JAIST Repository
https://dspace.jaist.ac.jp/

Title	音声中の感情認識のための新しい認識方略に関する研究
Author(s)	赤木，正人
Citation	科学研究費助成事業研究成果報告書：1-4
Issue Date	2013-05-15
Type	Research Paper
Text version	publisher
URL	http://hdl.handle.net/10119/11370
Rights	
Description	研究種目：挑戦的萌芽研究，研究期間：2010～2012，課題番号：22650032，研究者番号：20242571，研究分野：音声情報処理，科研費の分科・細目：知覚情報処理・知能ロボティクス

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２５年５月１５日現在

機関番号：13302
研究種目：挑戦的萌芽研究
研究期間：2010～2012
課題番号：22650032
研究課題名（和文）　音声中の感情認識のための新しい認識方略に関する研究

研究課題名（英文）　A study on new strategy of emotion recognition in speech

研究代表者
赤木　正人（AKAGI MASATO）
北陸先端科学技術大学院大学・情報科学研究科・教授
研究者番号：20242571

研究成果の概要（和文）：

本研究では，感情を基本因子ベクトル Arousal - Valence - Dominance の合成ベクトルとして表現するという新しい発想のもと，申請者らが提案している音声中の感情知覚モデルを感情音声認識に適用し，感情が複数含まれる音声からそれぞれの感情の程度までを推定する手法を提案した。評価の結果，感情空間へのマッピングについて提案法が最もヒトの特性に近く，認識精度も GMM を用いた手法と比較して本手法が認識率で大きく優れていることが確認できた。

研究成果の概要（英文）：

This study proposed a method of emotion recognition in speech, which can estimate not only the emotion itself but the degree of each emotion from speech that plural emotions are included in. This method represents each emotion as a resultant vector of the basic factor vectors, Arousal - Valence - Dominance. As the results of applying this method with our already proposed emotion perception model to emotion recognition in speech, the mapping of speech to the emotional space is the most correspondent to human responses. In addition, the recognition accuracy is also greatly excellent at the recognition rate compared with that by GMM.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	1,500,000	0	1,500,000
2011 年度	800,000	240,000	1,040,000
2012 年度	600,000	180,000	780,000
年度			
年度			
総　計	2,900,000	420,000	3,320,000

研究分野：音声情報処理
科研費の分科・細目：知覚情報処理・知能ロボティクス
キーワード：①音声認識　②感情音声　③音声知覚モデル　④感情基本因子　⑤対話解析

１．研究開始当初の背景

音声には大きく分けて言語情報（何を話しているか）と非言語情報（感情，個人性等）が含まれる。音声コミュニケーションではこれら両方が送受されている。このため，音声

対話の精緻な解析のためにはこれら双方を考慮する必要がある．特に人－人の対話解析に基づいて人－機械のインターフェースを構築しようとする場合，言語情報（音声認識）だけではなく，話し手の感情がどのように変化しているかという情報（感情認識）は重要な要素となる。

現在，感情認識の研究は，音声関係で権威ある国際会議（ICASSP，InterSpeech 等）で多く発表されるようになってきた．2009 年度の InterSpeech では，チュートリアルおよびスペシャルセッションで感情音声認識のセッションが組まれ，1 日以上このテーマが議論された．ところが，これらの研究では感情をカテゴリととらえ，従来型のパターン認識技術，すなわち音声認識・文字認識等で使用されてきた「入力を各感情カテゴリに振り分ける技術」（カテゴリ判別器）が用いられている。しかし，この方法が感情認識本来の目的を達成しているかどうか甚だ疑問である。なぜならば，人は，同じ感情（たとえば怒り）でも「少し怒っている」あるいは「かなり怒っている」というように感情の程度まで知覚している。また，一つの発話文から「怒っているけど悲しそうだ」などの複数の感情を知覚する。このため，機械による感情認識においても，複数の感情を同時にその程度までを含めて認識するシステムを構築する必要がある。

２．研究の目的

対話解析等で，送受された情報の内容をより精緻に解析するために対話者の感情の動きを自動的に捉えることが重要となっている。このため，解析手法の中心として音声からの感情認識の技術を確立することが求められており，近年多くの研究が成されている。これらの研究では，音声入力を各感情カテゴリに振り分けるための従来型のパターン認識技術が用いられているが，感情はそもそも従来のパターン認識が対象としているようなカテゴリ構造を持っていない。一つの発話文中においても感情の程度は変化し，また，複数の感情が含まれる場合もある。本研究では，感情を複数の基本因子ベクトルの合成ベクトルとして表現するという新しい発想のもと，研究代表者らが提案している音声中の感情知覚モデルを感情音声認識に適用し，感情が複数含まれる音声からそれぞれの感情の程度までを推定する手法を確立することを目的とする。

３．研究の方法

人が音声中の感情を知覚する場合，知覚された感情の程度は連続的に変化し，しかも複数感情が同時に知覚されることもありうる。このことは，感情認識においては，各感情は従来のパターン認識が対象としているような単純なカテゴリ構造を持っておらず，現有の感情認識システムのように感情をカテゴリとして捉えることはかえって感情認識の本質を捻じ曲げてしまうことを意味する。従来のパターン認識手法が得意とする入力を単一のカテゴリに振り分ける手法ではなく，新たな認識方略が必要となる。

この問題を解くための研究代表者らの提案は，「感情認識のために感情空間の再定義を行いこの空間上での認識手法を考案する」ことである。本研究では，従来の感情認識システムが感情をカテゴリとして捉えていたのとは異なり，感情空間は多数の感情基本因子ベクトルによって張られる連続した多次元空間として捉える（図１）。そして，音声に含まれる物理的音響特徴から個々の感情基本因子ベクトルへのマッピング手法を新たに提案し，感情基本因子ベクトルの合成ベクトルとして感情を表現する手法を考案する。

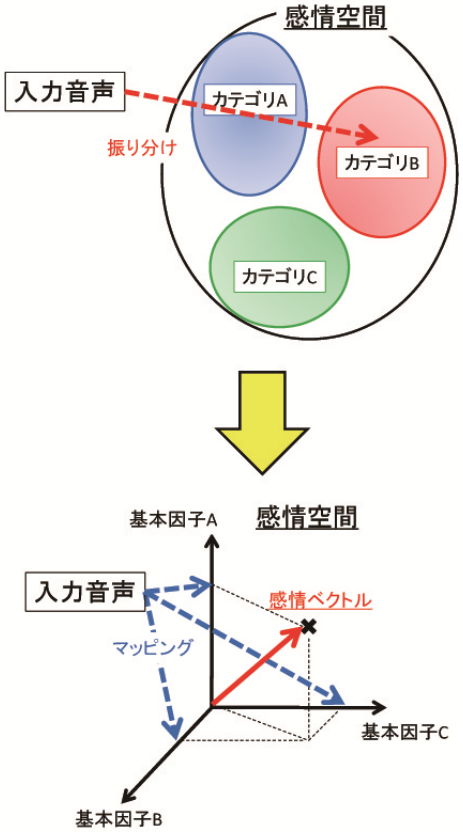

図１　感情空間の再定義および認識方略の変更。基本因子が張る空間として感情を定義。

具体的には，研究代表者らが提案している感情知覚モデル（三層構造感情知覚モデル：Huang and Akagi, Speech Communication 50, pp.810-828, 2008）を，表現豊かな音声の特質を扱う目的で，感情空間の表現として感情基本因子を付け加えることにより四階層構造（音響特徴量，温床表現語群，感情基本因子，感情）とする。感情基本因子としては，“怒り”，“恐れ”，“喜び”などのラベルではなく，感情の印象を表現できるActivation - Evaluation - Dominanceの3次元を採用する。

４．研究成果

(1) 感情空間の再定義

感情音声合成で用いていた三層構造感情知覚モデルに対して，表現豊かな音声の特質を扱う目的で，感情空間の表現として感情基本因子を付け加えることにより四階層構造（音響特徴量，温床表現語群，感情基本因子，感情）とした。感情基本因子としては，“怒り”，“恐れ”，“喜び”などのラベルではなく，感情の印象を表現できるActivation - Evaluation - Dominanceの3次元を採用した。この結果，感情を複数の基本因子ベクトルの合成ベクトルとしてより簡単に表現できるようになり，認識システムの構築が容易となった。

(2) 音響特徴の抽出および知覚モデルの改良

多数の音響特徴から感情基本因子 Arousal - Valence - Dominance の程度の推定を行うために，感情にかかわる適切な音響特徴を選択する手法について検討した。なぜならば，感情基本因子の程度の推定には，Fuzzy Interface System (FIS)を採用することが最も有効であることがわかったが，音響特徴によっては，感情基本因子の程度の推定に悪影響を及ぼすものも存在するからである。これらの検討により，特に従来難しいとされていたValenceについて精度の良い推定が行えるようになり，Arousal - Valence - Dominanceの3つの基本因子ベクトルの合成ベクトルとして感情の推定が行える土台ができた。

(3) 感情空間へのマッピングモデルの評価

提案している三層構造感情知覚モデルを用いて，推定された感情基本因子ベクトル Arousal - Valence - Dominance の組み合わせにより感情空間へのマッピングを行う手法について検討を行った。感情空間へのマッピングについて，聴取実験から得られたヒトの応答特性と比較した結果，従来手法よりもヒトの応答特性の模擬性能は高くなっており，三層構造感情知覚モデルとFISを組み合わせた場合に，最も性能が高いことが分かった。

(4) 感情認識実験

音声認識パイロットシステムの構築を行い，感情認識実験の精度を議論した。日本語およびドイツ語の感情音声に対して，本手法と従来手法であるGMMを用いた手法を適用した場合の認識精度を比較した結果，本手法が認識率で大きく優れていることが確認できた。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計3件）

[1] Dang, J., Li, A., Erickson, D., Suemitsu, A., Akagi, M., Sakuraba, K., Mienmatasu, N., and Hirose, K. (2010/11/01). “Comparison of emotion perception among different cultures,” Acoust. Sci. & Tech. 31, 6, 394-402 （査読あり）.

[2] Zhou, Y., Li, J., Sun, Y., Zhang, J., Yan, Y., and Akagi, M. (2010/10). “A hybrid speech emotion recognition system based on spectral and prosodic features,” IEICE Trans. Info. & Sys., E93D (10): 2813-2821 （査読あり）.

[3] 赤木正人(2010/08/01). “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, 66, 8, 393-398.（解説論文，査読なし）

〔学会発表〕（計7件）

[1] Elbarougy, R. and Akagi, M. (2013/03/01). “ Automatic Speech Emotion Recognition Using A Three Layer Model,” IEICE Tech. Report, SP2012-127 （大同大学，名古屋，愛知県）.

[2] Elbarougy, R. and Akagi, M. (2012/12/04). “Speech Emotion Recognition System Based on a Dimensional Approach Using a Three-Layered Model,” Proc. APSIPA2012 (CD-ROM), Hollywood, USA.

[3] Elbarougy, R. and Akagi, M. (2012/06/14). “Comparison of methods

for emotion dimensions estimation in speech using a three-layered model," IEICE Tech. Report, SP-2012-36 (NTT 研究所, 厚木, 神奈川県).
[4] Elbarougy R. and Akagi, M. (2012/02/25) "A Three-layered model for Automatic Speech Emotion Recognition using a Dimensional Approach," JSPS A3 Foresight Workshop, Ishikawa (粟津温泉, 石川県小松市).
[5] 赤木正人 (2011/10/02). "聴覚と音研究", 音響学会聴覚研究会資料, 41, 7, H-2011-104. (招待講演) (牛岳温泉リゾート, 富山県富山市)
[6] 赤木, 羽二生. (2011/03/09). "音声の知覚と認識 ー人は脳で音声を聞く. 機械は？ー", 日本音響学会平成 23 年春季研究発表会, 1-13-2 (招待講演) (早稲田大学, 東京).
[7] Akagi, M. (2010/11/29). "Rule-based voice conversion derived from expressive speech perception model: How do computers sing a song joyfully?" Tutorial, ISCSLP2010, National Cheng Kung University, Tainan, Taiwan.

〔図書〕(計０件)

〔産業財産権〕
○出願状況 (計０件)

○取得状況 (計０件)

〔その他〕
なし

６．研究組織
(1)研究代表者
赤木　正人 (AKAGI MASATO)
北陸先端科学技術大学院大学・情報科学研究科・教授
研究者番号：20242571

(2)研究分担者
鵜木　祐史 (UNOKI MASASHI)
北陸先端科学技術大学院大学・情報科学研究科・准教授
研究者番号：00343187

宮内　良太 (MIYAUCHI RYOTA)
北陸先端科学技術大学院大学・情報科学研究科・助教
研究者番号：30455852

李　軍鋒 (LI JUNFENG)
中国科学院・声学研究所・教授
研究者番号：50431466
2010 年 7 月 31 日まで

(3)連携研究者
なし