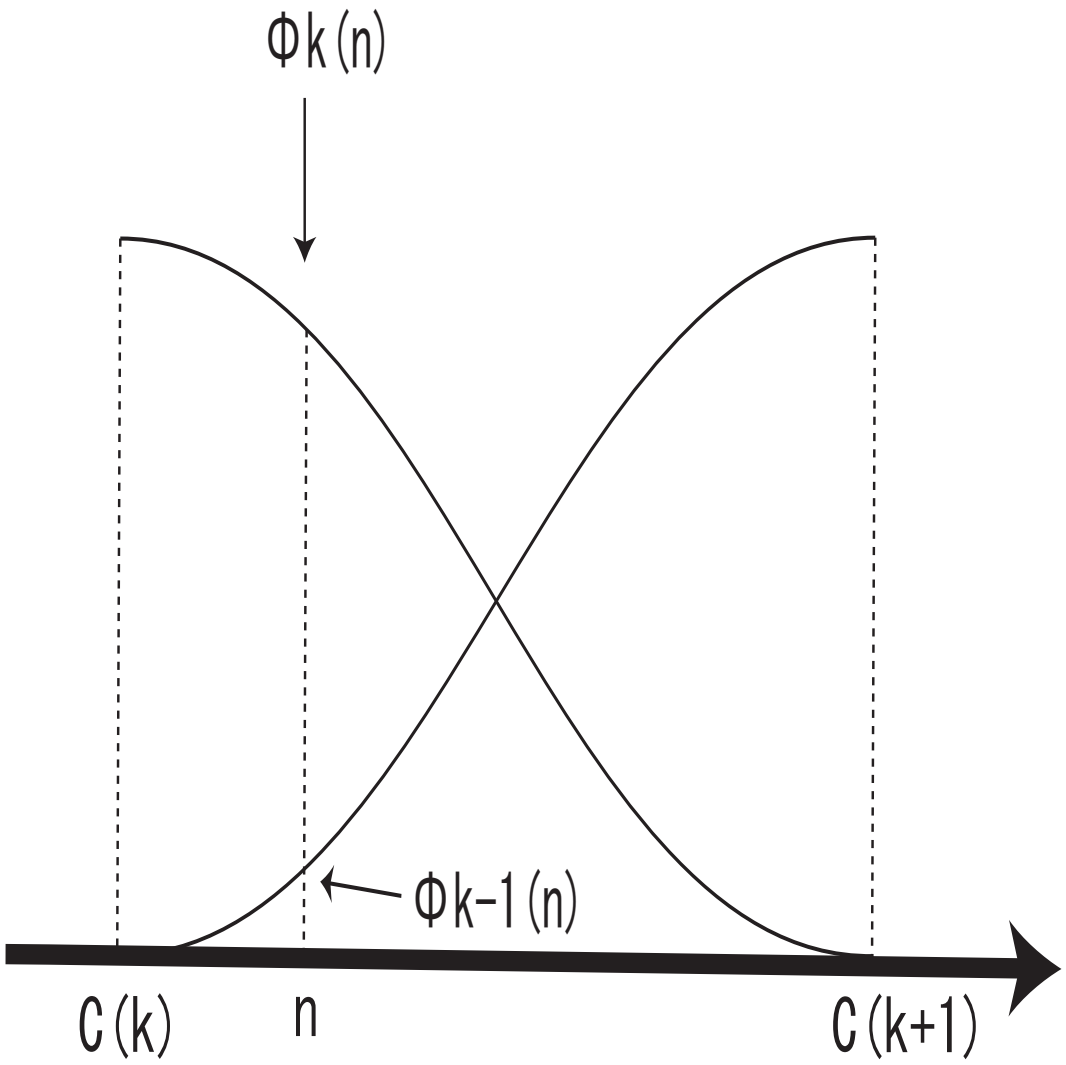

図 3.2: 2 つの隣接したイベント関数

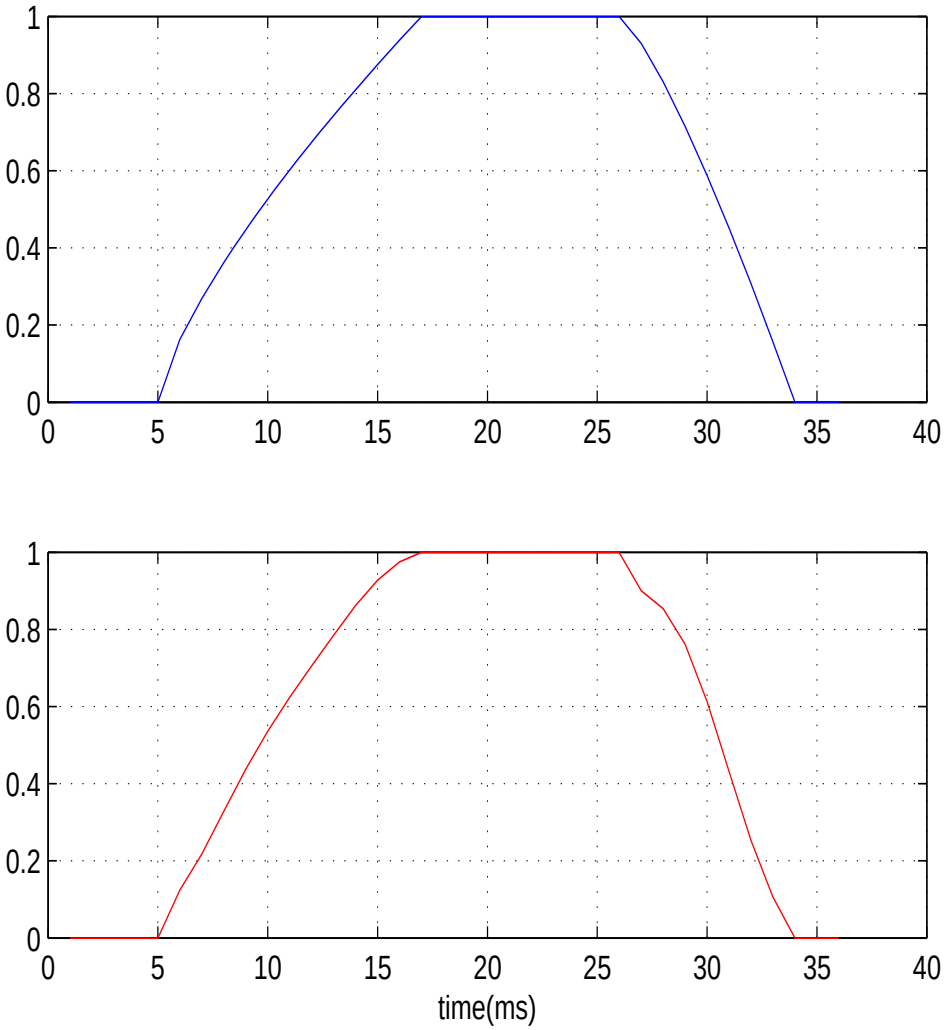

図 3.3: 非線形最小二乗法を用いてカーブフィッティングを行って作ったイベント関数（上)、MRTD で抽出されたイベント関数

e は ϕ の最大値を表しているので、 e は 1 である。そして、 Z が 0 のときに

$$X = c, \tag{3.16}$$

このとき、c はイベント関数の横幅すなわちターゲット間の距離を表している。$0 \leq \phi \leq 1$ のときにフィッティングを行っている。そして、 M はイベント関数の変化を表す。図 3.3 はこの式を用いてイベント関数を作ったものと、MRTD のイベント関数を表している。M と c のパラメータを動かすことで、MRTD の制約条件を崩さずかつ図 3.3 のようにイベント関数を自由に変形することが可能になった。

3.5 提案手法の評価

この節では、従来法である STRAIGHT と SFTR によってイベントを決定した MRTD を用いた手法との比較を行い提案法の評価を行う。イベント関数を変化を表している M とイベント関数の横の長さを表している c を変化させることによって、本当に動的成分を制御しているのかという点をここで評価し、提案手法によって音質が劣化していないか確かめるために主観評価実験を行った。

3.5.1 シミュレーション結果

この節では、音声信号がイベント関数の式 3.15 の M と c を変化させたときにどう変化するかを調べるために、静的特徴と先ほど決めた値 (動的特徴) を変化させ、決めた値 (動的特徴) だけを男声から女声の特徴へ変化させた。用いたパラメータとして、平均基本周波数、平均スペクトル包絡、平均ゲイン、平均非周期成分、さらに平均イベント関数を作成し、その後イベント関数の M と c をそれぞれ男声から女声に変化させ実際に動的特徴が変化しているか確かめた。音声データは連続発話音声の「誰にでもできるんじゃないかな」を用いた。

実行条件は下記の通りである。

- sampling 周波数：20Hz
- 分析窓長：40ms
- 分析シフト幅：1ms
- FFT 長：1024
- LSF 次数：60

音声データの特徴は下記の通りである。

- 男声の特徴:語尾を伸ばさない

- 女声の特徴:語尾を伸ばす

結果は図 3.4 図 3.5 に示す。図 3.4 は式 3.15 にある M を変化させて合成した音声である。青いラインが男声、赤いラインが女声の M の値を用いた結果である。上のパネルが基本周波数の動きを示していて、下のパネルがスペクトルの時間領域での動きを表している。この図の結果、下のパネルで青いラインと赤いラインの軌跡が変化していることから、式 の M を変化させるということは、スペクトルの時間変化を変化させるということに対応していることがわかる。次に図 3.5 は式 3.15 に示す c の値だけが違う。青いラインが男声のもの、赤いラインが女声のものである。その結果、後半部分にかけて変化が著しく出ていることがわかる。さきほどと同じように上のパネルは基本周波数の時間変動、下のパネルはスペクトルの時間変動を表していて、どちらも変化している。これは男声より女声のほうが語尾を伸ばすなどの特徴を含んでいるため、後半の部分に男声と女声で差がでている。この結果、 c を変化させることで、音韻長が変化することがわかった。以上のイベント関数の二つのパラメータを変化させた結果、 M と c のパラメータはイベント関数の変化を表す M はスペクトルの変化量、イベントターゲット間の距離である c は音韻長を表していることがシミュレーションによって明らかになった。

3.5.2 音質評価

提案手法の音質が従来法より劣化していないかを調べるために従来法によって作られた音声と、提案法によって作られた音声の音質を評価した。従来法より音質が劣化していると、実験に用いることができない。さらに北村らによって音質が知覚実験の結果に影響を与えるという知見もある [5]。そこで、音質が劣化していないか調べるために、主観評価実験を行った。

3.5.3 用いたの音声データ

用いた音声データベースは ATR 音声データベース C セットである [35]。

音声

男声、女声各 1 名が発話した 4 つの文章を用いて分析合成音声を作成した。一つは STRAIGHT と MRTD を用いてイベント位置を SFTR を用いて作成した音声であり、もう一つは提案手法である。実行条件は下記の通りである。

- sampling 周波数：20Hz
- 分析窓長：40ms
- 分析シフト幅：1ms

- FFT 長：1024
- LSF 次数：60

実験参加者は 22 才から 25 才までの大学院生男 8 名である。実験参加者には 5 段階の基準によって評価した (1:bad, 2:poor, 3:fair, 4:good, 5:excellent)。次の表に結果をのせる。

SFTR でイベントを設定した手法	ラベル情報を用いてイベントを設定した手法
2.208	3.925

表 3.1: MOS 評価の結果

実験の結果、従来法より提案手法のほうが音質がいいことが明らかになった。これはイベント位置を適正に配置することにより、従来法と比べてスペクトルの歪みが少なくなったと考えられる。さらに、イベント関数のモデル化による音質の劣化にはそれほど影響がないという結果になった。

3.6 まとめ

この章では声質変換モデルについて説明した。提案手法を用いることで、動的特徴を制御できることが明らかになった。さらに、音質の評価でも、従来法を上回る結果を得たことから、最初にあげた条件を満たすような声質変換モデルができたといえる。次の章以降は、このモデルを用いて、合成音声を作成していく。

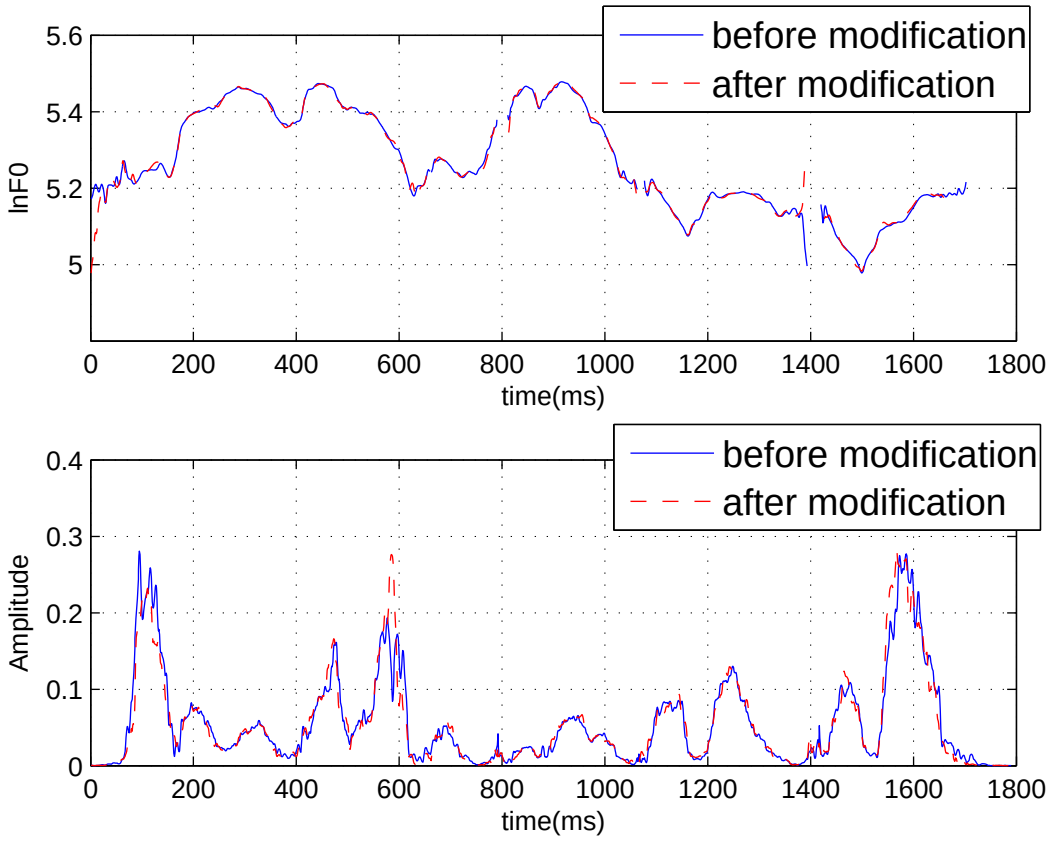

図 3.4: この図はスペクトルの変化量を男声から女声へ変形したものである。上のパネルはアクセントパターンを示している。そして下のパネルはスペクトルの変化パターンを示している。

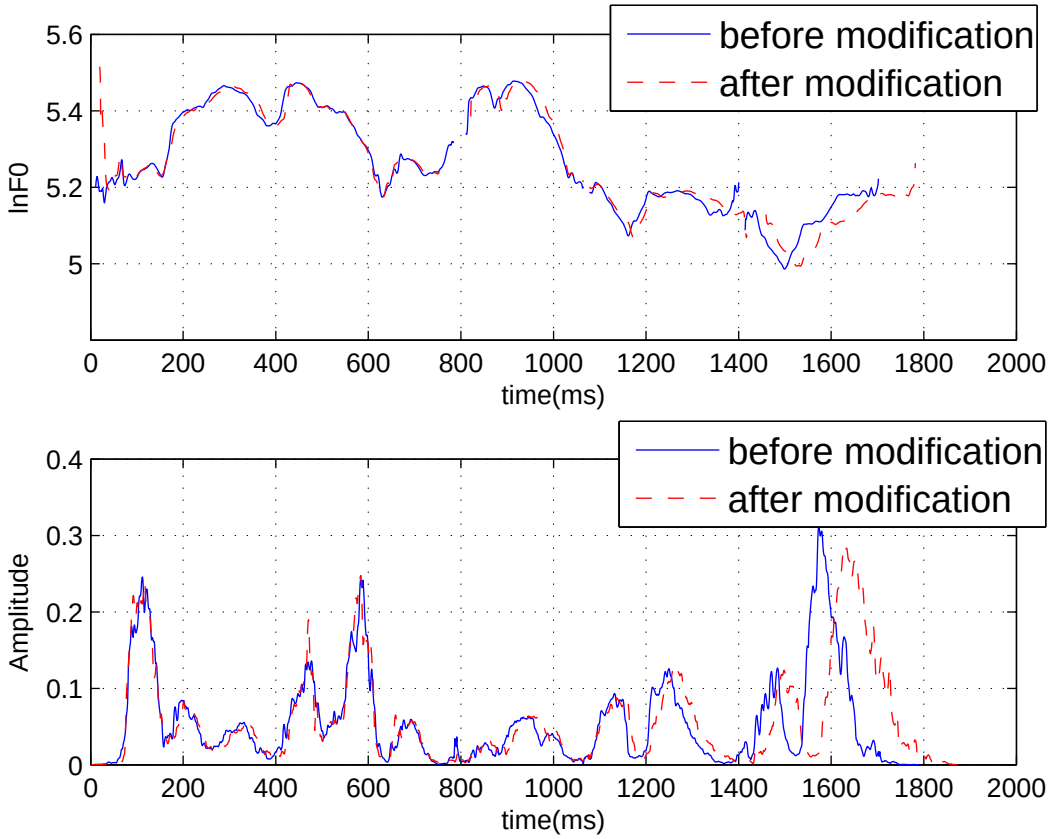

図 3.5: この図は音韻長を男声から女声へ変形したものである。上のパネルはアクセントパターンを示している。そして下のパネルはスペクトルの変化パターンを示している。

第4章　分析

4.1　目的

男声 · 女声音声から声質変換モデルの分析部を用いて各パラメータ値を得る．得られた各パラメータ値が男声 · 女声で違いがあるかどうかを確かめるために Kruskal の方法 (SPSS 13.0 for Windows) を用いた多次元尺度構成法 (MDS) で分析を行った．

4.2　分析する音声とパラメータ

本研究で使用した音声データベースは ATR 音声データベース C セットである。このデータベースの特徴は話者が 20 人と多いこと、アナウンサー、ナレーター、普通の人が入ったデータベースとなっている。アナウンサー、ナレーターより自然の発話データに近いデータベースとなっている。その中から男声、女声で各 4 名の音声を用いて声質変換モデルを用いてパラメータ値を抽出した。

音声データは連続発話音声の「だれにでもできるんじゃないかな」を用いた。実行条件は下記の通りである。

- sampling 周波数：20Hz
- 分析窓長：40ms
- 分析シフト幅：1ms
- FFT 長：1024
- LSF 次数：60

4.3　イベント位置におけるパラメータ値の分析

連続発話音声から抽出した各パラメータ値について各話者間での距離を求め、それを元に MDS 分析を行った。スペクトルパラメータの話者間の話者間の物理的な距離 $\mathbf{DIST_{CD}}$ については、LSF を LPC 係数に、LPC 係数を LPC ケプストラムに変換し、イベント位

置のケプストラム距離 ($\mathbf{CD}_k$) を求め、全イベント数 K の平均として求めた。

$$\mathbf{CD}_k = \sqrt{2\sum_{i=1}^{s}(s_{ik}^{(x)} - s_{ik}^{(y)})^2} \tag{4.1}$$

$$\mathbf{DIST_{CD}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{CD}_k \tag{4.2}$$

ここで、p はケプストラム次数 (60 次)、 $s_{ik}^{(x)}, s_{ik}^{(y)}$ は k 番目のイベントに対する話者 x と話者 y のケプストラムである。

また、基本周波数、スペクトル変化も同様に話者間の物理的距離 $\mathbf{DIST_p}$、$\mathbf{DIST_M}$ を求める。

$$\mathbf{DIST_p} = \sqrt{\sum_{i=1}^{K}(p_i^{(x)} - p_i^{(y)})^2} \tag{4.3}$$

$$\mathbf{DIST_M} = \sqrt{\sum_{i=1}^{K}(M_{(x)} - M_{(y)})^2} \tag{4.4}$$

ここで, $p_i^{(x)}, p_i^{(y)}$ は話者 x と話者 y の対数基本周波数, $M_{(x)}, M_{(y)}$ はスペクトルの変化を表す式 3.15 中のパラメータである.

ゲインのダイナミックレンジについても話者間の物理的距離 $\mathbf{DIST_g}$ を求める.

$$\mathbf{DIST_g} = (g^{(x)} - g^{(y)}) \tag{4.5}$$

$$g^{(x)} = max(g^{(x)}) - min(g^{(x)}) \tag{4.6}$$

$$g^{(y)} = max(g^{(y)}) - min(g^{(y)}) \tag{4.7}$$

$$\tag{4.8}$$

$g_i^{(x)}, g_i^{(y)}$ はゲインの最大と最小を引いたもので、それぞれのゲインのダイナミックレンジとなっている。

音韻長について話者間の距離 $DIST_t$ は, 連続発話音声の中の各音韻の継続長から次の式により求める (式 4.9).

$$\mathbf{DIST_t} = \sqrt{\sum_{i=1}^{K}(c_{(x)} - c_{(y)})^2} \tag{4.9}$$

$c_{(x)}, c_{(y)}$ は音韻長を表す式 3.15 中のパラメータである.

4.4 音声データ

用いた音声データは ATR 音声データベース C セットである [35]。詳細は表 4.1 に示す。

表 4.1: 用いた音声データ

音声データ	性別	平均基本周波数
04SFA15	male1	134.83Hz
09SFA15	male2	106.83Hz
11SFA15	male3	162.21Hz
18SFA15	male4	143.89Hz
02SFA15	female1	259.11Hz
03SFA15	female2	253.89Hz
06SFA15	female3	246.47Hz
07SFA15	female4	270.42Hz

表 4.2: Stress の評価

Stress	Stress の度合い
20%	あまりよくない (poor)
10%	まあまあ適合している (fair)
5%	良く適合している (good)
0.5%	非常に適合している (excellent)
0%	完全に適合している (perfect)

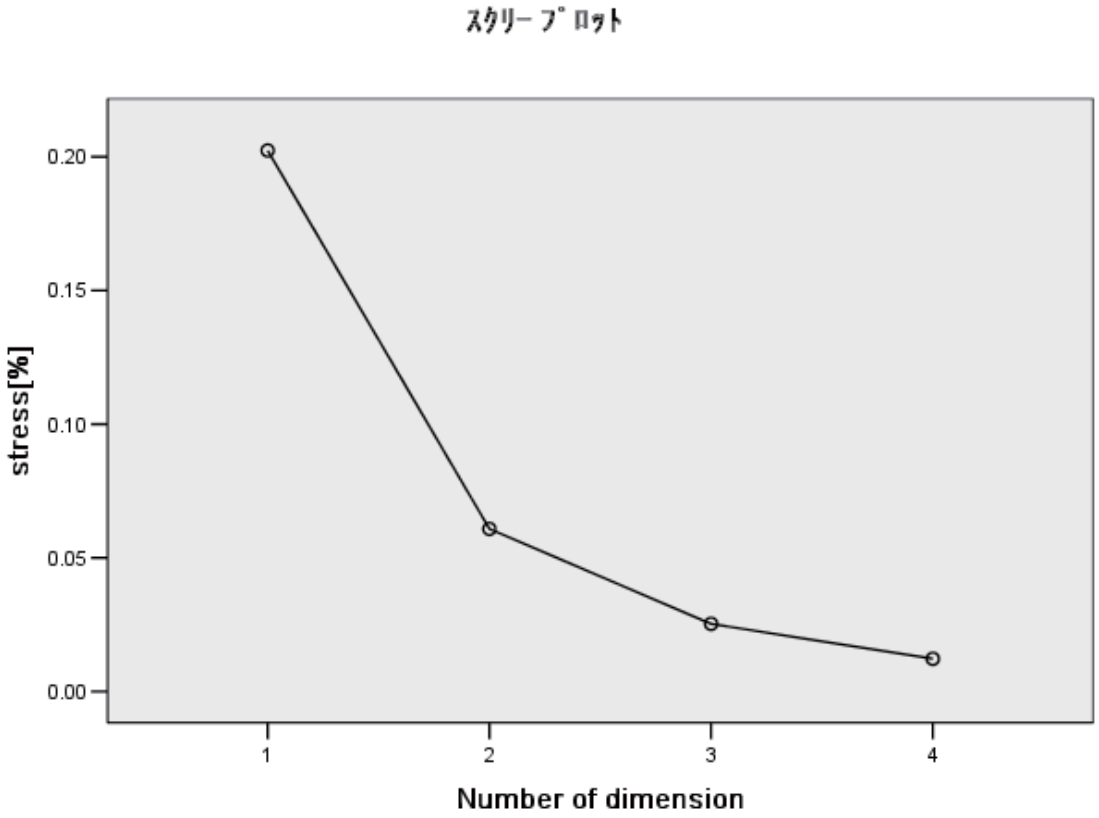

図 4.1: 次元と stress 値の関係

4.5 多次元尺度構成法 (MDS) を用いた分析

連続発話音声について、3 章で提案した声質変換モデルを用いて得られる静的特徴（平均基本周波数、スペクトル包絡、ゲインのダイナミックレンジ)、動的特徴 (基本周波数の変化、スペクトルの変化、音韻長) について、男声と女声で各音響特徴量を表すパラメータ値が男声と女声という分類で物理的に違いがあるかどうかを調べるために SPSS13.0 for windows により、Krusal の方法を用いた多次元尺度構成法 (MDS) を行った。この分析によって、男声女声に関して静的、動的における各パラメータ値の物理的距離を調べた。Stress とは、ある次元数における各音声データ間の心理的距離の適合度を示すものであり、表 4.2 に Stress の評価ついて示しておく [36]。表 4.2 より Stress の値は 10%以下であることが望ましい。

4.5.1 スペクトル包絡の MDS 分析

イベント位置のケプストラム距離を話者間で求め，多次元尺度構成法でプロットした．多次元尺度構成法の次元数が 3 の場合約 3%であった (図 4.3)．3 次元上にプロットしたものを，1-2 次元で見ることで，男声 · 女声の違いがはっきりわかる．これは男声と女声でスペクトル包絡が違うことを示している．

4.5.2 基本周波数の MDS 分析

スペクトル包絡と同様に基本周波数に対しても多次元尺度構成法で分析を行った。この分析における基本周波数は基本周波数ターゲットの距離を表している。stress 値は次元数が 2 次の場合、非常によく適合している 5%以下である約 1.5%である。分析の結果を図 4.4 示す。図 4.4 の結果から、男声と女声ではっきりと違いが現れていることから基本周波数に違いがあることがわかる。分析結果上でも基本周波数が男声と女声で異なっていることがわかる。

4.5.3 スペクトルの変化 (動的特徴) の MDS 分析

スペクトル変化に対しても多次元尺度構成法で分析を行った．分析の結果を図 4.5 に示す．stress 値は次元数が 2 次の場合，非常によく適合している 5%以下である約 4%である．この結果から，男声と女声ではスペクトルの変化が異なっていることがわかる．スペクトルの変化が男声と女声で違うという知見はこれまでない．

4.5.4 ゲインのダイナミックレンジの MDS 分析

ゲインのダイナミックレンジに対して多次元尺度構成法で分析を行った．分析結果を図 4.6 に示す．stress 値は次元数が 2 次の場合，非常によく適合している 5%以下である約 2%である．分析結果からゲインのダイナミックレンジにおいて男声と女声で違いが得られた．

4.5.5 音韻長の MDS 分析

分析したパラメータと同様に音韻長に対しても多次元尺度構成法で分析を行った。図 4.7 に分析結果を示す。分析の結果、音韻長では男声と女声とでは明確な違いが分析の結果が得られなかった。この結果は男声同士で同じ分析を行った鈴木ら [9] の結果と同じである。そして、櫻庭らが女声らしい話し方の要因の一つとしてあげている語尾を伸ばすというところに着目して、実際にこの音声データで語尾を伸ばしているかどうかを調査した。調査方法は音声の最後の発話部分の音韻の発話時間を調べた。結果を表 4.3 示す。表

4.3 が示すとおり、男声と女声で語尾の伸ばす時間に違いがでるという結果になった。これは男声が語尾を伸ばさない、女声が語尾を伸ばしているという結果を示している。音韻長全体では男声と女声では明確な違いが出なかったが、語尾に着目すると、伸ばす · 伸ばさないで男声と女声で違いが出るという結果が得られた。次にイベント関数の変化のない区間について男声と女声で変化のない区間の長さの違いがないかどうかを調査した。この区間は、音声中の話速に対応しており、この区間が短いと話速が早くなり、逆に長いと話速が遅くなるといった特徴が表れる。表 4.4 の結果から、男声と女声で話速に対しては違いが見られなかった。

4.5.6 まとめ

この章では、連続発話音声について、3 章で提案した声質変換モデルを用いて得られる静的特徴（平均基本周波数、スペクトル包絡、ゲイン)、動的特徴 (アクセントパターン、スペクトルの変化量、音韻長) について、男声と女声で各音響特徴量を表すパラメータ値が男声と女声という分類で物理的に違いがあるかどうかを調べるために MDS を用いて分析を行った。その結果、男声と女声ではスペクトル包絡、基本周波数、ゲイン、スペクトルの変化量に対して違いが見られた。分析結果から、これらのパラメータが男声 · 女声知覚に影響を与えていることが示唆される。音韻長の結果からは、男声 · 女声と比較しても優位な差が見られなかったが、語尾に着目して分析することで男声と女声で語尾の長さが違うという結果を得ることができた。話速に関する分析を男声と女声で行ったが違いが見られないという結果が得られた。分析の結果の違いが男声 · 女声知覚に影響を与えているか調べるために、5 章では聴取実験を行う。

表 4.3: 最後の音韻の長さ

音声データ	性別	音韻の長さ
04SFA15	男性	180ms
09SFA15	男性	122.5ms
11SFA15	男性	95ms
18SFA15	男性	105ms
02SFA15	女性	160ms
03SFA15	女性	190ms
06SFA15	女性	160ms
07SFA15	女性	225ms

表 4.4: イベント関数の変化のない区間の合計値

音声データ	性別	音韻の長さ
04SFA15	男性	212ms
09SFA15	男性	403ms
11SFA15	男性	328ms
18SFA15	男性	175ms
02SFA15	女性	323ms
03SFA15	女性	249ms
06SFA15	女性	384ms
07SFA15	女性	295ms

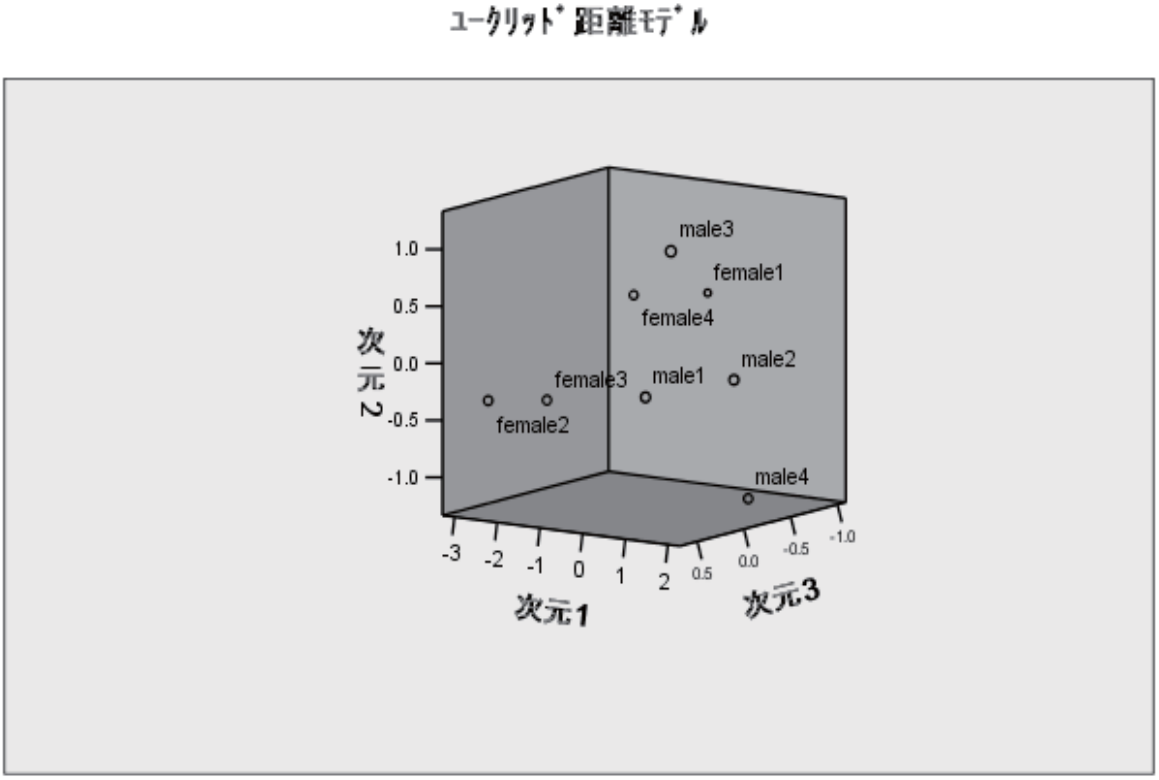

図 4.2: ケプストラム距離 (3 次元) の付置図

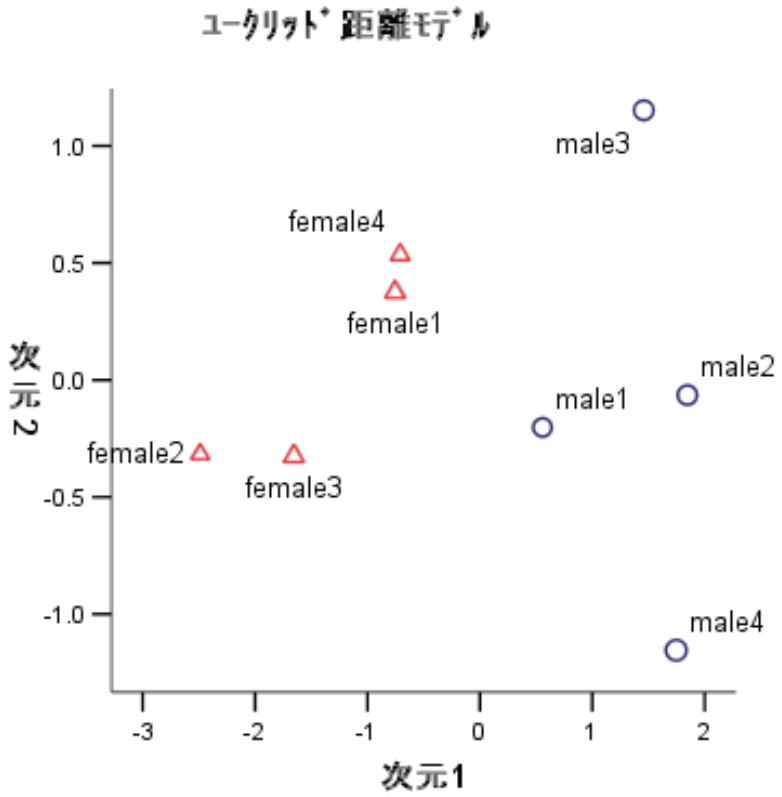

図 4.3: ケプストラム距離 (3 次元中の次元 1 ー次元 2 での) の付置図

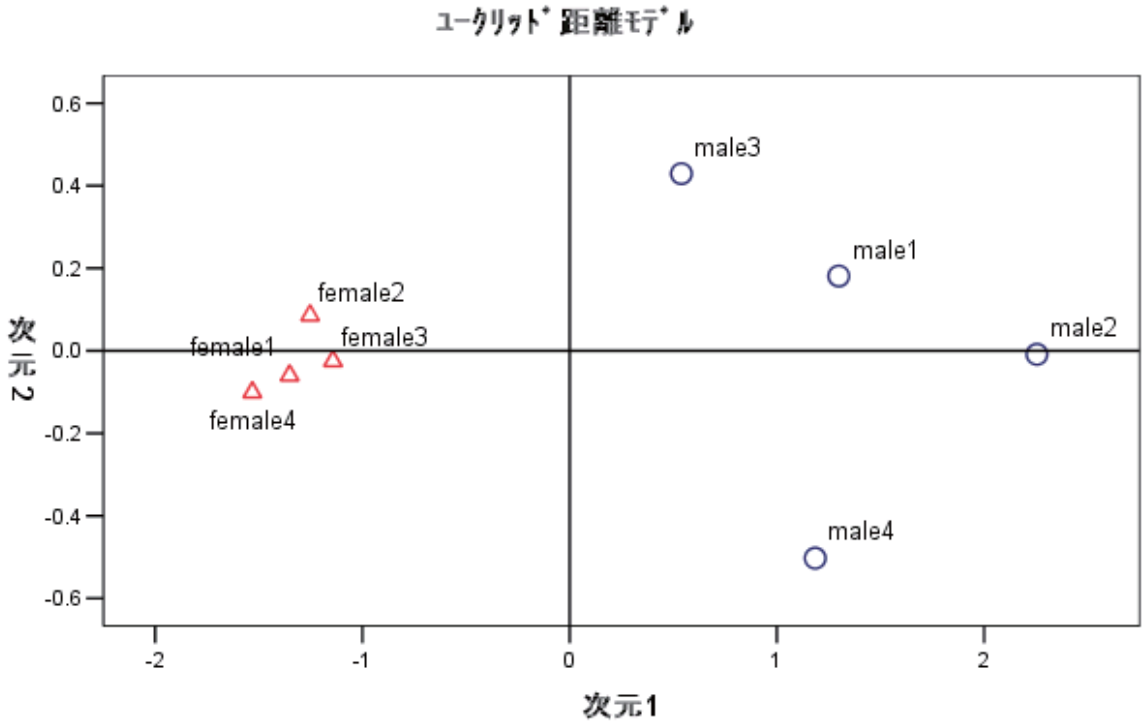

図 4.4: 基本周波数距離の付置図

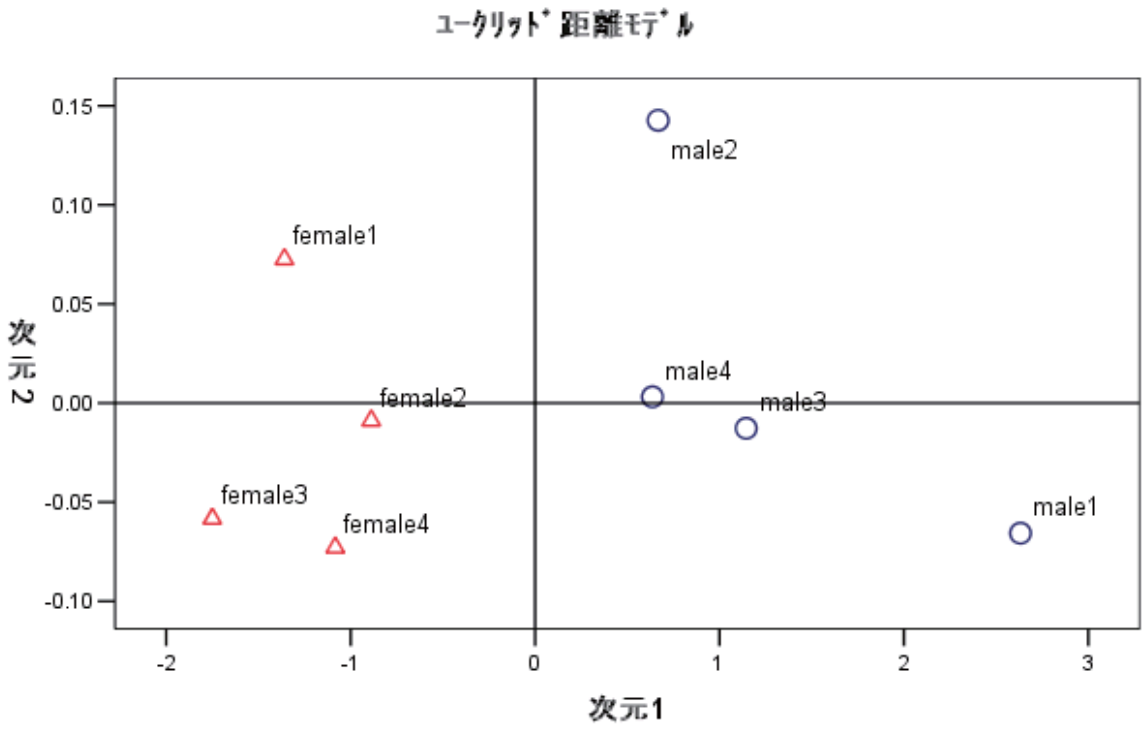

図 4.5: スペクトル変化距離の付置図

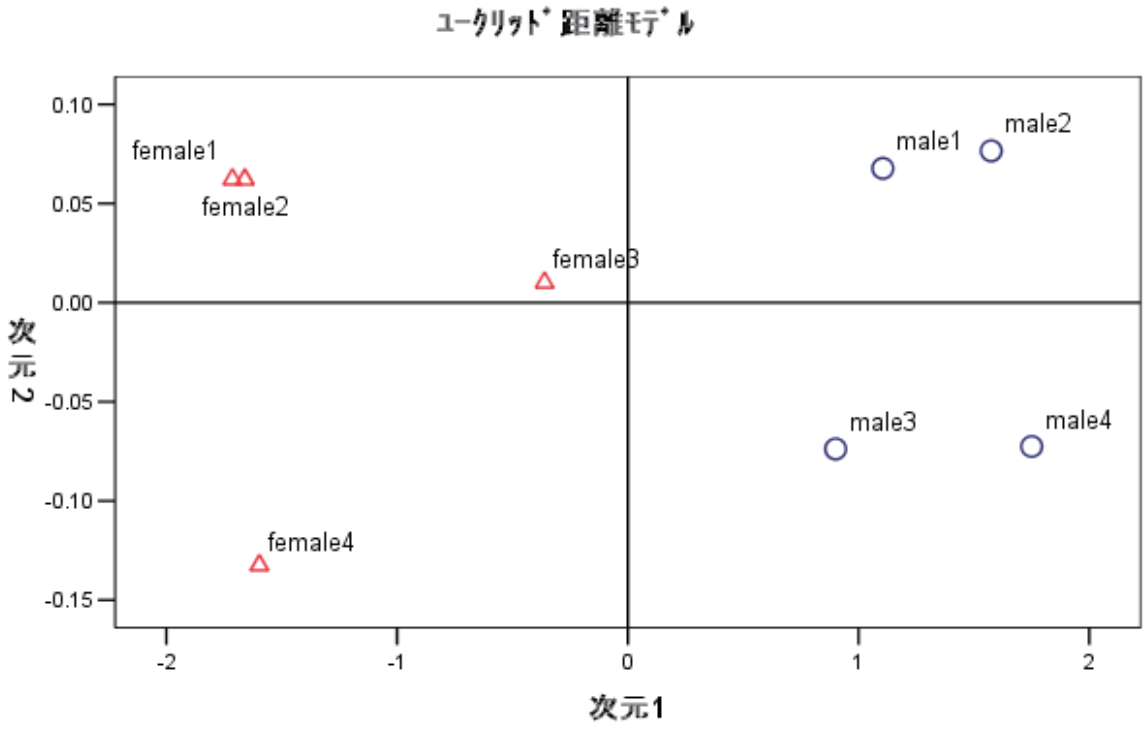

図 4.6: ゲインのダイナミックレンジ距離の付置図

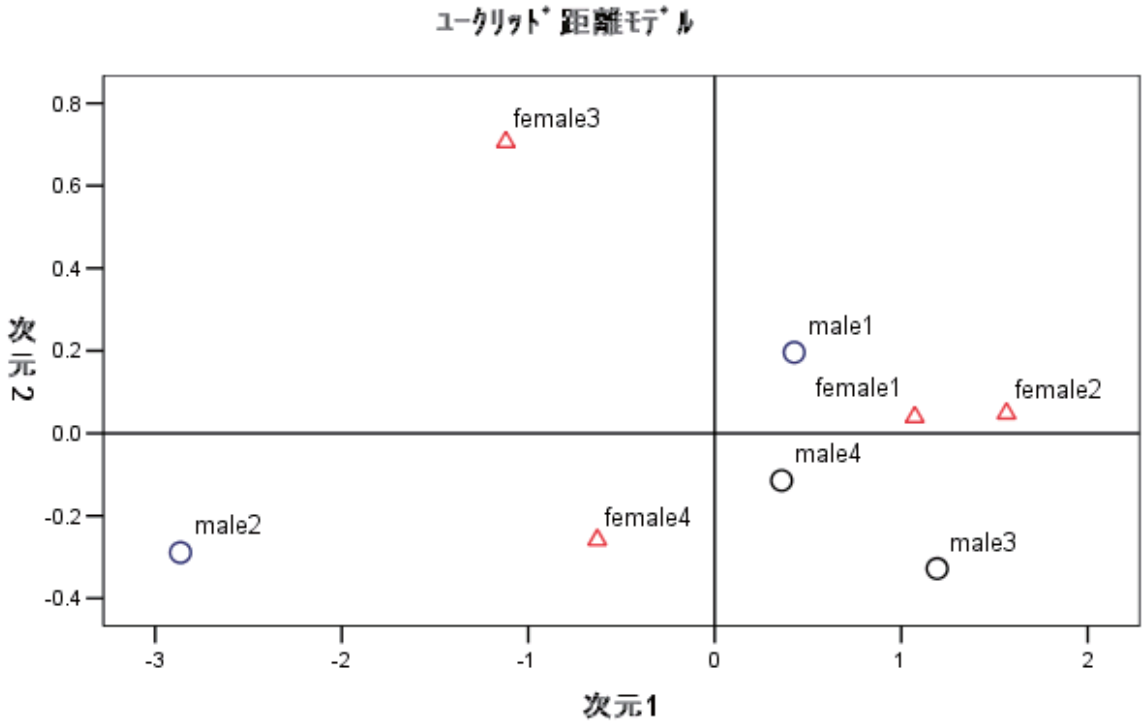

図 4.7: 音韻長距離の付置図

第5章　静的・動的特徴が男声・女声知覚に与える影響

5.1　目的

この章では、声質変換モデルによって、連続発話音声を静的特徴と動的特徴を分け、それぞれの音響特徴量が男声女声知覚にどのような影響を与えているか心理物理実験によって明らかにする。

5.2　実験１

実験１では、静的、動的特徴のパラメータが男声･女声知覚にどのような影響を与えているか明らかにすることを目的とするシェッフェの一対比較法を用いて,「男声･女声」について聴取実験を行った。

5.3　刺激音の作成

3章で提案したモデルを用いて、音声からパラメータを抽出する。用いたパラメータ値は、4章で分析を行ったパラメータ値である。分析結果から、基本周波数、スペクトル包絡、ゲイン、スペクトル変化量においては多次元尺度構成法の結果から男声と女声で違いが出るという結果が得られているので、男声･女声知覚に影響を与えているかどうかを調べるためにこれらのパラメータを採用した。そして、音韻長と話速と基本周波数の変化については、音韻長においては語尾に着目すると語尾を伸ばす、伸ばさないに差が出ることが分析により明らかになっており、基本周波数の変化と話速については話し方に起因する動的な特徴量として考え男声･女声知覚に影響を与えていると仮定し男声･女声知覚に与える影響を調べるためのパラメータとして聴取実験のパラメータとして採用した。今回聴取実験で扱うパラメータ値は平均基本周波数、スペクトル包絡、ゲイン、スペクトルの変化、音韻長、基本周波数の変化である。男声・女声の各入力音声は、提案モデルを用いることで上記のパラメータを抽出する。このパラメータ値を男声・女声で声質変換することで、あるパラメータ値は男声の特徴量、あるパラメータ値は女声の特徴量といった合成音を作ることが可能である。

5.3.1 音声データ

用いた音声データベースは ATR 音声データベース C セットである。用いた発話音声は「だれにでもできるんじゃないかな」を採用した。そして分析にも使用した男性 4 名、女性 4 名の音声データを用いた。

5.4 実験 1 の刺激音

実験 1 ではさきほど上げたパラメータ値が男声 · 女声知覚に与える影響を明らかにするためにシェッフェの一対比較法を用いた聴取実験を行う。実験 1 の刺激音は得られるすべてのパラメータの算術平均をとって作成した音声 (平均声) を元にして、平均声に対して各パラメータ一つと平均声のパラメータ値を交換することで、刺激音を作成する。交換するパラメータ値については、4 章の分析結果を用いて、一番男声・女声で平均からの距離のあるものを用いている。

5.4.1 静的な特徴を付加した刺激音

静的な特徴が男声 · 女声知覚に与える影響を調べるために、多次元尺度構成法での分析結果から、平均基本周波数、スペクトル包絡、ゲインのパラメータ値を各男声、女声のものと平均の各パラメータ値と交換したものが刺激音になる (図 5.1)。

- AV+F0(M):平均基本周波数を男声に変えたもの
- AV+F0(F):平均基本周波数を女声に変えたもの
- AV+SP(M):スペクトル包絡を男声に変えたもの
- AV+SP(F):スペクトル包絡を女声に変えたもの
- AV+Ga(M):ゲインを男声に変えたもの
- AV+Ga(F):ゲインを女声に変えたもの

5.4.2 動的な特徴を付加した刺激音

動的な特徴が男声 · 女声知覚に与える影響を調べるために、多次元尺度構成法での分析結果で差が表れたスペクトルの変化と音韻長 (語尾を伸ばす) とし、さらに動的特徴量として関係すると思われる話速と基本周波数の変化と語尾を上げたものと下げたものを刺激音とした。が刺激音である。動的特徴を付加させた刺激音を以下にまとめる (図 5.2)。

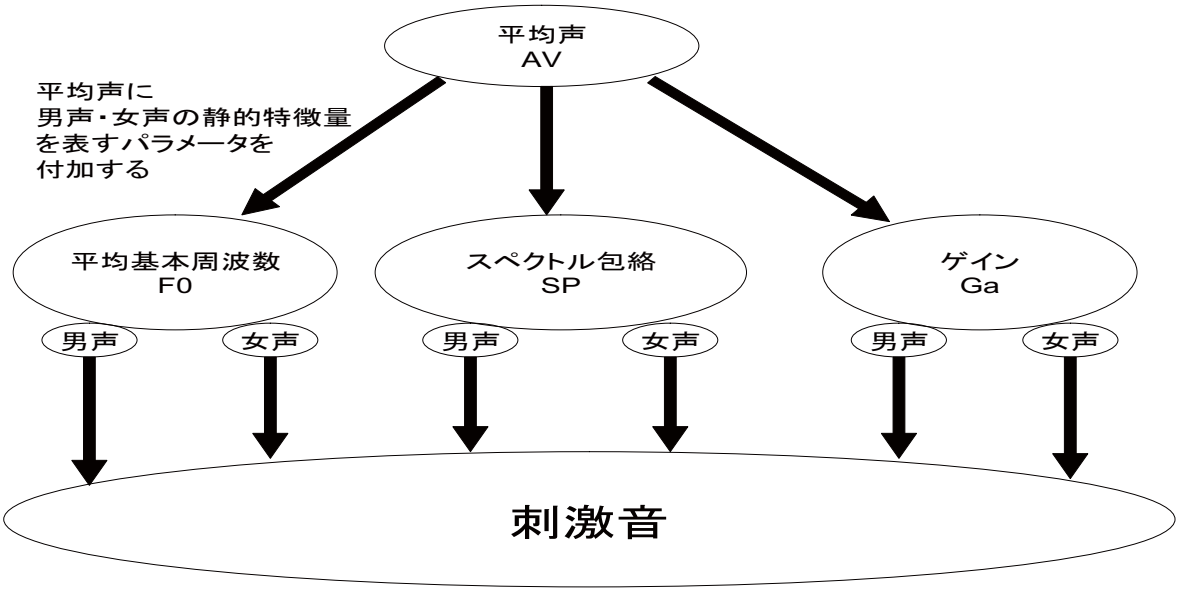

図 5.1: 実験 1 で用いる刺激音 (静的特徴)

- AV+DSP(M):スペクトルの変化を男声に変えたもの
- AV+DSP(F):スペクトルの変化を女声に変えたもの
- AV+DF0(M):基本周波数の変化を男声に変えたもの
- AV+DF0(F):基本周波数の変化を女声に変えたもの
- AV+DU(M):音韻長を男声に変えたもの
- AV+DU(F):音韻長を女声に変えたもの
- AV+G(M):語尾を下げたもの
- AV+G(F):語尾を上げたもの

5.4.3 実験手続き

- 実験参加者には次のような指示を与えて、男声 · 女声に関して評価してもらった。ヘッドホンから 2 つの音声を対にして流します。前の音声（A）と後の音声（B）を聴き比べて、後の音声が前の音声に比べて男声、女声に聞こえるということを下に

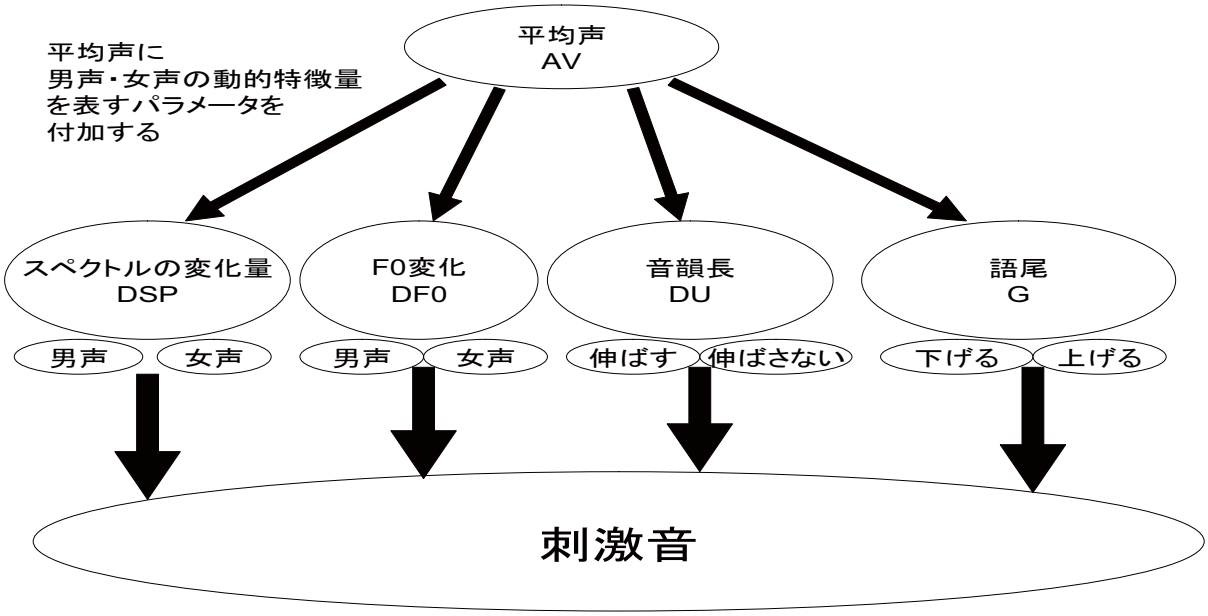

図 5.2: 実験 1 で用いる刺激音 (動的特徴)

記した 7 段階の評価尺度に従って判断してください。男声と聞こえる場合は正の値 (1 から 3) 女声と聞こえる場合は負の値 (-1 から-3) のあてはまる値にチェックしてください。前の音声と後の音声を聞いてどちらも同じ場合は 0 を選択してください。

評価は，“3. A が B と比べて非常に男声に聞こえる”，“2. A が B と比べて男声に聞こえる”，“1. A が B と比べて少し男声に聞こえる” ，“0. どちらも同じ (どちらともいえない)”，“-1. A が B と比べて少し女声に聞こえる”，“-2. A が B と比べて女声に聞こえる”，“-3. A が B と比べて非常に女声に聞こえる” の 7 段階である.

実験は一対比較法を用いて、最初に提示したものを A、後に提示したものを B とし、A が B に比べてどれくらい男声・女声らしいのかを 7 段階の評価尺度で実験参加者に回答させた。

5.4.4 実験参加者

被験者は正常な聴力を有する 22-25 歳の大学院生 7 名 (男性 7 名) である。実験参加者は過去にほかの聴取実験の経験がある。

5.4.5 刺激条件

実験で用いる刺激音は、先に示した 15 個の音声を 2 つずつ対にしたものである。刺激対の数は、1 つの音声データについて、刺激順序の違いも考慮した 210 対である。図 5.3 に刺激の呈示順序を示す。

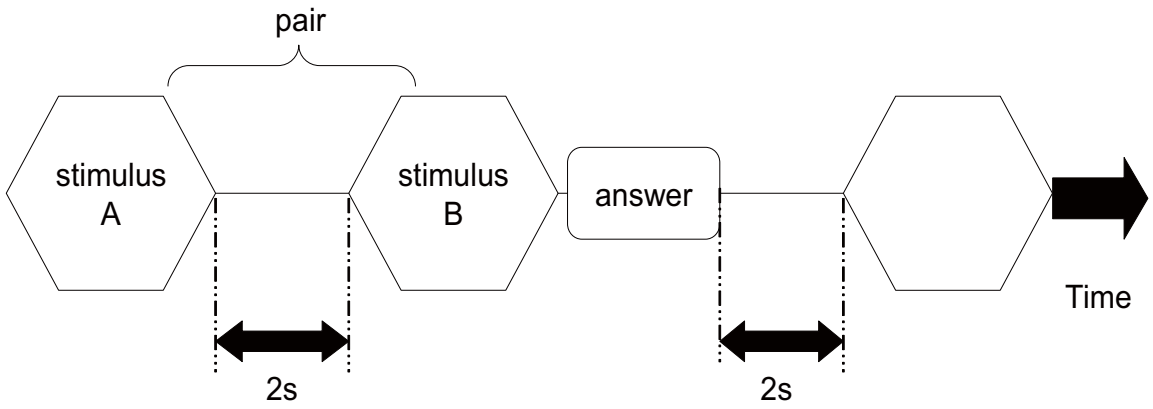

図 5.3: 刺激の呈示順序

5.4.6 実験環境

実験は、人のいない静かな部屋で行い、実験参加者にはヘッドホン (Sennheiser HDA200) を介して刺激音対を両耳に呈示し、PC ディスプレイ上の評価尺度をキーボードで入力させることで回答させた。また音圧レベルは実験参加者の聴きやすいレベルに設定した。そして表 2.1 には使用した実験機材を示す。

表 5.1: 実験機材

ノート PC	
ヘッドホンアンプ	STAX SRM-1/MK-2
ヘッドホン	Sennheiser HDA200

5.5 結果

上記の実験方法で得られたデータを、浦の変法によって処理した結果を表 5.2 に示す。また、表に示した母数の値に従って、15 の刺激の距離関係を直線上で示したものが図 5.4 になる。母数の値は、刺激がどれだけ男声 · 女声に聴こえたかを表す値であり、正の大きな値を示せば、その刺激が男声であることを示し、負の大きな値を示せば、その刺激が女声であることを示すことを表している。次の節では静的特徴と動的特徴に分けて結果を示す。

5.5.1 静的特徴量に関する結果

実験 1 の結果、平均基本周波数とスペクトル包絡の母数の値が、男声と女声で大きく異なるという結果が表れた。男声の平均基本周波数 (AV+F0(M)) では母数値が 1.62 と高い数値になっている。これは正の大きい値を示すことから、男声の特徴量から抽出した平均基本周波数は男声知覚に影響を与えていることが明らかになった。一方女声の平均基本周波数 (AV+F0(F)) ついては、母数の値が-1.39 という負の大きい値を示している。この結果は女声の特徴量から抽出した平均基本周波数は女声知覚に影響を与えていることが明らかになった。

次にスペクトル包絡について述べる。男声のスペクトル包絡 (AV+SP(M)) での母数値は 1.56 という平均基本周波数と並び高い数値となった。正の大きい値を示すことから、スペクトル包絡は男声知覚に大きな影響を与えていることが明らかになった。女声のスペクトル包絡 (AV+SP(F)) の母数値は-1.52 となり、負の大きい数値を表すことから、女声知覚に影響を与えている。

ゲインの結果は、男声 (AV+Ga(M)) と女声 (AV+Ga(F)) のパラメータを用いた場合にそれぞれ-0.025 と-0.004 という低い値になった。男声のパラメータ値と女声のパラメータ値は多次元尺度構成法の分析結果では違いが見られたが、実験 1 の結果からは男声 · 女声知覚には影響を与えていないことが明らかになった。以上で、静的特徴量についてまとめると、基本周波数とスペクトル包絡は多次元尺度構成法の分析結果でも違いが明らかになり、さらに男声 · 女声知覚に影響を与えていることが実験 1 によって明らかになった。一方で、ゲインについては多次元尺度構成法の分析では男声と女声で違いが見られたが、

表 5.2: 実験 1 の母数の推定

AV	0.033
AV+F0(M)	1.62
AV+F0(F)	-1.39
AV+SP(M)	1.56
AV+SP(F)	-1.52
AV+Ga(M)	-0.025
AV+Ga(F)	-0.004
AV+DSP(M)	-0.012
AV+DSP(F)	0.012
AV+DF0(M)	-0.25
AV+DF0(F)	0.071
AV+DU(M)	0.037
AV+DU(F)	-0.13
AV+G(M)	0.16
AV+G(F)	-0.067

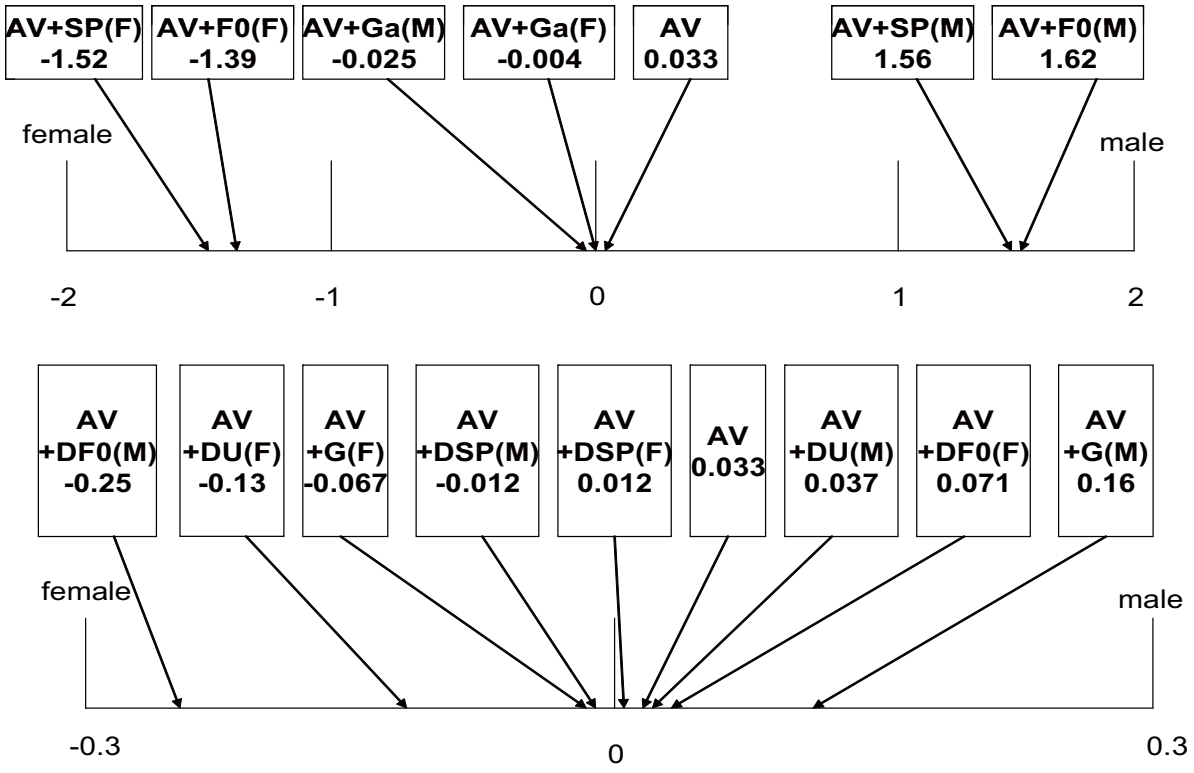

図 5.4: 実験 1 の結果．上のパネル：静的特徴の布置．下のパネル：動的特徴の布置．

実験 1 の結果からは優位な差が見られず、男声 · 女声知覚には基本周波数やスペクトル包絡に比べると影響を与えてないという結果が得られた。

5.5.2 動的特徴量に関する結果

動的特徴については、まず多次元尺度構成法での分析によって差が見られたスペクトルの変化からみていくことにする。スペクトルの変化は多次元尺度構成法での分析では明確な違いが見られた。しかし実験 1 の結果では、男声 (AV+DSP(M)) と女声 (AV+DSP(F)) でもっとも差異があったスペクトルの変化に関するパラメータ値を用いたが、母数の値はそれぞれ-0.012 と 0.012 という結果になった。この結果から、男声 · 女声知覚にスペクトルの変化はあまり影響を与えていないことが明らかになった。

次に話速を変化させた AV+S(M) と AV+S(F) については、母数値がそれぞれ、-0.009 と 0.024 であることから男声 · 女声知覚に影響あまり与えていないことが明らかになった。

次に語尾に着目することで、男声と女声で差が表れた音韻長の結果を見ていく。特徴として、男声の音韻長のパラメータ値を用いると語尾を伸ばさず、女声の音韻長のパラメータ値を用いると語尾を伸ばすといった特徴が付加される。AV+DU(M) と AV+DU(F)

の結果から、語尾を伸ばさない男声の母数値は 0.037 であり、語尾を伸ばすという特徴を持った女声の母数値は-0.13 という結果になった。

最後に基本周波数の変化と語尾を上げるという基本周波数の変動に寄与しているパラメータ値について結果をまとめる。まず、基本周波数を変化させている AV+DF0(M) と AV+DF0(F) の結果は、男声から抽出した基本周波数の変化である AV+DF0(M) の場合に母数値-0.25 をとることから、女声知覚に影響を与えることが明らかになり、逆に女声から抽出した基本周波数の変化である AV+DF0(F) の母数値 0.071 という結果が得られた。さらに語尾を下げる上げるといった基本周波数を変化させた AV+G(M) と AV+G(F) の結果をみていく。語尾を下げる処理をした音声である AV+G(M) の母数値は 0.16 と男声知覚に影響を与えている。そして語尾を上げる処理をした音声である AV+G(F) の母数値は-0.067 という女声知覚に影響を与えている。

5.6 実験 1 の考察

静的な特徴と動的な特徴を付加させた音声を用いて実験を行った結果、平均基本周波数とスペクトル包絡を付加することによって、男声 · 女声になっていることが静的特徴量の結果から示すことができる。平均基本周波数とスペクトル包絡が男声 · 女声で重要な要因であると Childers と Wu[15] や Kalatt ら [16] の先行研究でで言われてきたことからこの結果は先行研究の結果 [15][16] を支持するものである。男声 · 女声の声質変換手法において Nguyen と赤木 [22] が行った平均基本周波数とスペクトル包絡のフォルマントを変形することで声質変換を行った結果、女声から男声 100 %変化し、女声から男声は約 83 %という結果になったことから、平均基本周波数とスペクトル包絡が男声 · 女声知覚に与える影響は非常に強いことが言える。

次にスペクトルの変化量に関しては男声 · 女声知覚に影響を与えていないことがわかった。この結果は、個人性におけるスペクトルの変化に着目した北村ら [13] と Zhu[14] らと同じ結果が得られている。個人性に対してあまり影響を与えなかったスペクトルの変化量であったが、男声 · 女声知覚においてもあまり影響を与えないことが実験結果から明らかになった。

基本周波数の変化，語尾を上げる，語尾を伸ばすについては，男声 · 女声知覚に影響を与えている結果であるが，平均基本周波数やスペクトル包絡と比べると小さいため，実験 1 の結果からは基本周波数の変化，語尾，音韻長といった特徴が，男声 · 女声知覚に影響を与えているか依然不明である．

そして、ゲインは男声 · 女声知覚に与えている影響が少ないことが明らかになった。

以上の結果をまとめると男声 · 女声知覚に与える影響が強い物理量として、平均基本周波数、スペクトル包絡であり、あまり影響を与えない物理量として話速、スペクトルの変化量、ゲインであることが実験 1 から明らかになった。静的特徴に比べると動的特徴が男声 · 女声知覚与える影響が少ないことが明らかになった。

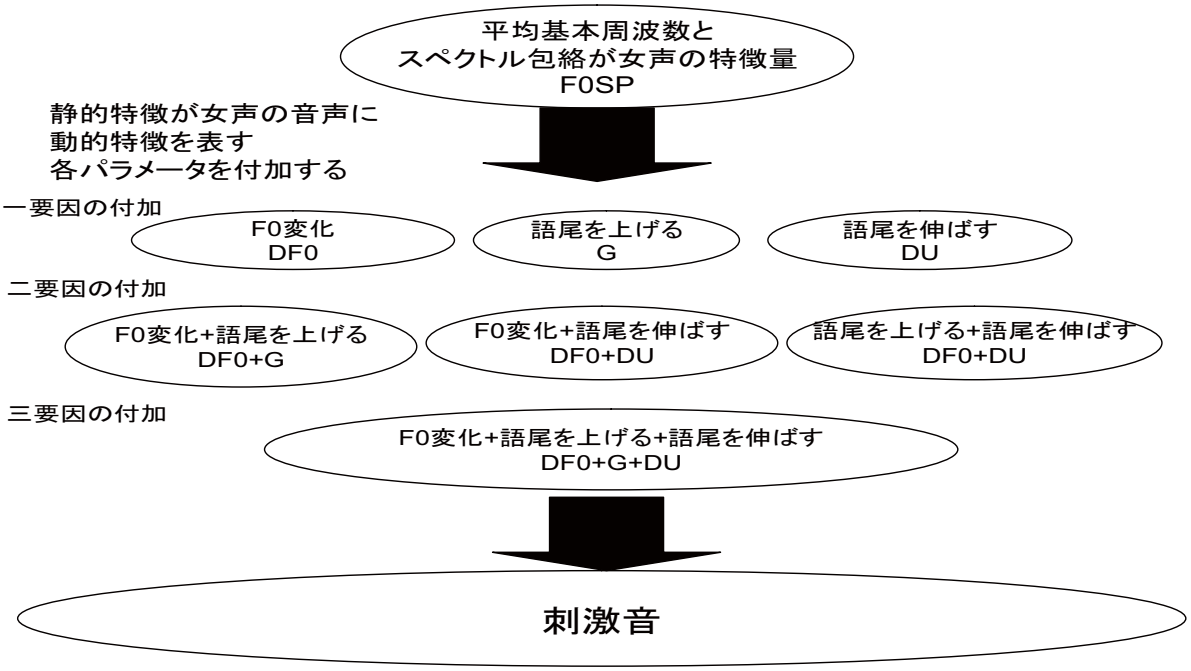

図 5.5: 実験 2 で用いる刺激音

5.7 実験 2 の目的

実験 2 では，実験 1 で影響の強い平均基本周波数とスペクトル包絡を固定し，実験 1 の結果から動的特徴の中で男声 · 女声知覚に影響を与えた “基本周波数の変化”，“語尾の変化”，“音韻長” といった特徴が知覚にどのような影響を与えているか調査することを目的とする．これら三要因は実験 1 の結果から男声知覚より女声知覚に影響を与えていたため，実験 2 は女声らしさに関して調査した.

5.8 実験 2 の刺激音

実験 2 で用いる刺激は図 5.5 示す。**刺激**: 静的特徴で強い影響を与えていた平均基本周波数とスペクトル包絡を女声のパラメータとして，その他のパラメータについては実験 1 と同様に平均声のパラメータを使用した (F0SP).
実験 2 ではこの音声 (F0SP) に，基本周波数の変化 (DSP)，語尾を上げる (G)，語尾を伸ばす (DU) という特徴を一つ付加したもの，二つ付加したもの，すべて付加したものをそれぞれ作成し，合計 8 種類の刺激を用いて実験を行った (図 5.5).

5.8.1 実験手続き

- 実験参加者には次のような指示を与えて、女声らしさに関して評価してもらった。ヘッドホンから２つの音声を対にして流します。前の音声（A）と後の音声（B）を聴き比べて、前の音声と後の音声を聞き比べてどちらが女声らしく聞こえるかということを下に記した７段階の評価尺度に従って判断してください。A が女声らしい聞こえる場合は正の値（1 から 3）B が女声らしいと聞こえる場合は負の値 (-1 から-3) のあてはまる値にチェックしてください。前の音声と後の音声を聞いてどちらも同じ場合は 0 を選択してください。

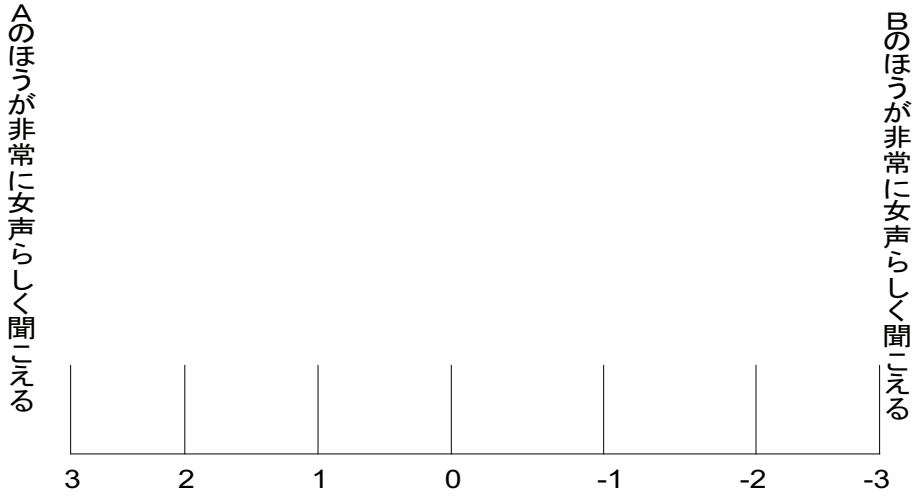

図 5.6: シェッフェの一対比較法で用いた女声らしさに関する７段階評価尺度

被験者のタスク:評価は，“3.A のほうが非常に女声らしい”，“2.A のほうが女声らしい”，“1.A のほうが少し女声らしい”，“0. どちらも同じ (どちらともいえない)”，“-1.B のほうが少し女声らしい”，“-2.B のほうが女声らしい”，“-3.B のほうが非常に女声らしい” の７段階とした.

表 5.3: 実験 2 の母数の推定

No.1	-0.42
No.2	-0.23
No.3	0.48
No.4	-0.57
No.5	0.45
No.6	-0.39
No.7	0.28
No.8	0.40

5.8.2 実験参加者

被験者は正常な聴力を有する 22-25 歳の大学院生 4 名 (男性 4 名) である。実験参加者は実験 1 にも参加している。

5.8.3 刺激条件

実験で用いる刺激音は、先に示した 8 個の音声を 2 つずつ対にしたものである。刺激対の数は、1 つの音声データについて、刺激順序の違いも考慮した 56 対である。

5.8.4 実験環境

実験は、人のいない静かな部屋で行い、実験参加者にはヘッドホン (Sennheiser HDA200) を介して刺激音対を両耳に呈示し、PC ディスプレイ上の評価尺度をキーボードで入力させることで回答させた。また音圧レベルは実験参加者の聴きやすいレベルに設定した。

5.9 実験 2 の結果と考察

聴取実験の結果を図 5.7 に示す. まず一つの特徴を付加した結果からみていくと，動的特徴が平均 (F0SP)，基本周波数の変化 (DF0)，語尾を上げる (G)，語尾を伸ばす (DU) では，-0.40， -0.23， 0.25， -0.45 という結果であった. 二つの特徴を付加した結果では，基本周波数の変化と語尾を上げる (DF0+G)，基本周波数の変化と語尾を伸ばす (DF0+DU)，

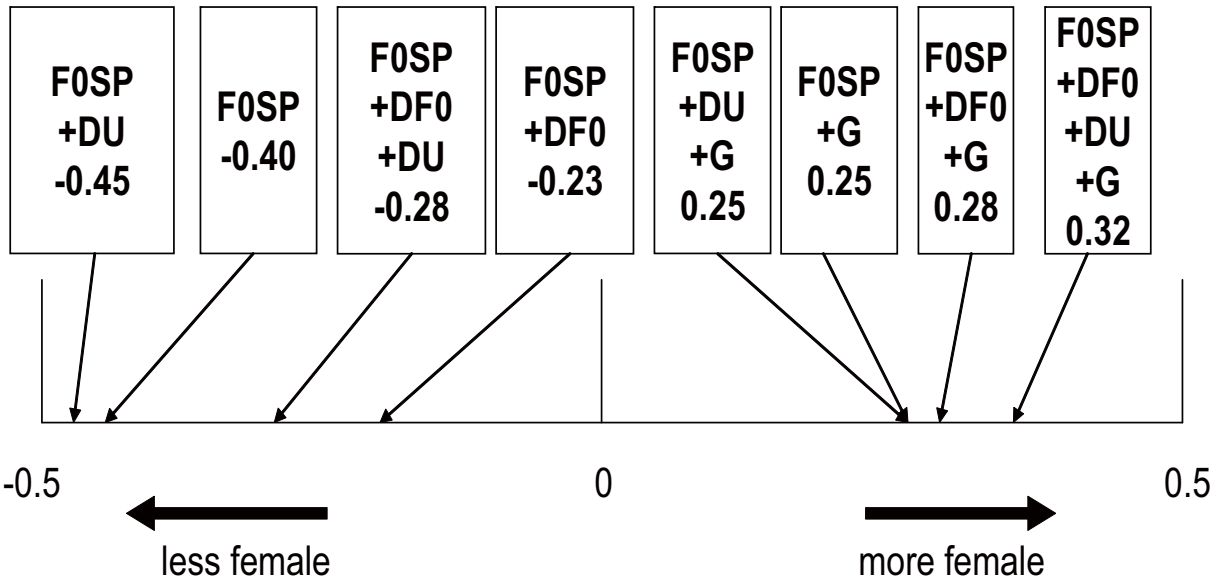

図 5.7: 実験 2 の結果

語尾を上げると語尾を伸ばす (G+DU) では，0.28， -0.28， 0.26 という結果であった. 三つの特徴を付加したもの (DF+DU+G) は 0.32 という結果であった．全体の傾向として，女声と判断された動的特徴を付加していくと，女声らしく知覚されるという結果が得られた. パラメータ値を平均したものより，動的特徴を女声にしたものがより女声らしく知覚されるということは，動的特徴が，付加的ながら，女声音声を知覚する上で手がかりとなっていることを示す結果である．このことは，櫻庭ら [19][17][18] が示した女声らしい話し方としてあげている抑揚をつける，語尾を上げる，語尾を伸ばすに対応しており，女声らしいと感じるためのパラメータ値として知覚に影響を与えている可能性を示唆するものである.

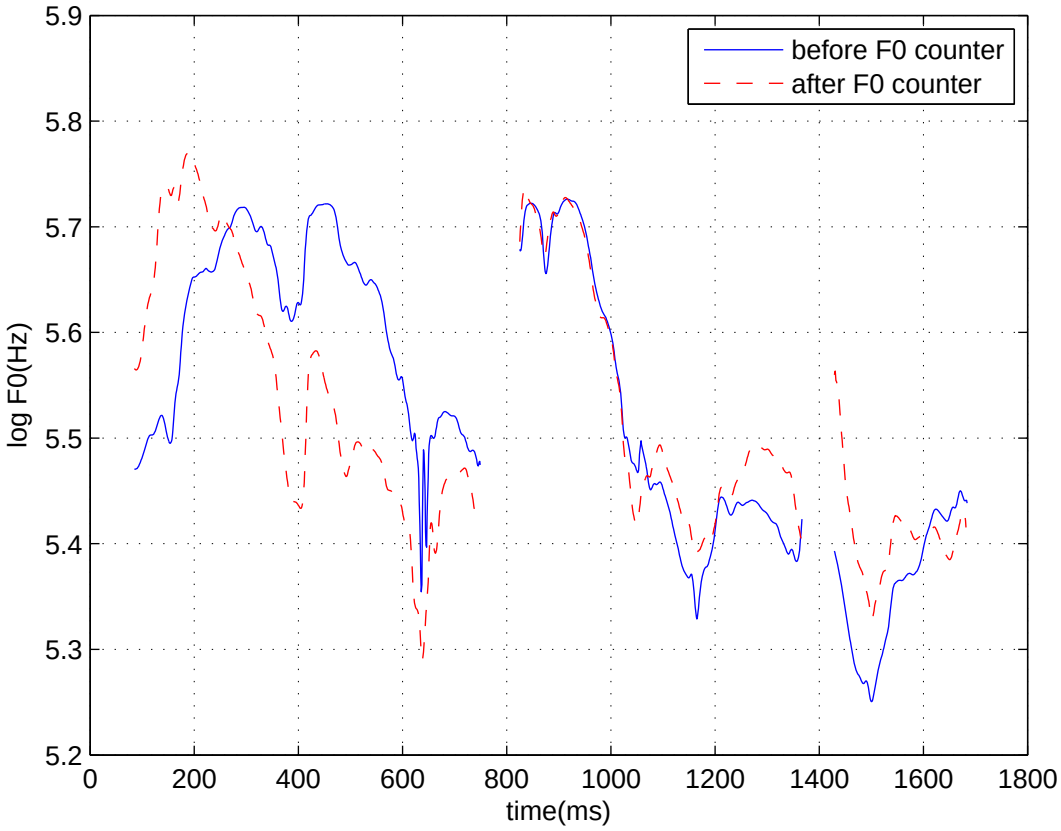

図 5.8: 青が変化前で赤が今回女声と知覚された F0 の変化パターン

第6章　全体の考察

本稿では，連続発話音声中に含まれる男声 · 女声知覚に寄与する音響特徴量を静的な特徴，および動的な特徴に分類し，これらがどのような順序で寄与しているかを明らかにすることを目的とした．この問題を解決するために，静的特徴と動的特徴が扱える声質変換モデルを用い，合成した音声を刺激として用いて，聴取実験により各特徴量がどのような順序で寄与しているか確かめた．結果および考察を整理する．

- 声質変換モデル

  今回用いた声質変換モデルではイベント関数のスロープの部分を非線形最小二乗法を用いたフィッティングを行った後、用いた式を使って、イベント関数を構築している。MRTD でのイベント関数とフィッティングを行ったイベント関数では、スロープをうまく表現できていない可能性がある。実験 1 の結果では、スペクトルの変化に対して変化がみられないような結果が得られているが、これはスロープがうまく表現できていないため変化が見られないような結果が出た可能性がある。そして声質変換モデルでは MRTD のイベント関数のスロープで表現できない部分によって誤差でてしまう課題が残っている。声質変換モデルのイベント関数のスロープの部分を MRTD のイベント関数のスロープに正確に近似できるように改善する必要がある。

- 静的特徴量

  今回静的特徴として用いたのは，平均基本周波数，スペクトル包絡，およびゲインのダイナミックレンジである．MDS 分析の結果と実験 1 から，平均基本周波数とスペクトル包絡について男声 · 女声知覚に影響を与えていることが示唆される結果を得た．これは先行研究 [15][16] を支持するものである．そしてゲインのダイナミックレンジについては，分析結果では違いが見られたが，実験の結果からは明確な違いが現れなかった．ゲインのダイナミックレンジについては男声 · 女声知覚に影響を与えていないことを示唆する結果であった．

- 動的特徴量

  動的特徴として，基本周波数の変化，スペクトルの変化，および音韻長に関して調査した．実験ではもう少し細分化し，語尾を上げる，語尾を下げる，語尾を伸ばす

を追加している．実験 1 の結果，静的特徴に比べると影響力が小さいものの，動的特徴のいくつかのパラメータ値は男声 · 女声知覚に影響を与えている可能性を示唆する結果が得られた. 具体的には音韻長，基本周波数の変化 (語尾を上げるを含む) といったパラメータが男声 · 女声知覚に影響を与えている可能性が見て取れた．スペクトルの変化については，話者知覚と同様にあまり影響を与えていないことを示唆する結果であった．実験 2 では基本周波数の変化と語尾の変化と音韻長を付加した音声がより女声に知覚されるかどうかを調査したところ，基本周波数の変化，特に語尾をあげる特徴を付加した音声がもっとも女声らしく知覚されることが明らかとなった.

- 各特徴量の寄与

  実験 1 と実験 2 の結果から男声 · 女声知覚には静的特徴である平均基本周波数とスペクトル包絡が大きな影響を与えており，次いで，動的特徴である基本周波数の変化と音韻長が影響を与えており，スペクトルの変化量とゲインのダイナミックレンジはあまり影響を与えていないことが明らかになった．今回音韻長全体を分析しても男声 · 女声で違いが見られなかったが，音韻長の特定の特徴に着目することで，男声 · 女声で差が出たことから音韻長に対する分析を細かく行うことが必要であろう.

第7章　結論

7.1　本論文で明らかになったことの要約

本論文では，連続発話音声中に含まれる男声・女声知覚に寄与する音響特徴量を静的な特徴，および動的な特徴に分類し，これらがどのような順序で寄与しているか明らかにするために、声質変換モデルを提案し、多次元尺度構成法を用いて分析を行った。そしてシェッフェの一対比較法を用いた聴取実験を行った。声質変換モデルでは、静的成分、動的成分を分析合成でき、音質のよい分析合成モデルができたといえる。そして、多次元尺度構成法での分析結果で男声 · 女声が基本周波数、スペクトル包絡、ゲイン、スペクトルの変化については違いがでたが、音韻長には違いがみられなかった。そして聴取実験の結果から、実験 1 で基本周波数、スペクトル包絡といった静的成分の影響が強く、動的成分も影響を与えていることが明らかになった。そして実験 2 では、静的成分を固定して動的成分を付加していくことで、動的成分のどの特徴量が女声らしさに影響を与えているか調査した。その結果から、語尾が動的成分の中で一番影響を与えていることが明らかになった。実験 1 と実験 2 の結果，男声 · 女声知覚には静的特徴である平均基本周波数とスペクトル包絡が大きな影響を与えており，次いで，動的特徴である基本周波数の変化と音韻長が影響を与えており，スペクトルの変化とゲインのダイナミックレンジはあまり影響を与えていないことが明らかになった.

7.2　今後の課題

今後の課題を以下に記す

- 平均声今回用いた平均声は男声と女声の算術平均を用いて作成した。そして、スペクトル包絡を平均にしたときに音質が劣化してしまう問題が残っている。そこで、Nguyen と赤木 [22] の手法を用いてスペクトルを平均した音声を作ることで、平均声の音質がよくなると考えられる。

- AP について

  今回非周期成分 (AP) については男声と女声で平均のものを用いており、さらに分析も行っていない。AP については女声の声帯の開き方などに影響を与えており、女声で違いがみられるため重要である。今後は、AP を考慮して分析する必要がある。

- 音声データの数

  今回実験に用いた音声は「だれにでもいいんじゃないかな」の一つだけであったため音声データに依存している可能性がある。さらに実際の会話音声を用いて実験を行うことでより動的な特徴が調べられると考えられる。そして大規模な聴取実験を行い得られた結果が一般的であるかどうか検証する必要がある。

- 男声らしいに関する調査

  実験2で行ったものは女声らしいについて調査したが，男声らしいについて調査していない．男声らしい特徴量はなんなのかということに対してさらに聴取実験を行い明らかにする必要がある．

謝辞

本研究を遂行するにあたり、数多くの貴重なご助言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赤木正人教授、鵜木祐史准教授、李軍鋒助教、並びに本学の教官の皆様に深く感謝致します。本研究を進める過程において、多大なアドバイスをくださり、熱心に御討論いただいた音情報処理学講座の皆様に深く感謝致します。また、ジョイントミーティングなどで熱心に御討論いただいた知能情報処理学講座の皆様に深く感謝いたしますまた、御多忙の中、聴取実験に参加いただいた皆様に深く感謝致します。最後に、2年間の研究生活を支えてくださった全ての皆様に深く感謝いたします。

参考文献

[1] 桑原尚夫, “個人性の音響的特徴量とその制御,” 音講論, 1-7-11, pp. 615-618, Oct 1993.

[2] 桑原尚夫, 大串健吾 “アナウンサーの声質とその音響的特徴,” 音声研究会資料, S82-38, Sep 1982.

[3] 齋藤毅, 北村達也, “3 連続母音に含まれる個人性情報の知覚要因,” 日本音響学会講演論文集, 2007, 1, 441-442 (2007).

[4] Tatsuya Kitamura, Masato Akagi, Speaker individualities in speech spectral envelopes, Journal of the Acoustical Society of Japan(E), 16, 283-289 (1995).

[5] 北村達也, 齋藤毅, “単母音の音響特徴量の変化が個人性知覚に与える影響,” 信学技報, 2007-03

[6] 北村達也,“ 物真似タレントによる物真似音声の分析,” 電子情報通信学会技術研究報告 (音声), 107, 282, 49-54 (2007).

[7] David R.R.Smith, Jennifer M. Fellowes and Dalia S. Nagel,‘On the perception of similarity among talkers,” Journal of the Acoustical Society of America,Vol.122,no. 6 pp3688-3696,2007.

[8] Robert E. Remez, Thomas C. Walters and Roy D. Patterson,“Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled,” Journal of the Acoustical Society of America,Vol.122,no. 6 pp3628-3639,2007.

[9] 鈴木教郎, 赤木正人, “文音声中に含まれる個人性情報の知覚,” 信学技報, 1999-03

[10] 齋藤毅, 後藤真考, “歌声の個人性知覚に寄与する音響特徴の検討,” 音講論, 3-Q-26, pp. 601-602, Sep 2007.

[11] 家永太郎, 赤木正人, “音声のピッチ周波数の時間変化パターンに含まれる個人性とその制御,” 信学技報, 1995-03

[12] M.Akagi and T.Ienaga, “Speaker individualities in fundermental frequency contours and its control,”J.Acoust.Soc.Jpn.(E)18,2(1997)

[13] 北村達也，赤木正人，北澤茂良，“連続音声スペクトル遷移に含まれる個人性,” 音講論, 3-8-9, pp. 389-390, Sep 1998.

[14] W.Zhu, H. kasuya “Perceptual Contributions of Static and Dynamic Features of Vocal Tract Characteristics to Talker Individuality ,” IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences Vol.E81-A No.2 pp.268-274

[15] D.G.Childers and K.Wu,“Gender recognition from speech. partI:coarse analysis,” Journal of the Acoustical Society of America,Vol.90,pp1828-1840,1991.

[16] D.H.Klatt and L.C.Klatt, “Analysis, synthesis and perception of voice quality variations among female and male talkers ,” Journal of the Acoustical Society of America,Vol.87, no. 2, pp820-857,1990.

[17] 櫻庭京子, 丸山和考, 峯松信明, 広瀬啓吉, 田山二朗, 今泉敏, 山内俊雄 “話者認識技術を用いた性同一症者の音声に対する男性度、女性度の自動推定とその臨床応用,” 信学技報, 2006-03

[18] 櫻庭京子, 丸山和考, 峯松信明, 広瀬啓吉, 田山二朗, 今泉敏, 山内俊雄 “男性から女性へ性別移行を希望する性同一性障害者の発話音声の分類に関する試案,” 信学技報, 2007-03

[19] 櫻庭京子，今泉 敏，広瀬啓吉， 新美成二，筧 一彦, ”女性と判定された性同一性障害者 (MtF) の声の基本周波数” 日本音響学会聴覚研究会資料 H-2003-16, Vol.33 (2)

[20] 藤野善行、”スペクトルと基本周波数のイベント操作による音声モーフィングに関する研究 ”、JAIST 修士論文、Febluary、2001

[21] 里地高典、”母音ターゲットスペクトル交換に基づく声質変換に関する研究 ”、JAIST 修士論文、Febluary、2004

[22] B. P. Nguyen and M. Akagi, “Spectral Modification for Voice Gender Conversion using Temporal Decomposition,” NCSP07, 2007.

[23] P. Zolfaghari and T. Robinson, “ Formant analysis using mixtures of Gaussians, ” Proc. of ICSLP, pp. 1229-1232, 1996.

[24] P. Zolfaghari, S. Watanabe, A. Nakamura, and S. Katagiri, “ Bayesian modelling of the speech spectrum using mixture of Gaussians, ” Proc. of ICASSP, pp. 553-556, 2004.

[25] 河原英紀, “聴覚の情景解析と高品質音声分析変換合成法 STRAIGHT,” 日本音響学会講演論文集, 1-2-1, pp.189-192, 1997.

[26] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigne “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction,” Possible role of a repetitive structure in sounds, Speech Communication, 27, pp.187-207 (1999).

[27] H. Kawahara, H. Katayose, A. de Cheveigne, R. D. Patterson, “Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity ,” Proc. EUROSPEECH’99, Volume 6, Page 2781-2784 (1999).

[28] H. Kawahara, A. de Cheveigne, H. Banno, T. Takahashi and T. Irino, “Nearly Defect-free F0 Trajectory Extraction for Expressive Speech Modifications based on STRAIGHT,” Proc. Interspeech2005, Lisboa, pp.537-540, Sept. 2005.

[29] P. C. Nguyen, and M. Akagi, “Improvement of the restricted temporal decompostion method for line spectral frequency parameters,” Proceedings of the 27th IEEE Internatinal Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002), Orlando, Florida, USA, pp. 265-268, May 2002. Vol.E86-D, No.3, pp.397-405, 2003.

[30] P. C. Nguyen, T. Ochi, and M. Akagi, “Modifired restricted temporal decomposition and its application to low bit rate speech cording,” IEICE Transactions on Information and Systems,Vol.E86-D, No.3, pp.397-405, 2003.

[31] P. C. Nguyen, M. Akagi and T.B.Ho,“Temporal decomposition: A promising approach to VQ-based speaker identification,” Proc. ICASSP, pp184-187, 2003.

[32] 板倉秀一,“音声工学,” 森北出版株式会社,2005

[33] B. S. Atal, “Efficient coding of LPC parameters by temporal decomposition,” Proceedings of the IEEE International Conference on Acoustics,Speech and Signal Processing(ICAASSP’83), pp.81-84, 1983.

[34] A.C.R.Nandasena and M.Akagi,“Spectral stability based event localizing temporal decomposition,” Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP’98), pp.957-969,1998.

[35] M. Abe, Y. Sagisaka, T. Umeda, and H. Kuwabara, “Speech database user manual,” ATR Technical Report, TR-I-0166, 1990.

[36] 天坂 格郎, 長沢 伸也, “官能評価の基礎と応用,” 日本規格協会, 2000.

[37] 林 知己夫, 飽戸 弘, “多次元尺度解析法” サイエンス社, 1976.

学会発表リスト

1. T.Shibata，M.Akagi，”A study on voice conversion method for synthesizing stimuli to perform gender perception experiments of speech,”
   Proc. NCSP08,180-183, Gold Coast, Australia, March 2008(to be appear).
2. 柴田武志，赤木正人，”連続発話音声中に含まれる男声女声知覚に寄与する音響特徴量,”
   電子情報通信学会技術研究報告, SP-2007-206,117-122, March 2008(発表予定).