

Title	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
Author(s)	青木, 祐介
Citation	
Issue Date	2009-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/8111">http://hdl.handle.net/10119/8111</a>
Rights	
Description	Supervisor: 赤木正人, 情報科学研究科, 修士

# 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

青木 祐介 (0710001)

北陸先端科学技術大学院大学 情報科学研究科

2009年2月5日

キーワード: 感情音声, 感情認識, 多層モデル, ルールベース, ファジィ推論システム.

## 1 序論

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 [1][2]。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す。

## 2 感情認識システムの実装

人間の知覚特性に則した音声からの感情認識を実現するため、本研究では、Huang and Akagi の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) を用いて感情認識システムを構築する。感情知覚多層モデルは、物理量である音響特徴量の層、心理量である感情カテゴリの層、そしてその二層を結ぶ形容表現である基本的心理特徴の層の三層で構成される。モデル化の際には、感情カテゴリ層と基本的心理特徴層の間を FIS を用いて接続した。また基本的心理特徴層と音響特徴量層の間を相関により接続した。

ファジィ推論は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 [3][4]。感情知覚多層モデルに基づき、複数の FIS を用いることで、非線形で曖昧な関係にあるモデルの層間を接続し、感情認識システムを構築する。

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。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 発話を使用した。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、基本的心理特徴の程度、感情カテゴリの程度を、音声分析と聴取実験によって収集する。音響特徴量の抽出には STRAIGHT を用いて音声分析を行い、基本的心理特徴及び感情カテゴリの評価値の収集のために聴取実験を行った [5]。

音声データから収集した音響特徴量を入力として、基本的心理特徴の程度を出力する初期 FIS を 17 種類構築した。また、基本的心理特徴の程度を入力として、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 Adaptive Neuro-Fuzzy Inference System による学習を行った。そして、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。

## 3 システムの検証

実装したシステムが、人間の知覚特性に則した感情認識を実現できているかどうかを検証するため、認識システムの性能について従来手法と比較し調査を行った。本研究では、三層構造モデルを用いることにより、音響特徴量層と感情カテゴリ層の間の知覚過程を表現した。二層を結ぶ基本的心理特徴層の挿入が、認識性能に与えた影響を確かめるため、比較対象として二層構造モデルを採用した。また、FIS が線形手法と比べ人間の知覚の曖昧さを表現出来ているかを確かめるため、比較対象として重線形回帰予測 (MRA) を用いた。これらの組み合わせにより実装システムの他に、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実装した。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ときの出力との関係を、ユークリッド距離と相関を用いて比較した。

## 4 結論

本研究では，人間の知覚特性に則した感情認識を実現するため，感情知覚多層モデルに基づき，FIS を用いて感情認識システムを実装した．

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．

## 参考文献

- [1] Chun-Fang Huang, Masato Akagi, “A Multi-Layer fuzzy logical model for emotional speech perception,” *Proc. EuroSpeech 2005*, pp. 417–420, Lisbon, Portugal, 2005.
- [2] Chun-Fang Huang, Masato Akagi, “A three-layerd model for expressive speech perception,” *Speech Commun.*, Vol.50, pp. 810-828, 2008.
- [3] 坂和正敏, “ファジィ理論の基礎と応用,” 森北出版, 1990.
- [4] J. S. R. Jang, C. T. Sun, E. Mizutani, “Neuro-Fuzzy and Soft Computing,” Prentice Hall, 1996.
- [5] 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.