

Title	怒りの感情音声における音響特徴量と感情知覚との関係に関する研究
Author(s)	平館, 郁雄
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1535
Rights	
Description	Supervisor: 赤木 正人, 情報科学研究科, 修士

修 士 論 文

怒りの感情音声における音響特徴量と
感情知覚との関係に関する研究

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

平館 郁雄

2002 年 3 月

修 士 論 文

怒りの感情音声における音響特徴量と 感情知覚との関係に関する研究

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 党建武 助教授

審査委員 小谷一孔 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

010095 平館 郁雄

提出年月: 2002 年 2 月

概 要

人間の最も基本的なコミュニケーション手段の一つである音声を、人間とコンピュータとのインターフェースとして用いることを考える場合に、言語情報だけでなく、パラ言語・非言語情報の伝達が行われれば、よりフレンドリなコミュニケーションが実現できると考えられる。その中でも、感情情報は人間らしく聞きやすい合成音声には欠かせない。

過去の感情音声の研究を見ると、全発声区間という大局的な部分における平均的な変化を分析したものばかりで、文章におけるアクセントの有無など局所的な部分を十分に分析したものはないという問題点がある。そこで本研究では、感情音声のアクセント部における音響特徴量と感情知覚との関係を調べることを目的とする。

本研究は「怒り」という感情を、激しい怒りを表す「Hot Anger」と、押し殺した怒りを表す「Cold Anger」とに分け、「Neutral」の音声との比較検討を行った。音響特徴量としては、基本周波数・パワー・持続時間・フォルマント周波数・スペクトルを扱い、アクセントにおける変化に注目して分析を行った。

その結果、各音響特徴量が「Neutral」「Cold Anger」「Hot Anger」の感情間において異なる傾向を示した。そして、分析結果から感情音声の合成音を作成するための規則を導いた。

目次

第1章	序論	1
1.1	はじめに	1
1.2	本研究の背景	2
1.2.1	感情音声を扱った過去の研究	2
1.2.2	過去の研究の問題点	3
1.3	本研究の目的	4
1.4	本論文の構成	4
第2章	過去の研究の分析結果	5
第3章	アクセントについて	7
第4章	感情音声データ	8
4.1	感情音声データ	8
4.2	聴取実験	9
4.2.1	目的	9
4.2.2	聴取実験	10
4.2.3	実験結果と考察	10
第5章	音声データの分析	16
5.1	目的	16
5.2	基本周波数の分析	16
5.2.1	分析手法	16
5.2.2	分析結果・考察	17
5.3	パワー変化率の分析	25
5.3.1	分析手法	25

5.3.2	分析結果・考察	25
5.4	持続時間の分析	31
5.4.1	分析手法	31
5.4.2	分析結果・考察	31
5.5	フォルマント周波数の分析	35
5.5.1	分析手法	35
5.5.2	分析結果・考察	36
5.6	スペクトルの分析	40
5.6.1	分析手法	40
5.6.2	分析結果・考察	41
第 6 章	全体のまとめ・考察	47
第 7 章	結論	52
7.1	本研究の成果	52
7.2	今後の課題	52
第 8 章	謝辞	54

目 次

3.1	日本語のアクセント型	7
4.1	a001 ~ a020 の認識率	11
4.2	b001 ~ b020 の認識率	12
4.3	c001 ~ c020 認識率	12
4.4	d001 ~ d020 認識率	13
4.5	e001 ~ e020 認識率	13
4.6	f001 ~ f020 認識率	14
4.7	g001 ~ g020 認識率	14
4.8	h001 ~ h020 認識率	15
4.9	i001 ~ i020 認識率	15
5.1	各感情における基本周波数の長時間平均	18
5.2	左図：基本周波数長時間平均における「Cold Anger」の「Neutral」に対する比、右図：基本周波数長時間平均における「Hot Anger」の「Neutral」に対する比	19
5.3	各感情におけるアクセントレベル上昇時の基本周波数音韻変化率	19
5.4	左図：基本周波数音韻変化率における「Cold Anger」の「Neutral」に対する比、右図：基本周波数音韻変化率における「Hot Anger」の「Neutral」に対する比	20
5.5	各感情におけるアクセントレベル上昇時の基本周波数時間変化率	21
5.6	左図：基本周波数時間変化率における「Cold Anger」の「Neutral」に対する比、右図：基本周波数時間変化率における「Hot Anger」の「Neutral」に対する比	21
5.7	各感情におけるアクセントレベル下降時の基本周波数音韻変化率	22

5.8	左図：基本周波数音韻変化率における「Cold Anger」の「Neutral」に対する比、右図：基本周波数音韻変化率における「Hot Anger」の「Neutral」に対する比	23
5.9	各感情におけるアクセントレベル下降時の基本周波数時間変化率	23
5.10	左図：基本周波数時間変化率における「Cold Anger」の「Neutral」に対する比、右図：基本周波数時間変化率における「Hot Anger」の「Neutral」に対する比	24
5.11	各感情におけるアクセントレベル上昇時のパワー音韻変化率	26
5.12	左：パワー音韻変化率における「Cold Anger」の「Neutral」に対する比、右：パワー音韻変化率における「Hot Anger」の「Neutral」に対する比	26
5.13	各感情におけるアクセントレベル上昇時のパワー時間変化率	27
5.14	左図：パワー時間変化率における「Cold Anger」の「Neutral」に対する比、右図：パワー時間変化率における「Hot Anger」の「Neutral」に対する比	28
5.15	各感情におけるアクセントレベル下降時のパワー音韻変化率	28
5.16	左：パワー音韻変化率における「Cold Anger」の「Neutral」に対する比、右：パワー音韻変化率における「Hot Anger」の「Neutral」に対する比	29
5.17	各感情におけるアクセントレベル下降時のパワー時間変化率	30
5.18	左図：パワー時間変化率における「Cold Anger」の「Neutral」に対する比、右図：パワー時間変化率における「Hot Anger」の「Neutral」に対する比	30
5.19	左図：「Cold Anger」の「Neutral」に対する文全体の持続時間比、右図：「Hot Anger」の「Neutral」に対する文全体の持続時間比	31
5.20	左図：「Cold Anger」の「Neutral」に対するアクセントの持続時間比、右図：「Cold Anger」の「Neutral」に対する非アクセントの持続時間比	32
5.21	左図：「Hot Anger」の「Neutral」に対するアクセントの持続時間比、右図：「Hot Anger」の「Neutral」に対する非アクセントの持続時間比	32
5.22	左図：「Cold Anger」の「Neutral」に対するアクセント母音の持続時間比、右図：「Cold Anger」の「Neutral」に対するアクセント子音の持続時間比	33
5.23	左図：「Cold Anger」の「Neutral」に対する非アクセント母音の持続時間比、右図：「Cold Anger」の「Neutral」に対する非アクセント子音の持続時間比	34
5.24	感情間のフォルマント周波数 [kHz] の関係 (左上)F1(右上)F2(左下) F3(右下)F4	37

5.25 「Neutral」と「Cold Anger」のフォルマント周波数（常用対数）の関係 .	38
5.26 「Neutral」と「Hot Anger」のフォルマント周波数（常用対数）の関係 . .	39
5.27 「Neutral」30～35ERB rate	41
5.28 「Cold Anger」30～35ERB rate	41
5.29 「Hot Anger」30～35ERB rate	42
5.30 各感情のアクセント低レベルにおけるスペクトル概形の傾向	43
5.31 合成音を作成する際の各感情のスペクトル	44
5.32 各感情のアクセント高レベルにおけるスペクトル概形の傾向	45
5.33 合成音を作成する際の各感情のスペクトル	46

表 目 次

4.1	聴取実験に使用した機器	10
5.1	「Neutral」に対する持続時間比の平均	34
5.2	フォルマント周波数分析条件	36
5.3	アクセント低レベルのスペクトル分析結果 (単位 : [dB])	42
5.4	アクセント高レベルのスペクトル分析結果 (単位 : [dB])	45

第1章 序論

1.1 はじめに

人間の最も基本的で日常的なコミュニケーション手段の一つとして音声がある。音声は伝達する情報として、言語的信息・パラ言語的信息・非言語的信息という3つの情報に分けることがある[1]。言語的信息とは言語記号が伝達する情報であり、話者による意識的な制御を受け、離散的な表現が可能である。パラ言語情報は対話の制御に関わる情報の他、話者の意図（強調の有無や程度など）や心的態度（疑い、関心、感心の有無等）に関する情報が含まれる。パラ言語的信息は、話者によって意図的に制御される点は言語的信息と同一であるが、単一の意味カテゴリーの内部で、量的かつ連続的な変化（例えば、ちょっとした「疑い」から強烈な「疑い」まで）が生じうる点において、離散的な言語情報から区別される。非言語情報は発話の言語的内容とは独立に存在し、話者の身体的ないし精神的状態に関するものをさす。身体的な状態としては、例えば年齢・性別・個人性・体調の良否など、また精神的状態としては、感情・気分・性格などが挙げられる。これらの状態のうち、あるものは離散的な範疇（例えば、性別や感情）をなすが、その多くは連続的な特徴をもち、また通常は話者が意図的に制御出来ないものである。但し、演劇や声帯模写などのように、意図的に模擬することは可能である。この中で、感情に関しては他者への伝達を意図的に行う場合がある。この場合には、感情自体は意図的に制御されないが、その表出・伝達は意図的に制御される。さらに、社会的・文化的要因によって、その表出に誇張・抑制・隠蔽、あるいは虚構などの意図が働くこともある。

近年、音声を人間とコンピュータとのインターフェースとして用いることが実用化されつつある。この際、コンピュータと音声でコミュニケーションをとる時に、ただ単に言語情報の伝達だけを扱うのではなく、前述のようなパラ言語・非言語情報の伝達が行われれば、よりフレンドリなコミュニケーションが実現できると考えられる。その中でも特に、感情の伝達は人間同士において重要な役割を担っており、人間らしく聴きやすい合成音声には感情情報が必要不可欠だと考えられる。

そこで、本研究では感情音声に焦点を当て、感情音声の音響特徴量と感情知覚との間の

関係を解明する。感情音声の音響特徴量と感情知覚とがどのような関係であるかを知ることが出来れば、それをフィードバックさせて合成音品質の向上にもつながるものと期待される。また、マンマシンインターフェースにおいても貢献をもたらし、さらには音声の本質の解明にとっても重要な知見をもたらすことが期待される。

1.2 本研究の背景

1.2.1 感情音声を扱った過去の研究

これまでの感情音声を扱った研究は、音響特徴量として基本周波数・振幅・持続時間といった韻律情報を中心に注目して行なわれている。そして、そのいずれの特徴量もある感情と「平静」とを比較して定性的あるいは定量的に調べられている。

平賀らは、基本周波数・振幅・持続時間の時系列変化に関する研究を行った [2]。その結果、言語情報を含んでも、そのサンプルに十分な感情表現がなされていれば、その時系列パターンには個人差も比較的許容出来る範囲に収まり、感情による特徴は明確にそのパターンに現れてくると報告している。これは、Noad らも感情を含むことによって音声に生ずる物理的変動の傾向が、性別・経歴などに依らないと報告している [3]。小池らは、韻律情報の分析から波形合成方式により合成音を作成し知覚実験を行うことにより、感情音声を表現できることを示した [4]。片江・木村は、声質と韻律の制御がどのように聴取者の感情判別に寄与するかを研究し、「喜び」は韻律の、「声を荒げた怒り」「悲しみ」は声質の寄与が大きいことを報告している [5]。重永は、韻律情報の他に母音のフォルマント周波数・声道断面積関数などの音韻情報の分析を行い、韻律情報と音韻情報の併用によって、高い感情判別率を実現することを示した [6]。

感情の程度について検討している研究も最近行われている。川波・広瀬は、感情のレベルを段階的に変化させた文章を含む模擬対話音声进行分析した [7][8]。大山・武田らは、「怒り」の度合いを平常・軽い怒り・怒り・激怒の4段階で表し詳細な分析を行い、基本周波数・最高音声パワー・平均発話速度・持続時間・呼気流量・声門下圧でそれぞれに異なる傾向が見られることを報告している [9][10][11][12]。

また、感情音声を調音運動の観点から分析・合成を行っている研究も見られる。鈴木・有泉は、声道模擬による音声生成モデルにより規則音声合成システムを構築し、分析に基づき基本周波数とフォルマント周波数の合成規則を導いた [13]。Erickson らは、感情によっては顎や舌端の位置の変化と F0 とフォルマント周波数の変化との関係に顕著な傾向

が出ることを報告している [14]。前川らも調音運動を観測することによって、「感心」と「疑い」との間の母音フォルマント周波数の差異が、舌の前後方向位置および唇間距離の異同によって生じていることを確認している [15]。さらに前川は、多次元尺度構成法と重回帰分析を用いることにより、音響特徴量の間に相関が存在することを報告している [16]。本研究の先行研究である杉本の研究においても音響特徴量間の関係の解明には重点を置いている。その研究において杉本は Lombard Effect を採り入れ、パワーと基本周波数を関連付けることにより「怒り」の認識率の向上を得た [17]。しかし、持続時間において過去の研究と異なる結果が出た。

以上、過去に行われてきた感情音声に関する研究を挙げたが、いずれの研究結果においても定性的な傾向について、それぞれの感情表現に対してほぼ一致した見解が得られている。

1.2.2 過去の研究の問題点

過去の研究にはいくつか問題点がある。まず一つは、過去の研究は単語や文章などについて全発声区間という大局的な部分における平均的な変化を分析したものばかりで、それについてはほぼ一致した見解が得られているが、文章におけるアクセントの有無など局所的な部分を十分に分析したものはないということが挙げられる。従来の研究から、音響特徴量の中でも基本周波数は感情知覚に最も影響を及ぼす重要なパラメータであり、しかもアクセント部において基本周波数は特に変動が大きいと考えられる。そこで、アクセント部という局所的な部分における音響特徴量の分析を行う必要がある。

次に、音声の生成機構は互いに関連しているため、それに伴い音響特徴量の間にも関連があると考えられるが、従来の研究では十分にわかっていないという問題点も挙げられる。

また、先行研究として杉本は「平静」「喜び」「怒り」「悲しみ」の感情について全発声区間の平均的な変化を分析し、従来の研究とほぼ一致した傾向を得たが、「怒り」については持続時間で従来の研究と異なる結果が出た。これは、「怒り」という感情の中にも「Hot Anger」「Cold Anger」というパワーや持続時間といった音響特徴量の変化の仕方が異なる感情があるにもかかわらず [18]、杉本は両方を同じ「怒り」として扱っているために感情判別の精度が落ちたという問題点が考えられる。

1.3 本研究の目的

前項で述べた過去の研究の問題点を踏まえて、本研究ではアクセント部における音響特徴量と感情知覚との関係を調べることを目的とする。そしてその際に、感情知覚の差異に伴う音響特徴量間の関係を検討する。また、「怒り」の感情を「Hot Anger」と「Cold Anger」に分け感情の分類を多くして、より精度の高い結果を期待する。

音響特徴量と感情知覚との間の関係を解明できれば、合成音声の品質向上につながると考えられる。また、音声対話システムにおいても、マンマシンインターフェースとして感情表現が可能な音声合成システムの実現が期待される。

1.4 本論文の構成

本論文の構成を以下に示す。

第1章では、本論文が対象としている研究分野の現状と問題点を指摘し、本論文の位置付けと目的を明らかにする。

第2章では、本研究で扱う基本周波数・パワー・持続時間・フォルマント周波数・スペクトルといった音響特徴量についての過去の研究の成果を示す。

第3章では、本研究で扱うアクセントについて説明する。

第4章では、本研究で用いる感情音声データを説明する。また、その感情音声データが本研究で分析するのに有用であるかを確認するため、聴取実験を行う。

第5章では、本研究で扱った基本周波数・パワー・持続時間・フォルマント周波数・スペクトルといった音響特徴量についてアクセント部における変化を中心に分析を行う。

第6章では、全体のまとめを行い、第7章にて本論文で得られた結果を要約し、今後の課題を示す。

第2章 過去の研究の分析結果

以下に、本研究で扱う音響特徴量に関する過去の研究の主要な分析結果を紹介する。

基本周波数についての分析結果

「怒り」「喜び」の基本周波数パターンは類似した形状を有しており、「平静」と比較して高く、ダイナミックレンジ・変化率も大きい[2]。最高基本周波数は怒りの度合いが大きくなるにつれて徐々に高くなる[10]。また、怒りの度合いが増すに従って、基本周波数の絶対値は大きくなっていくが、それに従って基本周波数パターンも変形する。「怒り」と「喜び」の基本周波数パターンは類似した特徴を示すが、「怒り」の方が文頭での上昇が急であり、文末まで比較的高い基本周波数を維持し、そして急激に降下する[8]。

「悲しみ」の基本周波数は平静と比較して低く、ダイナミックレンジ・変化率も小さい[2][8]。基本周波数の構造は、「喜び」「悲しみ」の感情表現において重要な役割を担っている[2][19]。

パワー (振幅・スペクトル) についての分析結果

「平静」の振幅パターンは文頭で大きく、その後徐々に減少していくかほぼ一定のままかの2通りとなっている[2]。「怒り」の振幅は「平静」と比較して全体的に非常に大きく[4]、振幅パターンのピークもかなり大きい[2]。また、「激怒」の場合には「平静」と比較してパワースペクトルにおける高域が増加する[9][10]。

「喜び」の振幅パターンは、「平静」とほぼ同じであるが、その大きさは「平静」より大きい。振幅パターンは大きく2つに分けられ、短文であれば語尾に向かって上昇傾向にあり、長文であれば包絡において大きなゆっくりとした抑揚が見られる。「悲しみ」の振幅は、そのレンジ・平均値ともに非常に小さく、語尾にいくに従って減少していく概形をとる[2]。

持続時間 (発話速度) についての分析結果

「怒り」では、文頭で発話速度が大きくなり、文の後半部で小さくなる傾向が見られる。これは、特に感情を込めずに発声した対話音声において見られる特徴である [7]。怒りの度合いが増加するに従って平均発話速度は早くなり、発話時間の短縮・伸長の度合いは大きくなる [10]。また、「怒り」の無音区間は非常に短い [2]。

「喜び」では、述部において持続時間が大きくなる傾向がある [7]。「悲しみ」は、発話速度が比較的遅い傾向がある。無音区間は非常に長く、話す際の「間」が感情を表現する際に非常に重要であることを示している [2]。

フォルマント周波数についての分析結果

調音運動の観点からフォルマント周波数を検討した研究がある。そこでは母音フォルマント周波数の差異が、舌の前後方向位置および唇間距離の異同によって生じていることが確認されている [15][16]。「平静」の F1 に対する「怒り」の F1 が 1.2 倍、「喜び」が 1.1 倍、「悲しみ」が 0.9 倍としているものもある [13]。また、F2 が「疑い」では高く、「感心」では低いという関係がある [15]。しかし、全体的に感情音声の研究でフォルマント周波数を扱っている研究は少ない。

上記のような研究結果が過去の感情音声の研究で行われてきたものであるが、これらは音響特徴量を全発声区間における長時間平均について検討したものがほとんどであり、アクセント部など局所的な部分を分析したものはない。そこで、本研究ではアクセント部において上述の音響特徴量がどのような振る舞いをするか分析する。次章において、本研究で扱うアクセント部を説明・定義する。

第3章 アクセントについて

本章では、本研究で扱うアクセントについて説明する。韻律情報の中で、イントネーションやアクセントは、合成音の高品質化の上で特に重要で、これには声の高さ・強さ・長さなどが関連するが、日本語では基本周波数がこれらを支配する直接的原因であるとされている [20]。基本周波数は基本イントネーション成分の上にアクセントによって決まる単語・文節固有のアクセント成分が重畳して、文全体のピッチパターンが定まる。日本語のアクセントは、高低の2レベルがあり、アクセントのあるモーラの直後にレベルが高から低に移る。これをアクセント核と呼ぶ。日本語の n モーラ単語には、図 3.1 に示すように、アクセント核が存在しない 0 型か、または核の位置が $k=1$ (モーラ目) から $k=n$ (モーラ目) までのいずれか 1 個所にある n 個のアクセント型が可能で、それぞれ k 型と呼ばれる。ちなみに、図でモーラ数 1 における「柄」と「絵」の違いや、モーラ数 2 における「端」と「箸」と「橋」の違いは、後に助詞「が」をつけるとわかる。本研究ではこのアクセント型を基準として全文章に適合させた。

モーラ数	1	2	3	4	5
0 型	柄	端			あたり前
1 型	絵	箸			アクセント
2 型		橋			お母さん
3 型					山桜
4 型					ほかけ舟
5 型					お正月

図 3.1: 日本語のアクセント型

第4章 感情音声データ

4.1 感情音声データ

本章では、本研究で音声波を分析する際に用いるための感情音声データを説明する。声優・演劇経験者は一般人に比べ、音声によって感情状態の表現を行う手法を的確に心得ている [2] ことがわかっているので、本研究で扱う感情音声データはプロの声優 (女性) から採取した計 179 (9 パターン、20 文章) サンプル ((株) 富士通研究所から貸与) を用いる。以下にその感情音声データのデータ形式・発声内容・発声リストを示す。

- データ形式

サンプリング周波数 22050Hz

16bit 量子化最大振幅は録音時に正規化されている

- 発声内容

「Neutral(平静)」が 1 発声、「Joy(喜び)」「Hot Anger(激しい怒り)」「Cold Anger(押し殺した怒り)」「Sadness(悲しみ)」の感情が各々 2 発声で計 9 パターンのデータを収録

a001 ~ a020	Neutral
b001 ~ b020	Joy[1]
c001 ~ c020	Joy[2]
d001 ~ d020	Cold Anger[1]
e001 ~ e020	Cold Anger[2]
f001 ~ f020	Sadness[1]
g001 ~ g020	Sadness[2]
h001 ~ h020	Hot Anger[1]
i001 ~ i020	Hot Anger[2]

e014 はデータ欠落

- 発声リスト

- 001 新しいメールが届いています
- 002 頭にくることなんてありません
- 003 待ち合わせは青山らしいんです
- 004 新しい車を買いました
- 005 知らないメールがあったら捨てて下さい
- 006 そんなの古い迷信ですよ
- 007 みんなからエールが送られたんです
- 008 手紙が届いたはずですよ
- 009 ずっと見えています
- 010 私のところには届いています
- 011 ありがとうございます
- 012 申し訳ございません
- 013 ありがとうは言いません
- 014 旅行するには二人がいいのです
- 015 気が遠くなりそうでした
- 016 こちらの間違いもございました
- 017 花火を見るのにゴザがいりますか
- 018 もうしないと言ったじゃないですか
- 019 時間通りに来ない訳を教えてください
- 020 サービスエリアで合流しましょう

4.2 聴取実験

4.2.1 目的

本研究で扱う感情は、話し手側ではなく聞き手側に存在するものを指す。そして本研究では、感情情報を音声データの分析によって得るので、音声データが本研究で分析するのに十分有用であるかを判断するため、聴取実験を行う。

4.2.2 聴取実験

聴取実験は、採取した音声データを1文章ずつ呈示し、その音声が「Neutral」「Joy」「Cold Anger」「Sadness」「Hot Anger」のどの感情のものなのかを判断させた。また、「Neutral」「Joy」「Cold Anger」「Sadness」「Hot Anger」のどの感情でもないと感じた場合は「その他」として、特に何かある感情が感じられた場合にはそれも回答させた。音声データは179サンプルあるので、60サンプルずつ3回に分けて実験を行った。各回とも、始めにサンプルa001「新しいメールが届いています」・a011「ありがとうございました」・a012「申し訳ございません」という「Neutral」の感情音声を練習として呈示し、その後本番として60サンプルをランダムに呈示した。被験者は大学院生9名であり、その全てが正常聴力を有する。被験者は防音室内でヘッドフォンにより受聴した。受聴はモノラルの両耳受聴である。聴取実験に使用した機器を表4.1に示す。

表 4.1: 聴取実験に使用した機器

機器	メーカー、機種
ヘッドフォン	SENNHEISER HDA 200
ヘッドホンアンプ	SANSUI AU α -907MR

4.2.3 実験結果と考察

実験結果を図4.1から図4.9に示す。図4.1から図4.9は、感情の各パターン（aからiまで）における認識率を表したもので、横軸が文章番号、縦軸が認識率である。棒グラフの下には認識されなかったサンプルについて、どの感情が選択されたかを示した。

各感情音声について結果を考察する。

「Neutral」「Hot Anger」の感情音声における認識率はほぼ全て100%で、全てのサンプルが80%以上の認識率だった（「Hot Anger」において「Joy」と認識されるものがあった）。

次に、「Joy」の感情音声において認識率が高く、9割が80%以上の認識率があった。認識されなかったものについては、感情パターンbで「Neutral」と認識されるものが目についた（感情パターンcでは文章によっては「Hot Anger」、「Cold Anger」、「Sadness」と認識されるものがあり言語情報に依存するところが見られた）。

「Cold Anger」の感情音声については、一番認識率に開きが見られた。認識率が80 %以上のサンプルは8割で、70 %以下のサンプルが1割あった。今回の聴取実験で特に認識率の低かったd001・e007・e009・e012は言語情報に依存したためと考えられる。また、他にうまく認識されなかったものについては「Neutral」と認識されたものが多く見られ、次に「Sadness」も見られた。

「Sadness」の感情音声は認識率100 %のサンプル数が5つの感情音声の中で一番少なかった。「Cold Anger」と同様に、80 %以上の認識率のサンプルは8割だった。うまく認識されなかったものについては、特に「Neutral」と認識されるものが見られた。「Cold Anger」と認識されるものもあったが、これは言語情報に依存したためと考えられる。

この聴取実験の結果、各感情において高い認識率のサンプルが多数存在しており、感情情報の抽出を目的とした分析にとって十分だと考えられる。

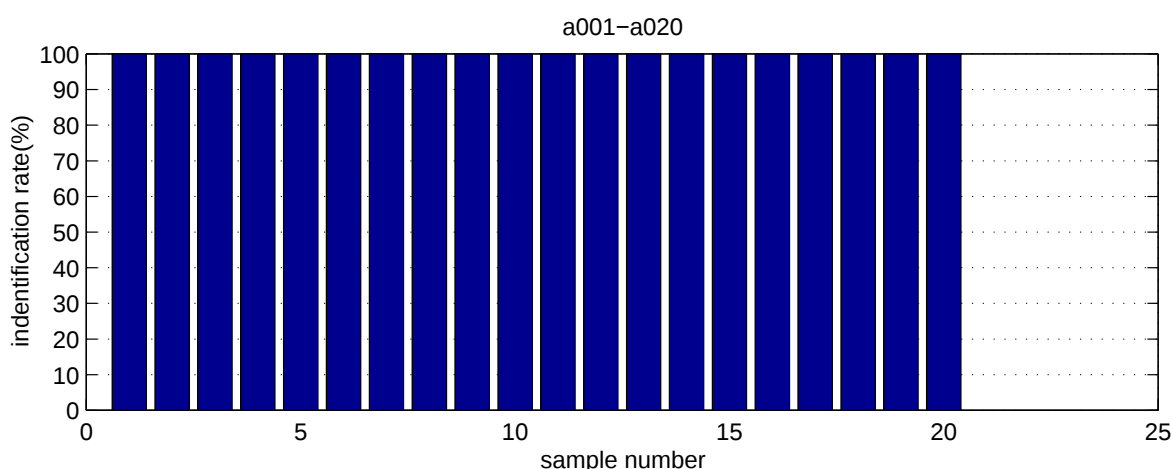
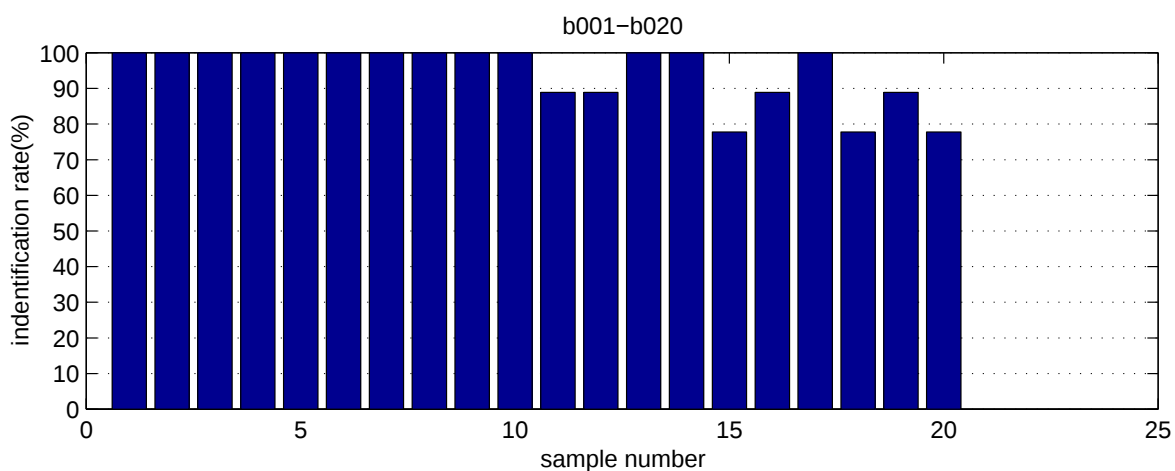
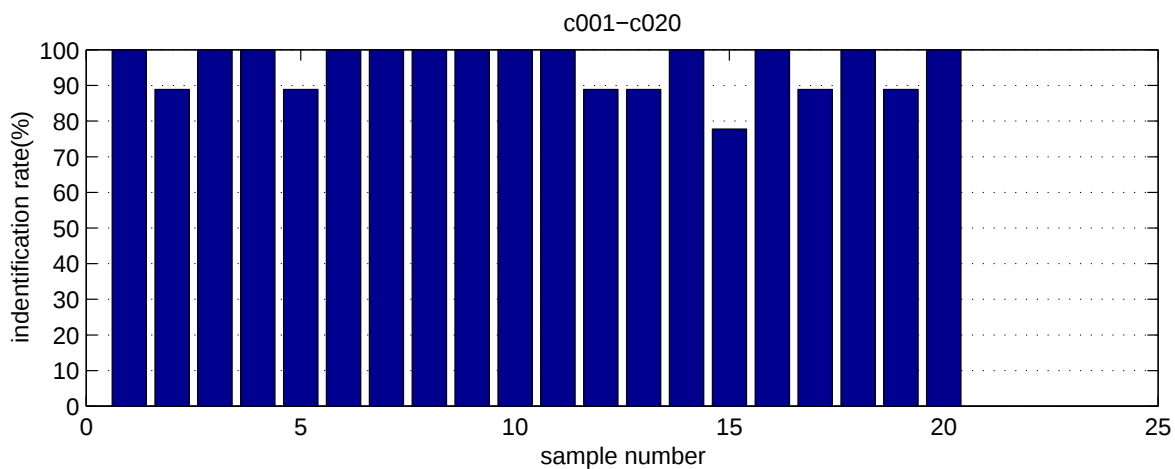


図 4.1: a001 ~ a020 の認識率



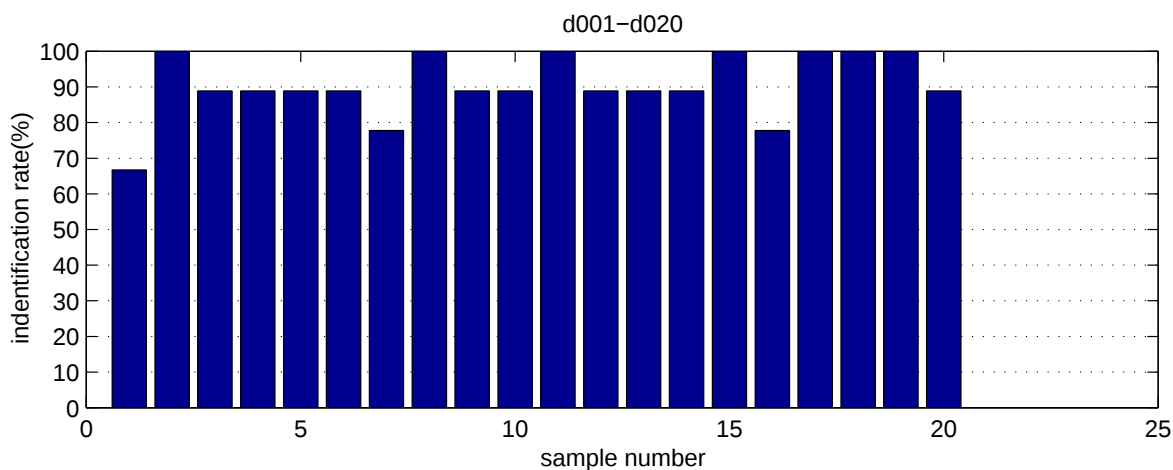
b011:その他 1 b012:平静 1 b015:平静 2 b016:その他 1
 b018:平静 1、その他 1 b019:Cold Anger1 b020:平静 2

図 4.2: b001 ~ b020 の認識率



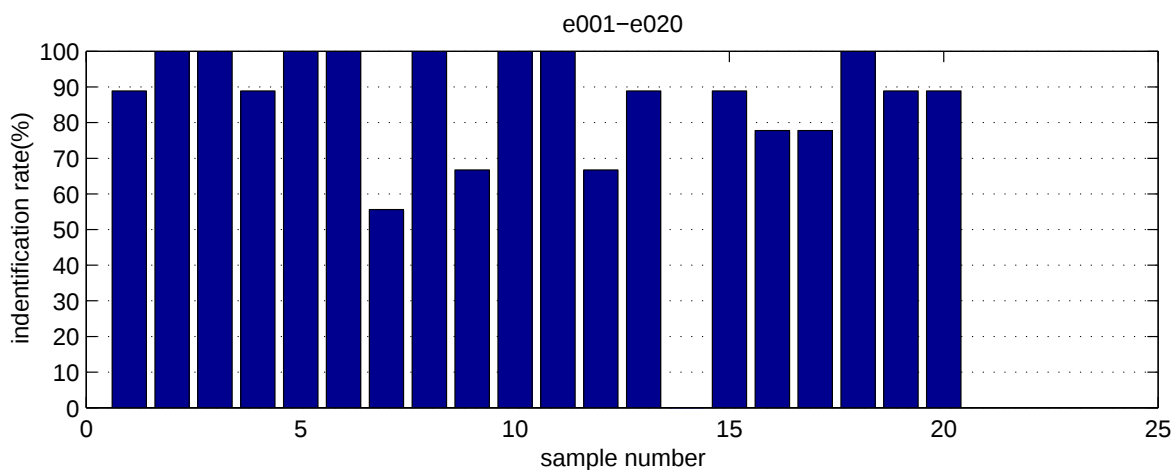
c002:Cold Anger1 c005:その他 1 c012:その他 1 c013:Hot Anger1
 c015:悲しみ 1、その他 1 c017:その他 1 c019:Hot Anger1

図 4.3: c001 ~ c020 認識率



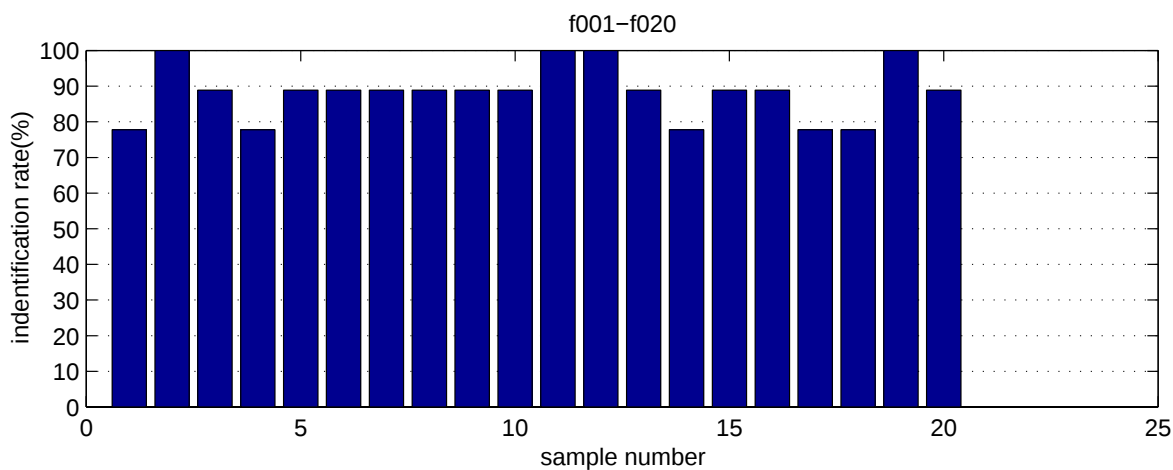
d001:平静 2、その他 1 d003:平静 1 d004:平静 1 d005:平静 1
 d006:悲しみ 1 d007:平静 1、その他 1 d009:平静 1 d010:その他 1
 d012:悲しみ 1 d013:平静 1 d014:その他 1 d016:平静 2
 d020:その他 1

図 4.4: d001 ~ d020 認識率



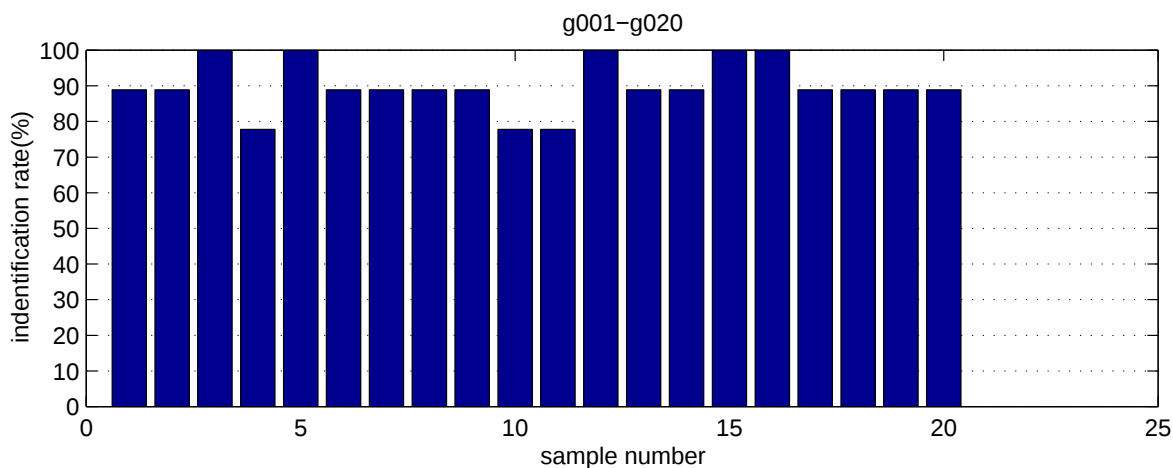
e001:その他 1 e004:その他 1 e007:悲しみ 3、その他 1
 e009:平静 1、悲しみ 1、その他 1 e012:Hot Anger1、悲しみ 1、その他 1
 e014:平静 1 e015:平静 1 e016:悲しみ 1、その他 1
 e017:平静 1、その他 1 e019:平静 1 e020:その他

図 4.5: e001 ~ e020 認識率



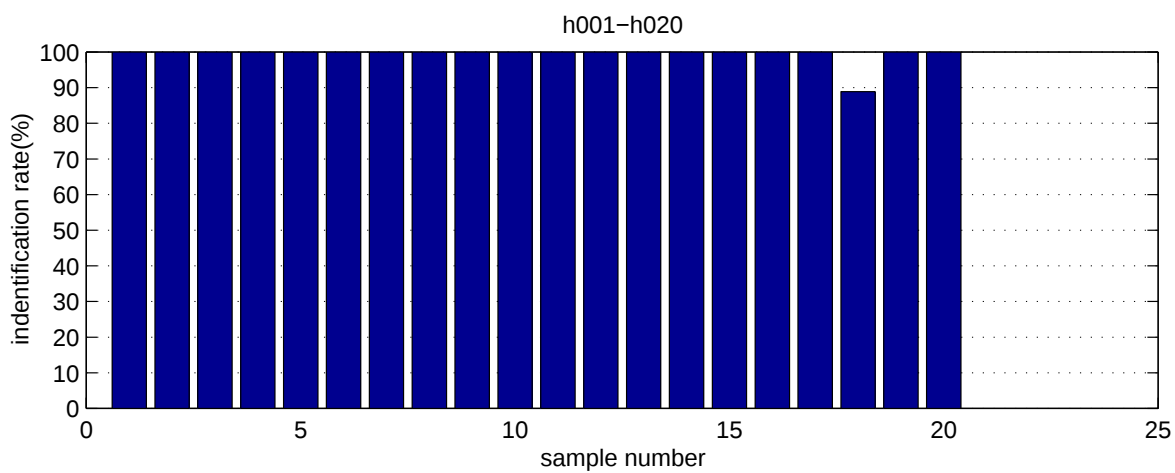
f001:平静 1、その他 1 f003:その他 1 f004:平静 1、その他 1 f005:その他 1
 f006:その他 1 f007:平静 1 f008:その他 1 f009:Cold Anger1
 f010:その他 1 f013:平静 1 f014:平静 1、その他 1 f015:その他 1
 f016:その他 1 f017:Cold Anger1、その他 1 f018:Cold Anger1、平静 1
 f020:平静 1

図 4.6: f001 ~ f020 認識率



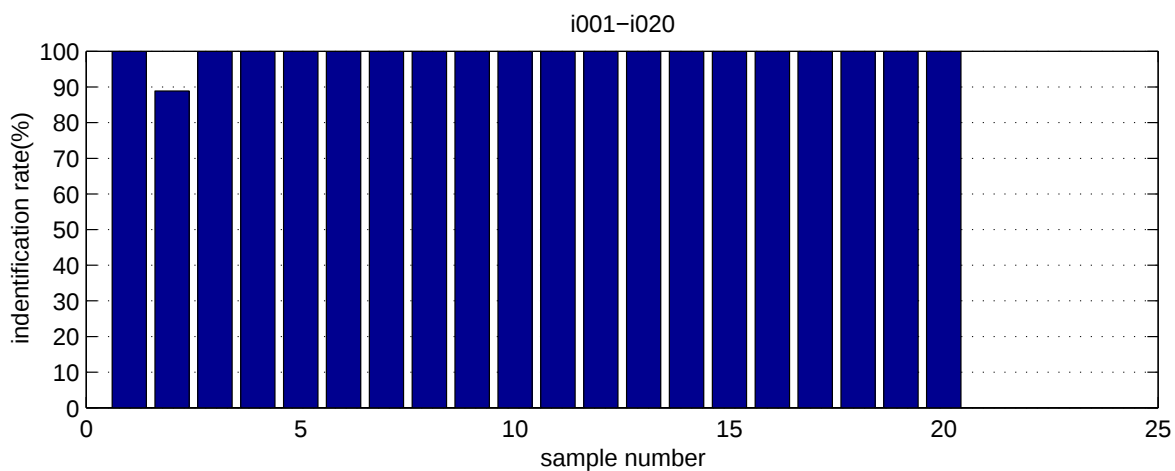
g001:平静 1 g002:平静 1 g004:平静 1、その他 1 g006:平静 1
 g007:平静 1 g008:平静 1 g009:平静 1 g010:平静 1、その他 1
 g011:平静 1、その他 1 g013:その他 1 g014:平静 1 g017:その他 1
 g018:平静 1 g019:悲しみ 1 g020:平静 1

図 4.7: g001 ~ g020 認識率



h018:喜び 1

図 4.8: h001 ~ h020 認識率



i002:喜び 1

図 4.9: i001 ~ i020 認識率

第5章 音声データの分析

5.1 目的

本研究では、音響特徴量と感情知覚との関係を解明することを目的としているため、感情音声データの分析を詳細に行うことは本研究にとって大変重要なことである。

過去の研究ではアクセント部という局所的な部分の分析が十分に行われていないため、本研究ではアクセント部に注目し音響特徴量の分析を行う。感情情報に関する音響特徴量としては、基本周波数・パワー・持続時間・フォルマント周波数・スペクトルを扱う。また、これら時間変化するパラメータを抽出・変換・合成するモデルを用いる必要がある。そこで、本研究ではこの音声分析変換合成系として高品質な合成音声を作成できる STRAIGHT(Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrogram)[21] を採用する。尚、音声データの分析はモーラ単位で行うが、後の分析で母音の中心時間が必要となるので、分析を行い易くするためにあらかじめサウンドスペクトログラムを用いて音素単位にセグメントする。

5.2 基本周波数の分析

5.2.1 分析手法

基本周波数は、文章の有声区間における長時間平均とアクセントにおける変化率を分析する。基本周波数は、STRAIGHT の TEMPO を用いて抽出する。

- 文章の有声区間における長時間平均

文章の有声区間における基本周波数長時間平均の分析手法としては、サンプルの有声区間における基本周波数を STRAIGHT の TEMPO で抽出して、その時間平均を求める。

- アクセントにおける音韻変化率・時間変化率

アクセントのレベルが低から高へ上昇する部分と、高から低へ下降する部分の基本周波数変化率を分析する。変化率は変化前のモーラに対する変化後のモーラの比を表す変化率（音韻変化率）と変化前のモーラから変化後のモーラへの時間に対する変化率（時間変化率）の2つを検討する。

変化前のモーラを a 、変化後のモーラを b とする。モーラ a, b の母音の中心時間を t_a, t_b 、その部分の基本周波数を $F(t_a), F(t_b)$ として、基本周波数の音韻変化率・時間変化率はそれぞれ以下の様に求める。

基本周波数の音韻変化率:

$$R_{Fp} = \frac{F(t_b)}{F(t_a)}$$

基本周波数の時間変化率 [Hz/sec] :

$$R_{Ft} = \left| \frac{F(t_b) - F(t_a)}{t_b - t_a} \right|$$

ここで、サンプルにおいて「でした」の「し」は「shi」ではなく「sh」に、「ます」の「す」は「su」ではなく「s」になっていたり、Hot Anger における「こちら」の「こ」など無声化したものは、基本周波数が無い部分が長いため分析の対象に含めなかった。しかし、有声音の箇所において STRAIGHT の TEMPO で検出出来なかった部分は、比較的基本周波数が無い部分が短かったため spline 関数で補間することによって基本周波数を求めた。

5.2.2 分析結果・考察

文章の有声区間における長時間平均

文章の有声区間における長時間平均を「Neutral」「Cold Anger」「Hot Anger」それぞれについてサンプル番号順に並べたものが図 5.1 である。図は横軸が「Neutral」「Cold Anger」「Hot Anger」のサンプル番号順に並べたもので、縦軸が基本周波数 [Hz] である。

この結果から、「Neutral」に対する「Cold Anger」「Hot Anger」の比をサンプルごとにとりヒストグラムで表したものが、それぞれ図 5.2 (左図・右図) である。これを見ると、「Cold Anger」は全てのサンプルで「Neutral」に対する比が 1 より小さくなっている。これは、「Cold Anger」の感情で文章における基本周波数の長時間平均が「Neutral」より小さいことを示している。逆に、「Hot Anger」は全てのサンプルで「Neutral」に対する比

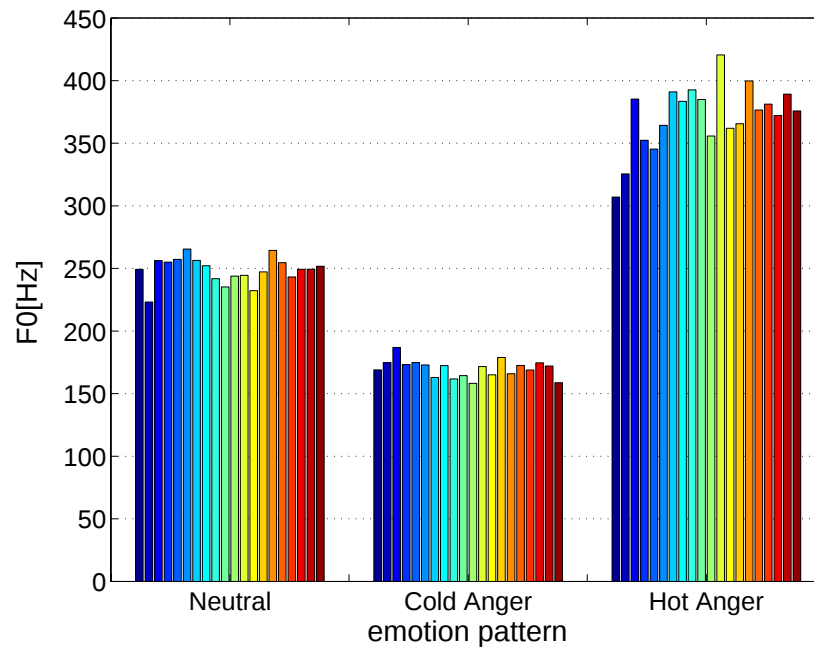


図 5.1: 各感情における基本周波数の長時間平均

が1より大きくなっており、「Hot Anger」の感情で文章における基本周波数の長時間平均が「Neutral」より大きいことを示している。これらより、基本周波数の長時間平均は「Cold Anger」「Neutral」「Hot Anger」という順で大きくなる傾向が見られる。

アクセントにおける音韻変化率・時間変化率

• アクセントレベル上昇時

アクセントレベルが上昇する時の各感情における基本周波数の音韻変化率を示したものが図 5.3 である (横軸は各感情におけるサンプルをサンプル番号順に並べたもの)。

この結果から、「Cold Anger」「Hot Anger」の音韻変化率の「Neutral」の音韻変化率に対する比をサンプルごとにとりヒストグラムにしたものが、それぞれ図 5.4(左図・右図)である。

図 5.4(左図)を見ると、比が1より小さい部分にサンプルが集まっており、「Cold Anger」の基本周波数音韻変化率が全体的に「Neutral」の基本周波数音韻変化率よりも小さい傾向にあることを示している。図 5.4(右図)は逆に比が1より大きい部分にサンプルが集まっており、「Hot Anger」の基本周波数音韻変化率が全体的に「Neutral」の基本周波数音韻変化率よりも大きい傾向にあることを示している。

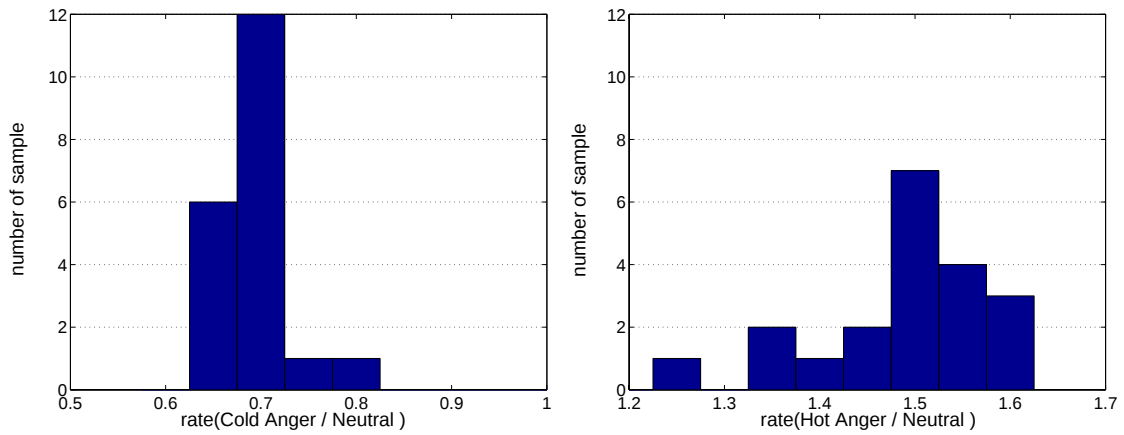


図 5.2: 左図：基本周波数長時間平均における「Cold Anger」の「Neutral」に対する比、
右図：基本周波数長時間平均における「Hot Anger」の「Neutral」に対する比

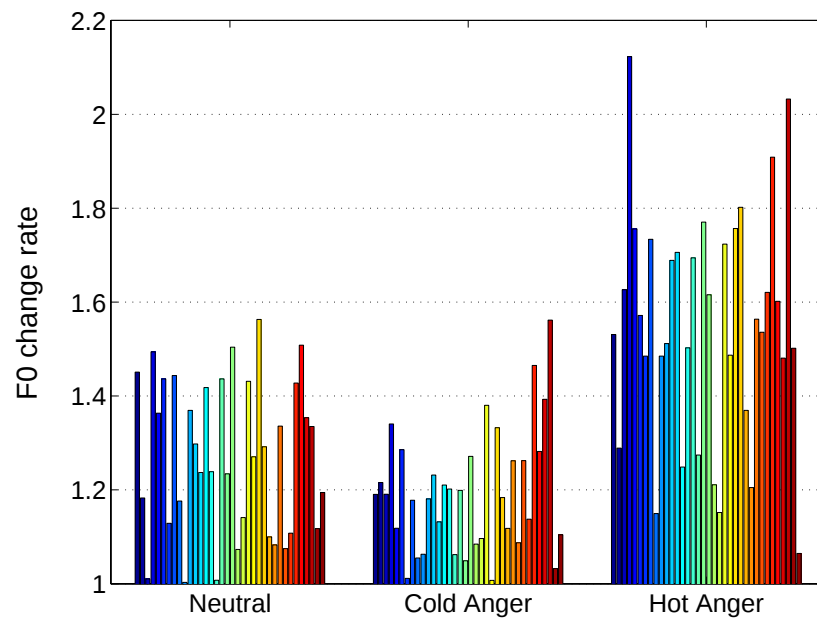


図 5.3: 各感情におけるアクセントレベル上昇時の基本周波数音韻変化率

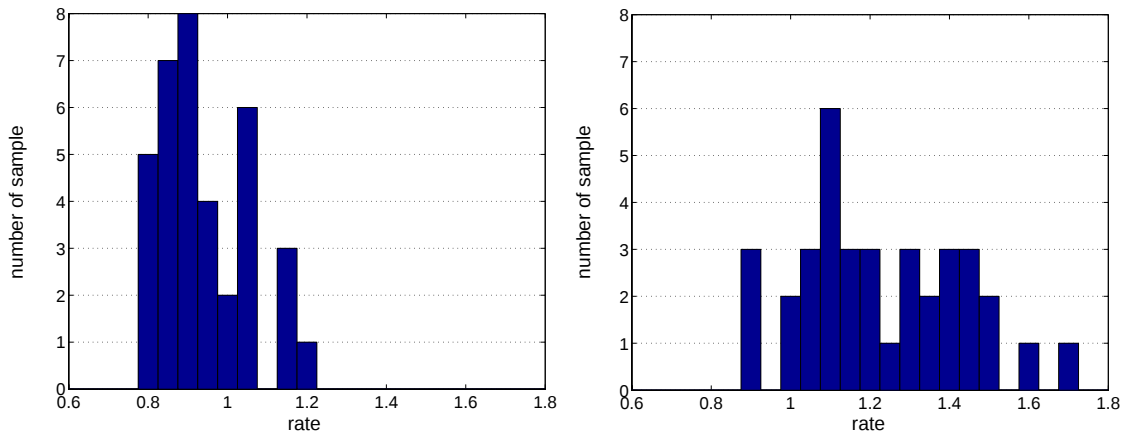


図 5.4: 左図：基本周波数音韻変化率における「Cold Anger」の「Neutral」に対する比、
右図：基本周波数音韻変化率における「Hot Anger」の「Neutral」に対する比

次に、アクセントレベルが上昇する時の時間変化率を感情別のサンプル番号順に並べたものが図 5.5 である。

この結果も音韻変化率の時と同様に、「Neutral」に対する「Cold Anger」「Hot Anger」の比をサンプルごとにとりヒストグラムで表したものが、それぞれ図 5.6(左図・右図)である。

図 5.6(左図)を見ると、比が 1 より小さい部分にサンプルが集まっていることから「Cold Anger」の基本周波数時間変化率は「Neutral」の基本周波数時間変化率と比較して全体的に小さいことが示されている。特に時間変化率比が約 0.5 の付近にサンプルが集中している傾向が見られる。逆に図 5.6(右図)から、「Hot Anger」の基本周波数時間変化率は「Neutral」の基本周波数時間変化率に比べて全体的に大きいことを示している。時間変化率比が約 2 の付近にサンプルが集中している傾向が見られるが、それより大きな比率を示すサンプルもあり、アクセントレベルが上昇する時に「Hot Anger」では急激に基本周波数が上がるということが言える。

以上から、アクセントレベル上昇時において基本周波数は音韻変化率・時間変化率ともに「Cold Anger」は「Neutral」より小さく、「Hot Anger」は「Neutral」よりも大きい傾向を示した。ただし、「Hot Anger」の変化率においてはサンプルによってばらつきが見られた。例えば、「Neutral」に対する音韻変化率比で「新しい車を買いました」という文章では「新しい」の「あ」から「た」が 1.2011、「車を」の「く」から「る」が 0.9771、「買いました」の「か」から「い」が 1.4805 という結果であった。このことから、文脈上で強調する部分の変化率が「Hot Anger」では特に大きいのではないかと考えられる。

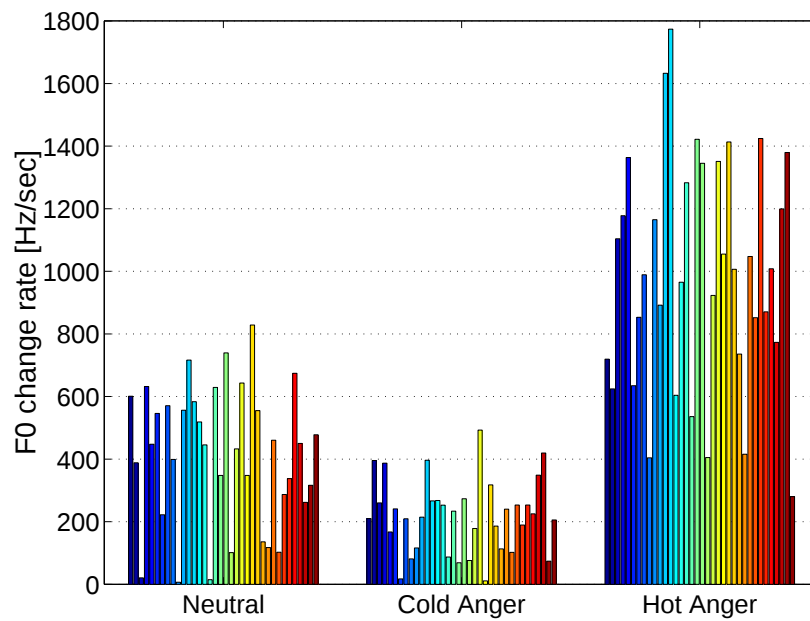


図 5.5: 各感情におけるアクセントレベル上昇時の基本周波数時間変化率

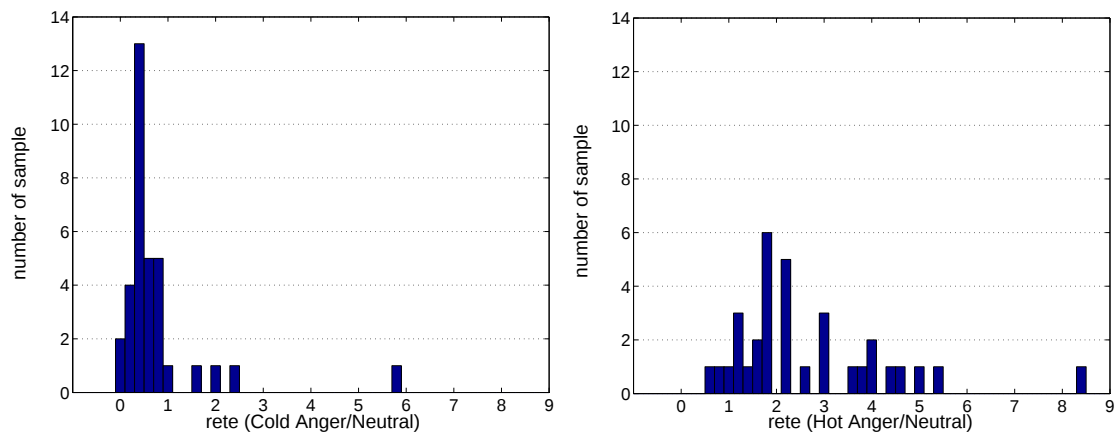


図 5.6: 左図：基本周波数時間変化率における「Cold Anger」の「Neutral」に対する比、
右図：基本周波数時間変化率における「Hot Anger」の「Neutral」に対する比

- アクセントレベル下降時

アクセントレベルが下降する時の各感情における基本周波数の音韻変化率を示したものが図 5.7 である (横軸は各感情におけるサンプルをサンプル番号順に並べたもの)。

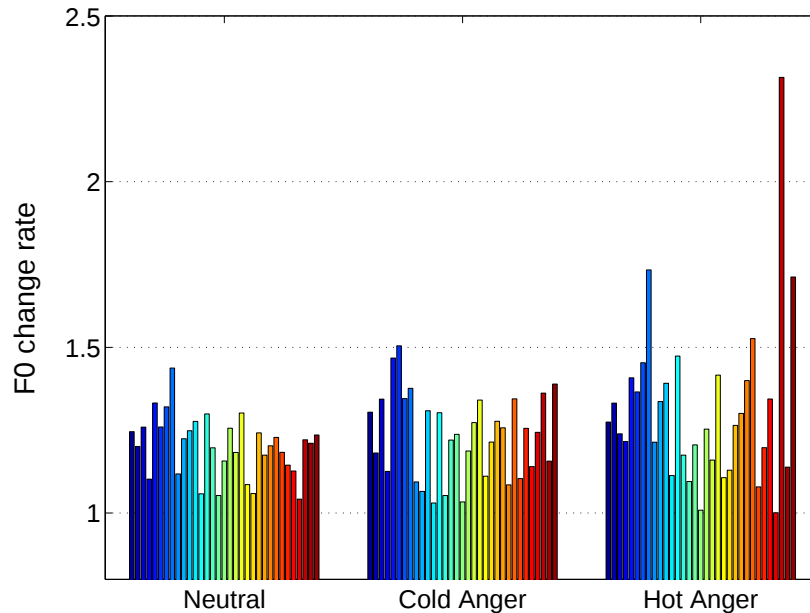


図 5.7: 各感情におけるアクセントレベル下降時の基本周波数音韻変化率

この結果から、「Cold Anger」「Hot Anger」の音韻変化率の「Neutral」の音韻変化率に対する比をサンプルごとにとりヒストグラムにしたものが、それぞれ図 5.8(左図・右図)である。

図 5.8(左図)を見ると、比が約 1 の部分にサンプルが集まっており、「Cold Anger」の基本周波数音韻変化率が全体的に「Neutral」の基本周波数音韻変化率と同じくらいにある傾向が見られる。図 5.8(右図)は比が 1 よりやや大きい部分にサンプルが集まっており、「Hot Anger」の基本周波数音韻変化率が全体的に「Neutral」の基本周波数音韻変化率よりも大きい傾向にあることを示している。しかし、アクセント上昇時と比較すると、その比率はあまり大きくない。これは、「Hot Anger」のアクセントレベル上昇時において基本周波数がかかなり高くなるが、その後、アクセント高レベル部分において大きく基本周波数が下がっているためだと考えられる。

次に、アクセントレベルが下降する時の時間変化率を検討する。アクセントレベル下降時の時間変化率を感情別のサンプル番号順に並べたものが図 5.9 である。

この結果から、「Cold Anger」「Hot Anger」の時間変化率の「Neutral」の時間変化率

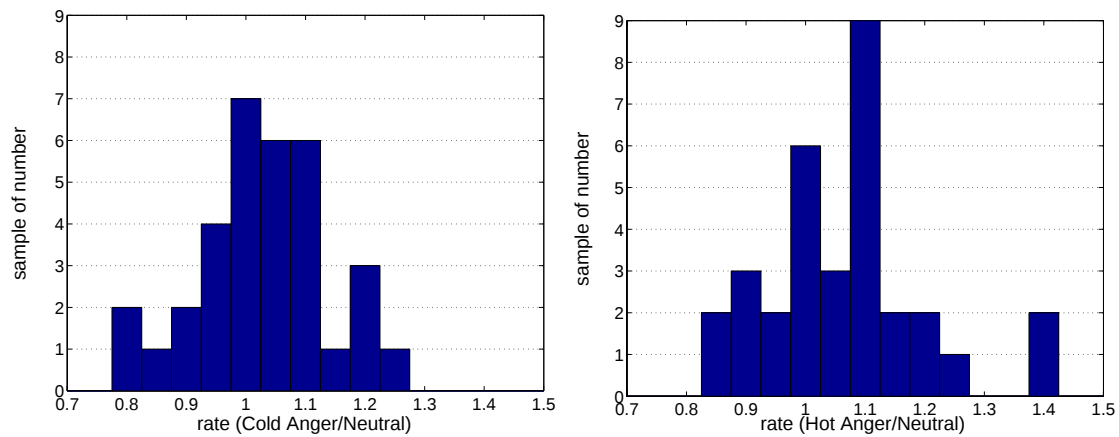


図 5.8: 左図：基本周波数音韻変化率における「Cold Anger」の「Neutral」に対する比、
 右図：基本周波数音韻変化率における「Hot Anger」の「Neutral」に対する比

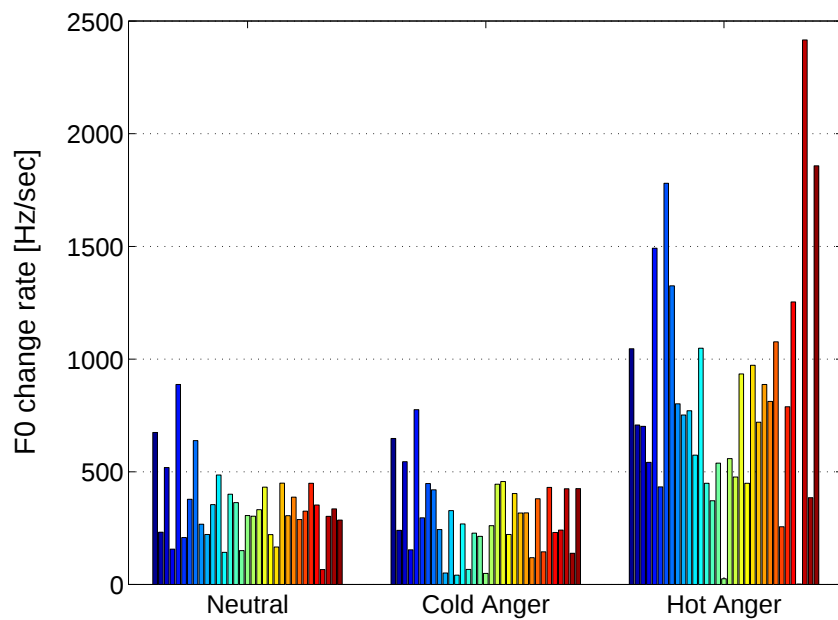


図 5.9: 各感情におけるアクセントレベル下降時の基本周波数時間変化率

に対する比をサンプルごとにとりヒストグラムにしたものが、それぞれ図 5.10(左図・右図)である。

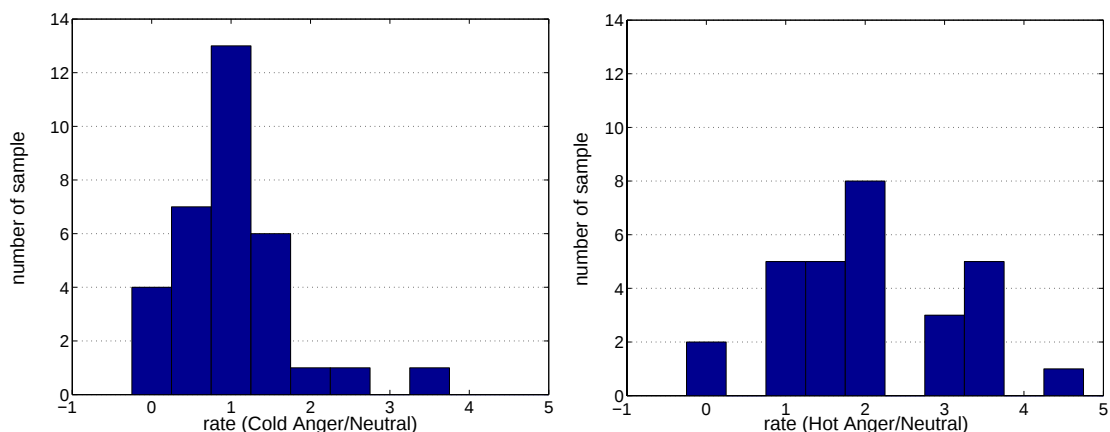


図 5.10: 左図：基本周波数時間変化率における「Cold Anger」の「Neutral」に対する比、右図：基本周波数時間変化率における「Hot Anger」の「Neutral」に対する比

図 5.10(左図)を見ると、比が約 1 の部分にサンプルが集まっていることから、アクセントレベル下降時に「Cold Anger」は「Neutral」とほぼ同じくらいの傾きで基本周波数が下がっていると言える。逆に図 5.10(右図)を見ると、「Hot Anger」は「Neutral」の約 2 倍の傾きで急激に、基本周波数が下がっていると考えられる。

以上より、アクセントレベル下降時において、基本周波数は音韻変化率・時間変化率ともに、「Cold Anger」と「Neutral」とは同じくらいである傾向が見られ、「Hot Anger」は「Neutral」と比較して大きい傾向が見られた。そして、「Cold Anger」「Hot Anger」ともにアクセントレベル上昇時と比べて、「Neutral」に対する比が小さいという傾向が見られた。

5.3 パワー変化率の分析

5.3.1 分析手法

パワーにおいては、アクセントのレベルが低から高へ上昇する部分と、高から低へ下降する部分の音韻変化率・時間変化率・を分析する。本来は、パワーの絶対値の比較もすれば良いのであろうが、録音時に最大振幅が正規化されているため比較不可能である。変化前のモーラを a、変化後のモーラを b とする。モーラ a, b の母音の中心を t_a, t_b 、その部分の振幅を $A(t_a), A(t_b)$ とする。また、有声区間における振幅平均を A とする。パワーの音韻変化率・時間変化率はそれぞれ以下の様に求める。

パワーの音韻変化率:

$$R_{Pp} = 10 \log \frac{A(t_b)^2}{A^2} - 10 \log \frac{A(t_a)^2}{A^2} = 10 \log \frac{A(t_b)^2}{A(t_a)^2}$$

パワーの時間変化率 [dB/sec] :

$$R_{Pt} = \frac{10 \log \frac{A(t_b)^2}{A^2} - 10 \log \frac{A(t_a)^2}{A^2}}{t_b - t_a}$$

5.3.2 分析結果・考察

• アクセントレベル上昇時

図 5.11 はアクセントレベルが上昇する時の各感情におけるパワーの音韻変化率を示したものである (横軸は各感情におけるサンプルをサンプル番号順に並べたもの)。この結果から、「Cold Anger」「Hot Anger」の音韻変化率の「Neutral」の音韻変化率に対する比をサンプルごとにとりヒストグラムにしたものが、それぞれ図 5.12(左図・右図)である。

図 5.12(左図)を見ると、比が約 1 の部分にサンプルが集中している。これは「Cold Anger」のパワー音韻変化率が「Neutral」のパワー音韻変化率とほとんど同じであることを表しているが、全サンプルの平均をとると比は 1.14 であるため、わずかに「Cold Anger」の音韻変化率が大きい傾向がある。図 5.12(右図)を見ると、比が約 2 の部分にサンプルが集中している。これは「Hot Anger」のパワー音韻変化率が「Neutral」のパワー音韻変化率より約 2 倍とかなり大きいことを表している。

次に、時間変化率も同様に検討する。図 5.13 は、アクセントレベルが上昇する時の各感情におけるパワーの時間変化率を示したものである (横軸は各感情におけるサンプルを

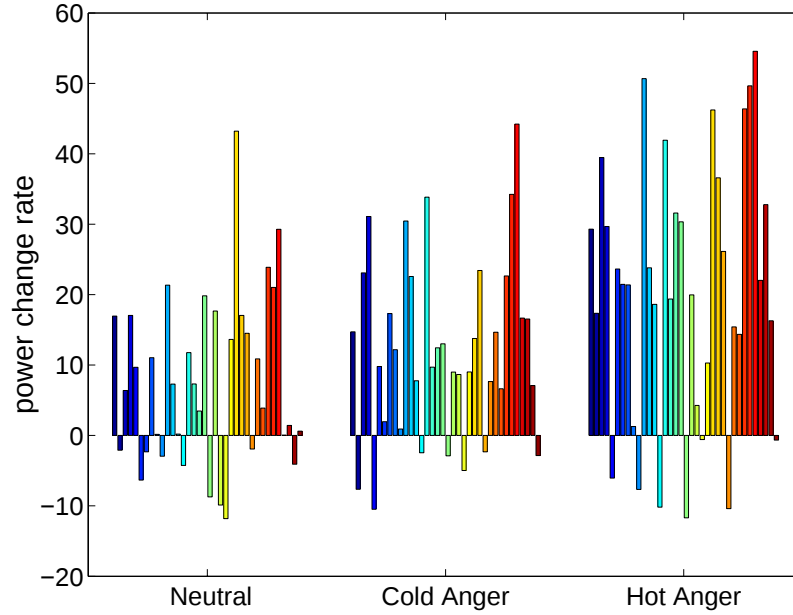


図 5.11: 各感情におけるアクセントレベル上昇時のパワー音韻変化率

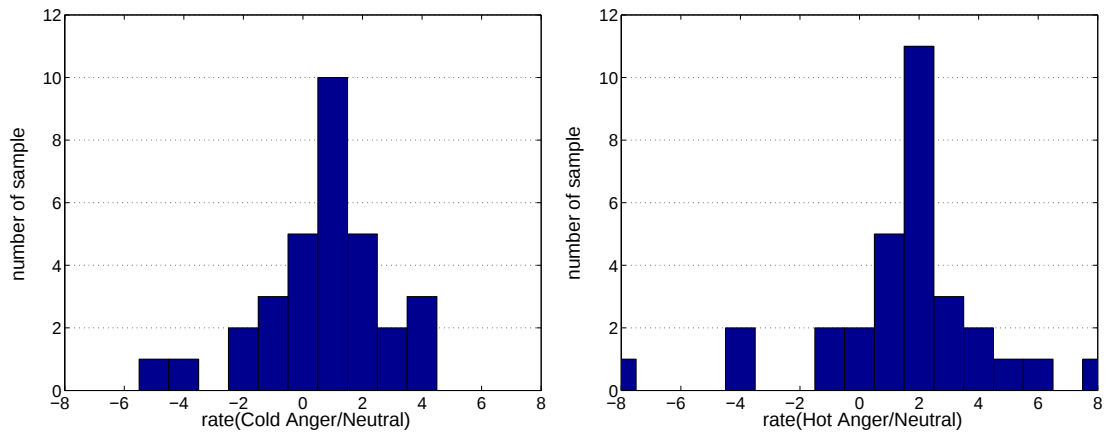


図 5.12: 左：パワー音韻変化率における「Cold Anger」の「Neutral」に対する比、右：パワー音韻変化率における「Hot Anger」の「Neutral」に対する比

サンプル番号順に並べたもの)。この結果から、「Cold Anger」「Hot Anger」の時間変化率の「Neutral」の時間変化率に対する比をサンプルごとにとりヒストグラムにしたものが、それぞれ図 5.14(左図・右図)である。

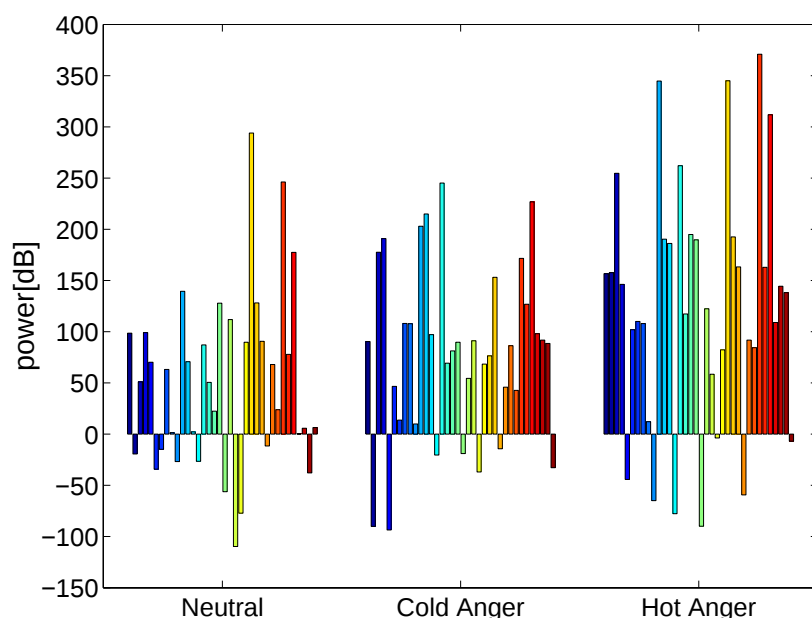


図 5.13: 各感情におけるアクセントレベル上昇時のパワー時間変化率

図 5.14(左図)を見ると、比が約 1 の付近にサンプルが集中している。これは音韻変化率の時と同様に、「Cold Anger」のパワー時間変化率が「Neutral」のパワー時間変化率とほぼ同じであるということを表している。図 5.14(右図)を見ると、比が約 2 の付近にサンプルが集中している。これは音韻変化率の時と同様に、「Hot Anger」のパワー時間変化率が「Neutral」のパワー時間変化率よりかなり大きいということを表している。

以上より、アクセントレベル上昇時においてパワーの音韻変化率・時間変化率は、「Cold Anger」がわずかに「Neutral」より大きいがほぼ同じくらいの傾向が見られ、「Hot Anger」は「Neutral」と比較してかなり大きな変化率を示した。

• アクセントレベル下降時

アクセントレベルが下降する時の各感情におけるパワーの音韻変化率を示したものが図 5.15 である (横軸は各感情におけるサンプルをサンプル番号順に並べたもの)。この結果から、「Cold Anger」「Hot Anger」の音韻変化率の「Neutral」の音韻変化率に対する比をサンプルごとにとりヒストグラムにしたものが、それぞれ図 5.16(左図・右図)である。

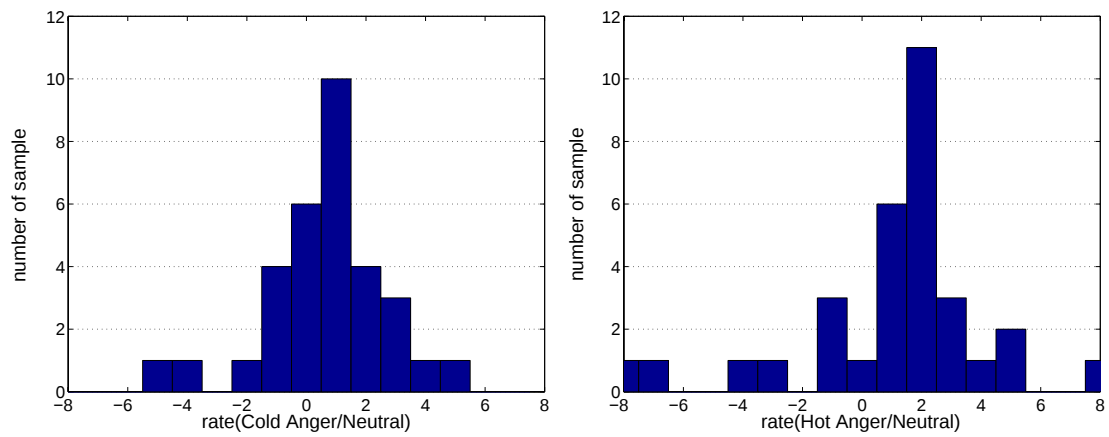


図 5.14: 左図：パワー時間変化率における「Cold Anger」の「Neutral」に対する比、右図：パワー時間変化率における「Hot Anger」の「Neutral」に対する比

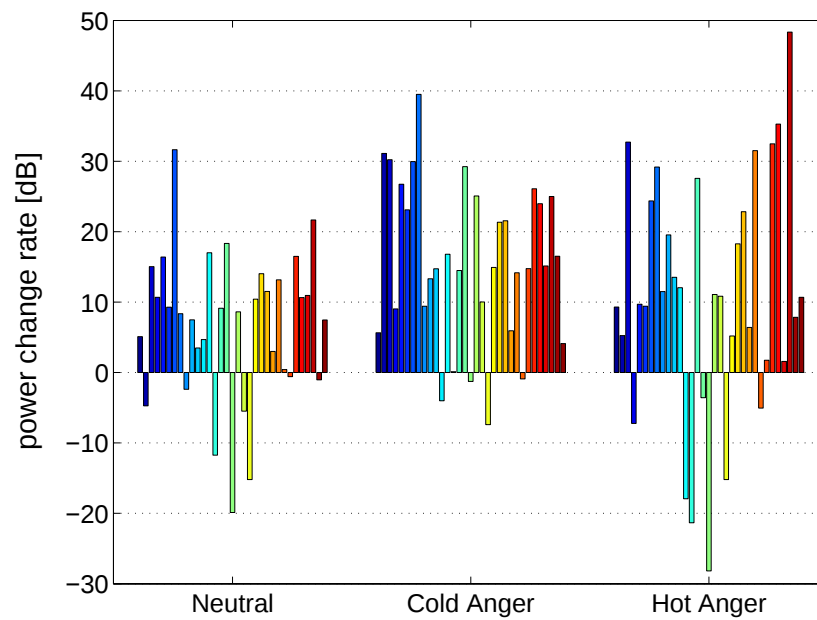


図 5.15: 各感情におけるアクセントレベル下降時のパワー音韻変化率

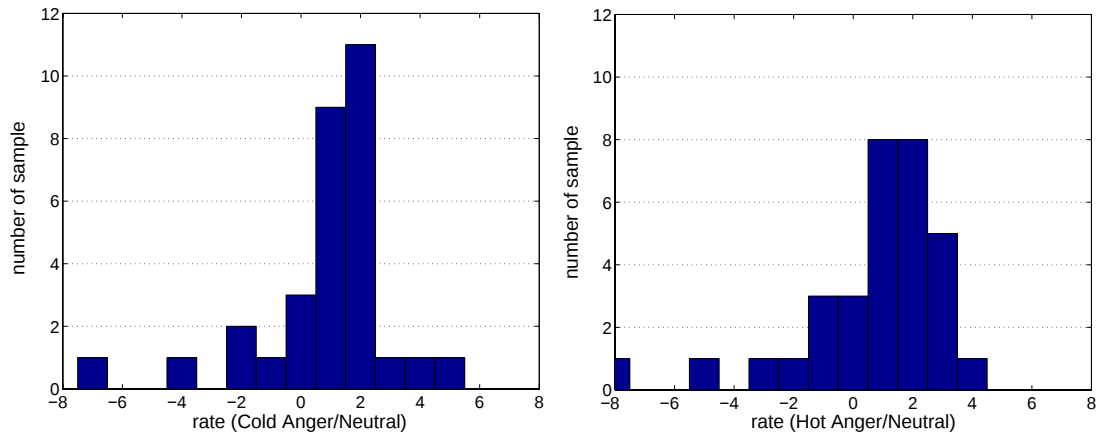


図 5.16: 左：パワー音韻変化率における「Cold Anger」の「Neutral」に対する比、右：パワー音韻変化率における「Hot Anger」の「Neutral」に対する比

図 5.16(左図)を見ると、比が 1 より大きい部分にサンプルが集中していることから、「Cold Anger」の音韻変化率が「Neutral」の音韻変化率よりも大きい傾向があることを示している。同様に、図 5.16(右図)を見ると、比が 1 より大きい部分にサンプルが集中していることから、「Hot Anger」の音韻変化率も「Neutral」の音韻変化率よりも大きい傾向があることを示している。

次に、アクセントレベル下降時におけるパワー時間変化率を検討する。アクセントレベルが下降する時の各感情におけるパワーの時間変化率を示したものが図 5.17 である(横軸は各感情におけるサンプルをサンプル番号順に並べたもの)。この結果から、「Cold Anger」「Hot Anger」の時間変化率の「Neutral」の時間変化率に対する比をサンプルごとにとりヒストグラムにしたものが、それぞれ図 5.18(左図・右図)である。

図 5.18(左図)を見ると、比が約 1 の付近にサンプルが集中している。これは音韻変化率の時と同様に、「Cold Anger」のパワー時間変化率が「Neutral」のパワー時間変化率とほぼ同じであるということを表している。図 5.18(右図)を見ると、比が約 2 の付近にサンプルが集中している。これは音韻変化率の時と同様に、「Hot Anger」のパワー時間変化率が「Neutral」のパワー時間変化率よりかなり大きいということを表している。

以上より、アクセントレベル下降時においてパワーの音韻変化率・時間変化率は、「Cold Anger」が「Neutral」よりわずかに大きい傾向が見られ、「Hot Anger」は「Neutral」と比較してかなり大きな変化率を示した。また、これらはアクセントレベル上昇時にも同様な傾向が見られた。

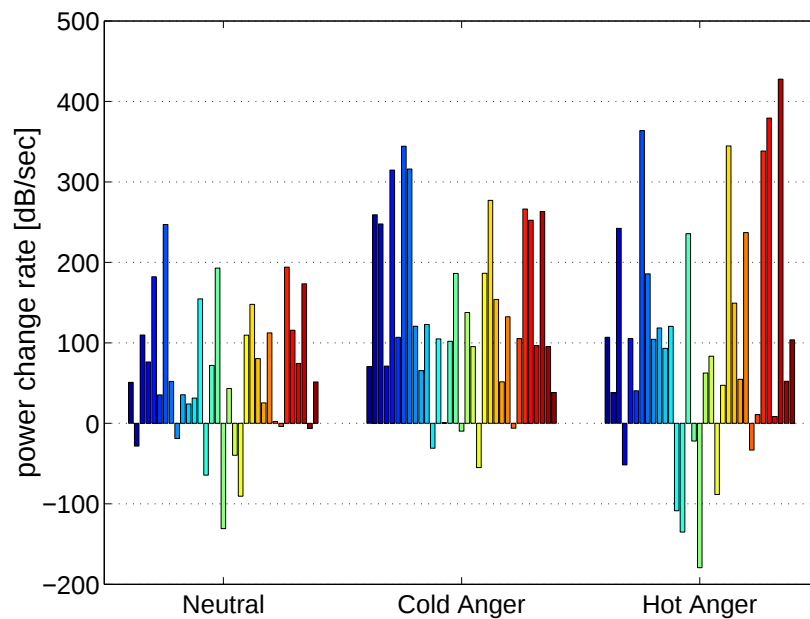


図 5.17: 各感情におけるアクセントレベル下降時のパワー時間変化率

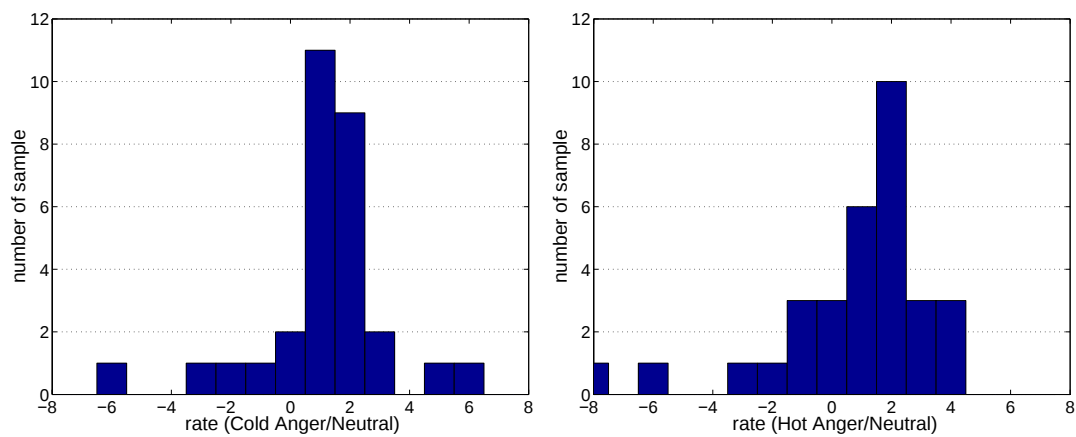


図 5.18: 左図 : パワー時間変化率における「Cold Anger」の「Neutral」に対する比、右図 : パワー時間変化率における「Hot Anger」の「Neutral」に対する比

5.4 持続時間の分析

5.4.1 分析手法

持続時間については、文全体、アクセント、アクセントの母音・子音、非アクセント、非アクセントの母音・子音をそれぞれ測定し、「Neutral」に対する「Cold Anger」「Hot Anger」の比を検討した。

5.4.2 分析結果・考察

まず、各文章の文全体について持続時間を測定して「Neutral」に対する「Cold Anger」「Hot Anger」の比を表したのがそれぞれ図 5.19(左図・右図)である(縦軸：サンプル数、横軸：対平静比)。

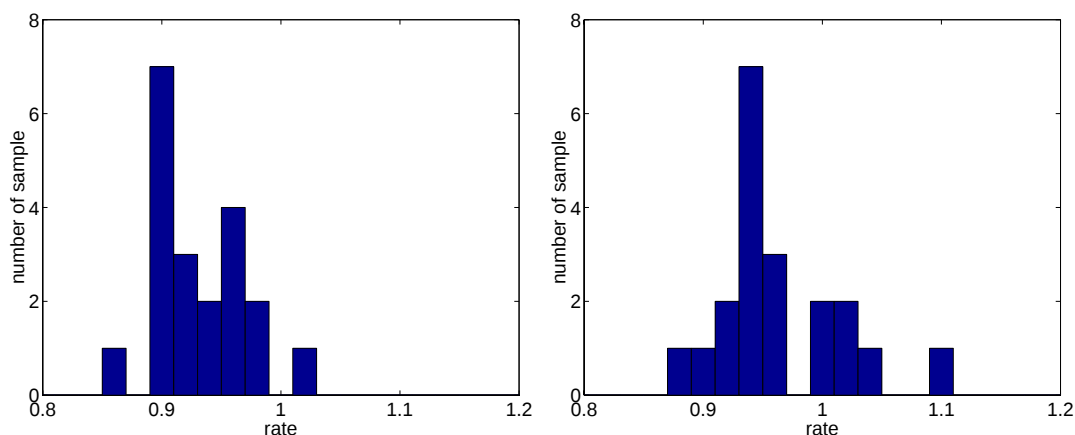


図 5.19: 左図:「Cold Anger」の「Neutral」に対する文全体の持続時間比、右図:「Hot Anger」の「Neutral」に対する文全体の持続時間比

図 5.19(左図) は、「Neutral」に対する比が 1 よりかなり小さい所にサンプルが集中している。これは「Cold Anger」における発話時間が明らかに「Neutral」より短いことを表している。図 5.19(右図) を見ると、同じく「Hot Anger」の発話時間は「Neutral」より短い傾向が見られるが、「Cold Anger」ほどではない。

次にアクセント・非アクセントにおける持続時間を検討した。図 5.20(左図・右図) は、それぞれ「Cold Anger」におけるアクセント・非アクセントの持続時間の「Neutral」に

対する比を表したものである（縦軸：サンプル数、横軸：対平静比）。

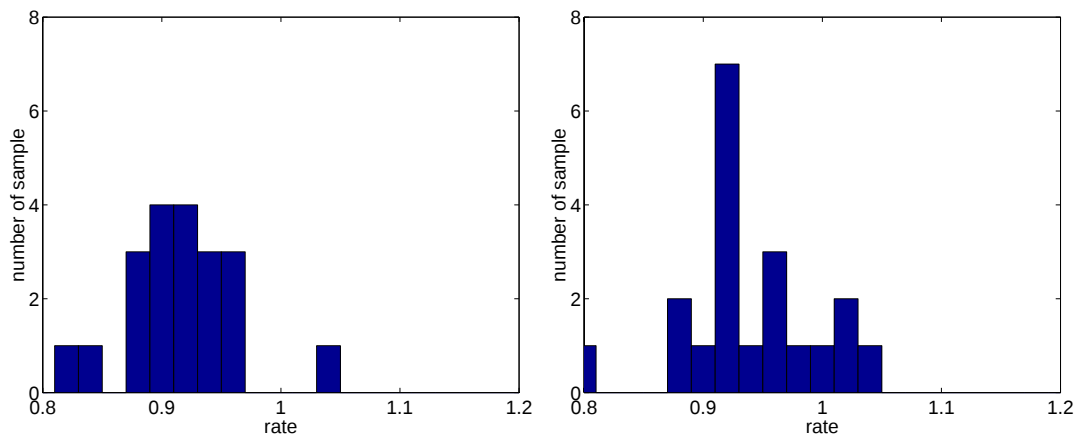


図 5.20: 左図:「Cold Anger」の「Neutral」に対するアクセントの持続時間比、右図:「Cold Anger」の「Neutral」に対する非アクセントの持続時間比

図 5.20(左図・右図)を見ると、二つの図においても「Neutral」に対する比が1より小さい所にサンプルが集まっている。これは、「Cold Anger」のアクセント・非アクセントの持続時間が「Neutral」より短いことを表しているが、図 5.19(右図)と比較すると図 5.20(左図)はやや左に、図 5.20(右図)はやや右にサンプルが集まっている。これは「Cold Anger」において、アクセントの持続時間はかなり短くなり、非アクセントはそれほど短くなっていないということを表している。同様に、「Hot Anger」におけるアクセント・非アクセント持続時間の「Neutral」に対する比を表したのが図 5.21(左図・右図)である。

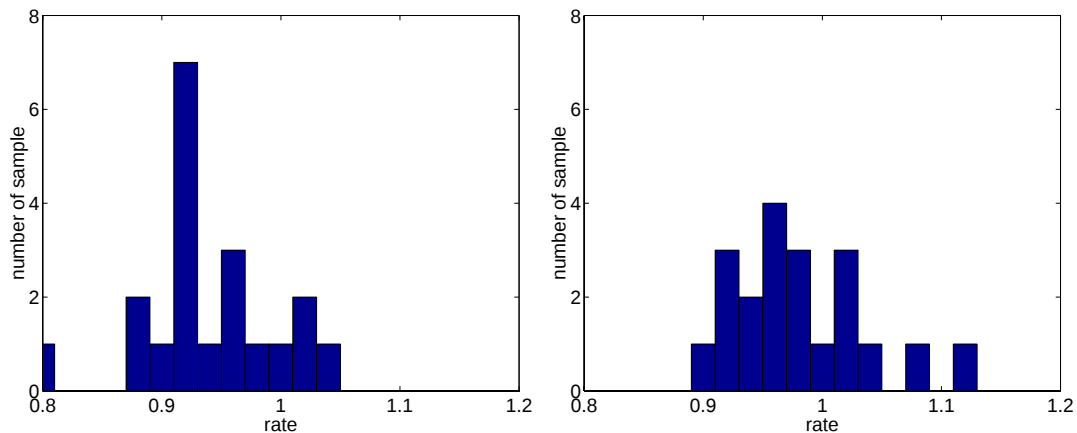


図 5.21: 左図:「Hot Anger」の「Neutral」に対するアクセントの持続時間比、右図:「Hot Anger」の「Neutral」に対する非アクセントの持続時間比

これを見ると「Hot Anger」のアクセント・非アクセントの持続時間も「Neutral」より短い傾向が見られる。さらに「Cold Anger」と同様にアクセントの持続時間はかなり短くなっているが、非アクセントの持続時間はそれほど短くなっていない。また、「Cold Anger」「Hot Anger」も「Neutral」に対して同じ様な傾向が見られるが、二つを比べると文全体・アクセント・非アクセントの3つとも「Cold Anger」が短いという傾向が見られる。

次にアクセント・非アクセントそれぞれにおける母音・子音の持続時間について検討した。図 5.22 はそれぞれ、「Cold Anger」アクセントにおける母音・子音の持続時間の「Neutral」に対する比を表した図である。

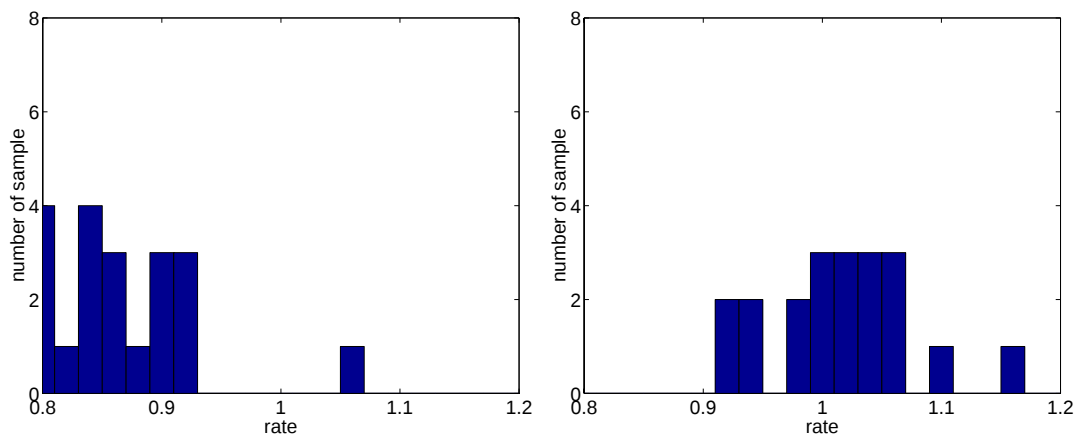


図 5.22: 左図:「Cold Anger」の「Neutral」に対するアクセント母音の持続時間比、右図:「Cold Anger」の「Neutral」に対するアクセント子音の持続時間比

これを見ると、アクセント母音に関して「Cold Anger」は「Neutral」に比べて持続時間がかなり短くなっている傾向が見られる。子音は比率が1のあたりにサンプルが集まっており、「Neutral」に比べてあまり変化していない傾向が見られる。図 5.23 はそれぞれ「Cold Anger」非アクセントにおける母音・子音の持続時間の「Neutral」に対する比を表した図である。

これを見ると、「Cold Anger」の非アクセントに関してもアクセントと同様に母音においては持続時間が「Neutral」より短くなっており、子音についてはほぼ同じという傾向が見られる。

また、「Hot Anger」についても同様にアクセント・非アクセントにおける母音・子音を測定したところ、「Cold Anger」のときと同じくアクセント・非アクセント共に母音の持続時間は「Neutral」に比べて短くなっているが、子音の持続時間は「Neutral」とほぼ同

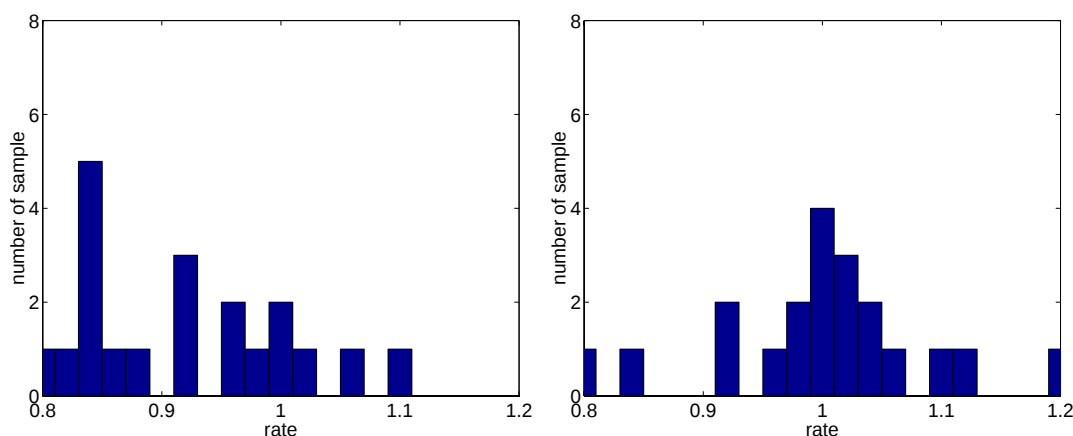


図 5.23: 左図 :「Cold Anger」の「Neutral」に対する非アクセント母音の持続時間比、右図 :「Cold Anger」の「Neutral」に対する非アクセント子音の持続時間比

じという傾向が見られた。ただし、「Cold Anger」と「Hot Anger」とで比較すると、母音の持続時間は「Cold Anger」の方が短い傾向が見られた。

また、全体的に見て「Cold Anger」「Hot Anger」の子音の持続時間は、「Neutral」とほぼ同じ傾向が見られたがサンプルによってはばらつきがあった。これは、/s/,/sh/といった摩擦音の持続時間が感情によってかなり変動があるためだと考えられる。

以上の分析から、ヒストグラムの平均値はそれぞれの特徴をよく表しているといえる。よって、この平均値を代表値として「Cold Anger」「Hot Anger」の「Neutral」に対する持続時間比を以下の表にまとめ、「Neutral」から「Cold Anger」「Hot Anger」の合成音を作成する際に、持続時間制御として適用する。

表 5.1: 「Neutral」に対する持続時間比の平均

Cold Anger				Hot Anger			
文全体				文全体			
0.93				0.96			
アクセント		非アクセント		アクセント		非アクセント	
0.91		0.95		0.94		0.98	
母音	子音	母音	子音	母音	子音	母音	子音
0.86	1.01	0.91	1.00	0.91	0.99	0.98	0.98

5.5 フォルマント周波数の分析

5.5.1 分析手法

フォルマント周波数

音源波形は、声道を通して開口端から放射されるとき、声道の持つ共鳴周波数の成分が特に強められ、特定の音色をもった音声信号として聴き取られる。この声道の共鳴周波数を、フォルマント周波数という。

フォルマント周波数は声道に起因し、基本周波数は声帯に起因するということにより生成機構が異なるため直接は連動していないが、音韻性を保つという点において連動していると考えられる。そこで、フォルマント周波数と基本周波数との関係を検討するという点でもフォルマント周波数の分析を行う必要がある。本研究では、LPC により予測パラメータからフォルマント周波数を推定する。

LPC によるフォルマント推定

音声波を周期 $\Delta T[\text{s}]$ で標本化した信号を $\{x_t\}(t: \text{は整数})$ とし、現在の標本値 x_t と、これに隣接する過去の p 個の標本値との間に、次の様な線形 1 次結合が成り立つと仮定する。

$$x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} = \varepsilon_t$$

これは線形予測係数を $\{\alpha_i\}$ 、線形予測残差を ε_t とする線形予測モデルということができる。この線形予測モデルの逆フィルタ

$$1 + \sum_{i=1}^p \alpha_i z^{-i}$$

に関して、伝達関数の極 $z_i : (i = 1, 2, \dots, p)$ は、予測多項式 $1 + \sum_{i=1}^p \alpha_i z^{-i} = 0$ を因数分解し、そうして得られた根として求まるが、根のうち複素数のものは 2 次の共振に対応し、その共振周波数と帯域幅は、次式で与えられる。

$$\begin{aligned} F_i &= \frac{\arg |z_i|}{2\pi\Delta T} [\text{Hz}] && : \text{フォルマント周波数} \\ B_i &= \frac{\log |z_i|}{\pi\Delta T} [\text{Hz}] && : \text{帯域幅} \end{aligned}$$

そして、スペクトル包絡の極のうち、比帯域幅 $\frac{B_i}{F_i}$ の小さいものを選ぶことにより、フォルマントを抽出できる。

フォルマント分析に対する線形予測法の利点

- 予測多項式の根を解くことによりフォルマントの中心周波数と帯域幅が精密に求められること
- どの極がどのフォルマントに対応するかを決定することは候補になる極の数が少ないので簡単（予測次数 p が先験的に定められているので、求められる複素共役極の最大数は $p/2$ である）
- LPC 分析では余分な極を除去するのはやさしい（余分な極の帯域幅は音声のフォルマントの帯域幅と考えられる値よりもはるかに大きい）

5.5.2 分析結果・考察

上述の方法で「Neutral」「Cold Anger」「Hot Anger」についてアクセントレベルが低から高へ上昇する時の高いモーラにおけるフォルマント周波数を分析する。尚、このLPC法によるフォルマントトラッキングはSpeechTools [22] を用いて行った。尚、分析は以下の条件で行った。

表 5.2: フォルマント周波数分析条件

サンプリング周波数	10[KHz]
窓関数	Hanning
フレーム長	30[msec]
シフト長	5[msec]
プリエンファシス	0.98
LPC 次数	12 次

分析結果を以下に示す。図 5.24 は、「Neutral」「Cold Anger」「Hot Anger」の感情におけるフォルマント周波数を第 1 から第 4 までの次数別に表したものである (横軸：各感情におけるサンプルをサンプル番号順に並べたもの、縦軸：フォルマント周波数 [kHz])。この結果から、まず「Neutral」と「Cold Anger」との関係を検討する。

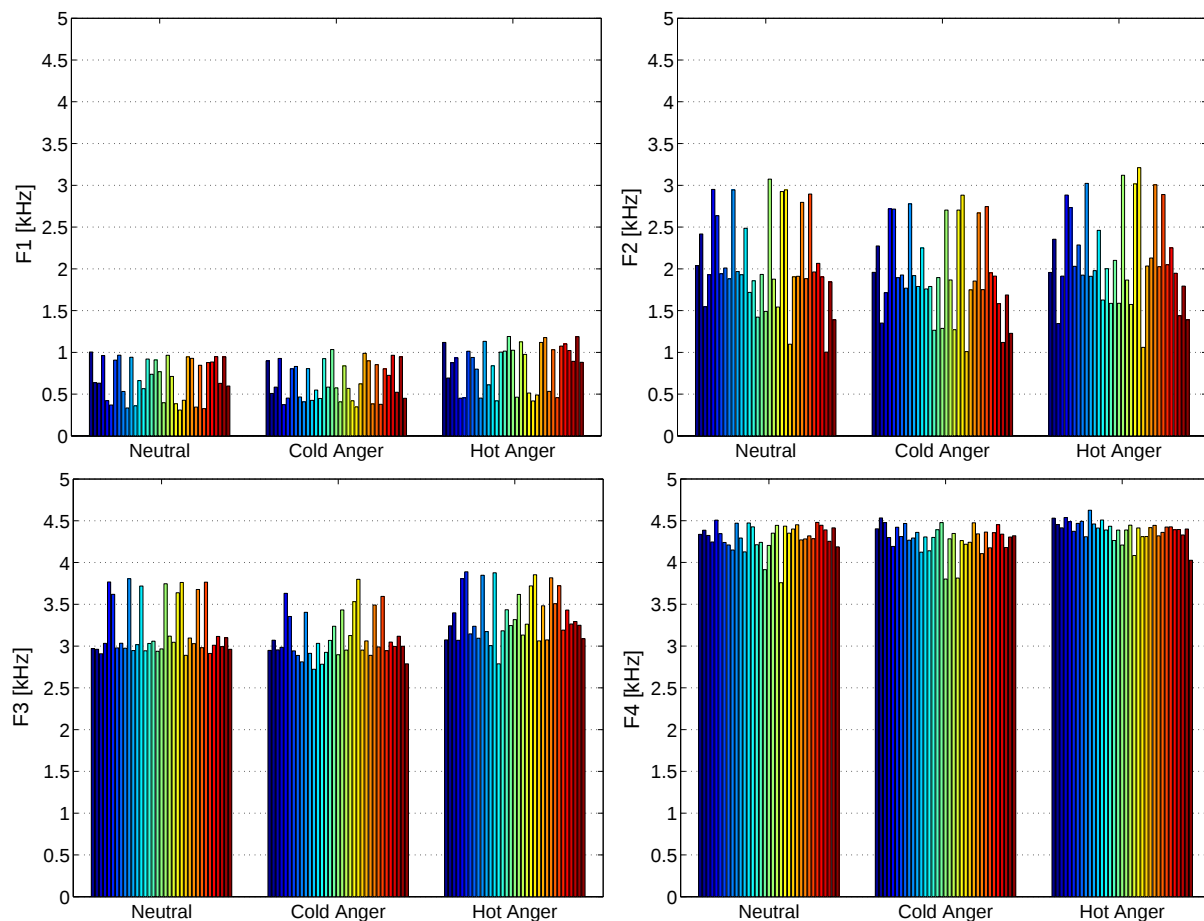


図 5.24: 感情間のフォルマント周波数 [kHz] の関係 (左上)F1(右上)F2(左下) F3(右下)F4

「Neutral」と「Cold Anger」のフォルマント周波数の関係をプロットしたものが図 5.25 である。図は、縦軸が「Neutral」におけるフォルマント周波数を常用対数で表示したもので、横軸は「Cold Anger」のフォルマント周波数を常用対数で表示したもので、プロットした点については「*」が第 1 フォルマント、「○」が第 2 フォルマント、「+」が第 3 フォルマント、「」が第 4 フォルマントを表している。そして、傾向を分かりやすくするため波線で $y = x$ を表示している。

この図を見ると、全体的にプロットした点は $y = x$ より上にある。これは、「Cold Anger」のフォルマント周波数が「Neutral」のフォルマント周波数より低いという傾向を示して

いる。中でも、F1 において最も「Neutral」からのずれが見られ、平均で 13 %の開きがあった。F2、F3、F4 とフォルマントの次数が高くなるにつれて 6 %、3 %、0.1 %のように、「Neutral」と「Cold Anger」との開きは小さくなっていき、F4 ではほとんど同じくらいになるという傾向も見られた。

ここで一般的に F1 ~ F3 は音韻に、高次のフォルマントは個人性に依存するといわれているので [23][24]、本研究では F1 ~ F3 をみることにする。そして、F1 ~ F3 のプロットされた分析結果から、最小自乗的に 2 次式で近似すると、 $y = 0.1147x^2 + 0.8988x + 0.0372$ という関係式が導かれる。これは図 5.26 で実線で表示した式であるが、この式を本研究で合成音を作成する際の「Cold Anger」フォルマント周波数制御に適用することにする。

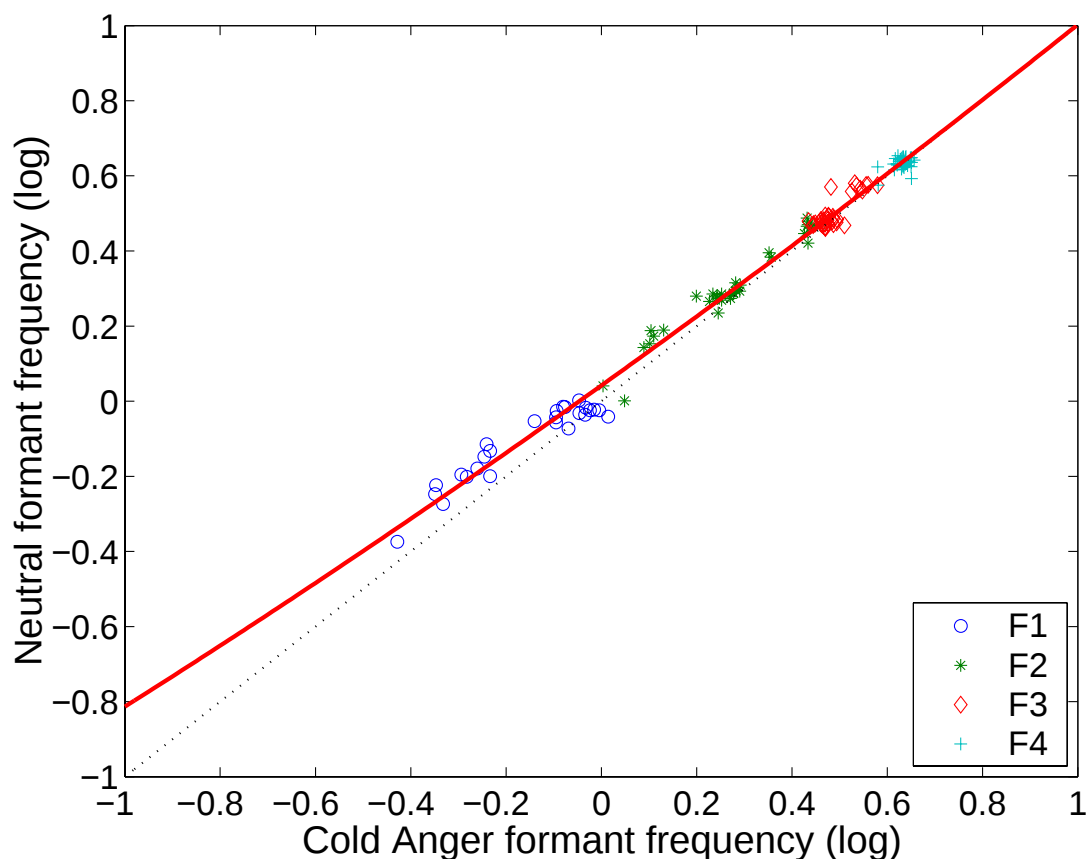


図 5.25: 「Neutral」と「Cold Anger」のフォルマント周波数（常用対数）の関係

同様に、「Neutral」と「Hot Anger」のフォルマント周波数の関係をプロットしたものが図 5.26 である。

この図を見ると、全体的にプロットした点は $y = x$ より下にある。これは、「Hot Anger」のフォルマント周波数が「Neutral」のフォルマント周波数より高いという傾向を示して

いる。中でも、F1において最も「Neutral」からのずれが見られ、平均で19%の開きが見られる。F2、F3、F4とフォルマントの次数が高くなるにつれて9%、5%、1%というように、「Neutral」と「Hot Anger」との開きは小さくなっていき、F4ではほとんど同じくらいになるという傾向も見られた。このフォルマントの次数が高くなるにつれて、「Neutral」との差が小さくなるというのは「Cold Anger」と同じ傾向である。

ここで、「Cold Anger」の時と同様に最小二乗的に2次式で近似すると、 $y = -0.0083x^2 + 1.1007x - 0.0668$ という関係式が導かれる。この式は、図5.26において実線で表示したもののだが、これを本研究で合成音を作成する際の「Hot Anger」フォルマント周波数制御に適用することにする。

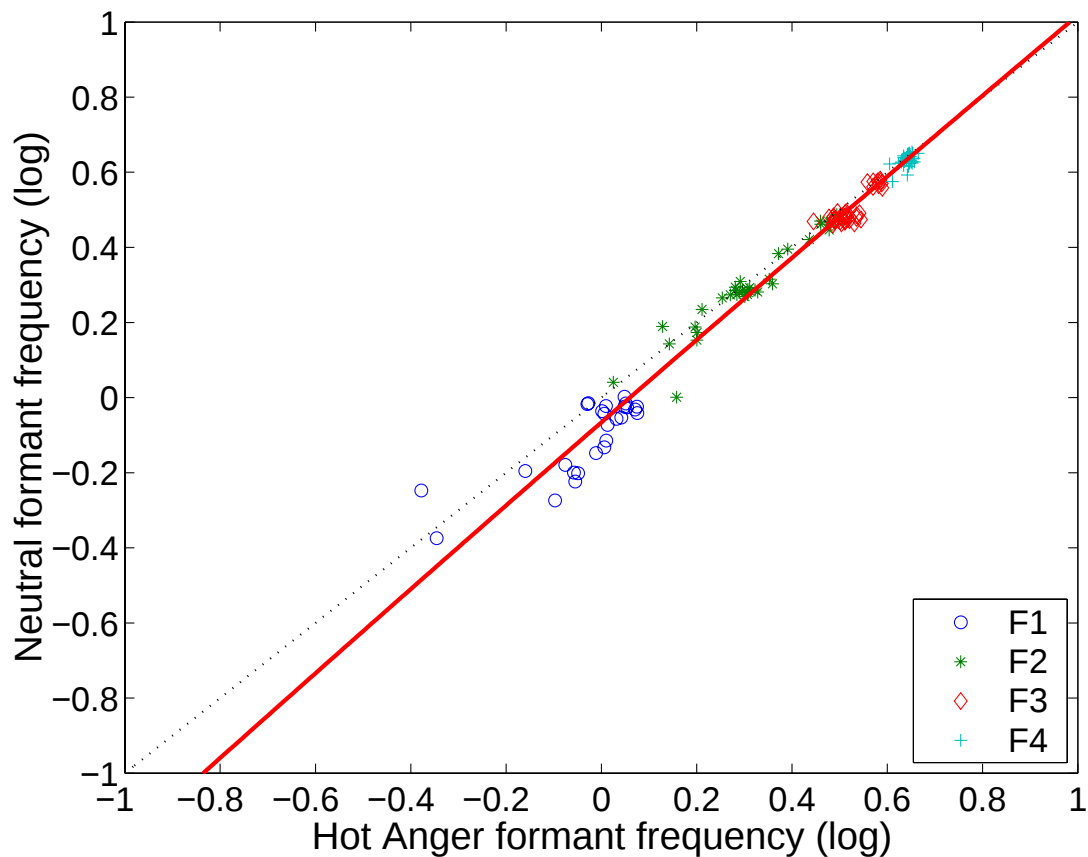


図 5.26: 「Neutral」と「Hot Anger」のフォルマント周波数（常用対数）の関係

5.6 スペクトルの分析

5.6.1 分析手法

興奮状態におけるスペクトル

人は興奮状態で発声すると、血圧が上がり、声門下圧が上昇し、声帯の緊張が強くなる [11]。そのような強い (パワーが大きい) 声の時、声門の閉鎖期間が長く、音源信号の波形は急峻な三角波に近づくので [25]、スペクトルの傾きは緩やかになり、声帯が緊張していないときと比較して高次の高調波が大きくなる。この様に、パワーとスペクトルの間には関連性があると思われる。「Cold Anger (押し殺した怒り)」「Hot Anger (激しい怒り)」という感情はどちらも怒っている感情であり、当然興奮してパワーが大きいので、「Neutral」のスペクトルと比較して高域成分が強調されているのではないかと考えられる。

スペクトルの抽出法

スペクトル分析において、パワースペクトルはSTRAIGHTを用いて振幅スペクトルを抽出し、各周波数成分を二乗して求めた。その際にSTRAIGHTで振幅スペクトルを抽出する部分としては、アクセントレベルが低レベルから高レベルへ移るときの低・高レベルモーラにおける母音区間とし、その区間内で時間平均をとったものを振幅スペクトルとして分析に用いる。

さらに、人間の内耳の基底膜の周波数特性を考慮した表現に対応させるため、周波数軸を ERB rate [26] に変換した。ERB rate は等価矩形帯域幅 (Equivalent Rectangular Bandwidth, ERB) を幅 1 として周波数軸を変形したものである。ERB と ERB rate はそれぞれ以下の式で求められる。

$$\begin{aligned} ERB &= 24.7(4.37F + 1) \\ ERBrate &= 21.4 \log_{10}(4.37F + 1) \end{aligned} \quad (5.1)$$

ここで、 F は周波数 (kHz) である。式 (5.1) は基底膜上の周波数マッピングを近似的に求めている [26]。

そして、「Neutral」「Cold Anger」「Hot Anger」という感情ごとのスペクトル概形の特徴をみるために、35 ERB rate (9.6574 kHz) までを $0 \sim 5$ ERBrate ($0 \sim 0.1631$ kHz) ・ 5

～ 10 ERBrate(0.1631 ～ 0.4423 kHz) ・ 10 ～ 15 ERBrate(0.4423 ～ 0.9203 kHz) ・ 15 ～ 20 ERBrate(0.9203 ～ 1.7395 kHz) ・ 20 ～ 25 ERB rate(1.7395 ～ 3.1420 kHz) ・ 25 ～ 30 ERB rate(3.1420 ～ 5.5440 kHz) ・ 30 ～ 35 ERB rate(5.5440 ～ 9.6574 kHz) という 7 つの帯域に分割してスペクトル全体の平均パワーを基準としたときの各々の帯域におけるパワーを調べる。

5.6.2 分析結果・考察

分析結果を以下に示す。例として「Neutral」「Cold Anger」「Hot Anger」の 30 ～ 35 ERB rate における帯域でのスペクトル分析結果を全サンプルのヒストグラムで表したものが、それぞれ図 5.27, 図 5.28, 図 5.29 である（横軸：パワー [dB], 縦軸：サンプル数）。

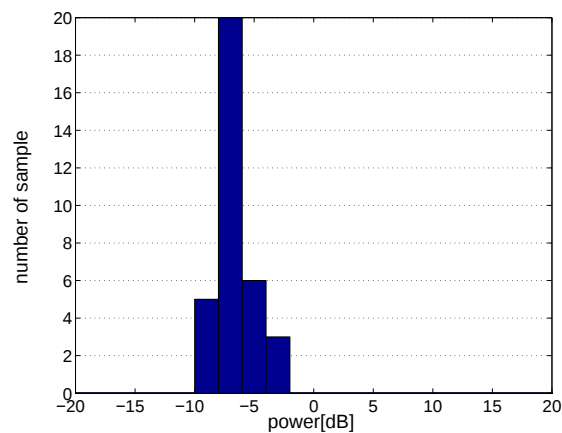


図 5.27: 「Neutral」 30 ～ 35ERB rate

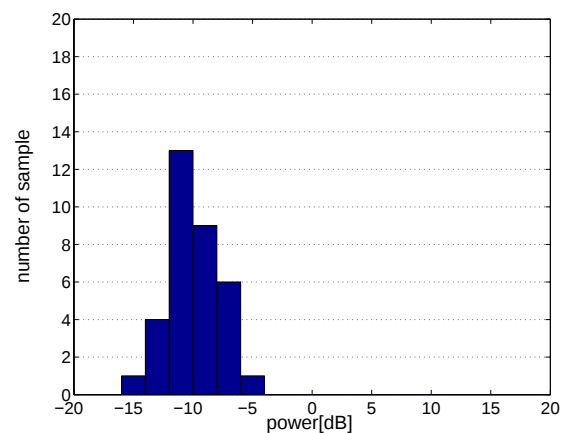


図 5.28: 「Cold Anger」 30 ～ 35ERB rate

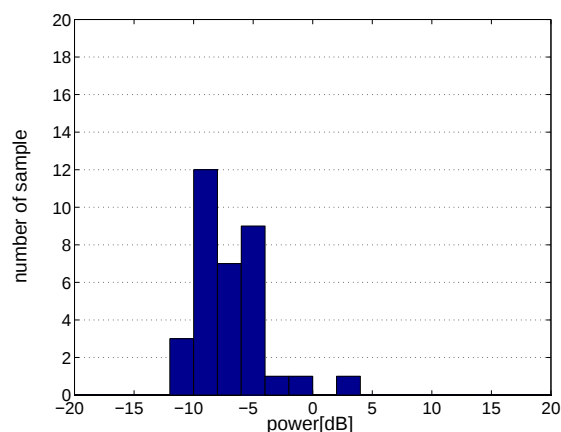


図 5.29: 「Hot Anger」 30 ~ 35ERB rate

この三つの図の全サンプルの平均値は、それぞれ「Neutral」が -6.6 [dB]、「Cold Anger」が -9.9 [dB]、「Hot Anger」が -6.8 [dB] となりヒストグラムから得られる傾向を良く表しているといえる。他の帯域でも同じことがいえるので、この全サンプルの平均値を本研究では各帯域でのスペクトル代表値として扱うことにする。この代表値を各帯域の中間値（例えば 30 ~ 35 ERB rate の場合 32.5 ERB rate）におけるパワースペクトル値として、各感情におけるスペクトル概形の傾向を検討する。

表 5.3 は、アクセント低レベルにおけるスペクトル分析結果をまとめたものである。カッコ内は、「Cold Anger」「Hot Anger」のそれぞれの帯域におけるパワーの「Neutral」に対する差を表したものである。

表 5.3: アクセント低レベルのスペクトル分析結果 (単位: [dB])

	0 ~ 5 ERB rate	5 ~ 10 ERB rate	10 ~ 15 ERB rate	15 ~ 20 ERB rate	20 ~ 25 ERB rate	25 ~ 30 ERB rate	30 ~ 35 ERB rate
Neutral	13.6	25.4	20.1	10.3	2.8	0.1	-6.8
Cold Anger	17.3 (+3.7)	26.2 (+0.8)	19.8 (-0.3)	12.4 (+2.1)	3.5 (+0.7)	1.6 (+1.5)	-8.6 (-1.8)
Hot Anger	1.7 (-11.9)	15.1 (-10.3)	16.1 (-4.0)	11.6 (+1.3)	2.7 (-0.1)	4.2 (+4.1)	-7.5 (-0.7)

このスペクトル概形の傾向を分かり易くするため、上記の値をプロットし、プロットした点をスプライン関数で補間したものが図 5.30 である。図は「○」が「Neutral」,「◇」が「Cold Anger」,「+」が「Hot Anger」を表している (横軸: ERB rate, 縦軸: パワー [dB])。

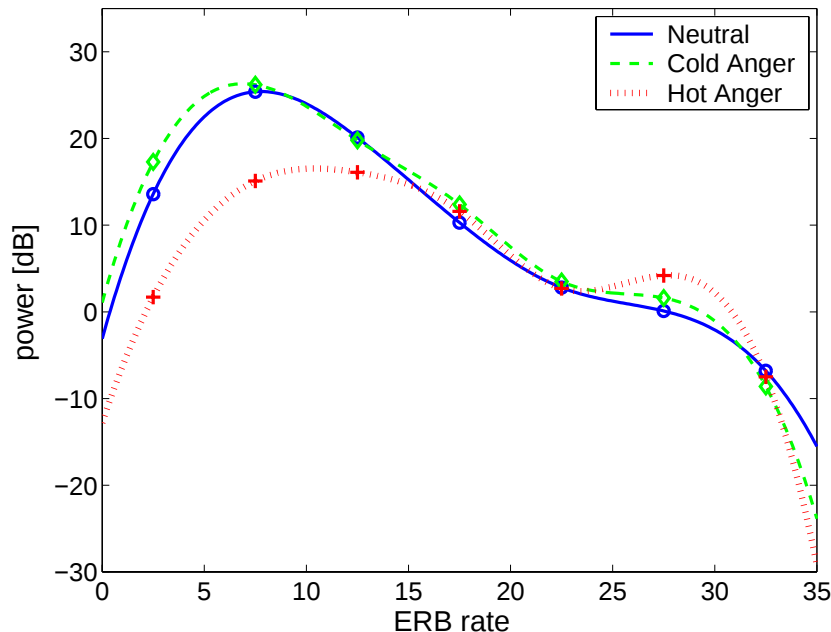


図 5.30: 各感情のアクセント低レベルにおけるスペクトル概形の傾向

これを見ると、2.5 ~ 7.5 ERB rate においては各感情のスペクトルの傾きはほぼ等しく、一様に増加している傾向が見られる。5 ~ 10 ERB rate を過ぎると、感情間で違いが見られ、「Neutral」「Cold Anger」は減少していき、「Hot Anger」は 10 ~ 15 ERB rate も同じくらいのパワーで 15 ERB rate 以上になると減少している。「Neutral」と「Cold Anger」は、ほぼ同じ様なスペクトル概形を示しているが、「Hot Anger」は 25 ~ 30 ERB rate において「Neutral」よりもパワーが強調される傾向が見られる。「Cold Anger」「Hot Anger」は、30 ~ 35 ERB rate において「Neutral」と比較するとパワーが強調されていないという傾向も見られる。

以上の傾向を踏まえて、本研究では最終的に合成音を作成する際に以下の手法をとる。まず、7.5 ERB rate まで「Cold Anger」「Hot Anger」のスペクトルを「Neutral」と同じに合わせる。そして 7.5 ERB rate 以上の帯域では、先にプロットした点に関して最小二乗的に近似した曲線を各感情のスペクトルとする。図 5.31 は近似式を表したもので、「Neutral」に関しては $y = 0.0007x^3 - 0.0254x^2 - 1.1727x + 35.8804$ 、「Cold Anger」

に関しては $y = -0.0008x^3 + 0.051x^2 - 2.3178x + 40.6549$ 、「Hot Anger」に関しては $y = 0.0004x^3 - 0.0544x^2 + 0.6973x + 23.7236$ という式が導かれ、合成音を作成する際には、この近似式を適用する。

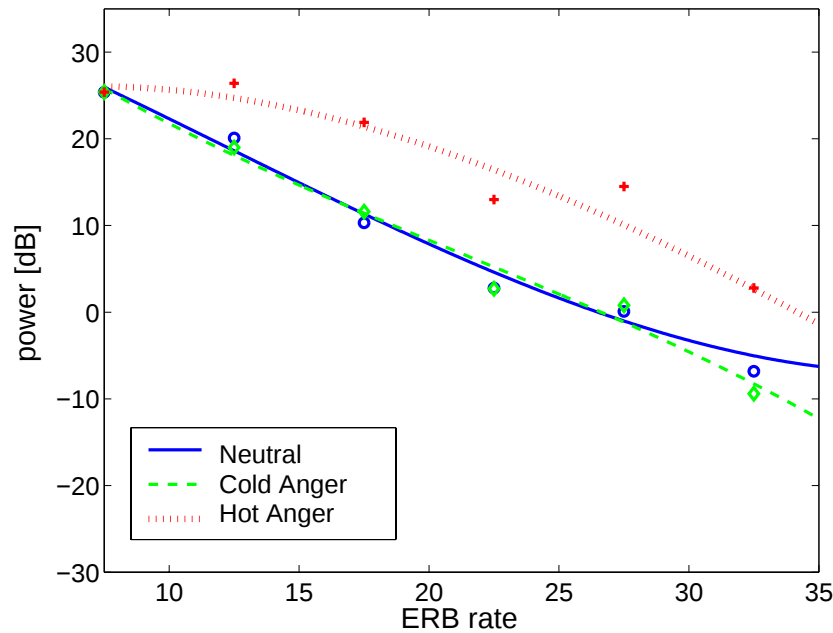


図 5.31: 合成音を作成する際の各感情のスペクトル

次に、アクセント高レベルにおけるスペクトル分析結果を表 5.4 に示す。カッコ内は、「Cold Anger」「Hot Anger」のそれぞれの帯域におけるパワーの「Neutral」に対する差を表したものである。

アクセント低レベルの時と同様に、このスペクトル概形の傾向を分かり易くするため、上記の値をプロットし、プロットした点をスプライン関数で補間したものが図 5.32 である。図は「○」が「Neutral」,「◇」が「Cold Anger」,「+」が「Hot Anger」を表している (横軸：ERB rate, 縦軸：パワー [dB])。

これを見ると、アクセント低レベルの時と同じく 2.5～7.5 ERB rate においては各感情のスペクトルの傾きはほぼ等しく、一様に増加している傾向が見られる。5～10 ERB rate を過ぎると、感情間で違いが見られ、「Neutral」「Cold Anger」は減少していき、「Hot Anger」は 10～15 ERB rate も増加している。「Neutral」と「Cold Anger」は、ほぼ同じ様なスペクトル概形を示しているが、25～30 ERB rate の帯域を中心に「Cold Anger」は「Neutral」よりもパワーが強調される傾向が見られる。「Hot Anger」も 25～30 ERB rate の帯域を中心に「Neutral」よりもパワーが強調される傾向が見られる。そしてアクセン

表 5.4: アクセント高レベルのスペクトル分析結果 (単位: [dB])

	0 ~ 5 ERB rate	5 ~ 10 ERB rate	10 ~ 15 ERB rate	15 ~ 20 ERB rate	20 ~ 25 ERB rate	25 ~ 30 ERB rate	30 ~ 35 ERB rate
Neutral	9.5	22.8	19.9	11.5	4.7	-1.0	-6.6
Cold Anger	14.2 (+4.7)	24.7 (+1.9)	19.2 (-0.7)	12.4 (+0.9)	7.6 (+2.9)	2.2 (+3.2)	-9.9 (-3.3)
Hot Anger	-7.8 (-17.3)	3.4 (-19.4)	12.5 (-7.4)	10.8 (-0.7)	6.4 (+1.7)	3.7 (+4.7)	-6.8 (-0.2)

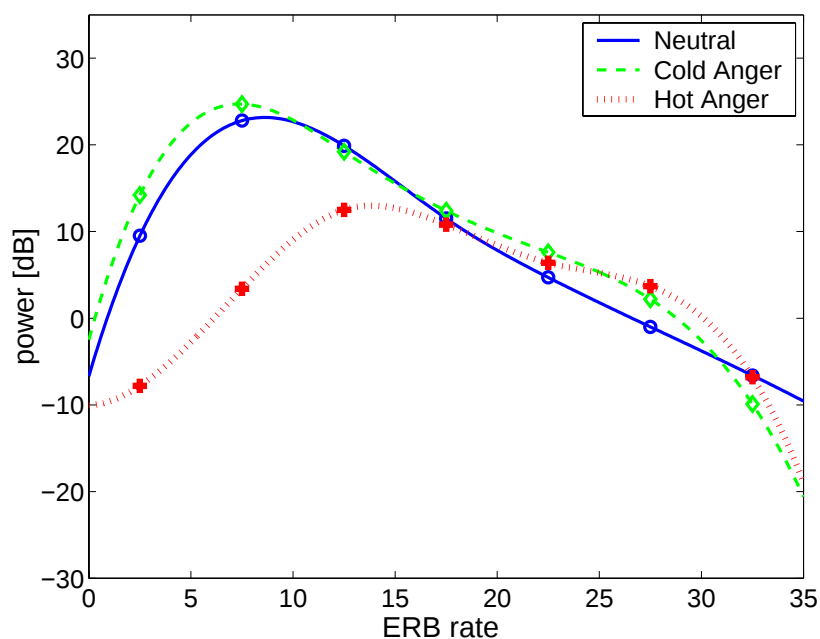


図 5.32: 各感情のアクセント高レベルにおけるスペクトル概形の傾向

ト低レベルの時と同様に、「Cold Anger」「Hot Anger」は、30～35 ERB rate において「Neutral」と比較してパワーが強調されなくなるという傾向も見られる。

以上の傾向を踏まえて、アクセント高レベルにおける合成音を作成する際にもアクセント低レベルと同様の手法をとる。図 5.33 は近似式を表したもので、「Neutral」に関しては $y = 0.002x^3 - 0.1197x^2 + 0.9477x + 21.9148$ 、「Cold Anger」に関しては $y = -0.0026x^3 + 0.1351x^2 - 3.2925x + 41.3271$ 、「Hot Anger」に関しては $y = 0.0021x^3 - 0.1982x^2 + 4.6898x - 1.5185$ という式が導かれ、合成音を作成する際には、この近似式を適用する。

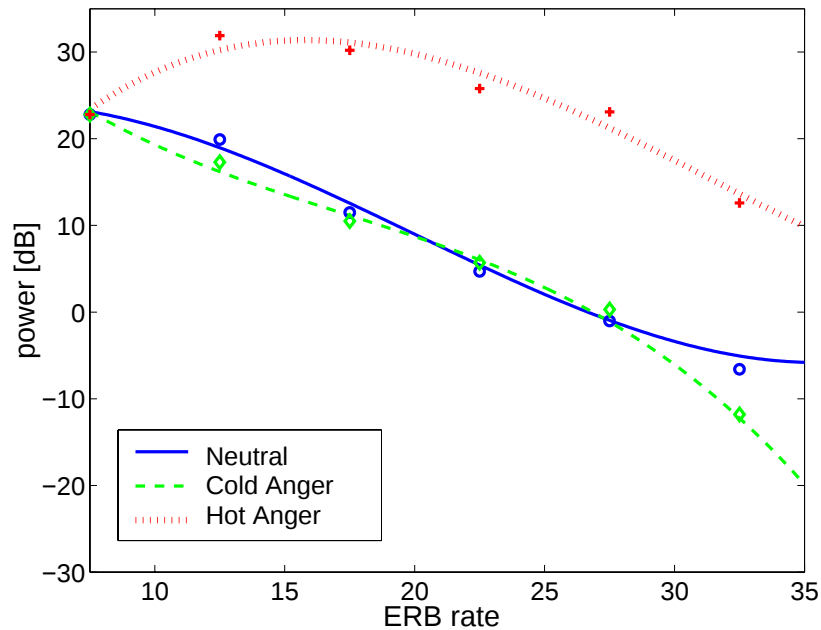


図 5.33: 合成音を作成する際の各感情のスペクトル

第6章 全体のまとめ・考察

本研究は、「Neutral」「Cold Anger」「Hot Anger」の感情について文章のアクセント部における音響特徴量の変化を分析した。音響特徴量としては、基本周波数・パワー・持続時間・フォルマント周波数・スペクトルを扱った。以下にそれぞれの音響特徴量の分析で得られた結果をまとめ、考察する。

基本周波数

基本周波数は、文章の有声区間における長時間平均、アクセントにおける時間変化率・音韻変化率について分析を行った。

- 文章の有声区間における長時間平均

「Cold Anger」「Neutral」「Hot Anger」の順で大きい傾向が見られた。

- アクセントレベル上昇時における音韻変化率・時間変化率

音韻変化率・時間変化率ともに「Cold Anger」「Neutral」「Hot Anger」の順で大きい傾向が見られた。「Hot Anger」はサンプルによってばらつきが見られた。これは、文脈上で強調する部分の変化率が「Hot Anger」で特に大きいためと考えられる。

- アクセントレベル下降時における音韻変化率・時間変化率

音韻変化率・時間変化率ともに、「Cold Anger」は「Neutral」と同じくらいの大きさで、「Hot Anger」は「Neutral」と比較して大きい傾向が見られた。「Neutral」に対する「Cold Anger」「Hot Anger」の比はともに、アクセントレベル上昇時と比較して小さい傾向が見られた。これは、「Cold Anger」「Hot Anger」のアクセントレベル上昇時から下降時にかけてのアクセント高レベル部分の基本周波数が「Neutral」に比べて急激に減少しているためだと考えられる。

これらの分析から、基本周波数に関して文全体の有声区間における長時間平均とアクセント上昇時における変化率は感情判別に際して大きな特徴であると考えられる。

また、分析結果より「Neutral」から「Cold Anger」「Hot Anger」の合成音を作成するときの基本周波数制御を導いた。文全体の基本周波数に関しては、「Neutral」に対して「Cold Anger」は0.68倍、「Hot Anger」に対しては1.5倍とする。アクセントレベル上昇時に関しては、「Neutral」の変化率に対して「Cold Anger」は0.95倍、「Hot Anger」に対しては1.23倍とする。アクセントレベル下降時に関しては、「Neutral」の変化率に対して「Cold Anger」は1.0倍、「Hot Anger」に対しては1.1倍とする。

パワー

パワーは、アクセントにおける時間変化率・音韻変化率について分析を行った。

- アクセントレベル上昇時における時間変化率・音韻変化率

音韻変化率・時間変化率ともに、「Cold Anger」は「Neutral」と比較してわずかに大きく、「Hot Anger」は「Neutral」と比較してかなり大きい傾向が見られた。

- アクセントレベル下降時における時間変化率・音韻変化率

音韻変化率・時間変化率ともに、「Cold Anger」は「Neutral」と比較してわずかに大きく、「Hot Anger」は「Neutral」と比較してかなり大きい傾向が見られた。

これらの分析から、「Hot Anger」においてパワーの寄与は大きいと考えられる。そして分析結果より、「Neutral」から「Cold Anger」「Hot Anger」の合成音を作成する際に、アクセントレベル上昇時においてパワーは「Neutral」の変化率に対して「Cold Anger」は1.14倍、「Hot Anger」は1.9倍とする。アクセントレベル下降時においてパワーは「Neutral」の変化率に対して「Cold Anger」は1.15倍、「Hot Anger」は1.3倍とする。

持続時間

持続時間は各感情について、文全体、アクセント、及びアクセントにおける母音・子音、非アクセント、及び非アクセントにおける母音・子音について分析を行った。

その結果、文全体においては「Cold Anger」「Hot Anger」「Neutral」の順で持続時間が短い傾向が見られた。「Cold Anger」「Hot Anger」は、アクセント・非アクセント部に

においても「Neutral」と比較して短い、非アクセント部はアクセント部ほど短くなっていない傾向が見られた。さらに、アクセント部・非アクセント部の母音においても「Cold Anger」「Hot Anger」は「Neutral」と比較して短くなっていたが、子音は「Neutral」とそれほど変わらない傾向が見られた。

このことから、「Cold Anger」「Hot Anger」は文全体においても「Neutral」と比較して短くなるが、その中でも非アクセントよりアクセントで、しかも特に母音で短くなると考えられる。そして、子音以外の全ての部分で「Cold Anger」は「Hot Anger」よりもさらに短縮されており、持続時間が「Cold Anger」という感情を知覚することにおいて重要な音響特徴量となっていると考えられる。また、子音は「Neutral」と「Cold Anger」「Hot Anger」とではほとんど変わらないが、摩擦音においては感情間でばらつきが見られた。このことから、感情間の持続時間差は母音の影響がかなり大きいと考えられる。

分析結果より、「Neutral」から「Cold Anger」「Hot Anger」の合成音を作成するための規則を導いた。「Neutral」から「Cold Anger」の音声を作成する際には、アクセントの母音は0.86倍、非アクセントの母音は0.91倍、子音については変化させない。「Neutral」から「Hot Anger」の音声を作成する際には、アクセントの母音は0.91倍、非アクセントの母音は0.98倍、子音については変化させない。

フォルマント周波数

「Cold Anger」「Hot Anger」のアクセント部におけるフォルマント周波数を「Neutral」のそれと比較した。

「Cold Anger」は「Neutral」と比較すると全体的に低く、特にF1はかなり低い傾向が見られた。フォルマントの次数が上がるにつれ、「Neutral」との開きは小さくなり、F4においてはほとんど同じくらいという傾向が見られた。

「Hot Anger」は、逆に「Neutral」と比較して全体的に高く、特にF1はかなり高い傾向が見られた。「Cold Anger」の時と同様に、フォルマントの次数が上がるにつれ、「Neutral」との開きは小さくなり、F4においてはほとんど同じくらいという傾向が見られた。よって、「Cold Anger」と「Hot Anger」とを判別するためには、低次のフォルマント周波数というものが大きく寄与するのではないかと考えられる。

また、分析より得られた結果から合成音を作成するための周波数変換マップを導いた。

スペクトル

「Cold Anger」「Hot Anger」についてアクセントレベル変化時のアクセント低・高レベルにおけるスペクトルを「Neutral」と比較した。

• アクセント低レベルにおけるスペクトル

「Cold Anger」は「Neutral」とほぼ同じスペクトル概形を示し、5～10 ERB rate までは、パワーが増加していき、5～10 ERB rate を過ぎると減少する。

「Hot Anger」は5～10 ERB rate まで増加し、10～15 ERB rate の帯域でも同じくらいのパワーを持続し、10～15 ERB rate を過ぎると減少する。25～30 ERB rate において「Neutral」と比較して、パワーが強調される傾向が見られた。

「Cold Anger」「Hot Anger」ともに30～35 ERB rate において、急激にパワーが減少する傾向が見られた。

• アクセント高レベルにおけるスペクトル

「Cold Anger」は「Neutral」とほぼ同じスペクトル概形を示し、5～10 ERB rate までは、パワーが増加していき、5～10 ERB rate を過ぎると減少する。25～30 ERB rate を中心に「Neutral」と比較してスペクトルが強調されるが、30～35 ERB rate において急激にパワーが減少する傾向が見られた。

「Hot Anger」は10～15 ERB rate まで増加し、10～15 ERB rate を過ぎると減少する。アクセント低レベルの時と同様、25～30 ERB rate において「Neutral」と比較して、スペクトルが強調される傾向が見られたが、アクセント低レベルの時よりさらに強調されている。30～35 ERB rate において急激にパワーが減少する傾向が見られた。

分析結果より、興奮状態のときに高域のパワーが強調されるのではないかという指摘は正しかったといえる。そして、高域といっても25～30 ERB rate でパワーが強調されることがわかった。また、分析結果からアクセント低・高レベルともに「Neutral」から「Cold Anger」「Hot Anger」の合成音を作成する際の近似式を導いた。

音響特徴量間の関係

音声の生成機構は互いに関連しているため、それに伴い音響特徴量の間にも関連があると考えられる。本研究においては、「Neutral」と比較して「Hot Anger」では全体の基本周波数が高く、アクセントにおける変化率も大きかった。また、フォルマント周波数も $F1 \cdot F2$ などで「Neutral」よりも大きく、スペクトルにおいては高域で強調される傾向が見られた。

これを音声の生成機構からの観点で考察する。まず「Hot Anger」は怒っているので興奮状態にある。興奮状態で発声すると、血圧が上がり、声門下圧が上昇し、声帯の緊張が強くなる。そして、声帯が引き伸ばされることにより声帯粘膜の張力が増加し、また、声帯が薄く軽くなるので振動数の増加がもたらされる。このようにして「Hot Anger」において基本周波数が高くなったものと考えられる。声帯の伸長には輪状甲状筋の収縮のみではなく、喉頭の枠組みである舌骨や甲状軟骨に付着する数多くの外喉頭筋も要因となっている。舌骨の上部にある筋群(舌骨上筋)は喉頭を引き上げる際に、舌骨が前方に移動し、結合組織の連結を介して甲状軟骨を前方に回転させ、声帯の伸長を助ける。この時、舌筋を含む数多くの外喉頭筋が運動する結果、声道の長さや舌の形状が変化し、母音のフォルマントも変動する。「Hot Anger」においてフォルマント周波数が「Neutral」と比較して高かったのはこのためと考えられる。

また、弱い声(パワーが小さい)は音源信号の波形は正弦波に近いが、強い声(パワーが強い)では1周期中で声門の閉鎖期間が長く、急峻な三角波状になる。すると、スペクトルの傾きは緩やかになり、声帯が緊張していないときと比較して高次の高調波成分が大きくなる。「Neutral」と比較して「Hot Anger」はアクセントにおいてパワー変化率が顕著に大きく、25~30 ERB rate という高域においてスペクトルが強調されていたのはこのためだと考えられる。一方、「Cold Anger」は「Neutral」と比較してパワーの変化率、スペクトル概形ともにほぼ同じ傾向が出たが、これもパワーとスペクトルが関連していることを表していると考えられる。このように、分析結果を生成機構の観点から考察してみると、音響特徴量の間には関連があることが考えられる。

第7章 結論

7.1 本研究の成果

本研究では、「Neutral」「Cold Anger」「Hot Anger」という感情音声のアクセントの有無における音響特徴量と感情知覚との関係を調べることを目的とした。そのために基本周波数・パワー・持続時間・フォルマント周波数・スペクトルといった音響特徴量について、主にアクセントにおける変化率の分析を行った。

その結果、基本周波数に関しては、有声区間の長時間平均とアクセントレベル上昇時に、パワーに関してはアクセントレベル上昇時に大きな特徴があることがわかった。

持続時間は、文全体において「Hot Anger」「Cold Anger」は「Neutral」よりも短いことがわかった。中でも、非アクセントよりアクセントで持続時間が短いことがわかった。そして、感情間で比較したとき、子音は持続時間が短縮・伸長せず、ほとんど母音で持続時間が変動していることがわかった。

フォルマント周波数は、全体的に「Cold Anger」は「Neutral」よりも低く、「Hot Anger」は「Neutral」よりも高いことがわかった。その中でも、特に低次のフォルマントは感情判別に大きな影響があることが示唆された。

スペクトルは、アクセントレベルが低から高へなると、「Neutral」と比較して「Cold Anger」「Hot Anger」とともに高域(25～30 ERB rate)で強調されることがわかった。

そして、分析結果から各音響特徴量について「Neutral」の音声から「Cold Anger」「Hot Anger」といった感情音声の合成音を作成するための規則を導いた。また、音響特徴量間の関係について相関性があることを示唆した。

7.2 今後の課題

分析結果が正しいかどうかを確認するため、聴取実験を行わなければならない。これにはまず、分析から得られた合成音作成規則を用いて、合成音を作成する。その際、いろいろなパターンを組み合わせることによって、聴取実験結果から、音響特徴量間の関係も考

察できると考えられる。そして、本研究では感情として「Neutral」「Cold Anger」「Hot Anger」を扱ったが、その他の「Joy」「Sadness」という感情についても同様の研究を行う必要がある。また、本研究では一人の話者について考慮したが、複数話者に対しても検討する必要がある。

第8章 謝辞

本研究を行うにあたり、終始多大なるご指導・御鞭撻を賜り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、鵜木 祐史 助手に深く感謝の意を表します。並びに本研究の過程において多くの御協力を賜りました、本学の教官の皆様に心より感謝致します。

本研究において用いた感情音声データを提供・利用許可して頂いた(株)富士通研究所の片江 伸之 氏並びに同所関係諸氏に深謝致します。また、本研究を進める上で有益な御議論・御助言をして頂いた赤木 研究室の伊藤 一仁 氏、石本 祐一 氏、西本 博則 氏に厚く御礼申し上げます。そして、日頃から有益な助言と激励を頂いた赤木研究室の平成 14 年度修了生、並びに諸先輩方、後輩達に深く感謝致します。

最後に大学院における研究生生活を温かく見守ってくれた、両親・姉に心から感謝致します。

関連図書

- [1] Fujisaki, H.(1997): Prosody, models, and spontaneous speech. In Sagisaka, Y. et al.(ed.), *Computing Prosody: Computational Models for Processing Spontaneous Speech*, Springer.
- [2] 平賀 裕, 斉藤 善行, 森島 繁生, 原島 博, “音声に含まれる感情抽出の一検討”, 信学技法, HC93-66, pp.1-8, Jan.1994.
- [3] J.E.H.Noad, S.P.Whiteside and P.D.Green, “A MACROSCOPIC ANALYSIS OF AN EMOTIONAL SPEECH CORPUS”, Proc.of Eurospeech, pp.517-520, 1997.
- [4] 小池 和仁, 斉藤 博昭, 中西 正和, “感情音声の合成”, 信学技法, NLC98-43,SP98-107,pp83-88, Dec.1998.
- [5] 片江 伸之, 木村 晋太, “感情音声合成における声質と韻律の制御の効果”, 日本音響学会講演論文集, pp.187-188, Sep.2000.
- [6] 重永 実, “感情の判別分析からみた感情音声の特性 (V) 特徴量の諸特性”, 信学技法, SP98-93,pp49-56, Nov.1998.
- [7] 川波 弘道, 広瀬 啓吉, “音声における態度・感情の韻律的特徴の分析”, 日本音響学会講演論文集, pp233-234, Sep.1997.
- [8] 川波 弘道, 広瀬 啓吉, “態度・感情音声における韻律的特徴の考察”, 信学技法, SP97-67,pp73-80, Nov.1997.
- [9] 大山 玄, 武田 昌一, 梶谷 綾香, 新美 成二, 今川 博, “日本語単語音声における怒り表現の生理的特徴の解析”, 日本音響学会講演論文集, pp.259-260, Oct.2001.
- [10] 大山 玄, 武田 昌一, 梶谷 綾香, 西澤 良博, “単語音声に含まれる感情表現の分析”, 文部省科学研究費補助金 特定領域研究 (B) 全体会議資料集, pp.19-24, Jul.2001.

- [11] 武田 昌一, 西澤 良博, 大山 玄, “「怒り」の音声の特徴分析に関する 1 考察”, 日本音響学会聴覚研究会資料, Vol. 31, No.2, H-2001-20, 信学技法, SP2001-164, pp.143-150, Mar.2001.
- [12] 武田 昌一, 大山 玄, 栃谷 綾香, 西澤 良博, “日本語単語音声における怒り表現の韻律的特徴の解析”, 日本音響学会講演論文集, pp.257-258, Oct.2001.
- [13] 鈴木 司, 有泉 均, “感情表現の分析と合成規則”, 信学技法, SP96-131, pp31-38, Mar.1997.
- [14] Donna Erickson, Arthur Abramson, Kikuo Maekawa, Tokihiko Kaburagi, “Articulatory characteristics of emotional utterances in spoken English”, ICSLP, Vol.2, pp.365-368, Oct.2000.
- [15] 前川 喜久雄, 籠宮 隆之, 菅田 雅彰, 鍋木 時彦, 岡留 剛, “調音運動から見たパラ言語情報の生成”, 日本音響学会講演論文集, pp.257-258, Oct.1999.
- [16] 前川 喜久雄, “パラ言語情報の知覚”, 文部省科学研究費補助金 特定領域研究 (B) 全体会議資料集, pp.1-4, Jul.2001.
- [17] 杉本 隆 “音声中の感情表現に関連する物理量とその制御に関する研究”, 北陸先端科学技術大学院大学修士論文, Feb.2000.
- [18] Klaus R. Scherer “Vocal Affect Expression: A Review and a Model for Future Research”, Psychological Bulletin, Vol.99, No.2, 143-165, 1986.
- [19] 大野 澄雄, 藤崎博也, “韻律的特徴による感情の表現”, 文部省科学研究費補助金特定領域研究 (B) 全体会議資料集, pp.97-100, Jul.2001.
- [20] 古井 貞熙, “音声情報処理”, 森北出版株式会社, 1998.
- [21] 河原 英紀, “聴覚の情景分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, Sep.1997.
- [22] 天白 成一, 平原 達也 “UNIX 音声研究用ツール -SpeechTools-”, 日本音響学会講演論文集, pp.331-332, Mar.1991.
- [23] 古井 貞熙, “デジタル音声処理”, 東海大学出版会, 1985.

- [24] 日本音声言語医学会編, “声の検査法”, 医師薬出版株式会社, 1979.
- [25] 白井 克彦, 田窪 行則, 前川 喜久雄, 窪園 晴夫, 本田 清志, 中川 聖一, “岩波講座 言語の科学 2 音声”, 岩波書店, 1998.
- [26] B.R.Glasberg, B.C.J.Moore, “Derivation of auditory filter shapes from notched-noise data”, Hearing Research, 47, pp.103-138, 1990.

学会発表リスト

平館 郁雄, 赤木 正人, “怒りの感情音声における音響特徴量の分析,” 日本音響学会音声聴覚研究会資料, 電子情報通信学会技術研究報告, SP2001, Mar 2002.