修 士 論 文

多層知覚モデルに基づく
音声中に含まれる感情の認識に関する研究

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

青木 祐介

2009 年 3 月

修 士 論 文

多層知覚モデルに基づく
音声中に含まれる感情の認識に関する研究

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　小谷一孔 准教授

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

0710001 青木 祐介

提出年月: 2009 年 2 月

概 要

人間は音声から言語情報だけでなく，感情，個人性といった非言語情報も知覚している．音声から言語情報以外の情報を知覚することで，人間は円滑なコミュニケーションをとっている．音声インターフェイスの需要が高まる中で，非言語情報も含む人間の知覚を模擬するアプリケーションが期待される．人間の知覚特性に則したアプリケーションが実現すれば，将来的には聴覚の補助やロボットなどでの活用が見込まれる．非言語情報知覚を表現するアプリケーションは，人間と計算機の音声コミュニケーションに更なる広がりをもたらすと考えられる．本研究では，非言語情報の中でも，言語などの事前知識がなくても相手の意図を知覚することが出来る，音声コミュニケーションにおいて有用な要素である感情に着目する．

ここで，現在の音声認識システムに目を向けると，入力音声から得られる物理量である音響特徴量を基に，音素·話者·感情などの要素について，該当する一つのカテゴリを選び出すという手法を採っている．確かに，音素や話者などでは，該当するカテゴリは一意である．しかし，感情に関しては該当するカテゴリは一意ではないと考えられる．なぜならば，人間は一つの音声からでも多様な感情をその強度も含めて知覚しているからである．また，人間は音声から物理量を正確に判断することは出来ず，むしろ曖昧な基本的印象により判断している．そのため，人間は明確な値を持つ音響特徴量から，曖昧な基本的印象による判断を経て，多様なしかも曖昧な感情の程度を知覚していると考えられる．しかし，このような過程は従来の認識システムでは考慮されていない．以上の点から，現在の感情認識システムでは，人間の知覚特性を十分に模擬出来ていないと考えられる．

人間の感情知覚特性を表現するモデルとして，Huang and Akagi は多層構造によるアプローチを取り，感情知覚多層モデルを構築した．本研究ではこのモデルを感情認識に応用し，人間の知覚特性に則した感情認識の実現を目指した．

本研究では，感情知覚多層モデルに基づき，ファジィ推論システム (Fuzzy Inference System: FIS) を用いて三層構造の感情認識システムを実装した．FIS は人間の知覚のような非線形で曖昧な関係を記述することが出来る数学的手法である．FIS を用いたことで，線形手法による実装と比較して認識精度の向上が見られた．このため，感情認識システムの実装手法として FIS が有効であることが確認された．三層構造モデルによる認識システムは従来の二層モデルによる認識システムと同等の認識精度であったが，二層の間に挿入された中間層は，物理量である音響特徴量や心理量である感情の程度と相反せず，知覚過程を説明するものであることが示唆された．

三層構造と FIS を用いたことで，人間の感情知覚の過程を表現する基本的印象の変化を明示することが可能となった．認識精度の比較から，このシステムによって従来のシステムよりも，人間の主観に近く，かつ知覚過程を説明する感情認識が可能であることが示された．本研究では，従来のシステムと比べ，より人間の知覚特性に則した感情認識を，感情知覚多層モデルと FIS を用いた認識システムの実装により実現出来た．

目 次

図目次

表 目 次

第1章　序論

1.1　はじめに

人間は音声から言語情報だけでなく，感情，個人性といった非言語情報も知覚している．音声から言語情報以外の情報を知覚することで，人間は円滑なコミュニケーションをとっている．音声インターフェイスの需要が高まる中で，非言語情報も含む人間の知覚を模擬するアプリケーションが期待される [1]．人間の知覚特性に則したアプリケーションが実現すれば，将来的には聴覚の補助やロボットなどでの活用が見込まれる．非言語情報知覚を表現するアプリケーションは，人間と計算機の音声コミュニケーションに更なる広がりをもたらすと考えられる．これまでに，非言語情報の知覚に関して様々な研究が行われている [2][3][4][5][6]．本研究では，非言語情報の中でも，言語などの事前知識がなくても相手の意図を知覚することが出来る，音声コミュニケーションにおいて有用な要素である感情に着目する．

ここで，現在の音声認識システムに目を向けると，入力音声から得られる物理量である音響特徴量を基に，音素·話者·感情などの要素について，該当する一つのカテゴリを選び出すという手法を採っている．確かに，音素や話者などでは，該当するカテゴリは一意である．しかし，感情に関しては該当するカテゴリは一意ではないと考えられる．なぜならば，人間は一つの音声からでも多様な感情をその強度も含めて知覚しているからである．また，人間は音声から物理量を正確に判断することは出来ず，むしろ曖昧な基本的印象により判断している．そのため，人間は明確な値を持つ音響特徴量から，曖昧な基本的印象による判断を経て，多様なしかも曖昧な感情の程度を知覚していると考えられる．しかし，このような過程は従来の認識システムでは考慮されていない．以上の点から，現在の感情認識システムでは，人間の知覚特性を十分に模擬出来ていないと考えられる．

そこで，本研究では従来の研究とは異なるアプローチで，より人間の知覚特性に則した感情認識の実現を目指す．

1.2　本研究の背景

前節で述べたように，音声中に含まれる感情は多様なしかも曖昧な程度を持っている．人間は音声中に含まれる音響特徴量から，曖昧な基本的印象による判断を経て，多様なしかも曖昧な感情の程度を知覚していると考えられる．計算機上でこのような情報を扱うためには，人間の感情知覚·生成機構を解明し，反映していく必要がある．

これまでの感情音声を扱った研究は，基本周波数·振幅·持続時間といった，韻律情報に起因する音響特徴量に着目している．そして，音響特徴量の変化により，感情の知覚に対する影響が明らかになっている [1][7][8]．また，音響特徴量の変化による感情音声合成についての研究も行われている [9][10]．人間の感情知覚を踏まえて，音声からの感情認識を行うためには，音響特徴量が感情知覚にどのような影響を及ぼすのかに着目した取り組みが必要となる．

そこで本節では，感情音声を扱った先行研究を示すとともに，これまでに行われてきた感情認識の研究について概説する．

1.2.1 感情音声に関する先行研究

これまでの感情音声を扱った研究は，先述のように音響特徴量に着目して行われている．このような研究の例を挙げると，平賀らは基本周波数·振幅·持続時間の時系列変化に関して研究を行った結果，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 [7]．林は発声時間とピッチ曲線による感情識別·同定について研究し，ピッチ曲線が感情伝達に重要であると報告している [8]．感情音声合成の分野では，平館と赤木は怒りの感情音声における音響特徴量と聴覚印象の関係を調査し，感情音声の合成のための規則を導き出した [9]．磯部らは声帯波を先鋭化させる非線形処理を行うことで，怒りの音声に含まれる濁りを付加した合成を行っている [10]．また，知覚認識に対する聞き手の影響を扱った研究としては，エリクソンと昇地は学童による知覚から，発話者の性差だけでなく，リスナーの性差や年齢差による感情知覚への影響を明らかにした [11]．また，沢村らは母国語の違いによる感情知覚への影響を明らかにした [12]．

文献 [7][8][9] から，どのような音響パラメータが感情知覚に影響するのかが，文献 [10] から，音響パラメータと感情知覚の間の関係は線形処理では表現しきれていないことが，文献 [11][12] から，感情知覚には聞き手の個人差が存在することがわかっている．以上の点を踏まえた上で人間の知覚を表現した感情認識を行う必要がある．

一方，Huang and Akagi は感情知覚のモデル化について，より良く音響特徴量と感情を結びつけるために多層構造によるアプローチを取り，感情知覚多層モデルを構築した [13][14]．このモデルは人間の多様なしかも曖昧な知覚特性を表現するため，音響特徴量の層と感情カテゴリの層の間に基本的心理特徴の層を仮定した三層構造を採っている．また，このモデルに基づいた感情音声の合成の可能性を検討した結果，良好な結果を得ている [13][14]．

1.2.2　感情認識に関する先行研究

ここでは，感情認識に関する先行研究を挙げる．白澤らは音声に含まれる話者の感情を多変量解析の手法を用いて判別した [15]．具体的には，音声データから韻律情報を抽出し，時間構造，振幅構造，ピッチ構造からなる特徴量を得た．そしてその主成分のマハラノビス距離によりを特徴量として感情の判別を行った．刀根らは韻律情報を用いて HMM に基づく感情モデルを構築し，動的な時系列パターンを扱うことで感情判別を行った [16]．廣瀬らは言語情報と感情情報を分離し，感情情報のみから特徴量を抽出し，サポートベクターマシンによって感情認識を行った [17]．森山と小沢は話し手や聞き手による感情の判断は内的，外的要因による影響を受けるため，第三者である観測者の知覚する情緒性という視点から，感情情報の評価を行った [18]．この研究では，言葉によるシステムの記述を行うため，ファジィ推論が用いられた．Lee and Narayanan は複数の感情を含む音声について最適なカテゴリ分けを行うために，ファジィ推論を用いて negative と non-negative に二分する判別を行った [19]．

ここで挙げた手法では，以下のような問題点が挙げられる．

- 感情認識における明確な特徴量が見つかっていないため，統計的手法による特徴量と感情の結び付けによる方法では有用な成果が出ていない．
- 感情を一意に定める方法を用いているため，多様な感情をその強度も含めて知覚することに対応できていない．

人間は音声から物理量を正確に判断することは出来ず，むしろ曖昧な基本的印象により判断していることを考慮すると，明確な値を持つ音響特徴量から多様なしかも曖昧な感情の程度を直接認識出来るとは考えにくい．そのため，人間の知覚過程を模擬することを考えたとき，従来の手法では模擬することが十分には出来ていないと言える．

1.3　本研究の目的

本研究では，人間の知覚特性に則して，音響特徴量から複数の感情の程度を認識することを目的とする．感情知覚多層モデル [13][14] に基づく音声からの感情認識システムの構築を試みる事で，多様なしかも曖昧な感情の認識を実現する．本研究で提案する感情認識システムの実現は，感情知覚･生成機構の解明につながり，マンマシンインターフェイスの感情表現の更なる向上が期待できる．

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，本研究において実装する感情認識システムの概要を述べる．
第 3 章では，感情認識システムの実装に用いる音声データについての分析を行う．
第 4 章では，感情認識システムの実装について述べる．
第 5 章では，実装システムの動作検証と他システムとの比較による評価を行う．
第 6 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．

第2章　感情認識システムの概要

2.1　はじめに

本章では，本論文で構築を試みる，人間の知覚特性に則した音声からの感情認識システムの構成を示し，認識システムを実現する過程で必要な，音響特徴量と感情カテゴリの間を結びつけるための方針を示す．前章の研究背景で紹介したように，これまでの感情認識の研究は，音響特徴量と感情カテゴリの直接的な結びつけにより行われている．しかし，人間は音声から物理量を正確に判断することは出来ず，むしろ曖昧な基本的印象により判断していることを考慮すると，明確な値を持つ音響特徴量から多様なしかも曖昧な感情の程度を直接認識出来るとは考えにくい．そのため，従来の手法では十分に人間の知覚過程を模擬することが出来ていないと言える．ここでは，本研究で採用する人間の感情知覚過程を表現したモデルである感情知覚多層モデルと，モデルの各層を結びつけるために使用するファジィ推論システムについて概説する．そして，実装する感情認識システムの構成について説明する．

2.2　感情知覚多層モデル

感情認識システムの実装のためには，音響特徴量と感情という高次の心理特徴を結びつける必要がある．しかし，人間の知覚特性を考慮すると，音響特徴と心理特徴の直接的かつ定量的な評価は困難であると考えられる．そのため，高次の心理特徴と音響特徴量の関係を，直接的かつ定量的な評価とは異なる，新たな枠組みで捉えた分析を行う必要がある．

そこで，心理量と物理量である音響特徴量の関係を階層構造で記述した知覚モデルが提案されている．Huang and Akagi は感情について表現した感情知覚多層モデル [13][14] を，齋藤らは歌声らしさについて表現した歌声らしさの知覚モデル [20] を構築した．本研究では，人間の感情知覚を模擬するために，人間の知覚を表現したモデルである，感情知覚多層モデルを基としてシステムを構築する．

初めに，感情知覚多層モデルが，知覚構造を表現するモデルとしてどのような構成になっているかを紹介する [13][14]．図 2.1 に感情知覚多層モデルの概要を示す．このモデルは，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した．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．具体的には，音響特徴量と感情カテゴリの間に，曖昧な印象を表現する基本

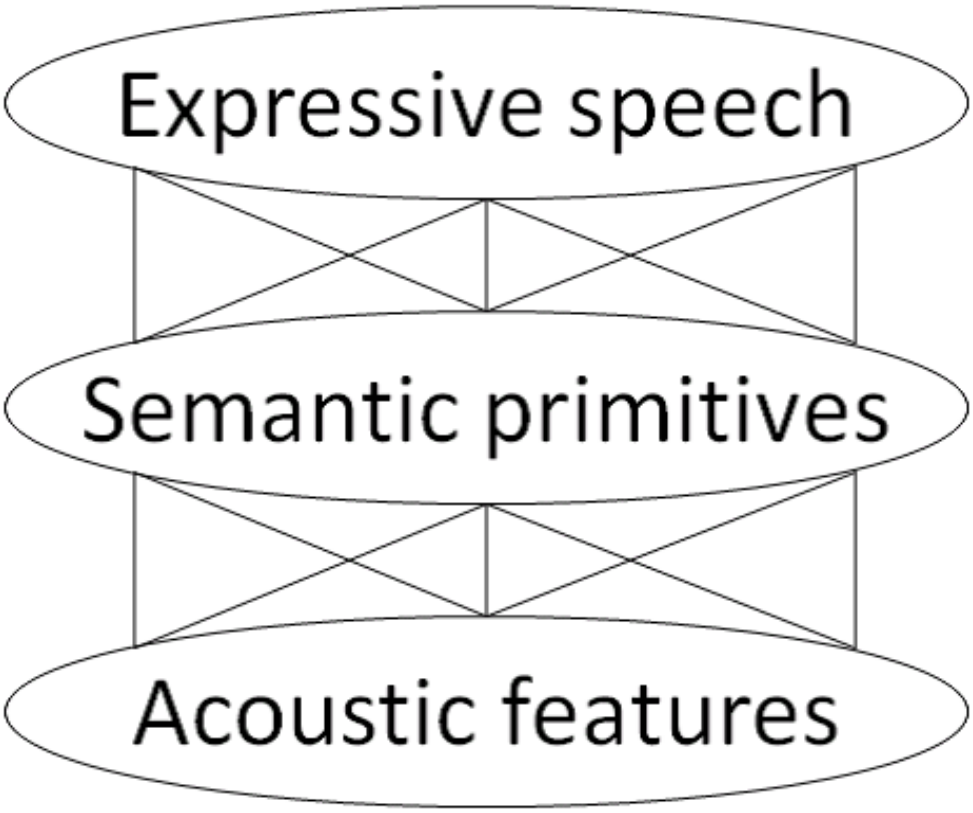

図 2.1: 感情知覚多層モデルの構成.

的心理特徴を想定した三層構造をとっている．そして基本的心理特徴の組み合わせにより感情知覚のモデル化を行っている．基本的心理特徴の印象の変化によって感情カテゴリ層での変化を判断することが出来る．また，基本的心理特徴によって曖昧な判断過程を明示している．

2.2.1 感情知覚多層モデルの構成要素

本節では，音響特徴量，基本的心理特徴，感情カテゴリの各層がどのような要素によって構成されているかを説明する．音響特徴量，基本的心理特徴，感情カテゴリの全ての要素が相互的に作用することで，人間の感情知覚を表現する．

感情カテゴリ

心理学分野において人間の基本感情について多く研究されている．基本感情は Anger, Joy, Sadness の 3 感情，Anger, Fear, Joy, Sadness, Boredom, Disgust の 6 感情など様々な定義があると報告されている [21]．一方で，これまでの感情音声の研究においては Anger,

Joy, Sadness の 3 感情を取り上げている例が多くみられる．それらを踏まえ，感情カテゴリは基準となる Neutral，基本 3 感情に含まれる Joy, Sadness, Anger を要素とされた．このうち，Anger は Cold-Anger と Hot-Anger で別の感情として感じられるため，分けて考える．感情知覚多層モデルにおける感情カテゴリは，これらの 5 種類により構成される．

基本的心理特徴

感情知覚多層モデルにおける基本的心理特徴は，人間が感情を含む音声を聴いた時に感じる基本的印象を表す [13][14]．音あるいは音声に対応する多くの形容詞から，感情カテゴリと関係する基本的心理特徴が選び出した．はじめに，候補として 60 種類の形容詞を選択した．このうち 46 種類は上田により音色の表現語として選ばれたものである [22]．この 46 種類は 166 人の聴取者によって音色を表現する形容詞として選びだした．しかしこの研究は音の音色に関するものであるため，音声知覚に対する表現後が十分ではなかった．そこで，さらに 14 種類の表現語を加えた上で，モデルの構築に適切な形容詞が選びだした．最初に，この 60 種類の形容詞から，モデルの構築に用いられた音声データを表現する上で多く用いられた 34 種類を選んだ．そして 34 種類の形容詞から，重回帰分析によって 17 種類の形容詞を選んだ．このモデルで用いられる基本的心理印象は，感情カテゴリ及び音響特徴量との関連が強く見られた bright, dark, high, low, strong, weak, calm, unstable, well-modulated, monotonous, heavy, clear, noisy, quiet, sharp, fast, slow の 17 種類によって構成される．これらの組み合わせにより感情知覚を表現する．

音響特徴量

計算機上で，音声信号の変化と感情の関係を扱おうとしたとき，音声信号から抽出することが出来る音響特徴量は重要な要素である．これまでの研究でも，音響特徴量の変化による影響が調査された結果として，感情知覚を表現する上で重要な要素であることが明らかとなっている．

したがってモデルの構築においては，どのような音響特徴量が心理量に影響を及ぼすのか調査する必要がある．モデルで使用する音響特徴量は従来の研究において影響力を持つとされた要素について再検討することで，導き出された．従来の研究では，音声の韻律情報が知覚に影響を及ぼすことがわかっている．したがって，ここではどのような韻律情報がモデルで扱われるかを説明する．

感情情報に関する音響特徴量としては，基本周波数 (F0)，パワーエンベロープ，パワースペクトル，時間長が扱われた．感情については，アクセントの影響が強いことが明らかになっているため，アクセントによる影響を踏まえて，音響特徴量を抽出した．

抽出された音響特徴量の中から，基本的心理特徴に強く影響する要素が選び出された．F0 の最大値，F0 の平均値，F0 上昇の傾きの平均値，第 1 句での F0 上昇の傾き，アクセント句でのパワーレンジの平均値，パワーレンジ，第 1 句でのパワー上昇の傾き，3 kHz 以上での平均パワーと全周波数での平均パワーの比，第一ホルマント周波数，第二ホル

マント周波数，第三ホルマント周波数，スペクトルの傾き，スペクトルの重心，文の時間長，子音の区間長，そして子音と母音の区間長の比の 16 種により構成される [13][14].

2.2.2　感情知覚多層モデルの層間接続

ここではモデルを構成する層間の結合について紹介する．各感情について強い影響力を持つ基本的心理特徴を選択し，それらの特徴と強い相関をもつ音響特徴量を選ぶことで，各感情の多層知覚モデルが構築された.

感情カテゴリ–基本的心理特徴間

人間は曖昧な判断により基本的心理特徴から感情を知覚していると考えられる．そのため，感情カテゴリと基本的心理特徴の層間を結合するためには，感情の判断という人間の曖昧な処理を考慮しなくてはならない．そこで，Huang and Akagi は各感情カテゴリと基本的心理特徴の結びつきについてファジィ推論システム (Fuzzy Inference System: FIS) を用いて調査した．感情ごとに基本的心理特徴を入力とする FIS を構築し，各基本的心理特徴の変動による影響を調べた．そして，最小二乗法を用いることで，変動による影響の近似直線を算出した．近似直線の傾きが正に大きい基本的心理特徴 3 種と，負に大きい基本的心理特徴 2 種を選択した.

こうして各感情について 5 つの基本的心理特徴を相互関係の強い接続として選別した．これにより，文献 [13][14] では人間の曖昧な判断に対応した層間要素の結合が実現されている.

基本的心理特徴–音響特徴量間

基本的心理特徴と音響特徴量の層間の結合では，人間が基本的心理特徴の変化を感じる時に，どんな音響特徴量が変化しているのかを調べる必要がある．そこで，文献 [13][14] では基本的心理特徴 17 要素と音響特徴量 16 要素の相関をとった．そして，相関の絶対値が 0.6 以上である関係が，強い影響力を持つとして選択された.

2.3　ファジィ推論システム

本研究では，物理量である音響特徴量と心理量である感情カテゴリの関係を適切に結びつける必要がある．人間は音声から感情を知覚する際，形容詞を組み合わせることで言語的に知覚を表現することが出来る．この言語による表現というのは，正確なものではなく寧ろ曖昧である．そのため，線形で正確な関係を扱う統計的手法による表現は感情知覚を扱う上では適切ではない.

このような非線形で曖昧な関係を記述することが出来る数学的手法として，ファジィ推論が挙げられる [23][24]．以下にファジィ推論の利点を挙げる．

- ファジィ推論は経験的知識のような人間の経験を数学的に扱うことが出来る．これは音声中の感情に対処する際に，モデルが行うべきことである．
- ファジィ推論は自然言語に基づく．それに加えてモデルの中で使用される自然言語は基本的心理特徴に等しい．
- ファジィ推論は複雑な非線形の関係をモデル化することが出来る．感情カテゴリと基本的心理特徴の間は複雑で非線形であるため，適切に働くと考えられる．

これらの点から，ファジィ推論を用いることで，例えば Sadness と感じる音声に対して，やや slow に感じる，適度に slow に感じる，かなり slow に感じるというような，基本的心理特徴の曖昧な言語によって表現される程度を扱うことが出来る．ファジィ推論を多入力 1 出力のシステムとしたものが FIS である．FIS を利用することで感情の多様性 · 連続性に対応した層間構造を構築することが出来る．

2.4 本論文で実装する感情認識システムの構想

Huang and Akagi によるモデルの構築においては，各層の間で影響力の強い要素が選択された [13][14]．しかし，感情を知覚する上で，影響力の弱い要素による影響も考慮する必要がある．そこで，本研究では，モデルの構成に用いられている全ての要素を用いた認識システムを実装する．

また，基本的心理特徴と音響特徴量の関係も，感情カテゴリと基本的心理特徴の関係と同様に，曖昧な判断により行われていると考えられる．そのため，FIS は基本的心理特徴–音響特徴量間でも有用であると考えられる，そこで，感情認識システムの実装においては全ての層間接続に FIS を用いる．

2.5 まとめ

本章では，人間の感情知覚特性に則した感情認識を実現するために構築する，音声からの感情認識システムに関する構成について説明した．人間の感情知覚の過程まで表現出来るように感情知覚多層モデルの考えに基づくことを説明した．また，人間の多様なしかも曖昧な知覚特性を表現するために，FIS を用いることを説明した．

第3章　音声データの分析

3.1　はじめに

実際に感情認識を行うために，感情を含んだ音声データが必要となる．音声データから抽出した音響特徴量を入力すると，複数の感情の強度が得られるシステムを構築するために，音声データの音響特徴量，基本的心理特徴，感情の値を，音声分析と聴取実験によって収集する．

音声データの条件として，感情カテゴリ層の感情を持って発話されていること，認識する上で発話の文章に依存しないことが挙げられる．本研究では音声データとして富士通感情音声データベースの 179 発話を使用する．このデータベースは発声文章は 20 種類あり，文章毎に 9 種類の発話が行われている．1 音声にデータ欠落があるため，全部で 179 種類の音声データとなる．発話は想定している 5 感情を意図して行われている．また，文章の種類も十分にあることから，認識に対する影響は少ないと考えられる．以上の点から本研究のシステム構築にあたり適切なデータベースであると考えられる．各発話について，音響特徴量の抽出，基本的心理特徴及び感情の評価値を収集し，得られた値を感情認識システムの構築に用いる．

3.2　本研究で扱う音声データ

本節では，本研究で音声波を分析する際に用いる，感情音声データについて説明する．声優･演劇経験者は一般人と比較し，音声による感情表現を的確に心得ていることがわかっている [7]．そのため，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 サンプル (9 パターン，20 文章) ((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･発話内容･発話リストを示す．

- データ形式

サンプリング周波数: 22050 Hz
16 bit 量子化
最大振幅は録音時に正規化されている

- 発話内容

表 3.1: 富士通感情音声データの発話内容.

UID	Expressive speech category
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)

表 3.1 に発話内容の一覧を示す.「Neutral (平静)」が 1 パターン,「Joy (喜び)」「Cold-Anger (押し殺した怒り)」「Sadness (悲しみ)」「Hot-Anger (激しい怒り)」の感情が各 2 パターンの計 9 パターンのデータが収録されている．UID は文字と数字のコードからなる．ただし，e014 はデータが欠落している．

- 発話文章

表 3.2 に発話文章の一覧を示す．ID は UID の数字コード部である．

3.3 音声データの分析

3.3.1 音響特徴量の抽出

感情認識システムの入力として音声データについての音響特徴量が必要となる．本研究では安定に基本周波数を抽出する事が出来る高品質な分析合成系 STRAIGHT[25][26] を用いる．STRAIGHT によって基本周波数，サウンドスペクトログラムに中に含まれる周期的駆動に起因する微細構造を組織的に取り除いた時間周波数表現を取り出す．取り出した値から F0 ，パワーエンベロープ，パワースペクトルに起因する音響特徴量を算出する．また，セグメンテーションを行い得られた情報から発話長に起因する音響特徴量を算出する．音響特徴量に関しては，Neutral の音声に対しどれだけ音響特徴量が変化したかを値として用いる．すなわち，Neutral についての値の平均をとりこれを基準とした比率をパラメータ値として用いる．

3.3.2 基本的心理特徴及び感情知覚の評価値の取得

図 4.1 における基本的心理特徴認識部の出力，及び感情カテゴリ認識部の入出力の値として，音声データから得られる基本的心理特徴と感情の主観評価値が必要となる．基本的

表 3.2: 富士通感情音声データの発話文章.

ID	Sentence
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	い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
009	ずっと見ています
010	私のところには届いています
011	ありがとうございました
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	こちらの手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

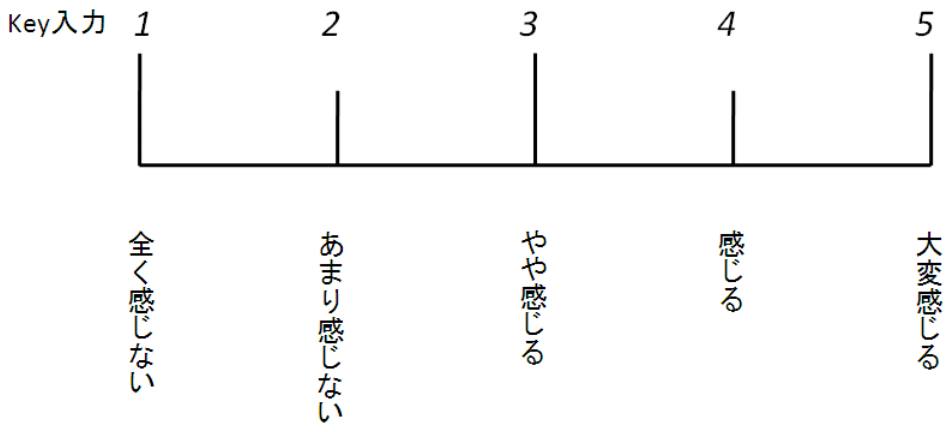

図 3.1: 聴取実験に用いた評価表.

心理特徴の印象評価値と各感情の強さの収集については，聴取実験によって行う．各発話に対し主観評価による聴取実験を行い，基本的心理特徴の印象評価値と各感情の強さを収集する．

実験手続き

聴取実験では，本研究において用いる 179 種類の音声データについて，多層知覚モデルに基づいた 17 種類の基本的心理特徴及び 5 種類の感情についてそれぞれの印象の程度を得る必要がある．実験参加者には次のような指示を与え，各基本的心理特徴，各感情の計 22 種類の印象について評価してもらった．

- ヘッドホンから音を流します。聴いたときの感想として、各感情に該当するかどうかか、下に記した５段階評価尺度に従って判断してください。該当すると感じたならば（３～５）に、該当しないと感じたならば（１～２）のキーを押してください

評価は，“1. 全く感じない”, “2. あまり感じない”, “3. やや感じる”, “4. 感じる”, “5. 大変感じる” の 5 段階である．

本実験で用いられた評価表を図 3.1 に示す．

実験参加者

実験参加者は正常な聴力を有する 20 代から 30 代の大学院学生 9 名 (男性 9 名) である．感情音声の認知において異文化間の影響が見られている [12] 事から母国語が日本語である者に限定した．

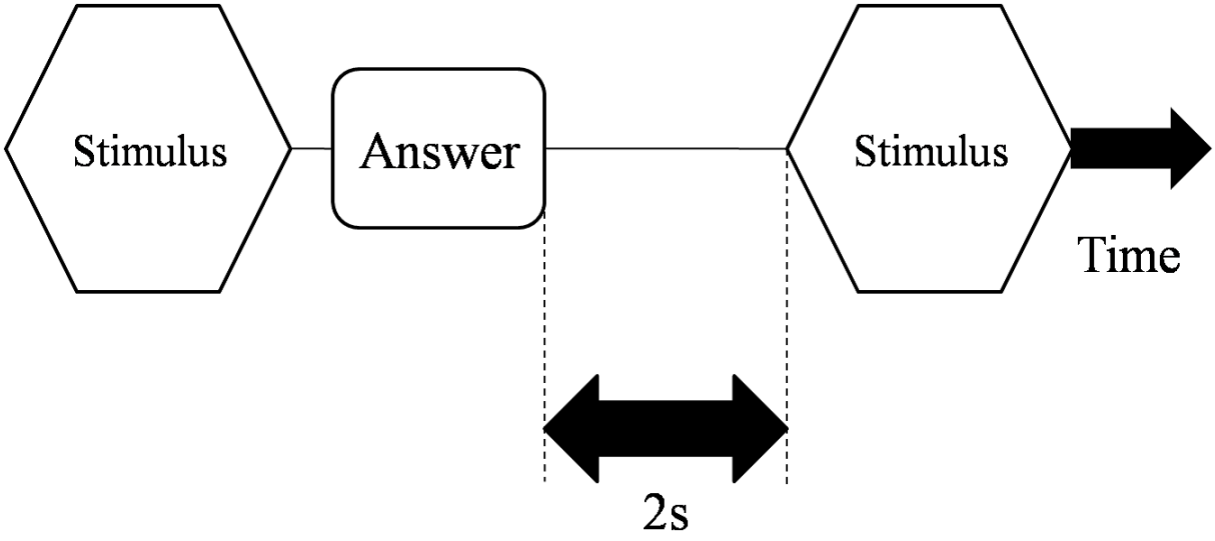

図 3.2: 刺激の呈示順序.

表 3.3: 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 HDA200
ヘッドフォンアンプ	YAMAHA DP-U50

刺激条件

実験で用いる刺激音は先に示した 179 個の音声である．全ての音声をランダムに呈示した．図 3.2 に刺激の呈示順序を示す．本実験は 1 セッションの練習を経て，1 印象に対し 1 セッション，合計 23 セッションによって構成される．

実験環境

実験参加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 3.3 に示す.

結果

はじめに，発話意図ごとに各感情の評価値の平均と分散を調査した結果を以下に示す. 図 3.3 は発話意図 Neutral の音声に対する各感情カテゴリの評価値，図 3.4 は発話意図 Joy の音声に対する各感情カテゴリの評価値，図 3.5 は発話意図 Cold-Anger の音声に対する各感情カテゴリの評価値，図 3.6 は発話意図 Sadness の音声に対する各感情カテゴリの評価値，図 3.7 は発話意図 Hot-Anger の音声に対する各感情カテゴリの評価値である．これ

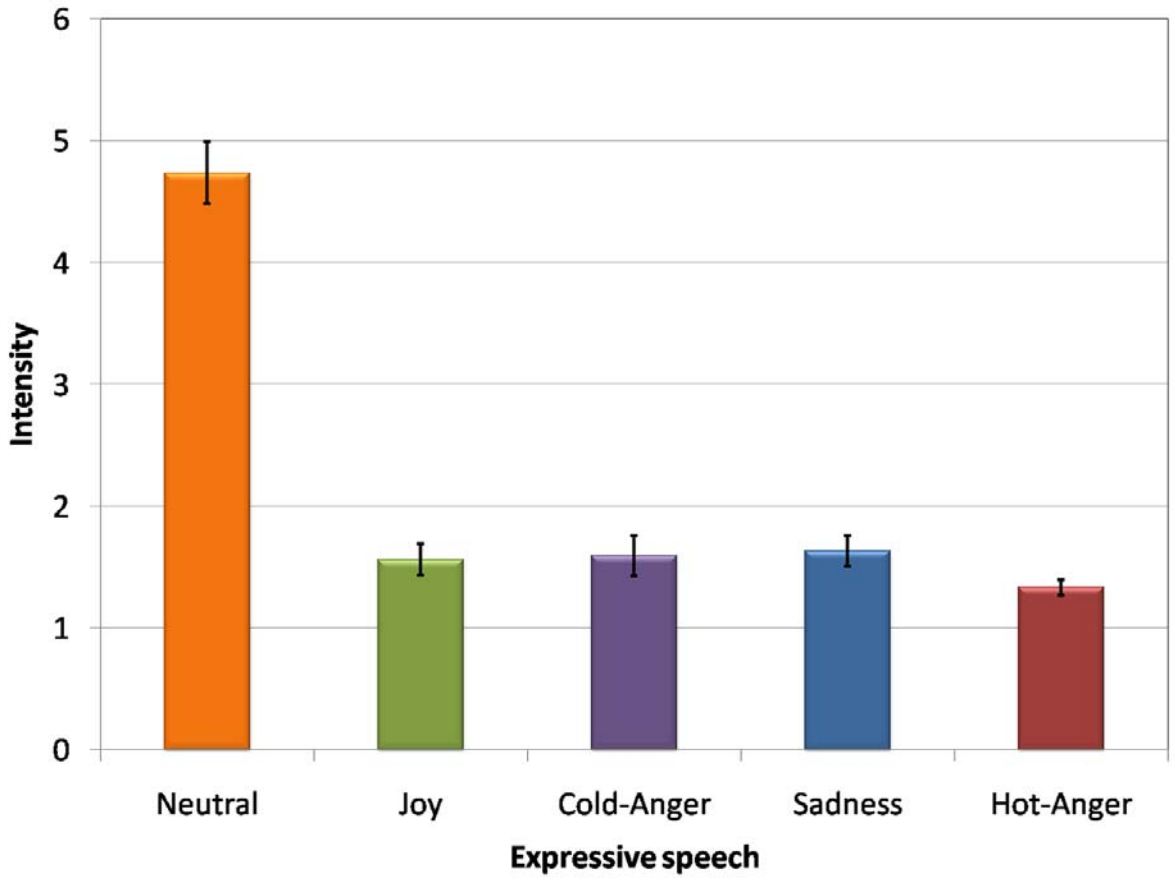

図 3.3: 発話意図: Neutral の音声データについての感情カテゴリの評価値の平均と分散.

らの結果からは，人間は音声から発話意図以外の感情も知覚している．この点から，人間の知覚特性に則した感情認識を行う上では，複数の感情の程度を表現することが必要であることが確認された．このため，人間の感情知覚を表現した認識システムの実現のためには，複数感情の程度の出力が必要であることが示唆された．

また，発話意図ごとに各基本的心理特徴の評価値の平均と分散を調査した結果を以下に示す．図 3.8 は発話意図 Neutral の音声に対する各基本的心理特徴の評価値，図 3.9 は発話意図 Joy の音声に対する各基本的心理特徴の評価値，図 3.10 は発話意図 Cold-Anger の音声に対する各基本的心理特徴の評価値，図 3.11 は発話意図 Sadness の音声に対する各基本的心理特徴の評価値，図 3.12 は発話意図 Hot-Anger の音声に対する各基本的心理特徴の評価値である．これらの結果から発話意図ごとに基本的心理特徴の印象程度に違いが見られる．このため，基本的心理特徴は感情の認識の際，判別要素として用いることが可能であることが示唆された．

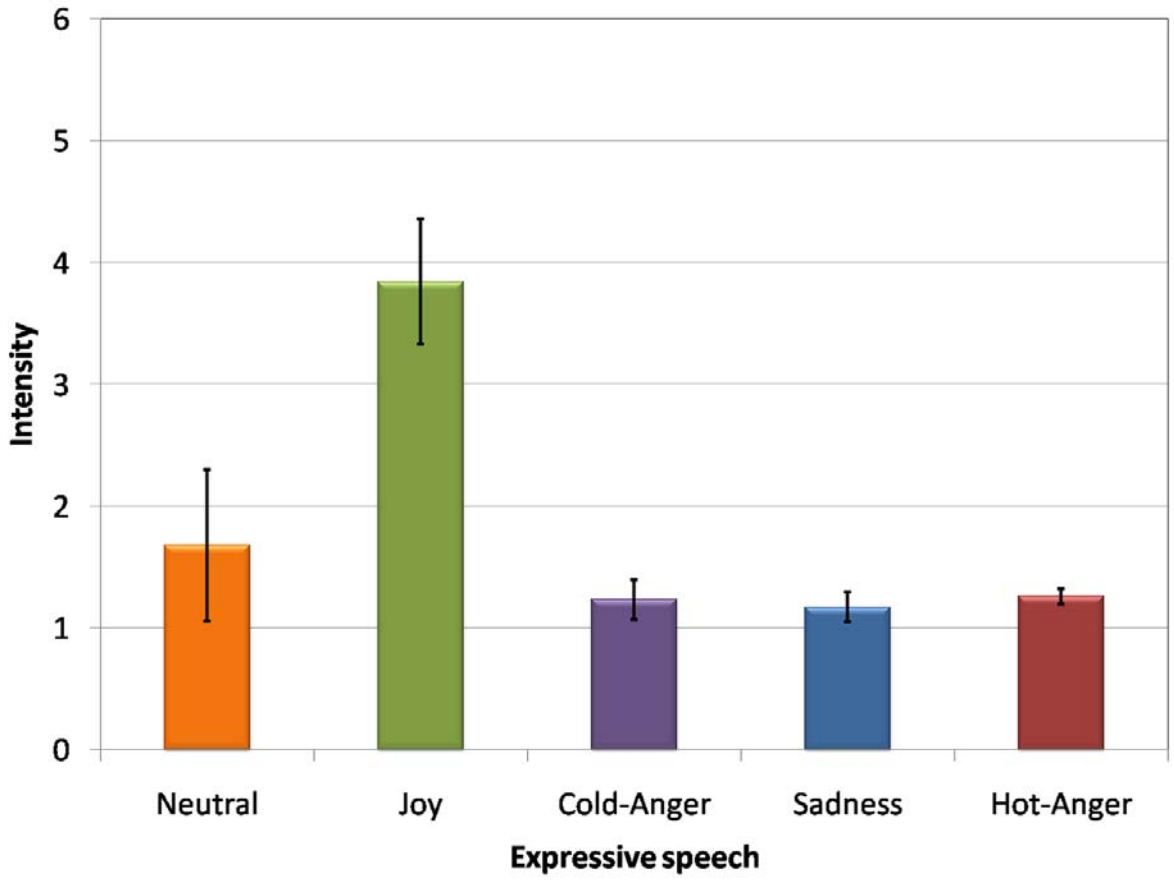

図 3.4: 発話意図: Joy の音声データについての感情カテゴリの評価値の平均と分散.

3.4 まとめ

この章では，本研究で使用する富士通の感情音声データについての紹介と，感情音声データからの音響特徴量の抽出，基本的心理特徴及び感情の主観評価値の収集を行った．音源ごとの各音響特徴量，各基本的心理特徴，各感情の値に差異が見られたことから，従来の認識では人間の知覚特性を表現出来ていないことが示唆された．これらの値を用いて感情認識システムを実装することで人間の知覚特性に則した感情認識を実現する．

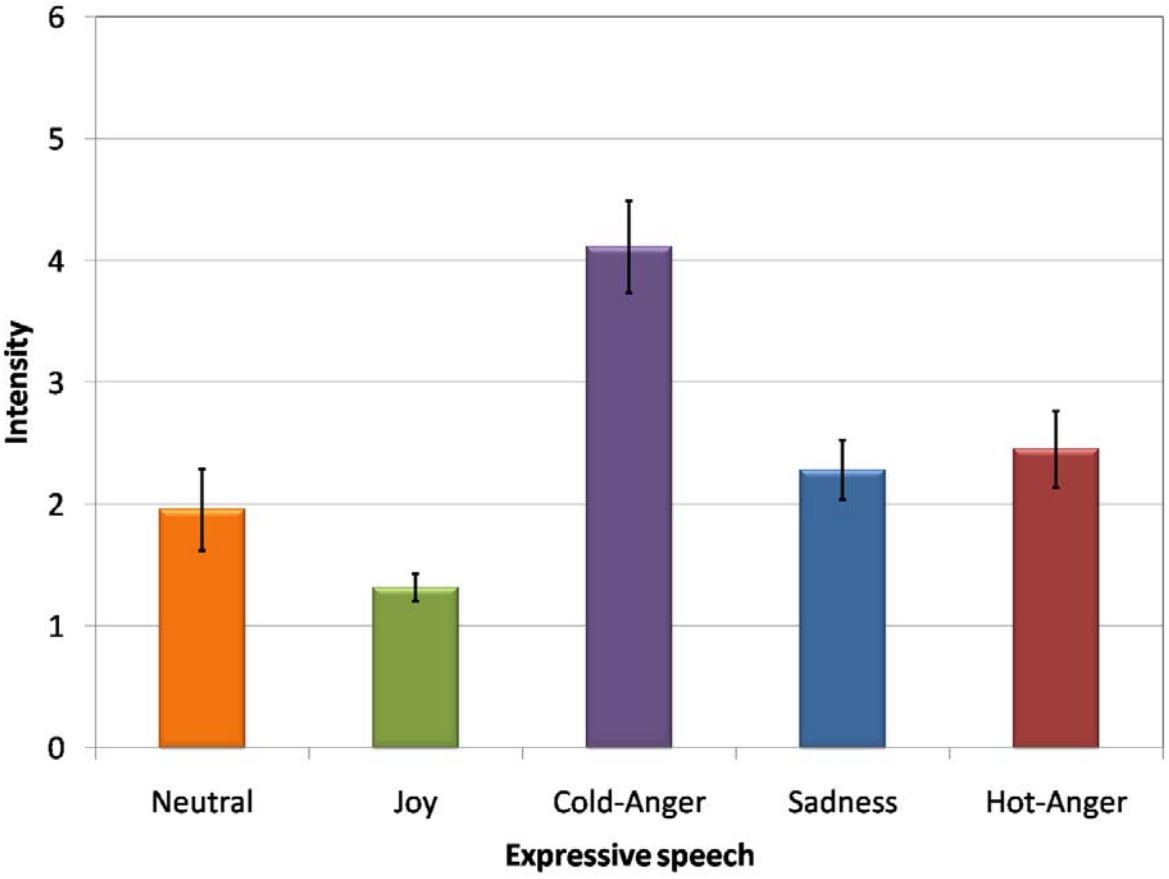

図 3.5: 発話意図: Cold-Anger の音声データについての感情カテゴリの評価値の平均と分散.

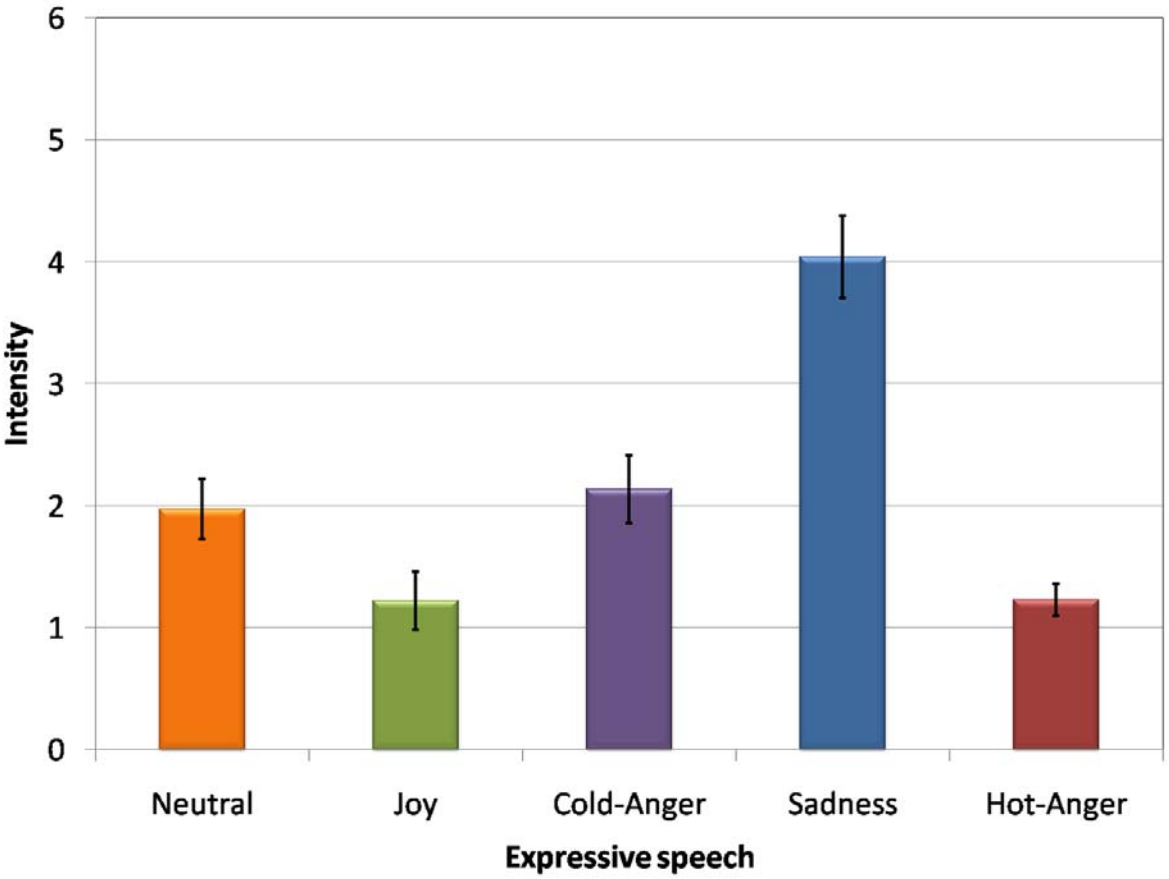

図 3.6: 発話意図: Sadness の音声データについての感情カテゴリの評価値の平均と分散.

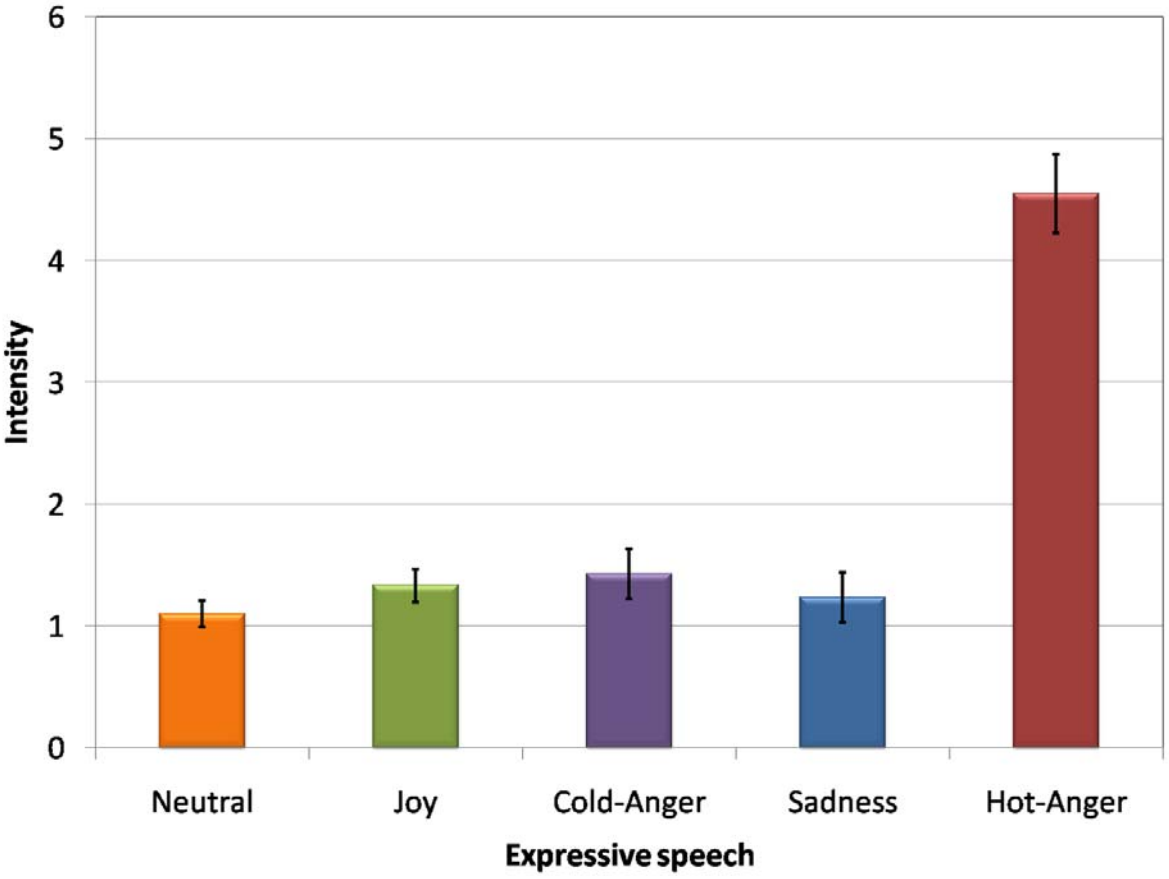

図 3.7: 発話意図: Hot-Anger の音声データについての感情カテゴリの評価値の平均と分散.

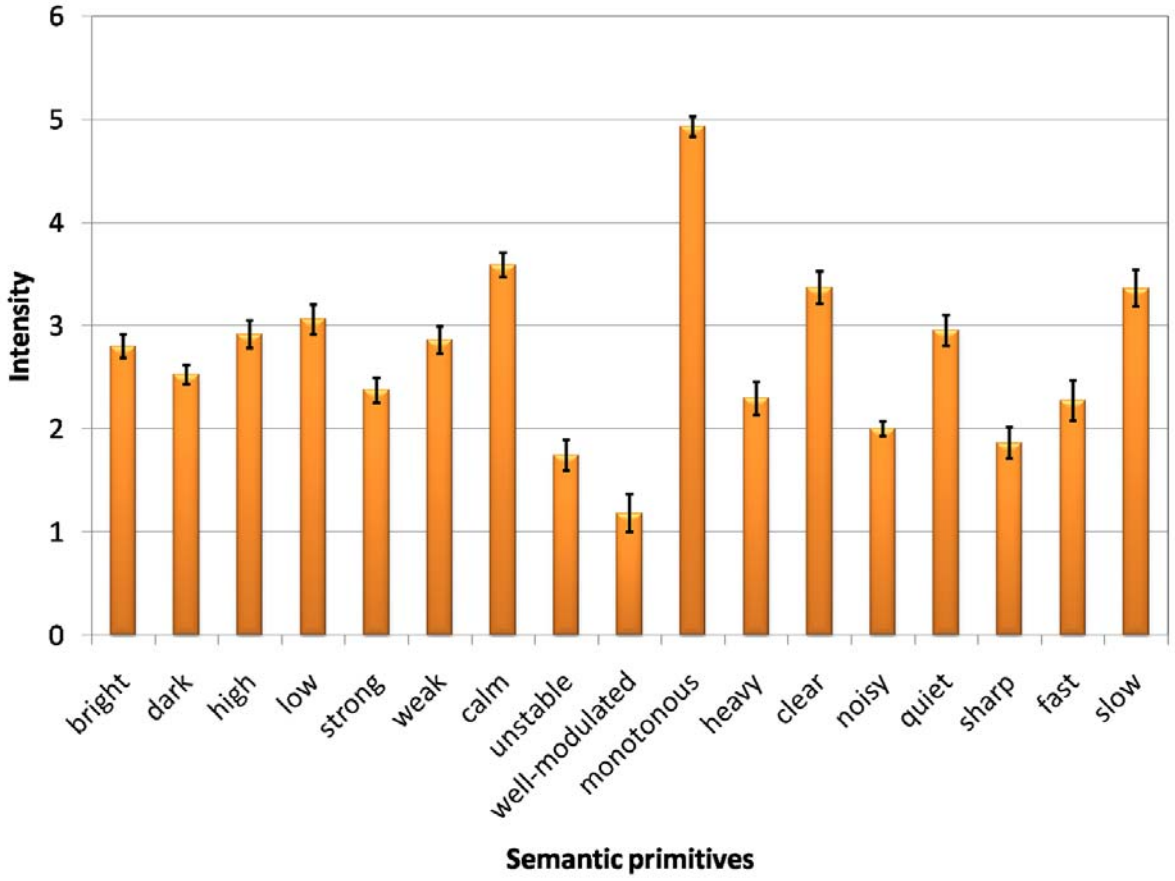

図 3.8: 発話意図: Neutral の音声データについての基本的心理特徴の評価値の平均と分散.

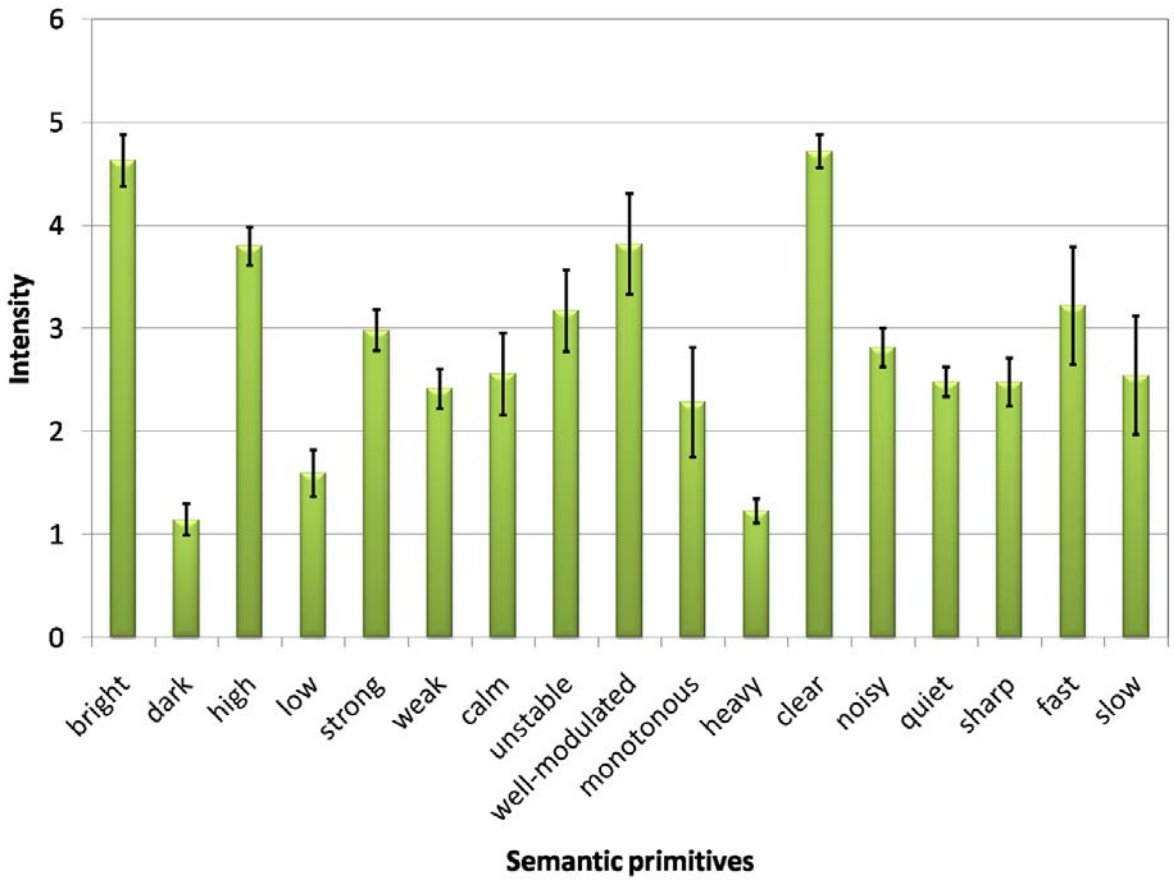

図 3.9: 発話意図: Joy の音声データについての基本的心理特徴の評価値の平均と分散.

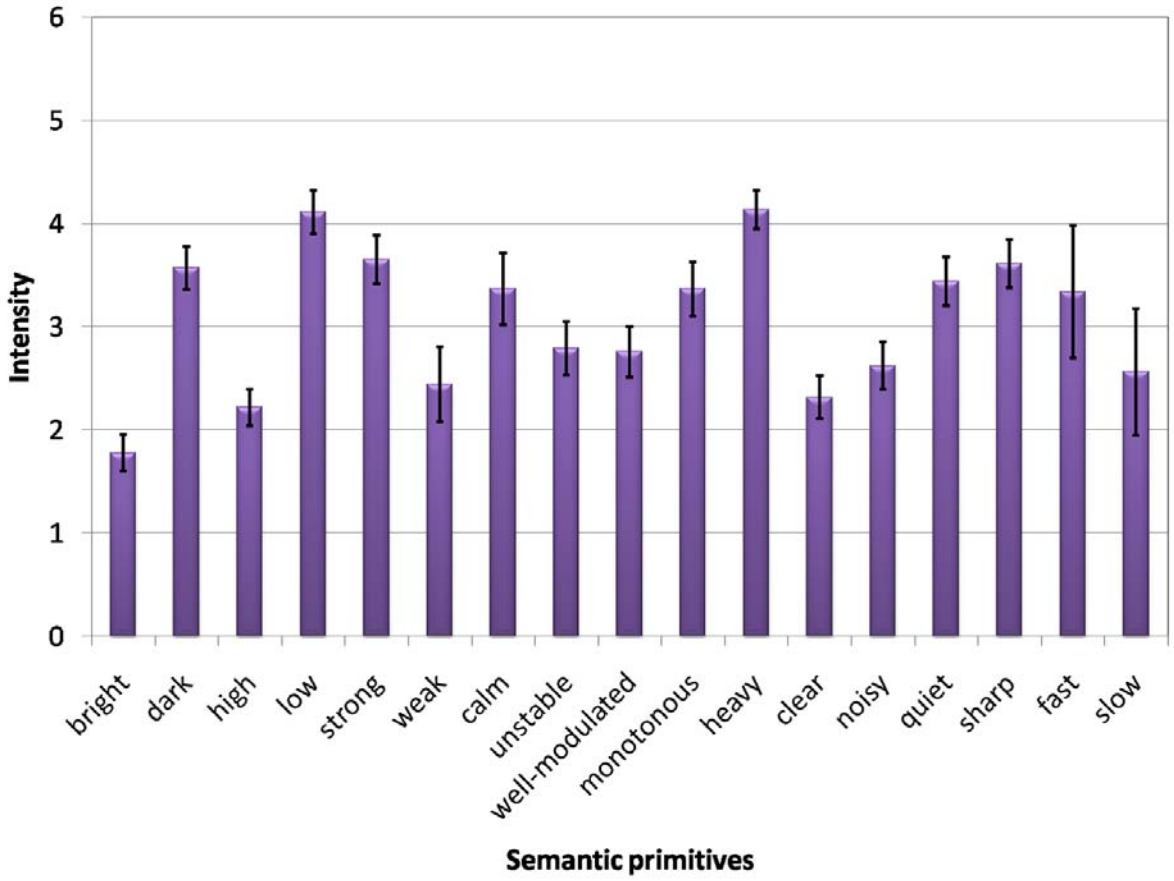

図 3.10: 発話意図: Cold-Anger の音声データについての基本的心理特徴の評価値の平均と分散.

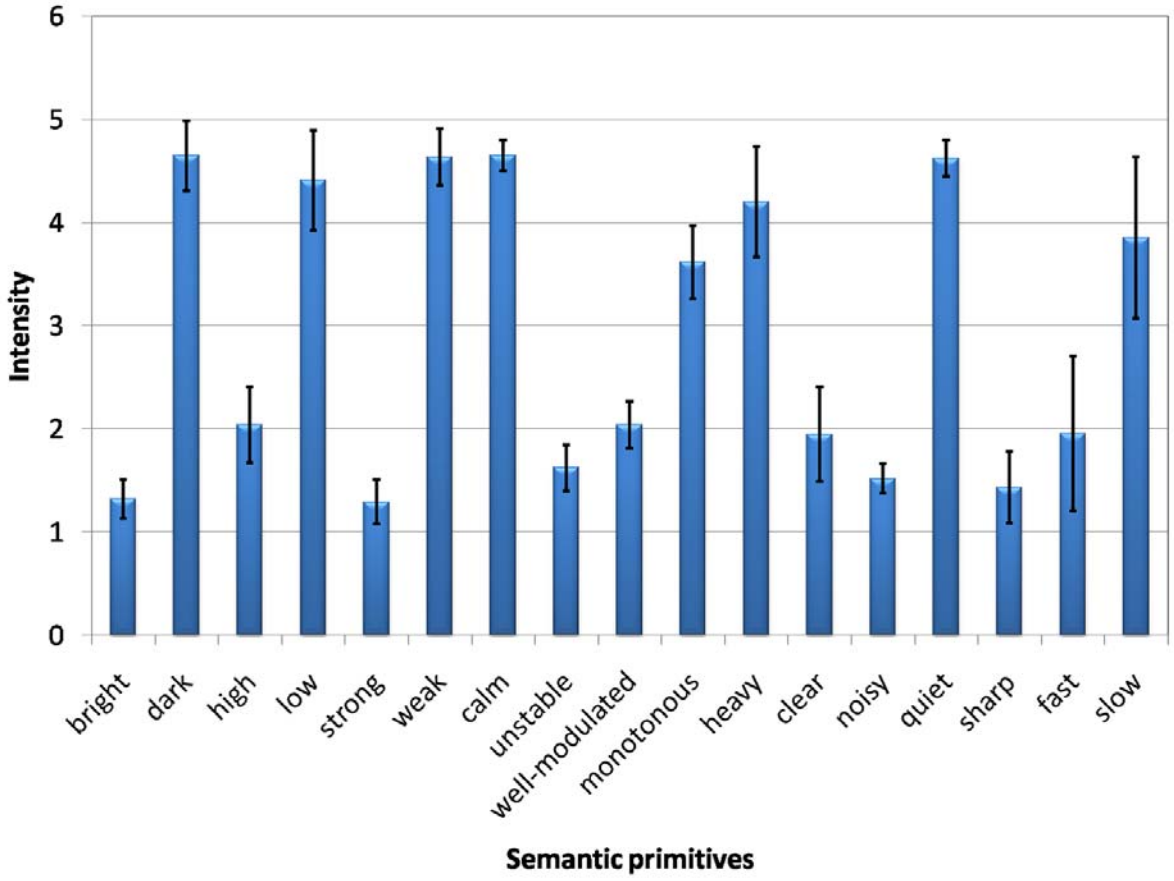

図 3.11: 発話意図: Sadness の音声データについての基本的心理特徴の評価値の平均と分散.

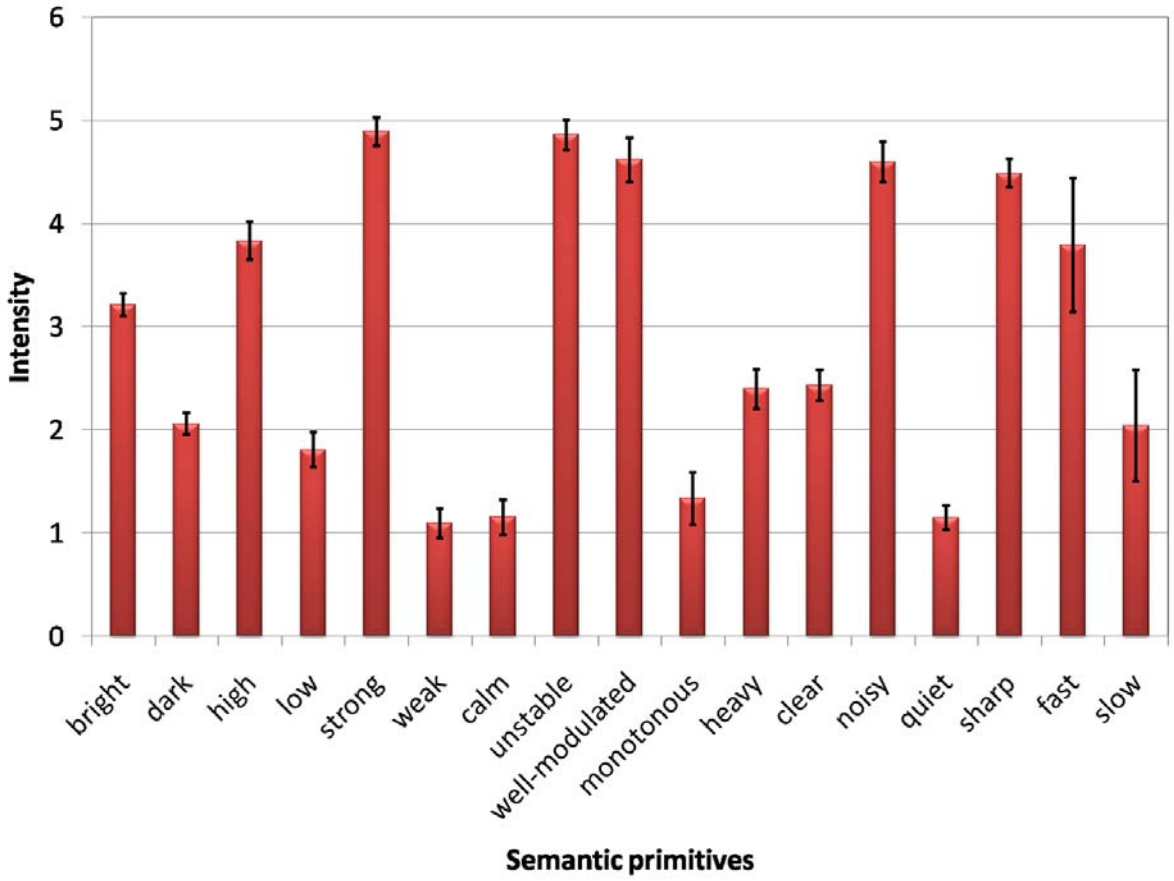

図 3.12: 発話意図: Hot-Anger の音声データについての基本的心理特徴の評価値の平均と分散.

第4章　感情認識システムの実装

4.1　はじめに

本章では，前章までに紹介した，感情知覚多層モデル，ファジィ推論システム，富士通感情音声データから収集した数値を用いることで感情知覚多層モデルに基づいた感情認識システムを実装する．提案認識システムでは従来の感情認識研究の持つ問題点を解決するため，入力に対し一つの感情カテゴリを出力するのではなく，感情の強さの程度を出力させる．また，我々は一つの発話音声から複数の感情をその程度も含めて同時に感じ取ることが出来ることから，複数感情の程度の認識を同時に行う，人間の知覚特性に則した認識システムを構築する．

はじめに，音声データから抽出した音響特徴量から，聴取実験によって得られた基本的心理特徴の評価値を認識する FIS を構築する．同様に，基本的心理特徴の評価値から，聴取実験によって得られた感情カテゴリの評価値を認識する FIS を構築する．

次に，構築したすべての FIS を，感情知覚多層モデルを基に組み合わせることで，本研究で目的としている人間の知覚特性に則した感情認識システムを実装する．

本研究では最終的に図 4.1 のようなシステムを構築する．

4.2　ファジィ推論システムの実装

感情認識システムの実装に際し，図 4.1 における認識部について，音声データから得られた値による入出力関係が満たされるように各層間を接続する必要がある．実装認識システムでは文献 [13][14] とは異なり，感情カテゴリと基本的心理特徴の間だけでなく，全ての層の結びつけにおいて Adaptive Neuro-Fuzzy Inference System (ANFIS) を用いる．本節では，音声データから得られた音響特徴量を入力すると，基本的心理特徴に基づいた印象評価値が出力される基本的心理特徴認識部と，基本的心理特徴の印象評価値を入力すると感情の強さが出力される感情カテゴリ認識部を，複数の FIS によって実装する [23][24]. FIS を用いることで内部処理を明示した上で，複数感情の程度を出力することが出来る．

しかし，全ての入出力を学習に用いてしまうと，未学習の音声に対しての認識性能を調査することが出来ない．そこで，富士通感情音声データに含まれる全ての音源の内約 8 割の音声データを用いて，初期 FIS を構築する．構築に使用していない音源を用いて認識することで，未学習の音源に対しての認識精度が明らかになる．発話意図，発話文章に偏りが無いように，音声データを 5 つのデータセットに分割した．4 つのデータセットで学

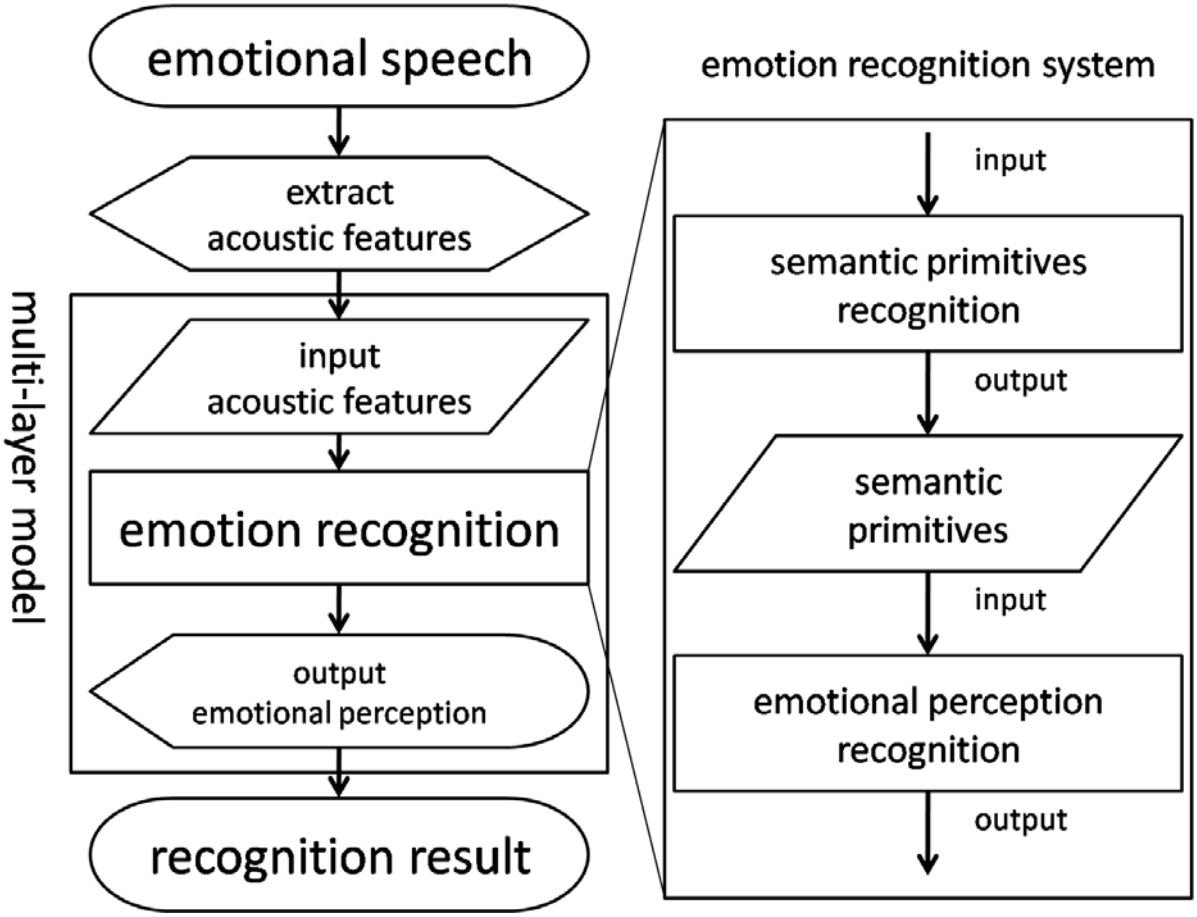

図 4.1: 感情認識システムのフローチャート.

習した FIS で，残りの 1 つのデータセットの認識を行う．また，ニューロ適応学習において過学習による認識誤りが起こらないように，FIS の適切な学習回数を検討する必要がある．Lee and Narayanan は FIS の確認誤差収束点を適切な学習回数としている [19]．本研究では Lee and Narayanan の手法に基づいて，適切な学習回数を設定するため全ての音声データを用いて確認誤差収束点を調査した.

図 4.2 に基本的心理特徴部の内，bright, dark, high, low, strong, weak に関する FIS の二乗平均平方根誤差を，図 4.3 に基本的心理特徴部の内，calm, unstable, well-modulated, monotonous, heavy, clear に関する FIS の二乗平均平方根誤差を，図 4.4 に基本的心理特徴部の内，noisy, quiet, sharp, fast, slow に関する FIS の二乗平均平方根誤差を，図 4.5 に感情認識部の各 FIS の二乗平均平方根誤差を示す．調査の結果，基本的心理特徴 FIS では 120 回程度で，感情カテゴリ FIS では 150 回程度で確認誤差が収束している．そこで，認識システムの実装では，これらの回数のニューロ適応学習を行った.

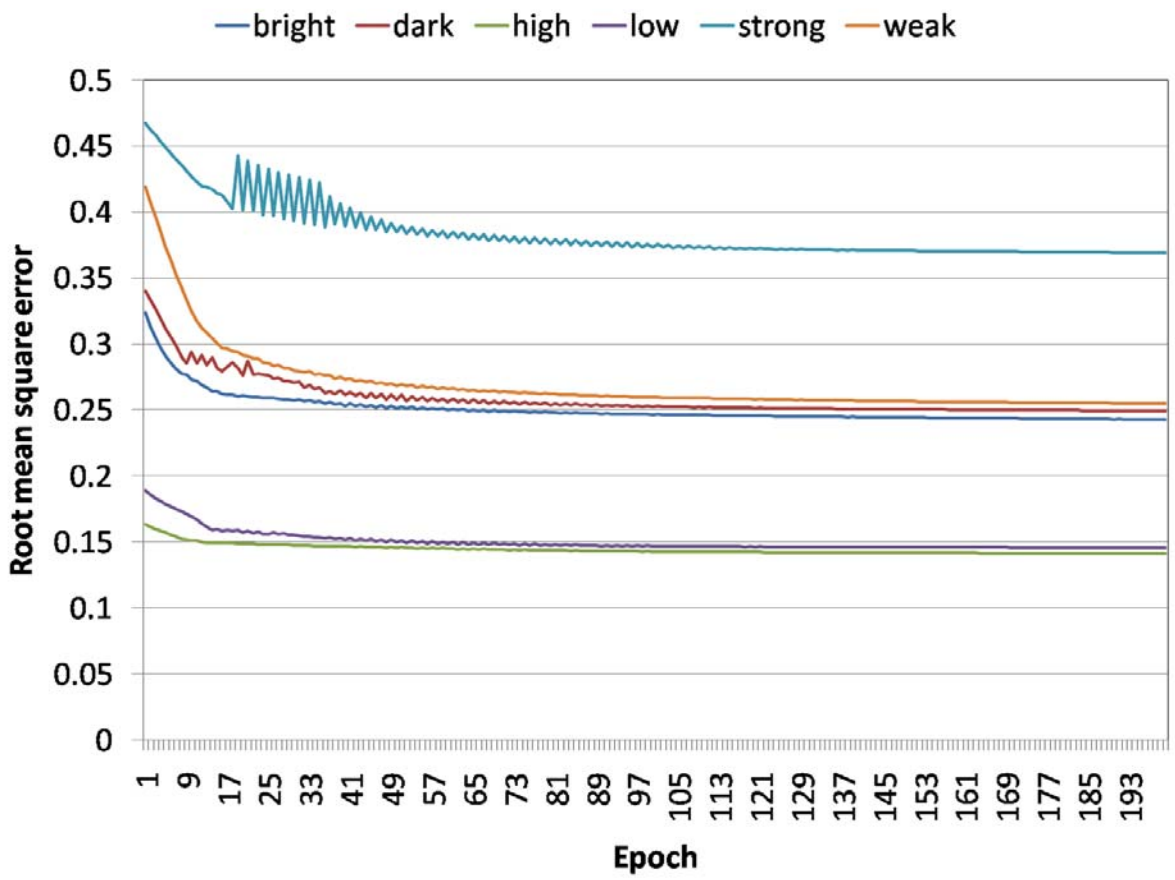

図 4.2: 基本的心理特徴部の内，bright, dark, high, low, strong, weak に関する FIS の二乗平均平方根誤差.

4.3 感情認識システムの実装

FIS は多入力 1 出力の構造を持っている．そのため，音響特徴量 16 項目を入力として基本的心理特徴 1 項目を出力する FIS が 17 種類，基本的心理特徴 17 項目を入力として感情カテゴリ 1 項目を出力する FIS が 5 種類，認識システムに必要となる．それぞれの FIS は，入力の値と出力の値を基に初期 FIS を構築し，学習を重ねることで理想の入出力関係を持つ FIS となる．

そして，構築した 22 種類の FIS を組み合わせることで感情認識システムの実装を行った．聴取実験による評価値のダイナミックレンジは 1.0~5.0 であるため，実装システムの各認識部の出力について，1.0 以下の出力は 1.0, 5,0 以上の出力は 5.0 に丸めた.

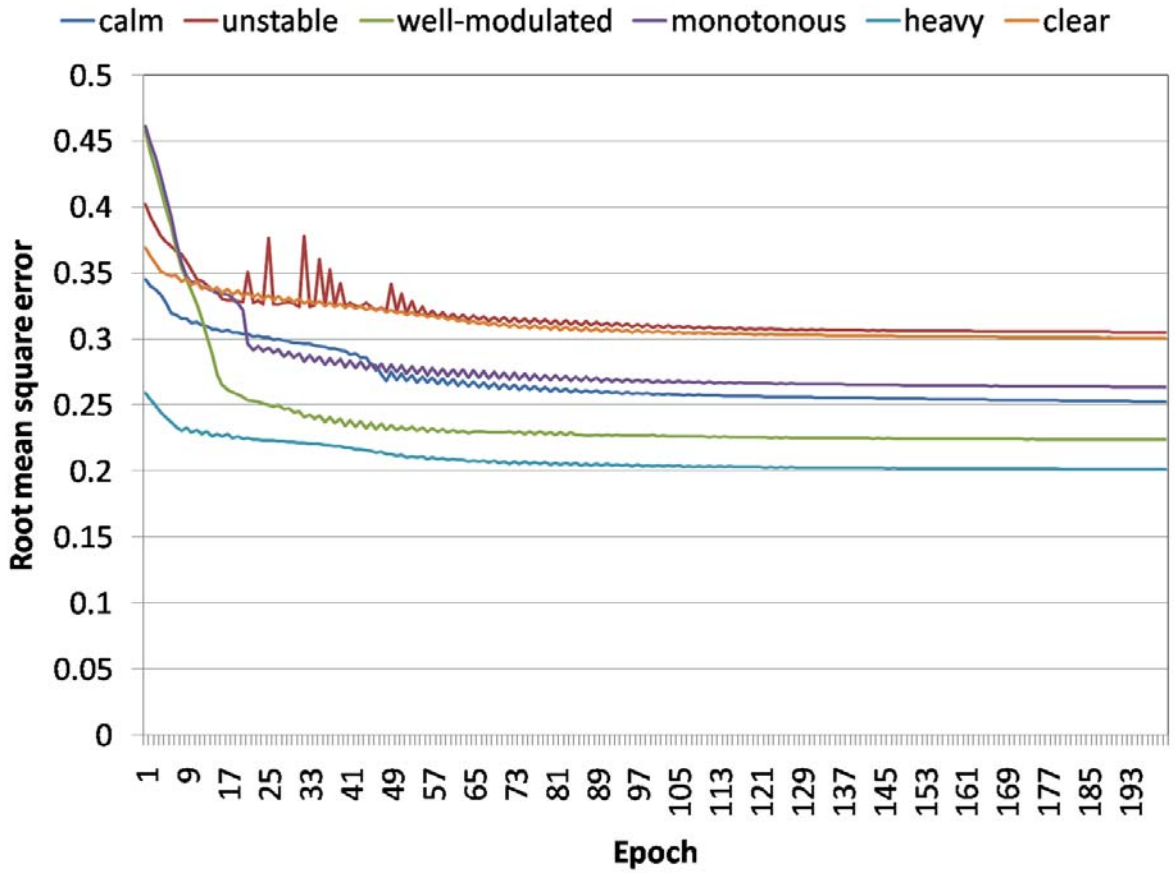

図 4.3: 基本的心理特徴部の内，calm, unstable, well-modulated, monotonous, heavy, clear に関する FIS の二乗平均平方根誤差.

4.4 まとめ

本章では，音響特徴量から基本的心理特徴を認識する FIS，基本的心理特徴から感情カテゴリを認識する FIS を構築した．そして構築した FIS を組み合わせることで，人間の知覚特性に則した感情認識を行うための三層構造を持つ認識システムを実装した．

はじめに，音声データから抽出した音響特徴量を入力として，基本的心理特徴を出力する初期 FIS を 17 種類構築した．また，聴取実験によって得られた基本的心理特徴を入力として，感情カテゴリの程度を出力する初期 FIS を 5 種類構築した．そして，これらの初期 FIS の入出力関係が，音声データの値と近づくよう ANFIS による学習を行った．学習回数については，過学習にならないように，クローズドデータによって調査を行った．そして，調査の結果から決定した学習回数を用いて FIS を最適化した．

次に，構築した FIS を組み合わせることで，感情認識システムの実装を行った．基本的心理特徴認識部の全ての出力が感情カテゴリ認識部の入力になるように実装した．各認識

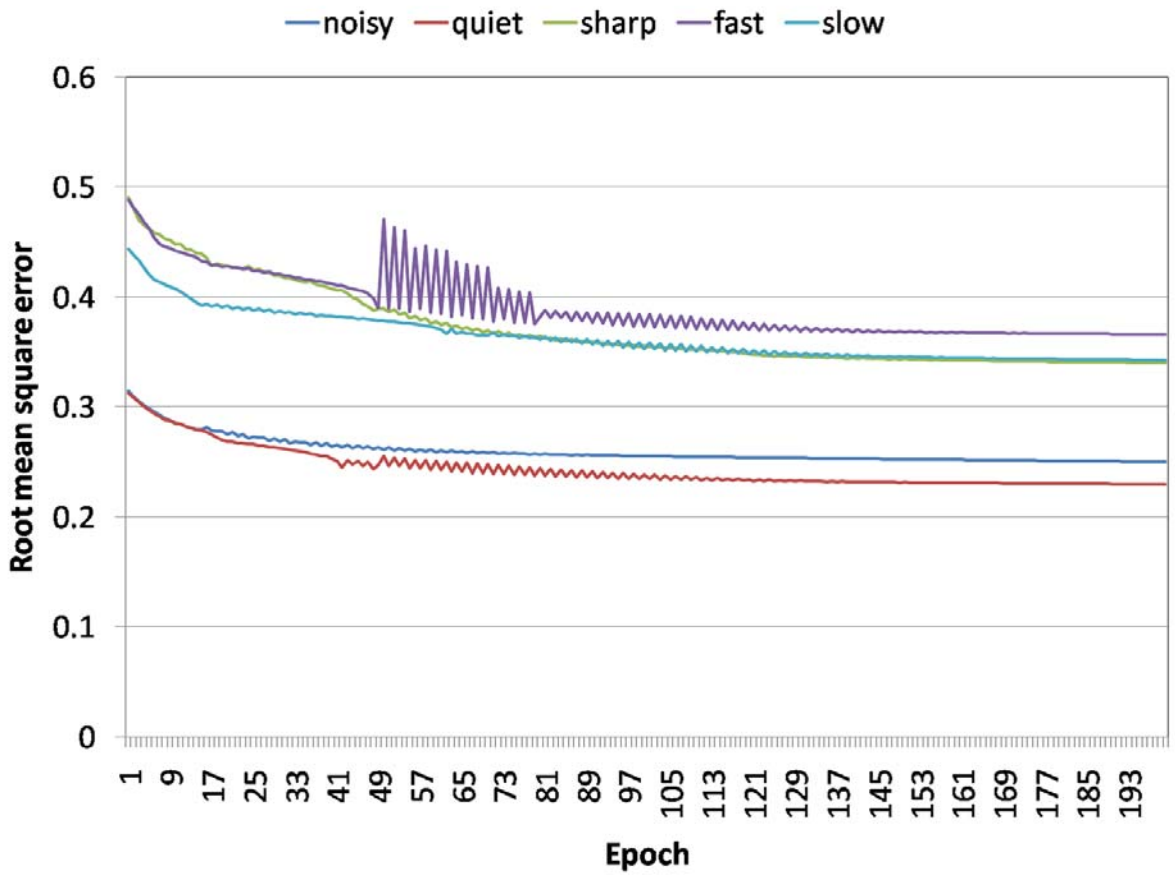

図 4.4: 基本的心理特徴部の内， noisy, quiet, sharp, fast, slow に関する FIS の二乗平均平方根誤差.

部について，聴取実験に沿った値が出力されるよう値を丸めた.

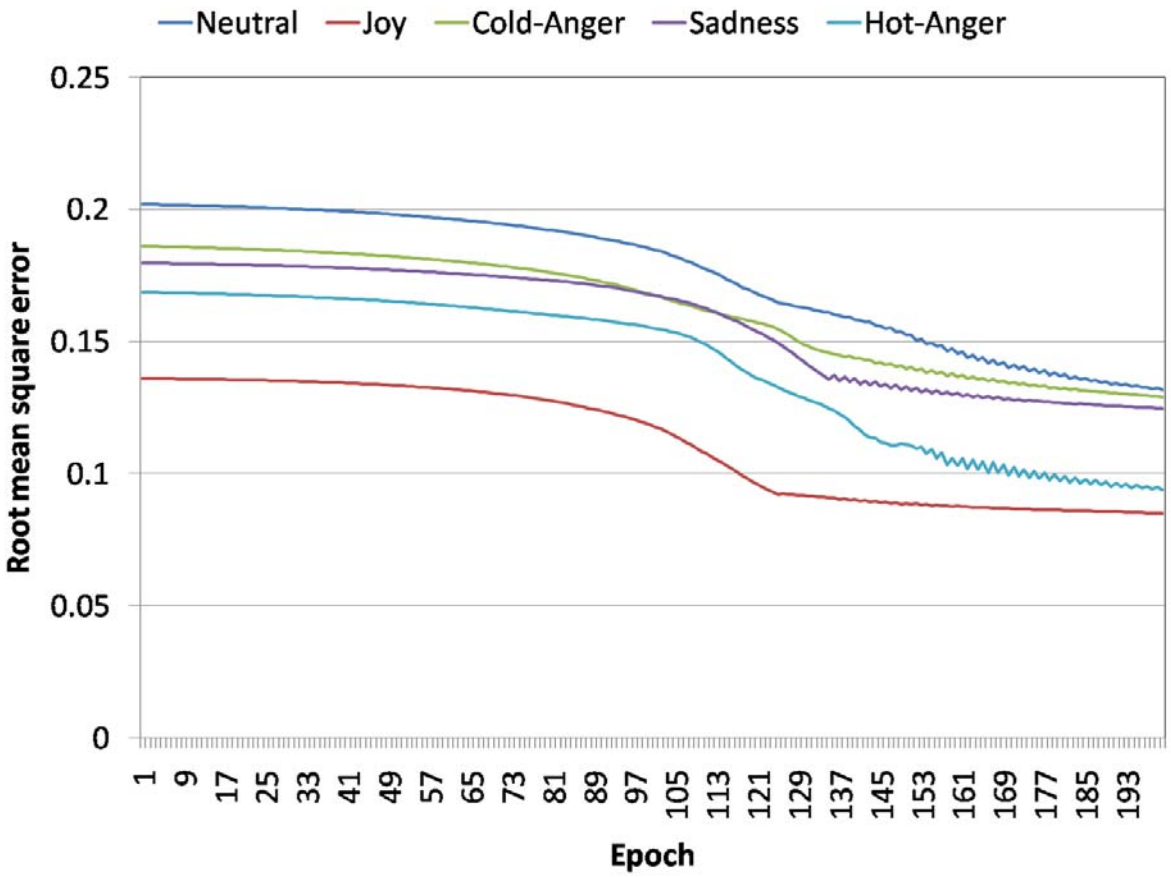

図 4.5: 感情認識部の各 FIS の二乗平均平方根誤差.

第5章　システムの評価

5.1　はじめに

前章までに，感情認識システムの実装を行った．しかし，実装したシステムが実際に人間の知覚特性に則した感情認識を行うのかどうかの確認はされていない．本研究では，多層知覚モデルを基として音声から感情認識することで，人間の知覚過程をより良く表現することを目的としている．そのため，実装されたシステムの認識性能について，多角的に確かめる必要がある．

はじめに，実装したシステムについて，音声データから抽出した音響特徴量を入力した際，期待通りの出力を得られるかどうかについて確かめる．

また，認識システムが従来の研究で用いられた感情認識システムと比較して有効であるかどうかを確かめる必要がある．比較による検証のために，従来用いられている線形手法によるシステム，二層構造システムを基に比較用の感情認識システムを実装する．

そして，それぞれの認識システムの精度について，ユークリッド距離と相関という二つの尺度によって評価を行う．

5.2　システムの動作確認

システムが目的に則して複数の感情の程度を認識出来ているかどうか確かめると共に，基本的心理特徴による中継がうまく働いているかどうかを確かめる必要がある．

実装したシステムの検証の一環として，認識実験によって実装したシステムの動作検証を行った．

まずはじめに，基本的心理特徴認識部について，各音声データの音響特徴量をシステムに入力することで，17 種類の基本的心理特徴の程度の出力が得られているか確認したところ，認識部の出力は 17 次元のベクトルによって基本的心理特徴を表現出来ていた．

また，感情認識部に基本的心理特徴認識部の出力を入力した結果，5 種類の感情の程度の出力が得られ，5 次元のベクトルによって多様性·連続性に対応した感情表現を行うことが出来ていた．

以上の結果から，システムは期待通りの認識構造を持つことが確認できた．次節以降の比較検証により，実装システムの認識精度について言及していく．

5.3 比較用認識システムの実装

本研究で実装したシステムがより良く人間の知覚特性を表現出来ているかを検証するため，システムの認識精度について従来手法によるシステムと比較する．従来の研究で用いられている線形手法と二層構造を，FIS と三層構造と組み合わせることで，実装システム以外に 3 種類の認識システムを実装する．

5.3.1 重線形回帰予測を用いた認識システム

本研究では，非線形で曖昧な関係にある人間の知覚を表現するため，FIS を用いて実装した．これに対し，FIS による実装が有効であるかを確かめる必要がある．そのため，従来手法で用いられる重線形回帰予測 (Multiple Regression Analysis: MRA) を用いて同様のシステムを実装し，FIS によるシステムと性能を比較することで，感情認識システムの実装手法としての FIS の有効性を確かめる．

5.3.2 二層構造モデルに基づく認識システム

本研究では，人間の知覚特性に則した感情認識を行うため，感情知覚多層モデルに基づいて認識システムを実装した．感情知覚多層モデルを用いることにより，音響特徴量と感情を結ぶ基本的心理特徴層が挿入された．二層を結ぶ層の挿入が，認識性能に与えた影響を確かめるため，従来の二層構造モデルに基づいた認識システムと比較する必要がある．

本研究で実装した認識システムと同様に，FIS を用いかつ二層構造である認識システムを構築する．このシステムと本研究で実装したシステムの性能を比較することで，多層構造による利点を明らかにする．

二層構造 FIS 認識システムの構築のためには，4.2 節と同様に FIS の確認誤差収束点を調査する必要がある．図 5.1 にクローズドデータによって得られた，音響特徴量から感情を認識する FIS の二乗平均平方根誤差を示す．この調査の結果から，二層構造 FIS システムを構成する各 FIS の学習回数を 150 回に定めた．

5.3.3 二層構造重線形回帰予測認識システム

比較検証において用いる，MRA と二層構造を組み合わせて，二層構造を持つ MRA による認識システムを実装する．このシステムとこれまでに紹介した 3 種のシステムを比較することで，認識システムの総合的な検証を行うことが出来る．

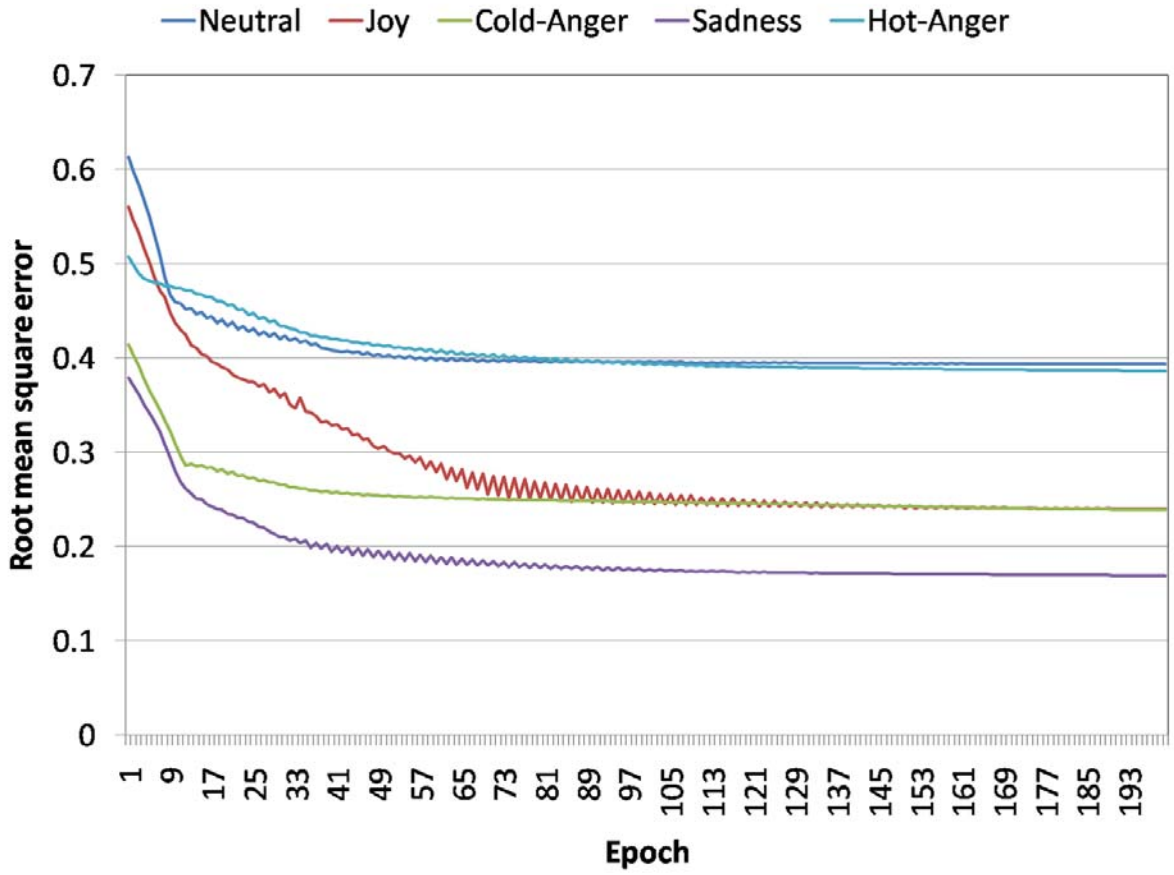

図 5.1: 音響特徴量–感情カテゴリ間 FIS の二乗平均平方根誤差.

5.4 各システム比較による評価

各認識システムの認識精度を議論する基準として，システムの出力値 (5 感情の強さの程度) が理想の出力値 (主観評価による 5 種類の感情の印象の程度) に近い値であるかどうかと，強さの程度の相対関係が類似しているかどうかの二点が重要となる．そこで，各感情の強さとシステムの認識出力を 5 次元のベクトルとして，それぞれの距離，及び，相関を評価対象とする．

5.4.1 ユークリッド距離による比較

本研究では，感情の程度を 5 次元のベクトルとして扱っている．そこで，聴取実験によって得たベクトルと，各認識システムの出力であるベクトルの間のユークリッド距離を算出し，その値によって認識性能を評価する．

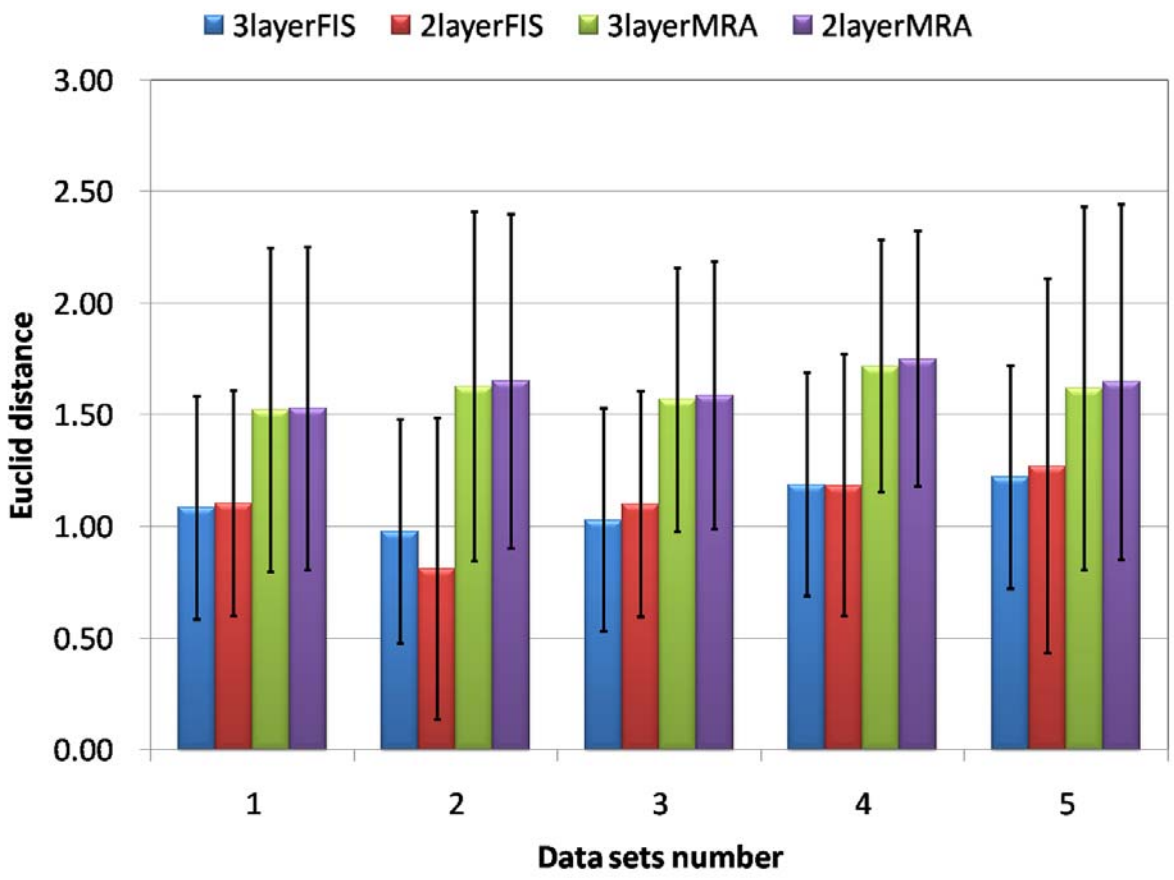

図 5.2: データセットごとの実験評価値と認識出力の間のユークリッド距離.

n 次元の 2 つのベクトル u= $(u_1, u_2, \ldots, u_n)$ と v= $(v_1, v_2, \ldots, v_n)$ の間のユークリッド距離 $d(\mathrm{u}, \mathrm{v})$ は次式によって算出される.

$$d(\mathrm{u}, \mathrm{v}) = \|\mathrm{u} - \mathrm{v}\| = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2} \tag{5.1}$$

算出されたユークリッド距離が短いほど，理想の出力に近い値が出力されているということになる．図 5.2 に個々のデータセットについての，図 5.3 に全ての音源についてのユークリッド距離による評価結果を示す.

感情認識システムの実装に際し，手法として FIS を用いた場合と MRA を用いた場合をユークリッド距離に関し比較すると，FIS を用いたシステムは MRA を用いたシステムよりも良い値が出ている.

モデルの層構造の違いによる影響を確認するため，三層構造を持つシステムと二層構造を持つシステムを比較すると，二層構造システムの方がばらつきが大きい傾向にあるものの，総合的に見ると構造の違いによる差は見られない．これらの結果について有意差検定

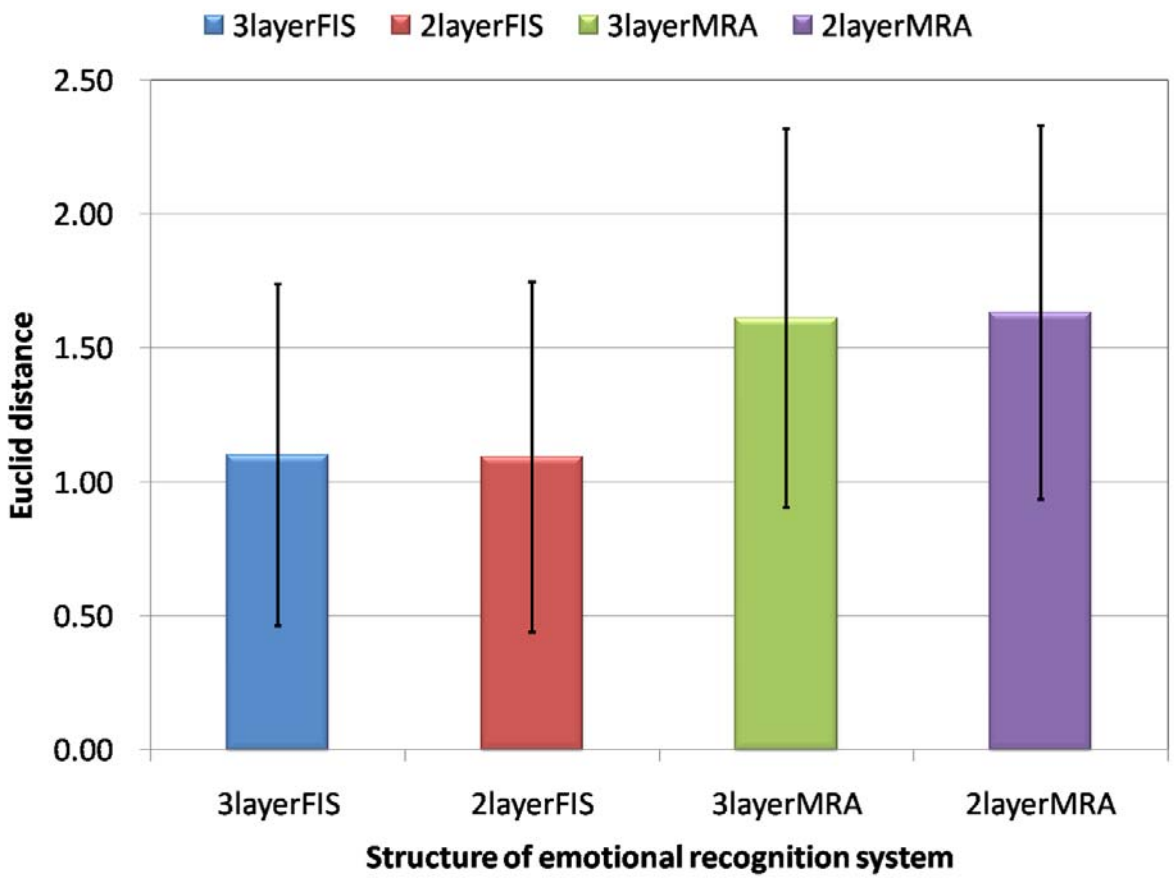

図 5.3: 全ての音源についての実験評価値と認識出力の間のユークリッド距離.

を行ったところ，有意水準 0.01 において三層構造と二層構造のシステムの間に有意差は見られなかった.

5.4.2 相関による比較

本研究では，人間の曖昧な知覚を表現するため，複数の感情の程度を同時に認識している．そのため，それぞれの感情の相対関係を表現出来ているかどうかという点が重要となる．そこで，聴取実験によって得た各感情のデータ列と，各システムの出力データ列の相関を調査する.

相関は 2 組の確率変数の間の類似性の度合を示す統計学的指標である．原則単位は無く，-1 から 1 の間の実数値を取る.

2 組の数値からなるデータ列 $(x, y) = \{(x_i, y_i)\}(i = 1, 2, \ldots, n)$ が与えられた時，次式によ

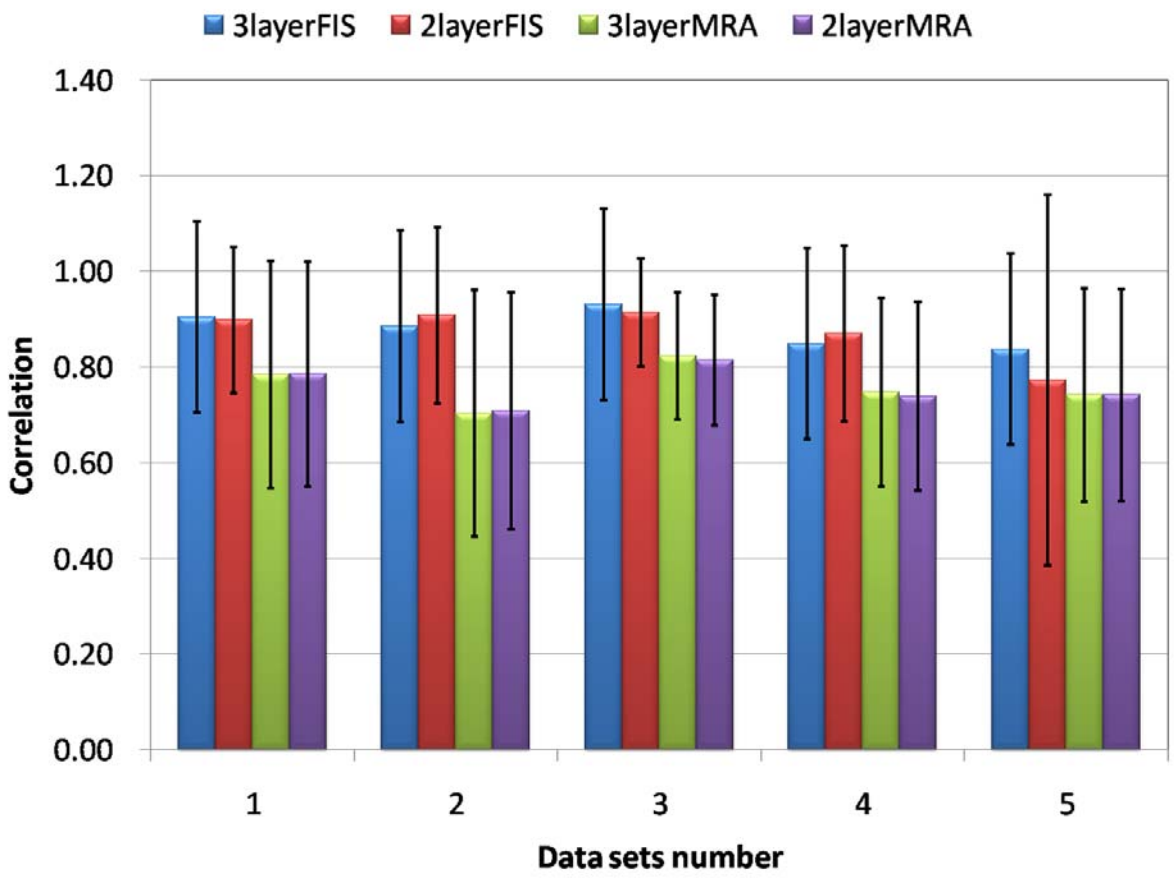

図 5.4: データセットごとの実験評価値と認識出力の間の相関.

り相関係数 R を求めることが出来る.

$$R = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{5.2}$$

ただし，$\overline{x}, \overline{y}$ はそれぞれデータ $x = \{x_i\}, y = \{y_i\}$ の相加平均である.

算出された値が 1 に近ければ近いほどそれぞれの感情の相対関係を表現出来ているということになる．算出されたユークリッド距離が短いほど，理想の出力に近い値が出力されているということになる．図 5.4 に個々のデータセットについての，図 5.5 に全ての音源についての相関による評価結果を示す.

感情認識システムの実装に際し，手法として FIS を用いた場合と MRA を用いた場合を相関に関し比較すると，FIS を用いたシステムは MRA を用いたシステムよりも良い値が出ている.

モデルの層構造の違いによる影響を確認するため，三層構造を持つシステムと二層構造を持つシステムを比較すると，二層構造システムの方がばらつきが大きい傾向にあるもの

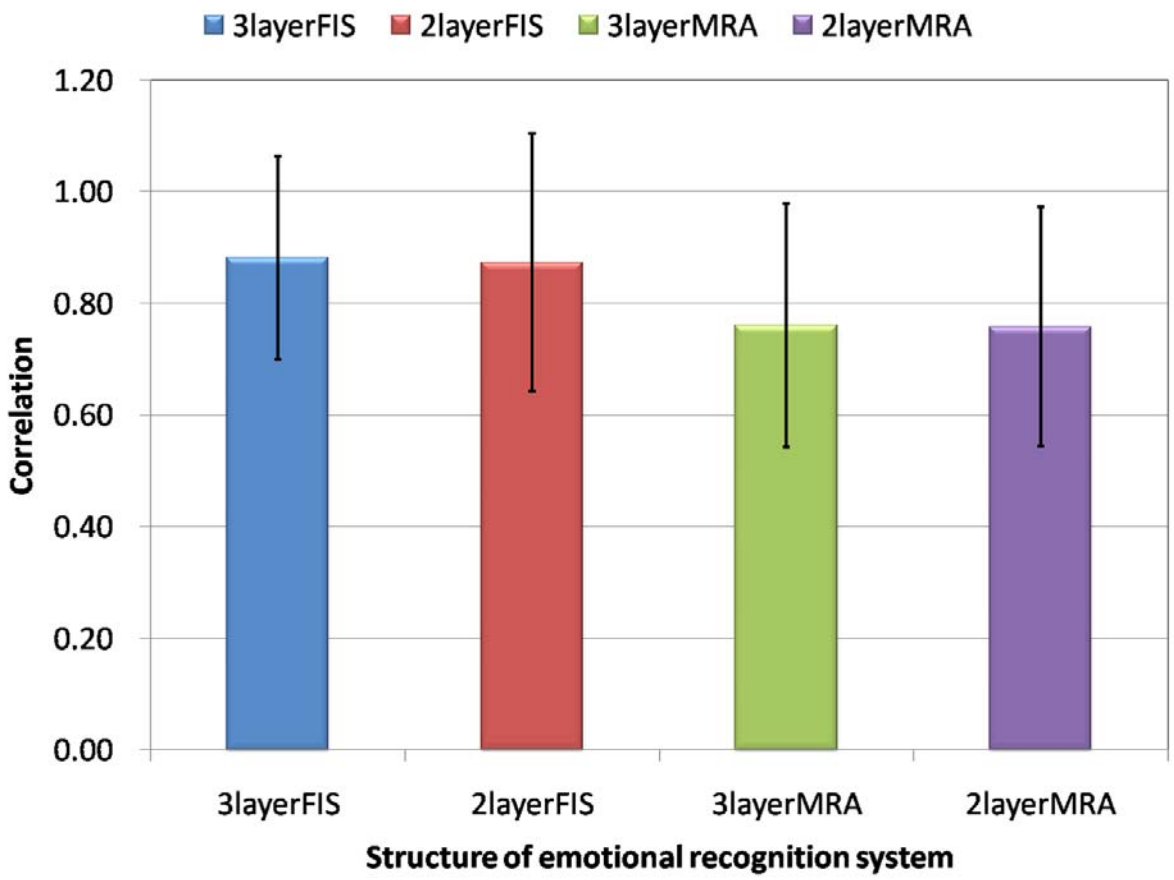

図 5.5: 全ての音源についての実験評価値と認識出力の間の相関.

の，総合的に見ると構造の違いによる差は見られない．これらの結果について有意差検定を行ったところ，有意水準 0.01 において多層構造と二層構造のシステムの間に有意差は見られなかった.

5.4.3 考察

比較実験の結果を検討するための考察を行う．三層構造と二層構造の認識精度が同等であることから，人間の知覚過程を説明しようとして基本的心理特徴を挿入した三層構造モデルでは，挿入した層は認識精度に対して悪影響を与えず，目的通り知覚の内部構造を説明する層となっている．すなわち，感情知覚多層モデルは基本的心理特徴層によって，二層構造における感情カテゴリと音響特徴量の間の内部処理を明示していると考えられる.

また，手法として FIS を用いた結果が MRA によるものよりも良好なものであったことから，FIS は，従来のシステムの持つ線形手法では十分に表現できていなかった，人間の曖昧な知覚をより良く表現出来る手法であることが確認された.

三層構造と FIS を用いたことで，従来の手法よりも人間の知覚特性に則した感情認識が実現できたと考えられる．

5.5 まとめ

本章では，実装した感情認識システムの認識性能について検証を行った．

はじめに，システムが動作するかどうかを確かめ，得られた出力に関する評価法を検討した．

そして評価を行うために，MRA による三層構造認識システム，FIS による二層構造認識システム，MRA による二層構造認識システムを構築した．

そして，聴取実験によって得られた評価値と，これらのシステムに音声データから抽出された音響特徴量を入力した出力の関係を，ユークリッド距離と相関を用いることで比較し，認識性能を検討した．

その結果，ユークリッド距離，相関ともに FIS を用いたシステムは MRA を用いたシステムよりも向上しており，感情認識システムを実装する上で，FIS は有効であると考えられる．

また，モデルの層構造による影響を確認するため，多層構造を持つシステムと二層構造を持つシステムを比較すると，ユークリッド距離，相関ともに構造の違いによる差は見られず，これらの結果について有意差検定を行ったところ，ユークリッド距離，相関ともに有意水準 0.01 において多層構造と二層構造のシステムの間に有意差は見られなかった．

認識結果から考察すると，感情知覚多層モデルは基本的心理特徴層によって，二層構造における感情カテゴリと音響特徴量の間の内部処理を明示していると考えられる．

三層構造と FIS を用いたことで，人間の感情知覚の過程を表現する基本的印象の変化を明示することが可能となった．認識精度の比較から，このシステムによって従来のシステムよりも，人間の主観に近く，かつ知覚過程を説明する感情認識が可能であることが示された．

第6章　結論

6.1　本論文で明らかになったことの要約

これまで，音声からの感情認識について様々な研究がなされてきたが，従来の感情認識は物理量から直接感情カテゴリにマッピングするという手法が主であった．しかし，人間は音声から物理量を正確に判断することは出来ず，むしろ曖昧な基本的印象により判断していることを考慮すると，明確な値を持つ音響特徴量から多様なしかも曖昧な感情の程度を直接認識出来るとは考えにくい．そのため，従来の手法では十分に人間の知覚過程を模擬することが出来ていないと言える．

本稿では，人間の感情知覚に則した感情認識を目的とし，感情を表現した知覚モデルである感情知覚多層モデルに基づく感情認識システムを構築した．そして，実装した感情認識システムが，従来手法よりも人間の知覚過程を模擬できているかどうかを確認するため，従来手法に基づいた感情認識システムを実装し，システムの認識性能の比較を行った．

その結果，実装した三層構造の認識システムは，従来の二層モデルに基づく認識システムと同等の認識精度を持っていた．このことから，感情知覚多層モデルを構成する基本的心理特徴層は，物理量である音響特徴量や心理量である感情の程度と相反するものではなく，むしろ知覚過程を説明するものであることが示唆された．このことから，認識システムの三層構造は人間の知覚特性に則したものであると言える．

また，実装に際し非線形で曖昧な関係にある人間の知覚を表現するため，FIS を用いたところ，線形手法によって実装したシステムと比較して性能の向上が見られた．このため，感情認識システムの実装手法として FIS が有効であることが確認できた．

三層構造と FIS を用いたことで，人間の感情知覚の過程を表現する基本的印象の変化を明示することが可能となった．認識精度の比較から，このシステムによって従来のシステムよりも，人間の主観に近く，かつ知覚過程を説明する感情認識が可能であることが示された．本研究では，従来のシステムと比べ，より人間の知覚特性に則した感情認識を，感情知覚多層モデルと FIS を用いた認識システムの実装により実現出来た．

6.2　今後の課題

今後の課題を以下に記す．

- 本研究では，感情について Neutral も含めた 5 次元の独立したベクトルとして扱った．しかし，Neutral を基準としてどのような違いがあるか調査したことを考慮する

と，5 次元ベクトルによる表現は適切であるとは言い難い．また，今回用いていない感情についても表現する必要があることから，出力における感情表現について検討する必要がある．また，未使用の感情についても認識出来るようにするためには，基本的心理特徴及び音響特徴量の要素について再考する必要がある．

- 本研究においては，音響特徴量として Neutral を基準とした比率を用いた．しかしながら，基準とした Neutral はあくまでも音声データベースのラベルに沿ったものであり，実際の聴取実験によって判断された Neutral とは必ずしも一致しない．そのため，基準の選び方を再考必要がある．

付 録A　別データベースによるシステムの検証

A.1　はじめに

本研究における実装システムについて，富士通感情音声データベースを用いての検証を行った．しかし，使用データベースは単一話者による発声データしか含まれておらず，異なる話者による発声データに対しての有効性は検証されていない．そこで，異なる話者による発声に対しての認識性能を確かめることで，実装システムのさらなる検証を行う．

A.2　別データベースからのデータ分析

本システムの有効性を確かめるため，他の音源による認識実験を行う．使用する音源はベルリン感情音声データベースのうち，女性話者である話者 16 によるものを選んだ [27]. 話者 16 による発話の中から Neutral, Joy, Sadness, Hot-Anger を含む発話文章 2 種類からなる 9 種類の発話音声を選んだ.

これらの音声について，3. 3 章と同様の音源分析と聴取実験を行った．被験者は母国語を日本語とし，ドイツ語が分からない 20〜30 代の日本人大学院学生 15 名 (男性 14 名, 女性 1 名) とする．実験の流れは 3. 3. 2 章に準ずるが，音声データ数が異なる．

A.3　別データベースに対する認識性能の検証

別データベースに対しても期待通りの感情認識が出来るのかを確かめるため，富士通感情音声データの全ての発話を学習に用いて構築した，多層構造 FIS システムに抽出した音響特徴量を入力し，認識を行った．聴取実験結果と，三層構造認識システム及び二層構造認識システムの出力から得られたユークリッド距離と相関による認識精度を表 A.1 に示す.

A.4　考察

別音源に対する三層構造認識システムの認識精度は，二層構造認識システムよりも高かった．このことから，別音源に対しても二層構造と同等以上の認識性能がある可能性が

表 A.1: ドイツ語データベースによる主観評価値とシステム認識値の比較.

	ユークリッド距離	相関
3layer	2.23	0.64
2layer	2.64	0.55

示唆された．一方で，富士通感情音声データによる検証と比べると認識率は低下した．

この原因を明らかにするため，個々の音源についての結果を調査したところ，9 種類の発話音声のうち 5 種類の発話音声については，富士通感情音声データを使用した際のシステムの認識精度と同等であった．しかし，残る 4 種類の発話音声については認識精度が低下している．この原因として，認識システムの基準となる評価値と，正規化した音響特徴量の比率が異なっている点が挙げられる．聴取実験の結果において，Neutral と Cold-Anger の間に 0.92 という高い相関が見られた．しかし，システムの出力においての相関では，0.76 に止まった事から，ベルリン感情音声データベースにおいては，富士通感情音声音声データにおける Cold-Anger の知覚とは異なる判断を行っていると考えられる．

参考文献

[1] 金澤博史, クリスマエダ, 竹林洋一, “計算機との対話のための非言語情報の認識と合成,” 信学論, Vol. J77–D–II, No.8, pp. 1512–1521, 1994.

[2] 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,” 音響講論 (秋), pp. 601–602, 2007.

[3] 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, SP2006–167, pp. 43–48, 2007.

[4] 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,” 信学技報, SP2007–206, pp. 117–122, 2008

[5] 中村友彦, 北村達也, 赤木正人, “fMRI を用いた歌声と話声における脳活動の差異の検討,” 音響研資, H–2008–108, 2008.

[6] 北村達也, “音声の個人性の生成と知覚,” 音響研資, H–2008–114, 2008.

[7] 平賀裕, 斎藤善行, 森島繁生, 原島博, “音声に含まれる感情抽出の一検討,” 信学技報, HC93–66, pp. 1–8, 1994.

[8] 林康子, “感動詞「ええ」におけるピッチ曲線と感情認知,” 信学技報, H98–61, pp. 65–72, 1998.

[9] 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,” 信学技報, SP2001–141, pp. 43–50, 2002.

[10] 磯部理沙子, 桐生昭吾, 武田昌一, 安田祐利, 真紀子, “声帯情報を用いた怒りの音声合成の試み,” 音響講論 (秋), pp. 347–348, 2008.

[11] エリクソン · ドナ, 昇地崇明, “日本人学童による感情音声の知覚,” 信学技報, SP2006–28, pp. 7–12, 2006.

[12] 沢村奏絵, 党建武, 赤木正人, Qiang Fang, Donna Erickson, 櫻庭京子, 峯松信明, 広瀬啓吉, “異文化間の感情音声の認知における共通要素についての検討,” 音響講論 (春), pp. 457–458, 2007.

[13] Chun-Fang Huang, Masato Akagi, “A Multi-Layer fuzzy logical model for emotional speech perception,” *Proc. EuroSpeech 2005*, pp. 417–420, Lisbon, Portugal, 2005.

[14] Chun-Fang Huang, Masato Akagi, “A three-layerd model for expressive speech perception,” *Speech Commun.*, Vol.50, pp. 810-828, 2008.

[15] 白澤敏行, 山村毅, 田中敏光, 大西昇, “音声に込められた感情の判別,” 信学技報, HIP96–38, pp. 79–84, 1997.

[16] 刀根優子, 荻原昭夫, 柴田浩, “音声対話システムのための HMM に基づく感情判別,” 信学技報, SP99–22, pp. 47–53, 2000.

[17] 廣瀬陽介, 平原誠, 永野俊, “教師付き独立成分分析による音声の感情認識,” 信学技報, NC2002–152, pp. 113–118, 2003.

[18] 森山剛, 小沢慎治, “ファジー制御を用いた音声における情緒性評価法,” 信学論, Vol. J82–D–II, No.10, pp. 1710–1720, 1999.

[19] C. Lee and S. Narayanan, “Emotion recognition using a data-driven fuzzy inference system,” *Proc. Eurospeech 2003*, pp. 157–160, Geneva, Switzerland, 2003.

[20] 齋藤毅, 辻直也, 鵜木祐史, 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,” 音響誌, Vol. 64, no. 5, pp. 267–277, 2008.

[21] Donna Erickson, “Expressive speech: Production, perception and application to speech synthesis,” Acoust. Sci. & Tech., Vol. 26, No. 4, pp. 317–325, 2005.

[22] 上田和夫, “音色の表現語に階層構造は存在するか,” 音響誌, Vol. 44, no. 2, pp. 102–107, 1988.

[23] 坂和正敏, “ファジィ理論の基礎と応用,” 森北出版, 1990.

[24] J. S. R. Jang, C. T. Sun, E. Mizutani, “Neuro-Fuzzy and Soft Computing,” Prentice Hall, 1996.

[25] 河原英紀, “聴覚の情景分析が生んだ高品質 vocoder: STRAIGHT,” 音響誌, Vol. 54, no. 7, pp. 521–526, 1998.

[26] H. kawahara, I. Masuda-Katsuse, A. Cheveigne, “Resturcturing Speech Representations Using a Pitch Adaptive Time-Frequency Smoothing and an Instantaneous-Frequency-Based F0 Extraction,” *Speech Commun.*, Vol.27, pp. 187-207, 1999.

[27] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss, “A Database of German Emotional Speech,” *Proc. EuroSpeech 2005*, pp. 1517–1520, Lisbon, Portugal, 2005.

謝辞

本研究を遂行するにあたり，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の意を表します.

本研究を遂行するにあたり，貴重な御助言をご指導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 鵜木祐史准教授，党建武教授，徳田功准教授，小谷一孔准教授，李軍鋒助教，末光厚夫助教に心より感謝致します.

本研究を遂行するにあたり，多大なる御助言と御協力を賜りました本学修了生である黄純芳氏に心より感謝致します.

本研究を遂行するにあたり，日頃から熱心な議論と多面にわたる御協力を賜りました，北陸先端科学技術大学院大学 音情報処理学講座の皆様，知能情報処理学講座の皆様，及び諸先輩方に厚くお礼申し上げます.

筆者が武蔵工業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武蔵工業大学 工学部 桐生昭吾教授，知識工学部 今井章久准教授に心より感謝致します.

本学での研究生活をおくるにあたり，貴重な御助言を賜りました本学修了生である桜井裕氏に心より感謝致します.

本学での研究生活をおくるにあたり，研究生活の心の支えとなってくれた親族，友人たちに心より感謝致します.

最後に，大学院での貴重な研究生活を与えてくれた両親に心から感謝し，お礼を申し上げます.

本研究に関する研究業績

国際会議

Yuusuke Aoki, Chun-Fang Huang and Masato Akagi, “An emotional speech recognition system based on multi-layer emotional speech perception model,” *Proc. Nonlinear Circuits and Signal Processing ’09, Hawaii, USA,* March 2009. (to appear)

口頭発表

青木祐介, 黄純芳, 赤木正人, “音声からの感情認識による感情知覚多層モデルの評価,” 日本音響学会 2009 年 春季研究発表会 講演論文集, 2–P–18, March 2009. (to appear)