

Title	音声の音響特徴量の動的成分が個人性知覚に与える影響に関する研究
Author(s)	出水田, 剛志
Citation	
Issue Date	2012-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/10425
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

修 士 論 文

音声の音響特徴量の動的成分が
個人性知覚に与える影響に関する研究

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

出水田 剛志

2012年3月

修士論文

音声の音響特徴量の動的成分が 個人性知覚に与える影響に関する研究

指導教員 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科学専攻

1010003 出水田 剛志

提出年月: 2012年2月

概要

音声コミュニケーションはヒトの基本的かつ基礎的な営みである。それに対して、“ヒトはどのように音声中から言語情報を得ているのか？”という問いに対する解答は未だ得られていない。解明に至らない一つの要因として、話者間での差異（個人性）が挙げられる。発話内容が同じであっても、ヒトにより音響特徴量は異なっている。ヒトは話者が異なっても、話者間に差異があるにも関わらず、共通の言語情報を聴取できる。これは、ヒトは個人性を正規化または適応することで言語情報を得ているからだと考えられる。一方、ヒトは個人性を利用して音声中から話者を判断することができる。しかし、“ヒトはどのように話者を判断しているのか？”という問いに対する明確な解答には未だ至っていない。どのように個人性を正規化または適応しているか？、あるいは、どのように話者を判断しているのか？という問題は音声科学の基本的課題である。これらのメカニズムを解明するためには、まず、音声中のどのような特徴量が個人性知覚要因となっているかを明らかにする必要がある。

個人性知覚に関する先行研究によって、様々な特徴量が個人性知覚へ寄与することが報告されてきた。これらの先行研究は、時間的に平均的な特徴量（静的成分）に関する研究と時間的に変動する特徴量（動的成分）に関する研究に大きく分類される。しかし、これまでの研究において、個人性知覚と動的成分の関係の調査が十分に行われて来たとは言いがたい。動的成分は発話器官の動きに由来し、その動きに関連する複数の音響特徴量に関連して変化していると考えられる。つまり、個人性知覚と動的成分の関係を調査する上で複数の音響特徴量を有機的に組み合わせる必要がある。複数の音響特徴量を有機的に集約する上で、声質や発話様式といった聴取印象に着目することは有益であると考えられる。聴取印象は形容詞（表現語）によって記述される。同じ非言語情報である感情音声や歌声などの分野においても、知覚と音響特徴量との関係が三層構造モデルを用いて構築されている。本稿では、個人性知覚（第一層）と聴取印象（第二層）、および音響特徴量（第三層）の関係を三層構造モデルによりトップダウン的に議論した。さらに、第三層の音響特徴量の変化が第一層の個人性知覚に与える影響をボトムアップ的に検証した。これらの結果を報告する。

個人性知覚（第一層）と聴取印象（第二層）の関係の解明には、まず、話者間の類似性評価実験に基づいて MDS により話者の知覚的距離空間を作成した。そして、次に、SD 法により求めた話者の聴取印象の強度とその空間との対応を重相関分析により調査することで求めた。その結果、「はきはき」が、個人性聴取印象の主要な因子であることがわかった。聴取印象（第二層）と音響特徴量（第三層）の関係の解明には、SD 法の結果と抽出した音響特徴量との間で相関分析を行うことで調査した。抽出した音響特徴量は基本周波数（ F_0 ）に関する特徴量、パワーに関する特徴量、スペクトルに関する特徴量、時間長に関する特徴量である。相関分析の結果、 F_0 の平均値、最大値、変化の傾きならびにスペクトル傾斜の最大値とダイナミックレンジが「はきはき」と相関があった。 F_0 の変

化の傾きとスペクトル傾斜のダイナミックレンジは動的な特徴量である。このことから、「はきはき」は動的成分に関わる聴取印象であることが確認できた。

さらに、分析の結果明らかとなった「はきはき」に関する三層構造モデルの検証をボトムアップ的に行った。検証を行う上で、 F_0 の変化の傾きとスペクトル傾斜のダイナミックレンジを変化させた刺激音を作成した。まず、作成した刺激音が第二層の聴取印象を制御できているかを検証した。その結果、 F_0 の変化の傾きとスペクトル傾斜のダイナミックレンジを変化させることで「はきはき」の強度を段階的に制御できていることを確認できた。次に、第三層の音響特徴量の変化が第一層の個人性知覚に与える影響を検証した。その結果、「はきはき」に関する音響特徴量の変化が個人性知覚に影響を与えた。よって、音声中の動的成分は個人性知覚に影響を与えていることが明らかとなった。また、聴取印象の強度が個人性知覚に影響を与えていることが示唆された。さらに、本研究で得られた知見や本研究で用いた手法は、“ヒトはどのように話者を判断しているのか？” “ヒトはどのように音声から言語情報を得ているのか？” という大きな課題の解明に繋がるものと考えられる。

目次

第1章	序論	2
1.1	本研究の背景	2
1.2	先行研究	2
1.3	本研究の目的	3
1.4	本論文の構成	6
第2章	三層構造モデルの概要	8
2.1	まえがき	8
2.2	三層構造モデルに関する先行研究	8
2.3	本研究の手法	9
2.4	STRAIGHT	9
2.5	まとめ	10
第3章	第一層と第二層の関係の分析	11
3.1	まえがき	11
3.2	音声データ	11
3.3	話者の知覚距離空間の作成	11
3.3.1	類似性評定実験	12
3.3.2	多次元尺度構成法による分析結果	12
3.4	個人性聴取印象の選定	16
3.4.1	SD法による聴取実験	16
3.4.2	重相関分析による分析結果	18
3.5	まとめ	18
第4章	第二層と第三層の関係の分析	19
4.1	まえがき	19
4.2	F_0 に関する特徴量	19
4.2.1	抽出した特徴量	19
4.2.2	結果と考察	21
4.3	パワーに関する特徴量	23
4.3.1	抽出した特徴量	23
4.3.2	結果と考察	23

4.4	スペクトルに関する特徴量	23
4.4.1	抽出した特徴量	23
4.4.2	結果と考察	24
4.5	時間長に関する特徴量	27
4.5.1	抽出した特徴量	27
4.5.2	結果と考察	27
4.6	まとめ	27
第5章	モデルの検証	28
5.1	まえがき	28
5.2	第三層が第二層に与える影響の検証（検証実験Ⅰ）	28
5.2.1	刺激音	28
5.2.2	実験協力者	34
5.2.3	実験手順	34
5.3	実験結果と考察	34
5.4	第三層が第一層に与える影響の検証（検証実験Ⅱ）	35
5.4.1	刺激音	35
5.4.2	実験協力者	35
5.4.3	実験手順	35
5.5	実験結果と考察	37
5.6	まとめ	39
第6章	全体考察	40
第7章	結論	42
7.1	本研究で明らかとなったことの要約	42
7.2	残された課題	42
	謝辞	45
	参考文献	46

目次

1.1	三層構造モデル	5
1.2	本論文の構成	7
3.1	実験中のモニターの様子	13
3.2	MDS の次元数とストレス値	14
3.3	実験中のモニターの様子	17
4.1	はきはきした話者 (1.5) と はきはきしていない話者 (-1.2) の F_0 パターン (括弧の中の値は「はきはき」の強度を示す)	20
4.2	聴取印象「はきはき」と F_0 の変化の傾きの関係	22
4.3	聴取印象「はきはき」とスペクトルの傾斜の変動幅の関係	26
5.1	刺激音 1 の合成手順	30
5.2	刺激音 2 の合成手順	31
5.3	刺激音 3 の合成手順	32
5.4	刺激音 4 の合成手順	33
5.5	検証実験 I の結果	34
5.6	刺激音の呈示の様子	35
5.7	実験中のモニターの様子	36
5.8	検証実験 II の結果	37
5.9	刺激音 1 と 4 の間での「はきはき」の強度の差と正答率の関係	38

表 目 次

3.1	実験に用いた機器	12
3.2	Stress の評価	14
3.3	6次元の空間における各話者の座標	15
3.4	表現語候補対	16
3.5	重相関分析の結果	18
4.1	「はきはき」と F_0 の関係	21
4.2	「はきはき」とパワーの関係	23
4.3	「はきはき」とスペクトルの傾斜との関係	25
4.4	「はきはき」と時間長の関係	27

第1章 序論

1.1 本研究の背景

音声コミュニケーションはヒトの基本的かつ基礎的な営みである。それに対して、“ヒトはどのように音声中から言語情報を得ているのか？”という問いに対する解答は未だ得られていない。解明に至らない一つの要因として、話者間での差異（個人性）が挙げられる。発話内容が同じであっても、ヒトにより音響特徴量は異なっている。ヒトは話者が異なっているにもかかわらず、話者間に差異があるにも関わらず、共通の言語情報を聴取できる。これは、ヒトは個人性を正規化または適応することで言語情報を得ているからだと考えられる[1][2][3]。一方、ヒトは個人性を利用して音声中から話者を判断することができる。しかし、“ヒトはどのように話者を判断しているのか？”という問いに対する明確な解答には未だ至っていない。どのように個人性を正規化または適応しているか？、あるいは、どのように話者を判断しているのか？という問題は音声科学の基本的課題である。これらのメカニズムを解明するためには、まず、音声中のどのような特徴量が個人性知覚要因となっているかを明らかにする必要がある。

1.2 先行研究

個人性知覚に関する先行研究によって、様々な特徴量が個人性知覚へ寄与することが報告されてきた。これらの先行研究は、時間的に平均的な特徴量（静的成分）に関する研究と時間的に変動する特徴量（動的成分）に関する研究に大きく分類される。静的成分は主に発話器官の形状に対応し、先天的な特徴である。一方、動的成分は発話器官の動きに由来し、後天的な特徴である。

音声の個人性知覚要因となる音響特徴量は、これまでに数多く調べられてきている。静的成分が個人性知覚に寄与すると報告している先行研究として、伊藤ら [4] は、スペクトル包絡と基本周波数 (F_0) と時間長が個人性知覚に与える影響を調査した。その結果、スペクトル包絡が最も重要であると報告した。また、スペクトル包絡の復元精度も個人性知覚に影響を与え、復元精度が低いときは基本周波数の情報が利用されることを示した。橋本ら [5] も、スペクトル包絡と F_0 と時間長の3要因が個人性知覚への寄与率を求めた。その結果、伊藤ら [4] と同様の結果を示している。北村ら [6, 7, 8] は単母音を用いた聴取実験を行い、スペクトル包絡のうち、高周波数帯域（高域）が個人性知覚に寄与していると報告している。また、Kitamura [9] らは、スペクトル包絡中の高域は発話中に形状が変化

しにくい下咽頭腔の形状に由来することを明らかにしている．Amino ら [10] は、鼻音が話者識別に有効であり、それは鼻腔や鼻咽腔などの調音器官の生理学的特徴が個人によって違うことにより由来していることを示した．

一方、動的成分が個人性知覚に寄与すると報告している先行研究として、Akagi ら [11] は、3 モーラ単語における F_0 の変化パターンが含む個人性情報について分析検討を行い、 F_0 の変化パターンが個人性知覚に寄与することを示している．大野ら [12] は文章音声を用いて、Akagi らの報告を検証している．Yang ら [13] は、フォルマント周波数の変化パターンが個人性知覚に寄与することを示している．

また、先行研究では動的成分に含まれる個人性情報量について、意見が分かれている．音声の個人性知覚において、動的成分より静的成分の方が多くの個人性情報を含んでいるという報告がある [14][15]．一方で、動的成分の方が個人性知覚に重要であることが報告されており [16]、話者認識の分野では動的成分の利用が話者識別の性能を向上させたという報告 [17] がある．これは、個々の研究で音声データセットが異なり、しかも、動的成分に含まれる個人性情報の有無や情報量の大小が話者依存であるためと考えられる．

このように、音声中の動的成分に含まれる個人性情報についての調査は十分に行われてきたとは言い難い．動的成分に含まれる個人性情報を明らかにすることは、個人性知覚要因を解明する上で有益である．

1.3 本研究の目的

本研究の目的は、音声中の動的成分が個人性知覚に与える影響を明らかにすることである．動的成分は発話器官の動きに由来し、その動きに関連する複数の音響特徴量が関連して変化していると考えられる．つまり、個人性知覚と動的成分の関係を調査する上で複数の音響特徴量を有機的に組み合わせる必要がある．複数の音響特徴量を有機的に集約する上で、声質や発話様式といった聴取印象に着目することは有益であると考えられる．聴取印象は形容詞（表現語）により記述される．

音声中の聴取印象に着目することで、個人性知覚要因の調査を行った先行研究 [18, 19] は既に行われている．木戸ら [18] は、声質をに関連する表現語を聴取実験により抽出し、聴取印象と音響特徴量との関連を決定木により調査している．また、山下ら [19] は、聴取印象の強度と各音響特徴量の関係を回帰分析により調査している．これらの研究では、聴取印象と音響特徴量の関係は分析されているが、合成音声などを用いての検証が行われおらず、実際に個人性知覚にどのような影響を与えるのか調査がなされていない．さらに、動的成分に対応する聴取印象についての検討は十分になされていない．

一方、同じ非言語情報である感情音声や歌声などの分野において、知覚と音響特徴量との関係が三層構造モデルを用いて構築されている [20, 21, 22]．これらの研究において、三層構造では、第一層である感情や歌声らしさに寄与する音響特徴量（第三層）を検討するために、中間層として、基本的心理特徴という聴取印象を導入している．第一層と第二層、第二層と第三層との関係をトップダウン的に調査することでモデルを構築し、さらに

ボトムアップ的にモデルの検証が行われている。また，三層構造により，複数の音響特徴量を組み合わせ，さらに聴取印象の強度の制御も実現している。

本研究では，三層構造モデルを用いることにより，動的成分が個人性知覚に与える影響を明らかにする。モデルを構築・検証することで，聴取印象に着目した個人性知覚に関する先行研究 [18, 19] で行われていない，分析により得られた音響特徴量が個人性知覚へ与える影響を検証することができる。また，三層構造を用いることで，個人性知覚と動的成分の関係を調査する上での問題点である，複数の音響特徴量を集約が可能となる。これにより，検討が十分になされていない，動的成分に対応する聴取印象における個人性知覚への影響が調査可能となる。

なお，動的成分の影響を調査するため，本研究では文章音声を対象とする。個人性知覚（第一層）と聴取印象（第二層），および動的成分に関連する音響特徴量（第三層）の関係を三層構造モデルによりトップダウン的に調査することで，個人性知覚と動的成分の関係を議論し，モデルを構築した。さらに，第三層の音響特徴量の変化が第一層の個人性知覚に与える影響をボトムアップ的に検証した。検証する上で，聴取印象の強度に対する個人性知覚への影響も調査した。本研究で用いた三層構造モデルを図 1.1 に示す。

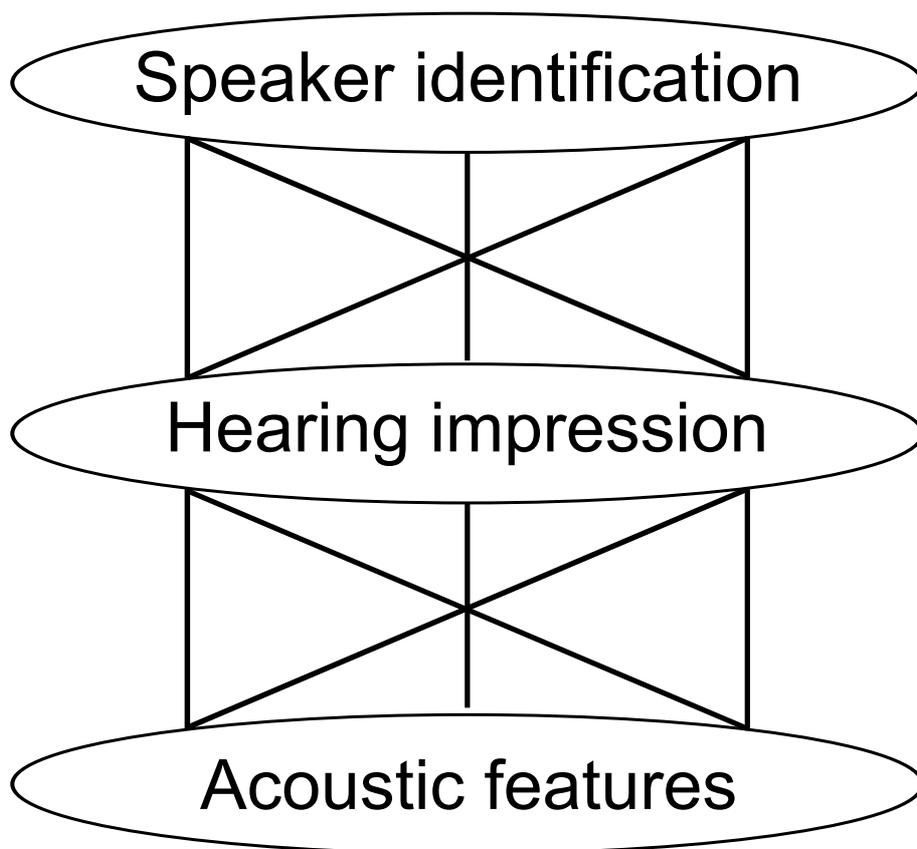


図 1.1: 三層構造モデル

1.4 本論文の構成

本論文は7章で構成される。

第1章では、本研究で対象とする研究分野の背景と問題点を述べ、本研究の目的を明らかにする。

第2章では、本研究で用いる三層構造モデルの概要について説明し、本研究の手法について述べる。

第3章では、三層構造モデルのうち、第一層（個人性知覚）と第二層（聴取印象）の関係をトップダウン的に分析した結果を述べる。

第4章では、三層構造モデルのうち、第二層（聴取印象）と第三層（音響特徴量）の関係をトップダウン的に分析した結果を述べる。

第5章では、第3,4章で構築された三層構造モデルをボトムアップ的に検証した結果を述べる。

第6章では、三層構造モデルをトップダウン的に分析し、ボトムアップ的に検証して得られた結果についての全体考察を述べる。

第7章では、本研究で得られた結果を要約し、今後の展望を述べる。

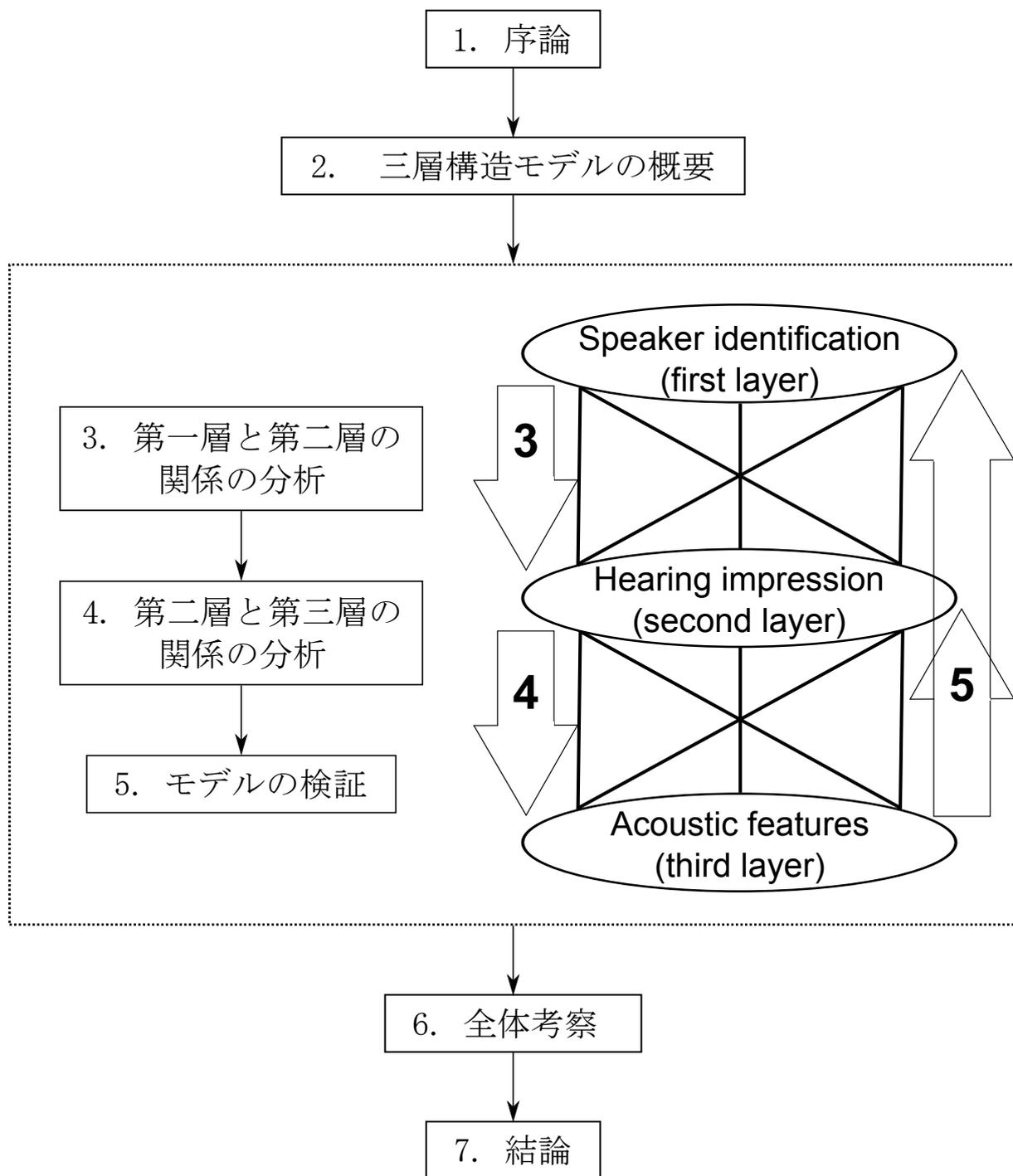


図 1.2: 本論文の構成

第2章 三層構造モデルの概要

2.1 まえがき

本研究の目的を達成するためには、三層構造モデルを分析構築し、検証を行う必要がある。まず、三層構造モデルを用いた先行研究を例に挙げる。次に、本研究でのモデルの分析構築、検証の手法を説明する。さらに、音響特徴量の分析や合成音声の作成を行う上で必要となる、音声分析合成系について説明する。

2.2 三層構造モデルに関する先行研究

Huangら[20, 21]は感情について表現した感情知覚多層モデルを、齋藤ら[22]は“歌声らしさ”を表現した歌声らしさの知覚多層モデルを提案している。

感情知覚多層モデルは、感情知覚における人間の知覚の曖昧さをモデル化するために構築されている。システムの構造として、音響特徴量から感情を知覚するまでの過程を表現するために、三層構造を採用している。第一層である感情と第三層である音響特徴量の間、中間層として聴取印象である基本心理特徴を導入したモデルとなっている。第一層の感情カテゴリとして、“Neutral,” “Joy,” “Sadness,” “Cold-Anger,” “Hot-Anger”の5つを用いており、第二層の基本心理特徴として17種類の形容詞を用いている。17種類の形容詞の選定は、感情知覚空間と形容詞との間での重回帰分析により行われている。第三層の音響特徴量には、 F_0 、パワー、スペクトル、時間長に関する特徴量が扱われている。Akagi[23, 24]は構築されたモデルを用いて感情認識を行った結果、聴取実験で得られた感情と一致した認識結果を得ている。

歌声らしさの知覚多層モデルは、「歌声らしさという聴覚印象が複数の基本的な心理的特徴の知覚に起因する」という仮説のもと構築されている。そのモデルは具体的に、第一層に“歌声らしさ”、第二層に“響き”と“揺れ”、第三層に F_0 のヴィブラートとそれに同期したホルマントの振幅変調成分、及び3kHz付近の顕著なスペクトルピーク成分と同帯域の強い高調波成分が扱われている。

これらの研究は、心理量である感情知覚や歌声知覚と物理量である音響特徴量の関係を、直接的に調査するのではなく、聴取印象を介して調査している。その結果、複数の音響特徴量が有機的に組み合わせられてモデルが構築されている。本研究では、心理量である個人性知覚と物理量である動的成分との関係を調査する。前章で述べたとおり、動的成分は発話器官の動きに由来し、その動きに関連する複数の音響特徴量が関連して変化してい

ると考えられる．そのため，複数の音響特徴量を有機的に組み合わせて議論できる三層構造モデルを用いることは有効である．そこで，三層構造モデルを用いて研究を進める．

2.3 本研究の手法

本研究では，個人性知覚（第一層）と聴取印象（第二層），および音響特徴量（第三層）の関係を三層構造モデルによりトップダウン的に調査することで，個人性知覚と動的成分の関係を議論する．個人性知覚（第一層）と聴取印象（第二層）の関係は，まず，話者間の類似性評定実験に基づいて多次元尺度構成法（MDS）により話者の知覚的距離空間を作成した．そして，次に，SD法により求めた話者の聴取印象の強度とその空間との対応を重相関分析により調査することで個人性聴取印象を求める．

聴取印象（第二層）と音響特徴量（第三層）の関係は，SD法の結果と抽出した音響特徴量との間で相関分析を行うことで調査する．

モデルの検証は，ボトムアップ的に行う．音響特徴量を変化させた合成音声を作成し，第三層の音響特徴量の変化が第二層の聴取印象に与える影響を確認し，さらに，第一層の個人性知覚に与える影響を調査する．

2.4 STRAIGHT

本研究では，分析を行うため音響特徴量を抽出し，さらに，検証実験を行うため合成音声の作成を行う．そのためには，音響特徴量を自由に操作でき，且つ高品質の合成音声を得られる音声分析合成系を用いる必要がある．音声分析合成系とは，音声生成過程のモデルに基づき，音声から声帯振動情報である基本周波数と声道形状情報であるスペクトル包絡を抽出し，これらを音声に再合成するシステムのことである．そこで，本研究では，音響特徴量を自由に操作でき，且つ高品質の合成音声を得られる音声分析合成系である STRAIGHT[30, 31] を採用する．

STRAIGHT 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アは基本周期，基本周波数を接点とする区分的線形関数による補間等価な時間周波数領域の平滑化を行うことにある．SPIKE 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となっている．本研究では，STRAIGHT バージョン 40-006b を使用した．

2.5 まとめ

第2章では，三層構造モデルを用いた先行研究について取り上げ，さらに，本研究の流れを説明した．三層構造モデルをトップダウン的に分析し，分析により構築されたモデルをボトムアップ的に検証することで，本研究の目的を達成する．さらに，STRAISHTについて説明し，STRAIGHTを用いることで，音響特徴量の分析や音声合成が可能になった．

第3章 第一層と第二層の関係の分析

3.1 まえがき

個人性知覚（第一層）に関連する聴取印象（第二層）を選定するため，話者の知覚的距離空間を作成し，その空間と聴取印象の対応を調査する．話者空間は多次元尺度構成法（MDS）を用いて，話者を知覚距離空間に布置することで求める．話者間の知覚的距離を求める必要があることから，多数話者の類似性評定実験を行った．また，話者空間と聴取印象の対応を調査するために，あらかじめ収集した表現語候補を用いて聴取印象を記述し，SD法により話者ごとの聴取印象を求めた．

3.2 音声データ

本研究では動的成分を対象とするため，文章音声を用いた．用いる音声データには，ATR音素バランス503文音声データベース [29]（Cセット）中の男性話者137名中から14名（108, 112, 119, 302, 403, 419, 506, 516, 517, 518, 603, 616, 702, 706）を選んだ．発話文は比較的短文である“冷房では冷えすぎが問題になる”を用いた．話者の選定にあたって，録音状態が悪いデータや発話中に間（ポーズ）を開けている話者や東京方言でない話者は除外した．間の長さや方言などが，個人性知覚に影響があることは想定されるが，間や方言に寄らない，個人性知覚に与える影響を調べるため，このような選定を行った．次に，予備実験により話者の聴取印象を求め，聴取印象に偏りがでないような話者セットを選定し，本研究で用いる音声データセットとした．振幅は最大値で正規化した．

3.3 話者の知覚距離空間の作成

話者空間は多次元尺度構成法（MDS）を用いて，話者を知覚距離空間に布置することで求める．話者間の知覚的距離を求める必要があることから，多数話者の類似性評定実験を行った．

表 3.1: 実験に用いた機器

Headphones	STAX SR-404
Amplifier	STAX SRM-1 MK-2
Audiointerface	YAMAHA DP-U50
Monitor	I·O DATA LCD-A152GS
PC	IBM ThinkPad X61

3.3.1 類似性評定実験

実験（以後すべての実験）は、北陸先端科学技術大学院大学 ライフサイエンス委員会で倫理審査を受けたのち、実施された。

刺激音

用いた刺激音は男性話者 14 名の単一文章である。

実験協力者

実験協力者は聴力に異常のない 22～28 歳の成人男性 10 名である。これらの実験協力者は、刺激音の話者の個人性を記憶していない。つまり、実験協力者にとって、刺激音の話者は未知話者である。

実験手順

実験は、実験協力者が、ランダムに呈示される 2 名が対となった音声を、5 段階（0：似ていない，1：あまり似ていない，3：やや似ている，4：似ている，5：同一人物）で評価した。順序効果を打ち消すため、順序を入れ替えた音声対も呈示する。それぞれの音声対は 2 回ずつ評価され、全試行回数は 392（=14 × 14 × 2）回であった。回答はモニター上のボタンをクリックすることにより行い、実験協力者は各々の聴き易いレベルで受聴し、聴き直しは許さなかった。実験は防音室で行い、表 3.1 に示す機器を用いた。実験中の実験中のモニターの様子を図 3.1 に示す。

3.3.2 多次元尺度構成法による分析結果

話者の知覚距離空間を求めるために、類似性評定実験の結果を入力として、Kruskal の方法（SPSS 14.0J for windows）を用いた MDS により分析を行った。Stress とは、ある次元におけるデータ間における心理距離の適合度をしらすものであり、表 3.2 に Stress の



図 3.1: 実験中のモニターの様子

評価について示す [32] . MDS の結果 , 6 次元のときストレス値が 10% を下回り (図 3.2) , その後収束したことから , 6 次元の知覚距離空間上に話者を布置した .

6 次元の空間における , 各話者の座標を表 3.3 に示す .

表 3.2: 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%	よく適合している (good)
0.5%	非常に適合している (excellent)
0%	完全に適合している (perfect)

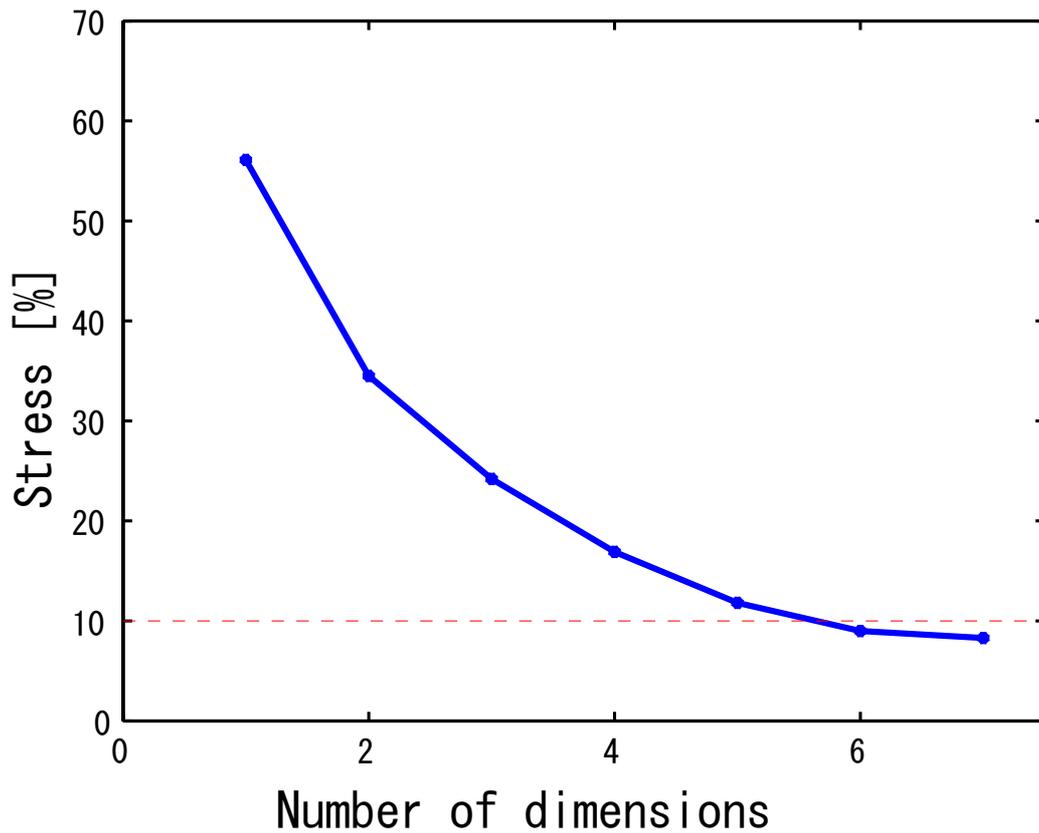


図 3.2: MDS の次元数とストレス値

表 3.3: 6次元の空間における各話者の座標

	1次元	2次元	3次元	4次元	5次元	6次元
108	1.29	-0.80	1.07	-1.07	-0.27	-0.90
112	0.56	1.66	0.53	-0.50	0.47	0.19
119	-2.02	-0.75	0.04	0.01	1.67	-0.31
302	0.21	1.57	-0.01	-1.23	-0.41	0.31
403	0.75	0.10	-2.06	-0.36	-1.07	-0.29
419	0.26	1.69	0.07	0	-0.39	-0.17
506	-1.45	-0.05	-0.47	-0.71	0.27	1.83
516	0.86	0.82	0.74	2.08	-0.80	0.19
517	1.48	-1.03	1.28	-0.93	0.87	-0.45
518	1.27	-2.20	-0.34	0.10	-1.00	0.94
603	0	-0.20	-2.30	0.35	0.81	-1.01
616	0.83	0	0.07	1.59	1.38	0.62
702	-1.98	-0.47	0.95	0.34	-0.81	0.29
706	-2.06	-0.33	0.42	0.35	-0.73	-1.24

3.4 個人性聴取印象の選定

話者空間と聴取印象の対応を調査するために、収集した表現語候補を用いて聴取印象を記述し、SD法により話者ごとの聴取印象を求めた。

3.4.1 SD法による聴取実験

表現語候補の収集

まず、先行研究 [27, 28, 19] により収集された表現語を採用した。また、発話様式に関する表現語が不足していたため、自由記述によるアンケートを用いての収集も行った。アンケートは協力者 5 名に音声の一つずつ呈示し、「どんな話し方でしたか？」という問いに対して、音声を一度聴取するたびに答えてもらった。収集した表現語のうち、話者の性格や職業を表しているもの（やさしい声、アナウンサーっぽい声など）は除外した。また、用いた音声データが男性のみであったため、表現語のうち「女性的な-男性的な」は除外した。収集した表現語候補を表 3.4 に示す。

表 3.4: 表現語候補対

低い	—	高い
かすれた	—	澄んだ
落ち着かない	—	落ち着いた
弱々しい	—	迫力のある
若い	—	老けた
細い	—	太い
張りのない	—	張りのある
鼻声でない	—	鼻声
歯切れの悪い	—	歯切れのよい
滑舌の悪い	—	滑舌の良い
ていねいでない	—	ていねいな
ゆったりしていない	—	ゆったりした
淡々としていない	—	淡々とした
はきはきしていない	—	はきはきした
元気のない	—	元気のある

刺激音

用いた刺激音は男性話者 14 名の単一文章である。



図 3.3: 実験中のモニターの様子

実験協力者

実験協力者は聴力に異常のない22～28歳の成人男性10名である。これらの実験協力者は、刺激音の話者の個人性を記憶していない。つまり、実験協力者にとって、刺激音の話者は未知話者である。

実験手順

実験協力者には、表現語候補3.4ごとに各話者の音声を7段階（-3, 3: 非常に, -2, 2: かなり, -1, 1: やや, 0: どちらでもない）で評価させた。ランダムに呈示される各音声は表現語候補ごとに3回ずつ評価を行い、全試行回数は630（=14 × 15 × 3）回であった。実験は表現語ごとに分けて行い、評価する表現語候補の順番は実験協力者の間でランダム化した。実験協力者には評価尺度に慣れてもらうための練習課題をそれぞれの表現語候補について行ってもらう。1つの表現語に対しての評価を終えたら、適宜休憩を入れてもらった。回答はモニター上のボタンをクリックすることにより行い、実験協力者は各々の聴きやすいレベルで受聴し、聴き直しは許さなかった。実験は、類似性評定実験同様、防音室で行い、表 3.1 に示す機器を用いた。実験中の実験中のモニターの様子を図 3.3 に示す。

表 3.5: 重相関分析の結果

表現語候補	相関係数
ゆったりとした	0.89
高い	0.86
はきはきした	0.85
落ち着いた	0.82
淡々とした	0.78
元気のある	0.78
澄んだ	0.78
迫力のある	0.78
歯切れのよい	0.77
ていねいな	0.77
張りのある	0.76
鼻声	0.72
太い	0.66
滑舌の良い	0.63
老けた	0.60

3.4.2 重相関分析による分析結果

求められた知覚距離空間とSD法による聴取実験により得られた結果から重相関分析により、各聴取印象と知覚距離空間の関係の分析を行った。重相関分析の結果を表3.5に示す。知覚距離空間との相関が最も高かった聴取印象は、「ゆったり（相関係数： $r = 0.89$ ）」であり、続いて「高い（ $r = 0.86$ ）」、「はきはき（ $r = 0.85$ ）」、「落ち着いた（ $r = 0.82$ ）」であった。これらは個人性聴取印象の主要な因子と考えることができる。さらに、「はきはき」は抑揚に関する聴取印象と考えることができるので、動的成分に対応する可能性がある。以上の理由から、「はきはき」に着目する。

3.5 まとめ

個人性知覚（第一層）に関連する聴取印象（第二層）を選定するため、話者の知覚的距離空間を作成し、その空間と聴取印象の対応を調査した。その結果、動的成分に対応する可能性がある「はきはき」が個人性聴取印象の主要因であることが分かった。

第4章 第二層と第三層の関係の分析

4.1 まえがき

聴取印象「はきはき」(第二層)に対応する音響特徴量(第三層)の分析を行う。分析は、小節3.3.2でのSD法の結果と抽出した音響特徴量との間での相関分析を行った。音響特徴量は、 F_0 に関する特徴量、パワに関する特徴量、スペクトルに関する特徴量、時間長に関する特徴量をそれぞれ抽出した。各特徴量の抽出には、STRAIGHTを用いることにより算出を行った。 F_0 の抽出誤りは目視により修正を行った。ラベリングは、音声処理ソフトウェア WaveSurfer[33]上でスペクトログラムを参考にしつつ行った。

4.2 F_0 に関する特徴量

4.2.1 抽出した特徴量

前章で述べたとおり、「はきはき」は抑揚と関係があることが予想される。はきはきした話者とはきはきしていない話者の F_0 パターンを比較したところ、はきはきしている話者は抑揚が大きく、はきはきしていない話者は抑揚が小さいことが確認できた(図4.1)。そこで、 F_0 の平均値、最大値、最小値に加えて、抑揚の大きさに対応する特徴量である、 F_0 の傾き[20, 21]を抽出した。以下に、抽出した音響特徴量の詳細について記述する。また、下線を引かれている音響特徴量は時間的に変動する特徴量を表していることから、動的成分として定義する。

- F_0 に関する音響特徴量
 - 平均値：文章全体における F_0 の平均値
 - 最大値：文章全体における F_0 の最大値
 - 最小値：文章全体における F_0 の最小値
 - 上昇の傾き：各フレーズにおける F_0 の上昇を1次回帰により求めた傾き
 - 下降の傾き：各フレーズにおける F_0 の下降を1次回帰により求めた傾き
 - 変化の傾き：上昇、下降に対する F_0 の変化の傾き絶対値

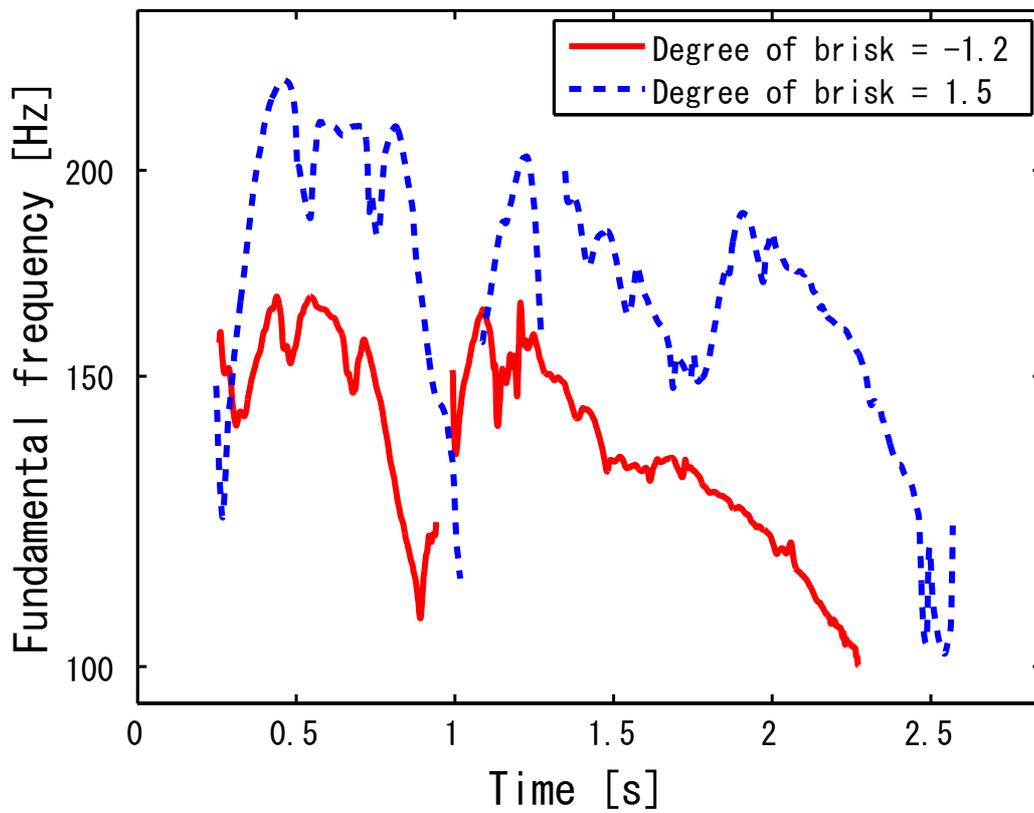


図 4.1: はきはきした話者 (1.5) と はきはきしていない話者 (-1.2) の F_0 パターン (括弧の中の値は「はきはき」の強度を示す)。

表 4.1: 「はきはき」と F_0 の関係

平均値	0.69
最大値	0.82
最小値	-0.29
上昇の傾き	0.71
下降の傾き	-0.77
変化の傾き	0.78

4.2.2 結果と考察

相関分析の結果を表 4.1 に示す。表 4.1 より、聴取印象「はきはき」は F_0 の平均値（相関係数 $r = 0.69$ ）、最大値（ $r = 0.82$ ）、 F_0 の傾き（ $r = 0.71, -0.77, 0.78$ ）と相関があることが示された。 F_0 の傾きの大きさに伴い、 F_0 の最小値、最大値は影響を受ける。そのため、平均値、最大値に相関が見られたのは、 F_0 の変化傾きが相関していることによる間接的な相関だと考えられる。「はきはき」と F_0 の変化の傾きの関係を図 4.2 に示す。

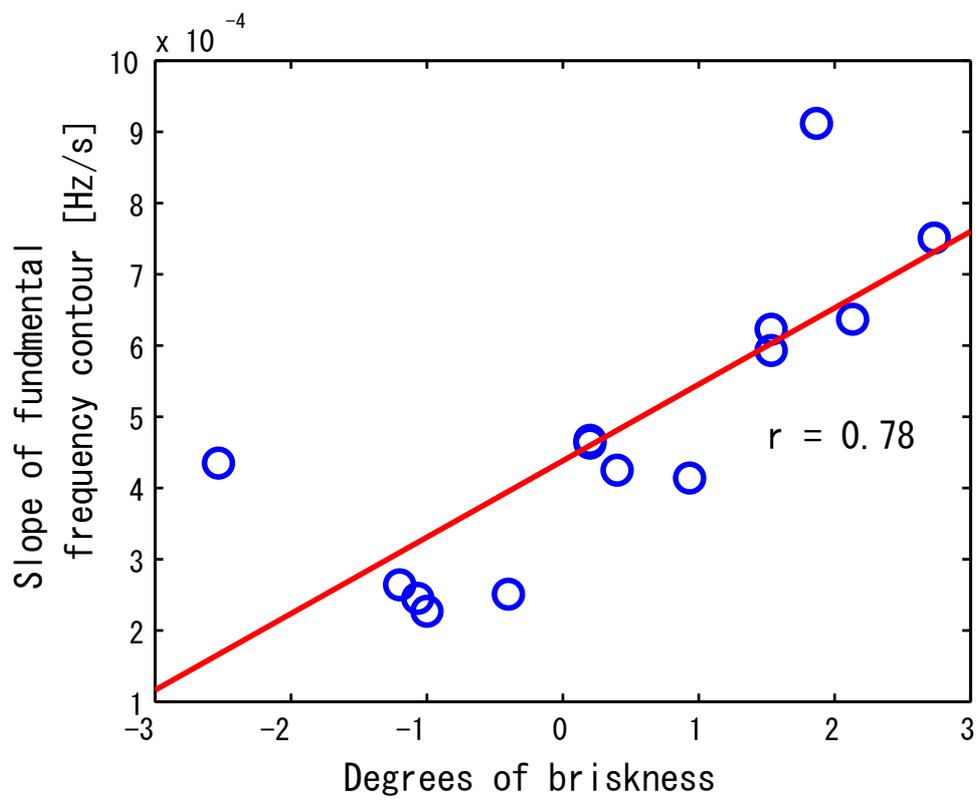


図 4.2: 聴取印象「はきはき」と F_0 の変化の傾きの関係

表 4.2: 「はきはき」とパワーの関係

平均値	-0.38
変動幅	0.14
上昇の傾き	0.18
下降の傾き	-0.06

4.3 パワーに関する特徴量

4.3.1 抽出した特徴量

抑揚に対応して、パワーも変化している可能性があるため、パワーに関する特徴量を抽出した。以下に、抽出した音響特徴量の詳細について記述する。同様に、下線を引かれている音響特徴量は時間的に変動する特徴量を表していることから、動的成分として定義する。

- パワーに関する音響特徴量
 - 平均値：文章全体及び各フレーズにおけるパワーの平均値
 - 変動幅：文章全体におけるパワーの変動幅
 - 上昇の傾き：句頭におけるパワーの上昇の傾き
 - 下降の傾き：句末におけるパワーの下降の傾き

4.3.2 結果と考察

相関分析の結果を表 4.2 に示す。表 4.2 より、聴取印象「はきはき」はパワーに関する特徴量のいずれとも相関がないことが示された。感情音声における分析結果 [20, 21] においては、 F_0 とパワーは相関するという結果が得られている。しかし、今回用いた音声は感情や意図を含まない Neutral な音声のみであったため、相関がなかったと考えられる。

4.4 スペクトルに関する特徴量

4.4.1 抽出した特徴量

抑揚に対応して、スペクトルも変化している可能性があるため、スペクトルに関する特徴量を抽出した。また、スペクトルからさらに、フォルマント周波数 ($F_1 \sim F_3$) とスペクトルの傾斜を抽出した。以下に、抽出した音響特徴量の詳細について記述する。同様に、下線を引かれている音響特徴量は時間的に変動する特徴量を表していることから、動的成分として定義する。

- スペクトルの時間変化に関する音響特徴量
 - 変化量：各音素間でのスペクトル変化量の平均
 - 変化の傾き：各音素間でのスペクトル変化の傾きの平均
- フォルマントに関する音響特徴量
 - 平均値：文章全体におけるフォルマント周波数の平均
 - 変化量：各音素間でのフォルマント周波数の変化量
- スペクトルの傾斜に関する音響特徴量
 - 平均値：文章全体におけるスペクトルの傾きの平均
 - 最大値：文章全体におけるスペクトルの傾きの最大値
 - 最小値：文章全体におけるスペクトルの傾きの最小値
 - 変動幅：文章全体におけるスペクトルの傾きの変動幅

スペクトルの変化量は隣接する音素間のスペクトル距離を表す。各音素における音素中心点の STRAIGHT スペクトルから求めた、STRAIGHT ケプストラム係数（30 次）を用いて、以下の式から音素間の距離を算出した。

$$DSP = \sqrt{\sum_{n=1}^{30} (C_n^{(t1)} - C_n^{(t2)})^2} \quad (4.1)$$

ここで、DSP はスペクトルの変化量、 t_1, t_2 は連続する 2 つの音素の時間的位置、 N は音素数、 C は STRAIGHT ケプストラム係数を表している。

スペクトルの変化速度は隣接する音素間のスペクトル変化の傾きを表し、以下の式から求めた。

$$SSP = \frac{\sqrt{\sum_{n=1}^{30} (C_n^{(t1)} - C_n^{(t2)})^2}}{t_2 - t_1} \quad (4.2)$$

ここで、DSP はスペクトルの変化の傾きを表している。

4.4.2 結果と考察

相関分析の結果を表 4.3 に示す。表 4.3 より、聴取印象「はきはき」はスペクトルの変化量 ($r = 0.66$) と変化の傾き ($r = 0.61$)、スペクトルの傾斜の最大値 ($r = 0.78$) と変動幅 ($r = 0.82$) と相関があることが示された。スペクトルの変化量と変化の傾きに相関が認められたのは、スペクトルの傾斜が相関があることによる間接的な相関であることが考えられる。また、スペクトルの傾斜の変動幅に伴い、スペクトルの傾斜の平均値、最大

表 4.3: 「はきはき」とスペクトルの傾斜との関係

音響特徴量	相関係数
スペクトルの時間変化	
変化量	0.66
変化の傾き	0.61
フォルマント	
平均値 (F1)	0.11
変化量 (F1)	0.49
平均値 (F2)	-0.15
変化量 (F2)	0.39
平均値 (F3)	0.24
変化量 (F3)	0.48
スペクトルの傾斜	
平均値	0.58
最大値	0.78
最小値	-0.29
変動幅	0.82

値は影響を受ける。そのため、スペクトルの傾斜の最大値に相関が見られたのは、変動幅が相関していることによる間接的な相関だと考えられる。スペクトルの傾斜の平均値は、やや相関がみられた ($r = 0.58$)。

Fant [34] の音源-フィルタ理論に基づけば、スペクトルの傾斜は、声帯の振動特性、口唇からの放射特性と対応している。「はきはき」とスペクトルの傾斜の変動幅が相関があったことから、 F_0 の変化に対応して、声帯の振動特性が変化し、さらに、口唇の開閉を大きくしていることに放射特性が変化が大きくなっていると考えられる。「はきはき」とスペクトルの傾斜の変動幅の関係を図 4.3 に示す。

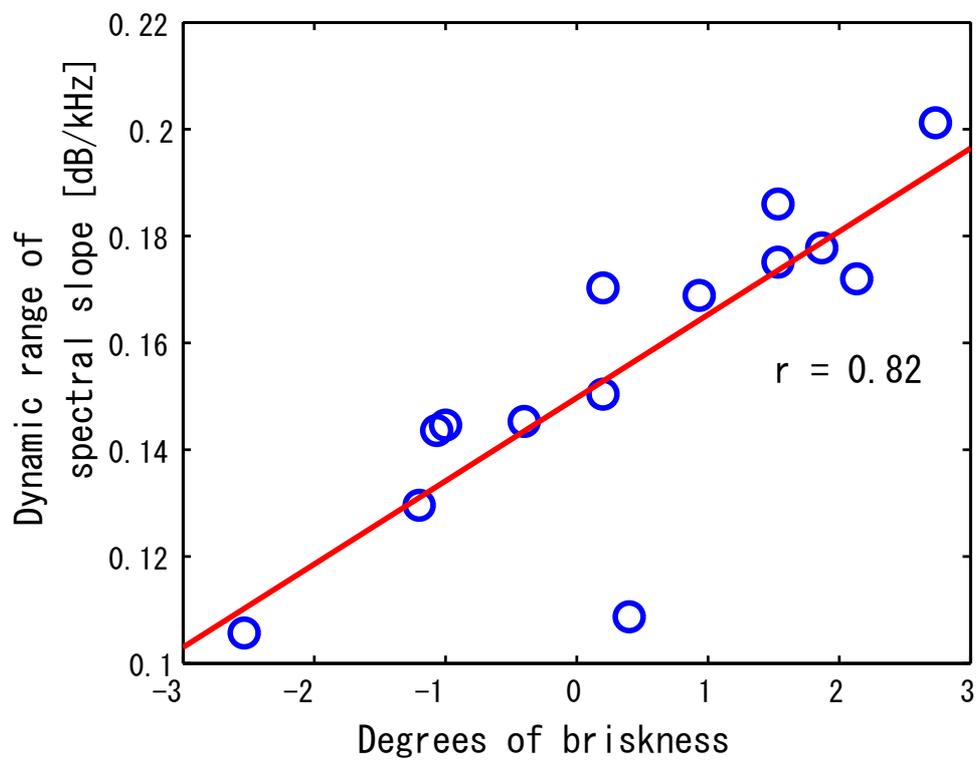


図 4.3: 聴取印象「はきはき」とスペクトルの傾斜の変動幅の関係

表 4.4: 「はきはき」と時間長の関係

音響特徴量	相関係数
時間長	0.05
子音長	0.15

4.5 時間長に関する特徴量

4.5.1 抽出した特徴量

抑揚に対応して、時間長も影響を受けている可能性があるため、以下に示す特徴量を抽出した。

- 時間長に関する音響特徴量
 - － 時間長:文章全体の時間長
 - － 子音長:子音の区間長

4.5.2 結果と考察

相関分析の結果を表 4.4 に示す。表 4.4 より、聴取印象「はきはき」はパワーに関する特徴量のいずれとも相関がないことが示された。先行研究 [35] において、抑揚の変化に伴う時間長の変化が報告されている。しかし、パワーに相関がなかった時と同様、今回用いた音声セットにより、相関が得られなかったものと考えられる。

4.6 まとめ

聴取印象「はきはき」(第二層)に対応する音響特徴量(第三層)の分析を行った結果、三層構造モデルを用いることにより、 F_0 の変化の傾きとスペクトルの傾斜の変動幅の 2 つの音響特徴量が関連しあっており、これらは「はきはき」と相関があることが明らかとなった。これらは、動的特徴である。よって、「はきはき」は動的特徴に関連する聴取印象であることが明らかとなった。

第5章 モデルの検証

5.1 まえがき

前章までの分析により、「はきはき」が個人性聴取印象の主要な因子であり、 F_0 の変化の傾きとスペクトルの傾斜の変動幅と対応していることから、動的成分に対応する聴取印象であることが確認できた。そこで、三層構造モデルの検証をボトムアップ的に行うことで、これらの動的成分の変化が個人性知覚に与える影響を調査した。検証を行う上で、4種類の刺激音を作成した。まず、検証実験 I では、第三層の音響特徴量の変化が第二層の聴取印象に与える影響を検証した。 F_0 の変化の傾きとスペクトルの傾斜の変動幅を変化させることで、「はきはき」の強度が変化することを確認する。次に、検証実験 II では、第三層の音響特徴量の変化が第一層の個人性知覚に与える影響を検証した。

5.2 第三層が第二層に与える影響の検証（検証実験 I）

第三層の音響特徴量の変化が第二層の聴取印象に与える影響を検証するために、 F_0 の変化の傾きとスペクトルの傾斜の変動幅を制御した合成音を作成した。作成した合成音を用いて、 F_0 の変化の傾きとスペクトルの傾斜の変動幅を制御により、「はきはき」の強度が変化することを検証する。

5.2.1 刺激音

刺激音は 14 名の話者のうち 3 名 (108, 419, 702) の音声データを用い、STRAIGHT [31] により作成した。3 名の話者は、小節 3.4.1 における SD 法による聴取実験において、それぞれ「はきはき」している (話者 702, 1.5)、していない (話者 108, -1.2)、どちらでもない (話者 419, 0.2)、と評価された話者である (括弧の中の値は「はきはき」の強度)。刺激音は、それぞれの話者において、

刺激音 1 STRAIGHT 分析合成音 (図 5.1) ,

刺激音 2 X F_0 の変化の傾きを「はきはき」の強度が X になるように制御した合成音 (図 5.2) ,

刺激音 3 X スペクトルの傾斜の変動幅を制御した合成音 (図 5.3) ,

刺激音 4 X F_0 の変化の傾きとスペクトルの傾斜の変動幅の両方を制御した合成音（図 5.4）,

の 4 種類を作成した。

刺激音 2-4 の制御は、それぞれ「はきはき」の強度が -2, 0, 2 となるように制御した。また、制御による音質の劣化が実験に悪影響を及ぼす危険性を回避するため、SN 比 25 dB のピンクノイズをそれぞれの刺激音に付与した。振幅は最大値で正規化した。

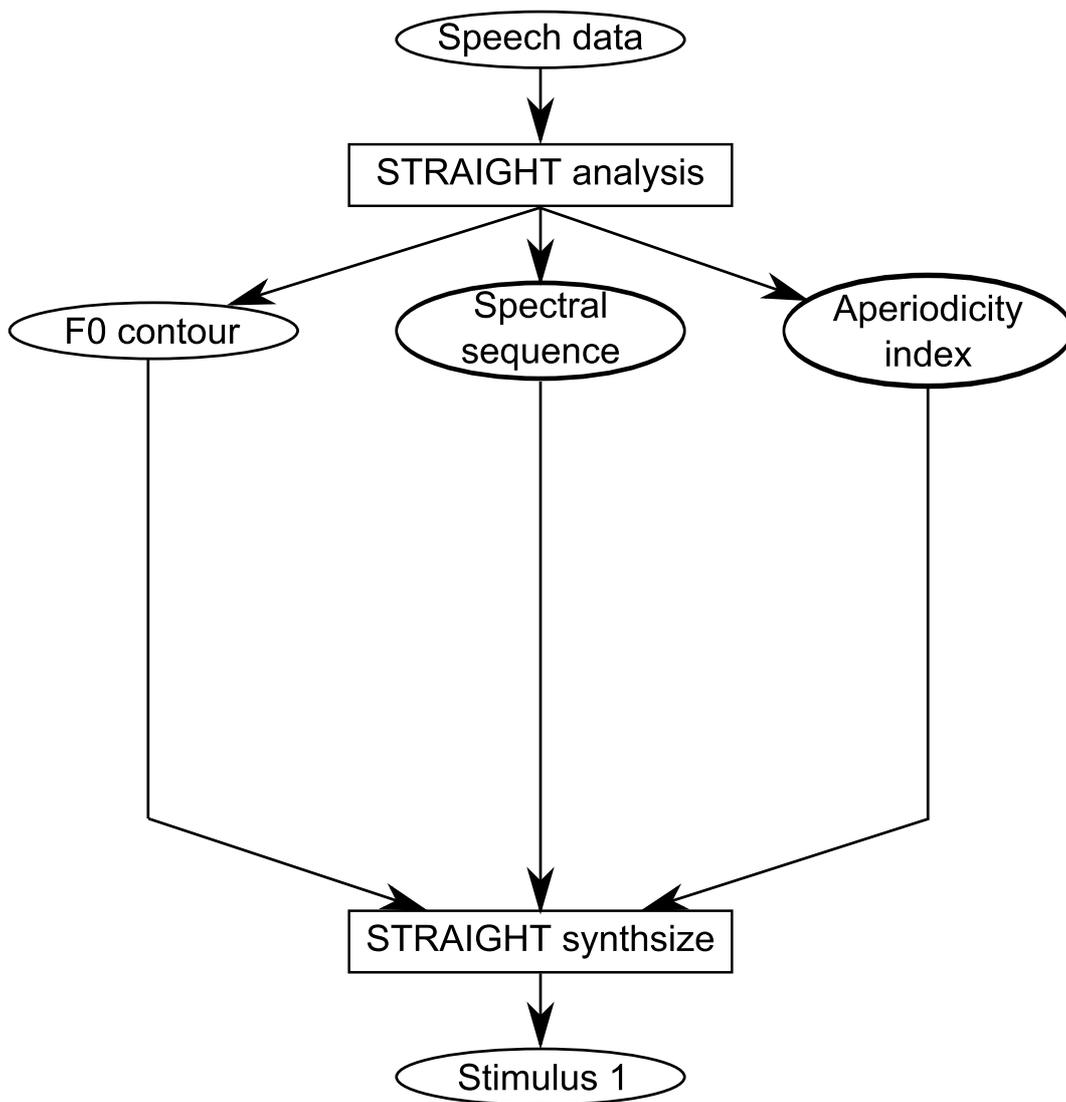


図 5.1: 刺激音 1 の合成手順

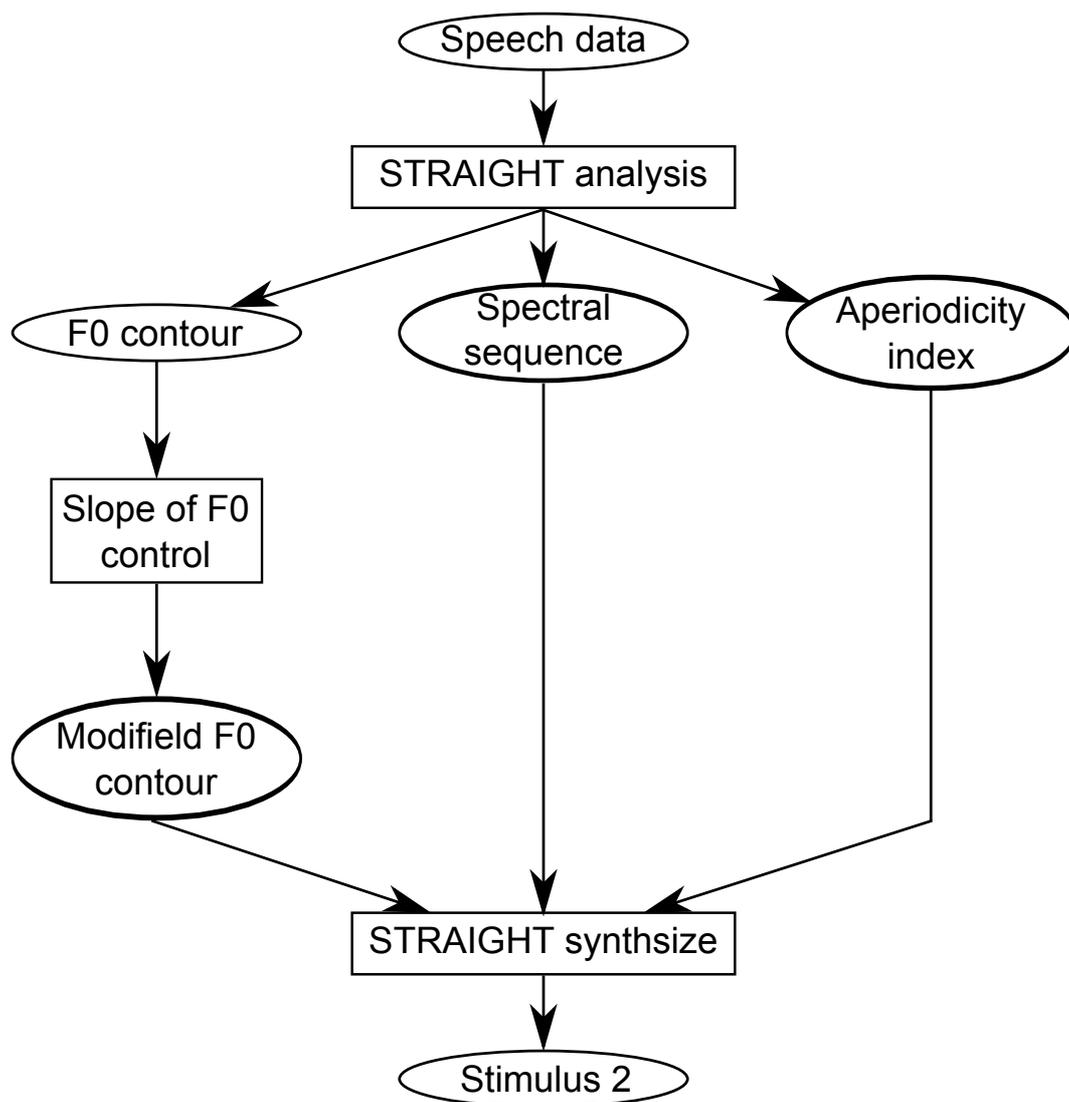


図 5.2: 刺激音 2 の合成手順

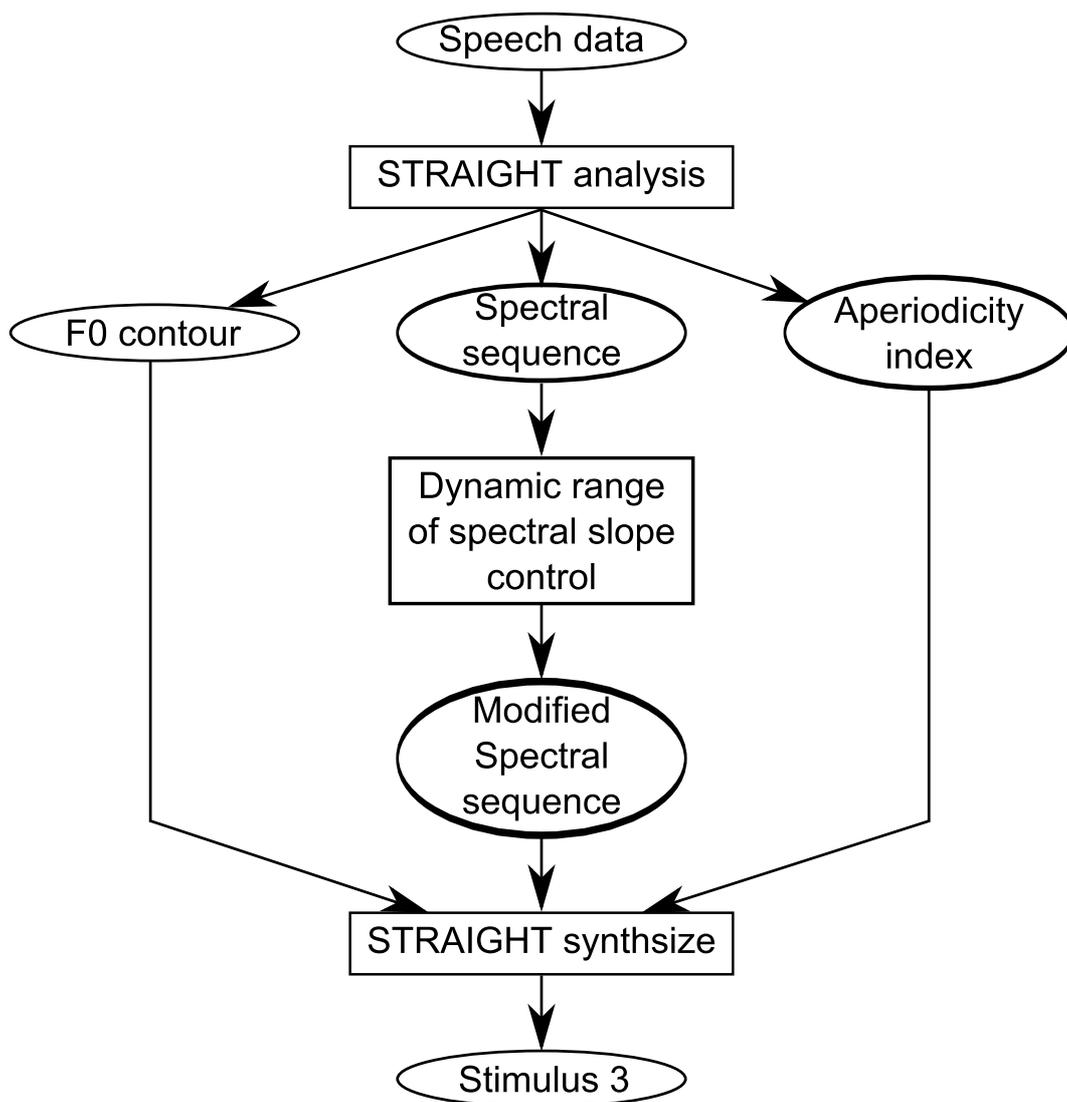


図 5.3: 刺激音 3 の合成手順

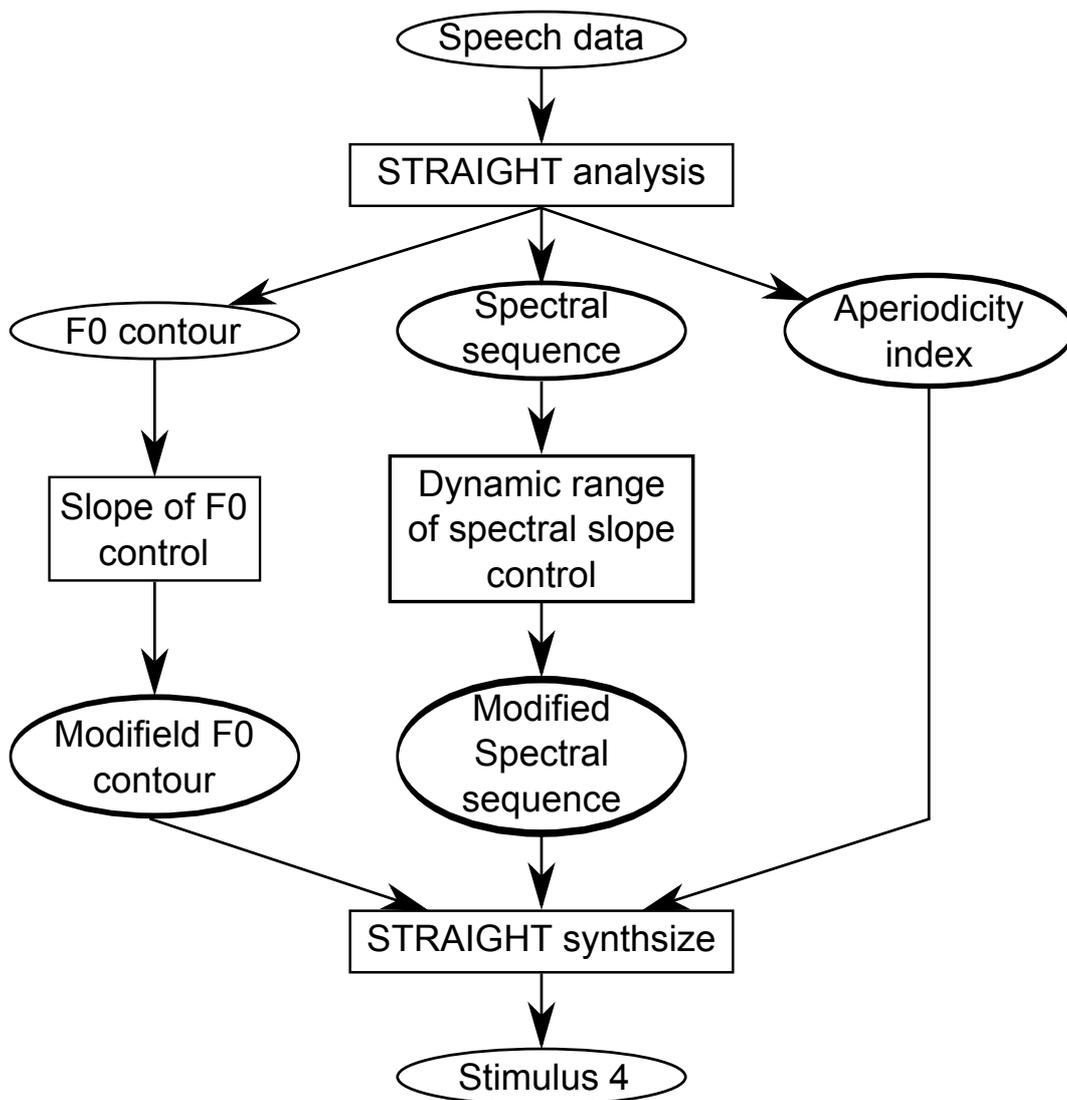


図 5.4: 刺激音 4 の合成手順

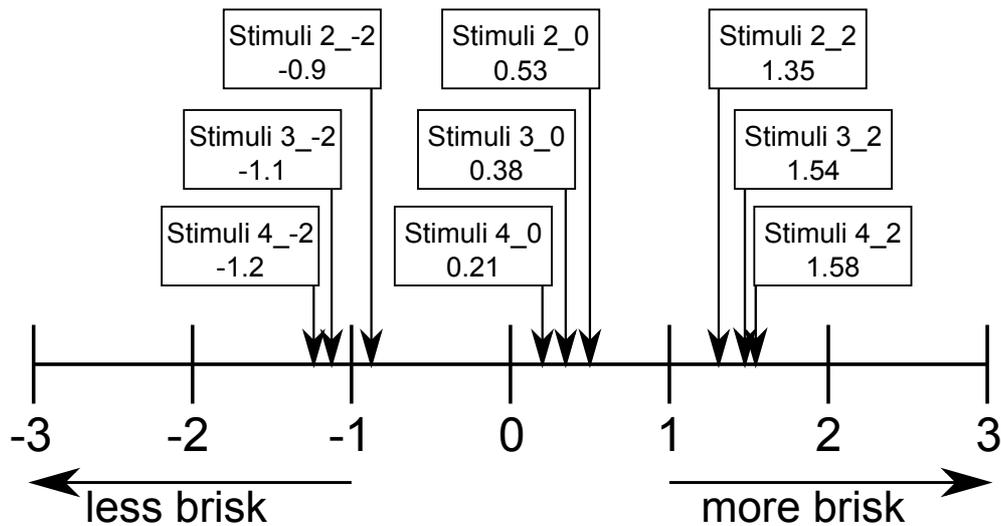


図 5.5: 検証実験 I の結果

5.2.2 実験協力者

実験協力者は聴力に異常のない 22~28 歳の成人男性 8 名である。これらの実験協力者は、刺激音の話者の個人性を記憶していない。つまり、実験協力者にとって、刺激音の話者は未知話者である。

5.2.3 実験手順

SD 法により実験を行った。作成した刺激音 1-4 を 8 名の実験協力者に呈示し、「はきはき」について評価してもらった。各刺激音は 3 回ずつ評価を行い、全試行回数は 90 (= 3 × 10 × 3) 回であった。その他の実験方法は小節 3.4.1 での SD 法による実験と同様である。

5.3 実験結果と考察

刺激音 2-4 における実験結果を図 5.5 に示す。値は被験者間および話者間で平均化している。図 5.5 から、 F_0 の変化の傾き、スペクトルの傾斜の変動幅を制御することで、「はきはき」の強度の制御が可能であることが示された。また、 F_0 の変化の傾き、スペクトルの傾斜の変動幅をそれぞれ制御（刺激音 2,3）するよりも、両方を制御（刺激音 4）するほうが意図した強度に近づけられることが確認できた。これは、 F_0 の変化の傾きとスペクトルの傾斜の変動幅が関連しているからだと考えられる。



図 5.7: 実験中のモニターの様子

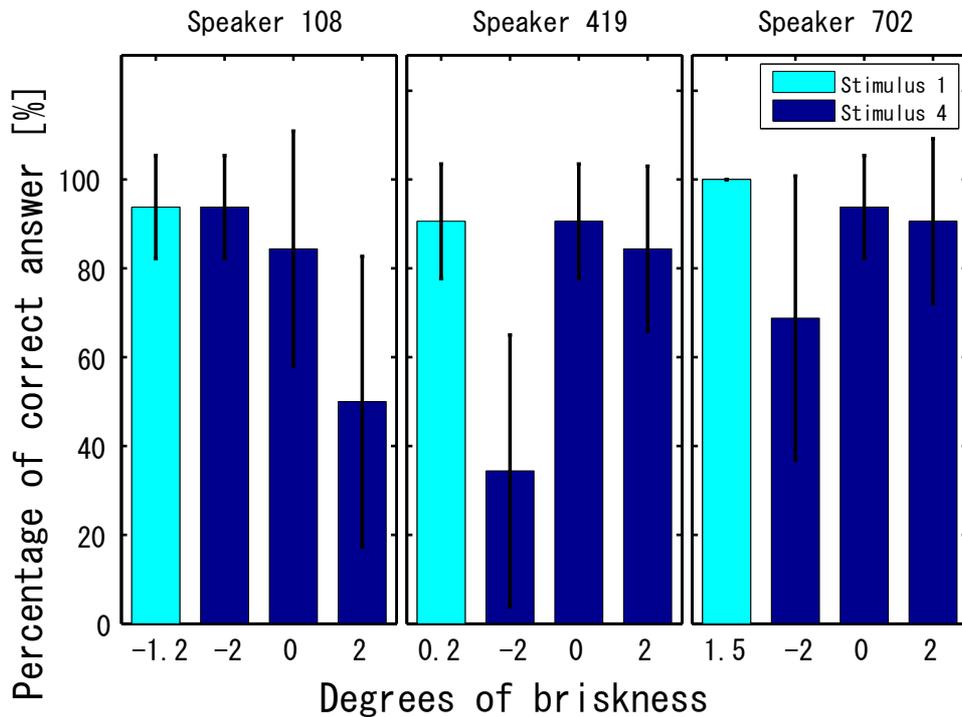


図 5.8: 検証実験 II の結果

5.5 実験結果と考察

実験結果を図 5.8 に示す．値は被験者間で平均化している．図 5.8 から，いずれの話者においても，刺激音 A が刺激音 1 のときは正確に話者を判断できていることが示された．また，刺激音 A が，話者 108 においては，刺激音 4.2，話者 419，702 においては，刺激音 4-2 のとき，正答率の減少が確認された．これらの刺激音は，元の音声（刺激音 1）から「はきはき」の強度が大きく離れている音声である．刺激音 1 と 4 の間での「はきはき」の強度の距離と正答率との関係を図 5.9 に示す．図 5.9 から強度の距離がと正答率に相関があること（相関係数： $r = 0.74$ ）が確認できた．このことから「はきはき」の強度を大きく変化させると，異なる話者として知覚されることが示された．つまり，音声中の動的成分は個人性知覚に影響を与えていることが明らかとなった．

刺激音 4 は， F_0 の変化の傾きとスペクトルの傾斜の変動幅を制御している． F_0 の変化の傾きを制御することで， F_0 の平均値も変化している． F_0 の平均値が個人性知覚に影響を与えることは先行研究 [5, 8] によって示されている．本研究での結果はこれらを支持するものである．また， F_0 の変化パターンが個人性知覚に影響を与えることは既に示されているが [11]，本研究において，パターンのうち，変化の傾きが重要であることが示唆された．さらに，スペクトルの傾斜が個人性知覚に影響を与えることは北村ら [8] によって示されているが，本研究では，スペクトルの傾斜の変動幅も重要であることを示した．

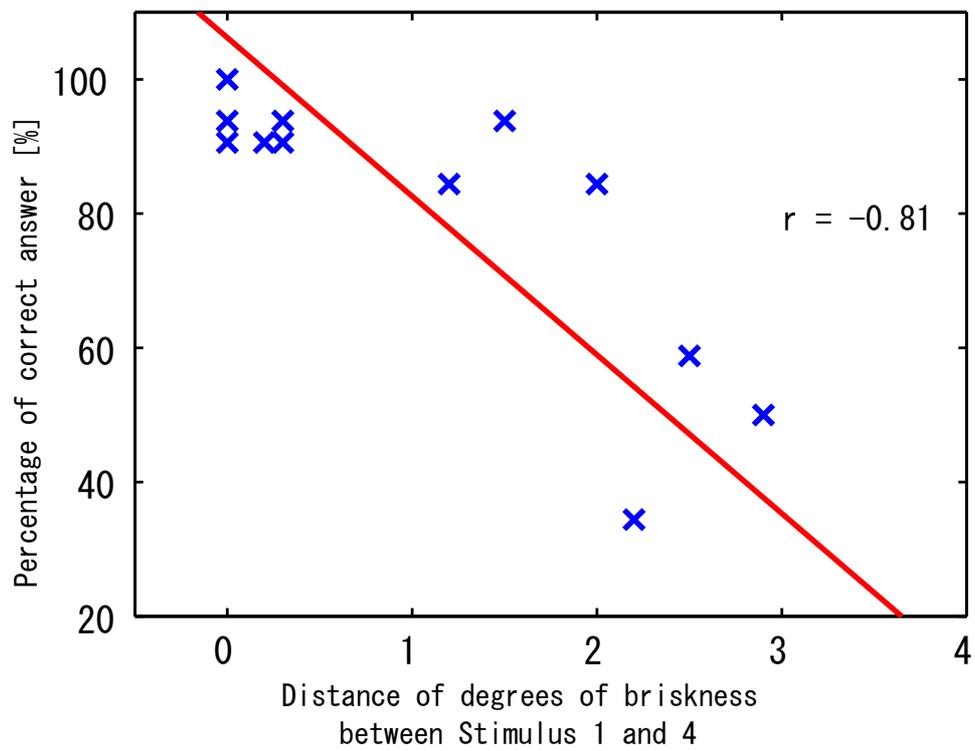


図 5.9: 刺激音 1 と 4 の間での「はきはき」の強度の差と正答率の関係

5.6 まとめ

第三層の音響特徴量が第一層の個人性知覚に与える影響をボトムアップ的に検証した結果、「はきはき」に関わる音響特徴量の変化が個人性知覚に影響を与えた。よって、音声中の動的成分は個人性知覚に影響を与えていることが明らかとなった。また、聴取印象の強度が個人性知覚に影響していることが示唆された。

第6章 全体考察

本研究は、音声中の動的成分が個人性知覚に与える影響を明らかにすることを目的とした。そのために、聴取印象に着目し、三層構造モデルを用いた。個人性知覚（第一層）と聴取印象（第二層）、および音響特徴量（第三層）の関係を三層構造モデルによりトップダウン的に調査することで、個人性知覚と動的成分の関係を分析した。さらに、分析により構築されたモデルをボトムアップ的に検証した。

個人性知覚（第一層）と聴取印象（第二層）の関係は、まず、話者間の類似性評定実験に基づいてMDSにより話者の知覚的距離空間を作成した。そして、次に、SD法により求めた話者の聴取印象の強度とその空間との対応を重相関分析により調査することで個人性聴取印象を求めた。次に、聴取印象（第二層）と音響特徴量（第三層）の関係は、SD法の結果と抽出した音響特徴量との間で相関分析を行うことで求めた。その結果、「はきはき」が個人性聴取印象の主要な因子であり、 F_0 の変化の傾き、スペクトルの傾斜の変動幅といった動的成分と相関があることが明らかになった。このことから、「はきはき」は動的成分に関わる個人性聴取印象であることが確認できた。

モデルの検証は、 F_0 の変化の傾き、スペクトルの傾斜の変動幅を変化させた合成音声を用いた聴取実験により行った。まず、第三層の音響特徴量の変化が第二層の聴取印象に与える影響を確認し、さらに、第一層の個人性知覚に与える影響を検証した。その結果、「はきはき」に関わる音響特徴量の変化が個人性知覚に影響を与えた。よって、音声中の動的成分は個人性知覚に影響を与えていることが明らかとなった。また、聴取印象の強度が個人性知覚に影響していることが示唆された。

「はきはき」は動的成分に関連する聴取印象であり、発話器官の形状に依存しない、後天的な特徴である。はきはきした声は、 F_0 の変化の傾き、スペクトルの傾斜の変動幅と相関があったことから、抑揚をつけて、口唇の開閉を十分に行う発話様式であると考えられる。このような特徴が個人性知覚に影響を与えること検証により明らかにした例はない。

また、聴取印象の強度を段階的に制御した合成音声を用いた聴取実験により、聴取印象の強度の変化が個人性知覚に影響していることが示唆された。Izumidaら[36]は、話者間で音響特徴量を入れ替えて音声をを用いた、実験協力者による話者識別実験を行っている。その結果、話者の組合せにより、話者の識別率に大きな違いがあることを示している。Izumidaら[36]の研究では、話者間で音響特徴量を入れ替えた際に、もとの話者から聴取印象が変化したのか検討していない。個人性知覚に関連する従来研究の多くは、個人性知覚要因を特定するために、話者間で特徴量を入れ替えた合成音声を用いているが、本研究により、その合成音声がもとの話者から聴取印象がどのくらい変化したのか検討する

必要性を示した .

第7章 結論

7.1 本研究で明らかとなったことの要約

本研究は、音声の中の動的成分が個人性知覚に与える影響を明らかにすることを目的とした。そのために、聴取印象に着目し、三層構造モデルを用いた。個人性知覚（第一層）と聴取印象（第二層）、および音響特徴量（第三層）の関係を三層構造モデルによりトップダウン的に調査することで、個人性知覚と動的成分の関係を分析した。さらに、分析により構築されたモデルをボトムアップ的に検証した。その結果、以下のことを明らかにした。

- 聴取印象「はきはき」個人性聴取印象の主要な要因の一つである。
- 聴取印象「はきはき」は、 F_0 の変化の傾きとスペクトルの傾斜の変動幅と相関のある、動的成分に関連する聴取印象である。
- 聴取印象「はきはき」は、抑揚をつけて、口唇の開閉を十分に行う発話様式である。
- F_0 の変化の傾きとスペクトルの傾斜の変動幅を制御することで、個人性知覚が変化する。
- 音声の中の動的成分は個人性知覚に影響を与えている。
- 聴取印象の強度が個人性知覚に影響している可能性がある。

7.2 残された課題

個人性知覚要因の全容を解明する上で、残された課題を以下に記す。

音質の改善

今回、STRAIGHT 分析合成系を用いて、音響特徴量を制御し合成音声を作成した。その際、音質が劣化してしまう問題が残っている。そこで、Nguyen ら [37] が提案している、Temporal Decomposition (TD) と Gaussian Mixture Model (GMM) を用いた手法で合成することで、音質の改善が行えると考えられる。

ファジィ推論システムの導入

今回、回帰分析によりモデルの構築している。一方で、Huangら [20, 21] は感情知覚多層モデルにおいて、ヒトの知覚特性に則したモデルの構築を目指して、ファジィ推論システム (Fuzzy Inference System: FIS) を導入している。個人性知覚における三層構造モデルにおいても、ヒトの知覚特性を則するため、FIS による構築の必要がある。

第4フォルマント周波数の分析

本研究では、文章音声を用いたため、第4フォルマント周波数 (F4) 以上の共振周波数に関しては抽出が困難であり、検討していない。北村ら [38] は単母音を用いた研究において、 F_0 の変化に伴い、口腔や咽頭腔、口唇の体積、声道長が変化していることを報告している。咽頭腔の体積や声道長は F4 との関連が指摘されているので [39, 40]、F4 の分析も今後必要である。

非周期成分の分析

今回、非周期成分 (AP) については分析、制御を行っていない。「はきはき」は声帯の振動特性と関連のあるスペクトルの傾斜の変動幅と相関があった。このことから、今後 AP を考慮して分析する必要がある。

話者空間を構成する聴取印象のリストの解明

動的成分に絞って研究を行ったため、話者空間を構成する聴取印象の分析は行っていない。今後、三層構造モデルをもとに個人性の研究を行う上で、話者空間を構成する聴取印象のリストを決定する必要がある。さらには、聴取印象間の関係の分析も必要である。

未知話者と既知話者での違い

本研究での聴取実験は、音声データが未知話者となるような実験協力者により行われた。本研究で得られた知見が、聴取者にとって個人性を熟知した話者 (既知話者) でも同様な結果が得られるのか、または、用いる聴取印象や音響特徴量が変化するのか、さらには、それらに対する分解能が変化するのか、といった検討も今後必要である。

データ数

今回、検証実験 II では、3 名の話者で行ったため、音声データに依存している可能性は否定できない。そこで、多数話者を用いた聴取実験を行うことで、本研究で得られた知見が一般性があるのか検討する必要がある。

本研究で得られた知見は、個人性知覚要因の一端に過ぎないが、以上、一連の課題を遂行することで、個人性知覚要因の全容の解明、更には、個人性知覚メカニズムの解明・音声知覚の解明に繋がるものである。これらの研究を遂行することは、個人性知覚メカニズムの解明、さらには音声科学の発展に繋がるものである。そして、本研究で得られた知

見や本研究で用いた手法は，“ヒトはどのように話者を判断しているのか？” “ヒトはどのように音声から言語情報を得ているのか？” という大きな課題の解明に貢献することを願う。

謝辞

本研究を進めるにあたり，多大なる御指導ならびに御鞭撻を賜りました赤木 正人 教授に深く感謝致します。

本研究を進めるにあたり，日頃から熱心な御指導ならびに御鞭撻を賜りました鷓木 祐史 准教授に心より感謝致します。

本研究を進めるにあたり，日頃から熱心に御討論頂き，また御助言を賜りました宮内 良太 助教に心より感謝致します。

本研究を進めるにあたり，熱心に御討論頂き，また御助言を賜りました党 建武 教授，末光 厚夫 助教，川本 真一 助教に心より感謝致します。

筆者が本学で研究する機会を支援して頂き，ならびに甲南大学在学中から今日に至るまで多大なる御指導と御助言を賜りました甲南大学 知能情報学部 北村 達也 准教授に深く感謝致します。

本研究を進めるにあたり，日頃から熱心な議論と激励をいただき，さらには，公私にわたり本学での生活を支えてくださった，博士後期課程の木谷 俊介氏，濱田 康弘氏，森田 翔太氏，久保 理恵子氏に心より感謝致します。

また，本研究を進めるにあたり，日頃から熱心な議論と激励をいただきました，音情報処理分野の諸先輩方，及び諸氏に熱く御礼申し上げます。

本研究における聴取実験のために，貴重な時間を割いて頂きました実験協力者の方々に感謝の意を表します。

最後に，本学での研究生活を支え，温かく見守ってくれた両親に心から感謝致します。

参考文献

- [1] 古井貞熙: 人と対話するコンピュータを創っています—音声認識の最前線—, 角川学芸出版, 2009.
- [2] 加藤 和美, 筧 一彦, “音声における話者への適応性の検討,” 日本音響学会誌, Vol. 44, No. 2, pp. 180–186, 1988.
- [3] 北村 達也, 出水田 剛志, 橋 亮介, “声の類似性から個人性知覚を探索,” 日本音響学会 2011 年秋季研究発表会講演論文集, pp. 253–256, 2011.
- [4] 伊藤 憲三, 斉藤 収三, “音声の音響的特徴パラメータが個人性の知覚に及ぼす影響,” 信号処理学会論文誌, J65-A, pp. 101–108, 1982.
- [5] 橋本 誠, 北川 敏, 樋口 宜男, “音声の個人性知覚に影響を及ぼす音響的特徴の定量的分析,” 日本音響学会誌, Vol. 54, No. 3, pp. 169–178, 1998.
- [6] T. Kitamura, M. Akagi, Speaker, “individualities in speech spectral envelopes,” *J. Acoust. Soc. Jpn. (E)*, Vol. 16, No. 5, pp. 283–289, 1995.
- [7] 北村 達也, 赤木 正人, “単母音の話者識別に寄与するスペクトル包絡成分,” 日本音響学会誌, Vol. 53, No. 3, pp. 185–191, 1997.
- [8] T. Kitamura and T. Saitou, “Effects of acoustic modification on perception of speaker characteristics for sustained vowels,” *Acoustical Science and Technology*, Vol. 28, No. 6, pp. 434–437, 2007.
- [9] T. Kitamura, K. Honda and H. Takemoto, “Individual variation of the hypopharyngeal cavities and its acoustic effects,” *Acoustical Science and Technology*, Vol. 26, No. 1, pp. 16–26, 2005.
- [10] K. Amino, T. Sugawara, T. Arai, “Idiosyncrasy of nasal sounds in human speaker identification and their acoustic properties,” *Acoustical Science and Technology*, Vol. 27, No. 4, pp. 233–235, 2006.
- [11] M. Akagi and T. Ienaga, “Speaker individuality in fundamental frequency contours and its control,” *J. Acoust. Soc. Jpn. (E)*, Vol. 18, No. 2, pp. 73–80, 1997.

- [12] 大野 弘, 赤木 正人, “文音声中の基本周波数変化パターンに含まれる個人性の検討,” 電子情報通信学会技術研究報告, 音声, Vol. 97, No. 586, pp. 89–96, 1998.
- [13] B. Yegnanarayana, S. P. Wagh and S. Rajendran, “A speaker verification system using prosodic features,” Proc. of ICSLP 94, 1994.
- [14] 北村達也, 赤木正人, 北澤茂良, “スペクトル遷移パターンが個人性知覚に与える影響について,” 日本音響学会聴覚研究会資料, H-98-97, pp. 1–8, 1998.
- [15] 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,” 電子情報通信学会技術研究報告, 音声, Vol. 98, No. 639, pp. 39–46, 1999.
- [16] H. Kasuya, W. Zhu, M. Matsuda and C. Yang, “Voice quality conversion based on an ARX speech analysis-synthesis method and its application to the study of speaker individuality,” J. Acoust. Soc. Am., Vol.100, No. 4, pp. 2600, 1996.
- [17] S. Furui, “Research on individuality features in speech waves and automatic speaker recognition techniques,” Speech Communication Vol. 5, No. 2, pp. 183–197 (1986).
- [18] 木戸 博, 箕輪 有希子, 粕谷 英樹, “声質表現語の音響関連量に関する非線形分析:決定木による方法,” 日本音響学会誌, Vol. 58, No. 9, pp. 586–588, 2002.
- [19] 山下 泰樹, 松本 弘, “成人の読上げ音声における声質評価値と音響関連量の分析,” 日本音響学会誌, Vol. 62, No. 12, pp. 856–864, 2006.
- [20] C-F. Huang and M. Akagi, “A Mukti-Layer fuzzy logical model for emotional speech perception,” Proc. EuroSpeech 2005, pp. 417–420, Lisbon, Portugal 2005.
- [21] C-F Huang and M. Akagi, “A three-layerd model for expressive speech perception,” Speech Commum, 50, pp. 810–828 2008.
- [22] 齋藤 毅, 辻 直也, 鶴木 祐史, 赤木 正人, “歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析,” 日本音響学会誌, Vol. 64, No. 5, pp. 267–277, 2008.
- [23] M. Akagi, “Analysis of production and perception characteristics of non-linguistic information in speech and its application to inter-language communications,” Proc. APSIPA 2009, Sapporo, pp. 513–519, 2009.
- [24] 赤木 正人, “音声に含まれる感情情報の認識—感情空間をどのように表現するか—,” 日本音響学会誌, Vol. 66, No. 8, pp. 393–398, 2010.
- [25] 粕谷 英樹, 楊 長盛, “音源から見た声質,” 日本音響学会誌, Vol. 51, Vol11, pp. 869–875, 1995.

- [26] 阿部 匡伸, “発話様式のバリエーション,” 日本音響学会誌, Vol. 51, No. 11, pp. 882–886, 1995.
- [27] 木戸 博, 粕谷 英樹, “通常発話の声質に関連した日常表現語の抽出,” 日本音響学会誌, Vol. 55, No. 6, pp. 405–411, 1999.
- [28] 木戸 博, 粕谷 英樹, “通常発話の声質に関連した日常表現語–聴取評価による抽出–,” 日本音響学会誌, Vol. 57, No. 5, pp. 337–344, 2001.
- [29] 阿部 匡伸, 匂坂 芳典, 梅田 哲夫, 桑原 尚夫, “研究用日本音声データベース利用解説書,” Tech. Rep. ATR, TR-I-0166, 1990.
- [30] 河原 英紀, “聴覚の情景分析が生み出した高品質 VOCODER: STRAIGHT,” 日本音響学会誌, Vol. 54, No. 7, pp. 521–526, 1998.
- [31] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigne, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based on F0 extraction: Possible role of a repetitive structure in sounds,” *Speech Commun.*, Vol. 27, pp. 187–207, 1999.
- [32] 大串 健吾, 中山 剛, 福田 忠彦: 画質と音質の評価技術, 昭晃堂, 1991.
- [33] WaveSurfer, <http://www.speech.kth.se/wavesufer/>
- [34] G. Fant: Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations, Mouton, 1970.
- [35] 前川 喜久雄, 北川 智利, “音声はパラ言語情報をいかに伝えるか,” 認知科学, Vol. 9, No. 1, pp. 46–66, 2002.
- [36] T. Izumida, T. Kitamura, “Study of perceptual factors for speaker identification focusing on perceptual similarity of speaker characteristics,” *Acoustical Science and Technology*, Vol. 32, No. 5, pp. 216–219, 2011.
- [37] B. P. Nguyen and M. Akagi, “A flexible spectral modification method based on temporal decomposition and Gaussian mixture model,” *Acoustical Science and Technology*, Vol. 30, No. 3, pp. 170–179, 2009.
- [38] 北村 達也, パーハム モクタリ, “ F_0 変化に伴う声道形状変化の観測,” 電子情報通信学会技術研究報告, 応用音響, Vol. 104, No. 715, pp. 25–28, 2005.
- [39] 竹本 浩典, 本多 清志, 正木 信夫, 島田 育廣, 藤本 一郎, “3次元MRI動画データに基づく声道下部構造のモデル化,” 日本音響学会秋季研究発表会講演論文集, pp. 281–282, 2003.

- [40] 本多 清志, 北村 達也, 竹本 浩典, 藤田 覚, パーハム モクタリ, “下咽頭腔の共鳴特性,” 日本音響学会秋季研究発表会講演論文集, pp. 235-236, 2004.

本研究に関する研究業績

国際会議

- T. Izumida and M. Akagi, “Study on hearing impression of speaker identification focusing on dynamic features,” *Proc. 2012 RISP International Workshop on Non-linear Circuits, Communications and Signal Processing*, pp. 401–404, Hawaii, USA, March 2012.

研究会

- 出水田 剛志, 赤木 正人, “聴取印象に着目した音声の個人性知覚に関する基礎研究,” *日本音響学会聴覚研究会資料*, Vol. 41, No. 7, pp. 551–554, 富山, October 2011.

口頭発表

- 出水田 剛志, 赤木 正人, “音声の個人性に関連する表現語の検討,” *電気関係学会北陸支部連合大会*, G-7, 福井, September 2011.
- 出水田 剛志, 赤木 正人, “音声の動的成分に着目した個人性聴取印象の検討,” *日本音響学会 2012 年春季研究発表会*, 1-R-1, pp. 423–426, 神奈川, March 2012.

その他の研究業績

論文

- T. Izumida and T. Kitamura, “Study of perceptual factors for speaker identification focusing on perceptual similarity of speaker characteristics,” *Acoustical Science and Technology*, Vol. 32, No. 5, pp. 216–219, September 2011.

研究会

- 出水田 剛志, 井上 賢治, 林 亮太郎, “複数のアルゴリズムを用いたマルチカーエレベーター制御,” 電子情報通信学会技術研究報告, Vol. 110, No. 370, pp. 17–20, 山口, January 2011.

口頭発表

- 出水田 剛志, 北村 達也, “個人性の類似性に着目した個人性知覚要因の検討,” 日本音響学会 2010 年春季研究発表会, 1-R-15, pp. 623–624, 東京, March 2010.
- 北村 達也, 出水田 剛志, 橘 亮輔, “声の類似性から個人性知覚を探る,” 日本音響学会 2011 年秋季研究発表会, 1-8-10, pp. 253–256, 島根, September 2011.