

# 顔表情・音声・会話文のマルチモーダルな感情推定とアクトロイドへの応用

前田浩貴<sup>†</sup>任福継<sup>‡</sup>西出俊<sup>‡</sup><sup>†</sup> 徳島大学 工学部<sup>‡</sup> 徳島大学 大学院ソシオテクノサイエンス研究部

## 1. はじめに

近年、人間と似た動作が可能なロボット技術の発展により、人間の感性をコンピュータ上で処理する研究が盛んに行われるようになってきた。コンピュータで感情を扱う研究例としては、人間の感情を機械的に認識する研究 [1]、顔表情画像からの感情推定の研究 [2]、形態素単位で文の持つ感情表現を推定する研究 [3] などが挙げられる。これらの研究では単一のモダリティを対象としており、複数のモダリティを統合した感情認識研究例は少ない。本研究では画像、音声、言語情報のモダリティを統合することで感情認識を行うことを目的としており、本稿ではこれまで構築したシステムについて報告する。

## 2. 提案手法

提案手法全体の概略図を図 1 に示す。目標とする全体のシステムに対し、これまで構築した部分を赤枠で示す。本研究では画像、音声、言語情報の三つのモダリティを統合することで感情推定を行う。画像はカメラから取得した顔表情画像を、音声はマイク入力された音声データを、言語情報はマイク入力された会話文を入力とする。各モダリティの処理結果をモーダル値として出力し、それらを統合することが本システムの目標である。各モダリティの処理方法について次節で説明する。

### 2.1 顔表情に基づく感情推定手法

顔表情に基づく感情推定は Ekman が開発した Facial Action Coding System (FACS) [4] に基づき、Duo らが開発した目、鼻、口、眉の動きを表現する 38 個の特徴点に基づく表情認識手法を用いる [5]。本手法では Web カメラから取得した画像に対して OpenCV による顔検出を行い、特徴点の座標情報を取得する。特徴点の座標情報をもとに特徴量を計算し、LIBSVM を用いて感情推定を行う。また、現在は未実装だが、推定した感情結果に対し、認知科学に基づいた重みづけを行ったものをモーダル値とする予定である。

### 2.2 音声に基づく感情推定手法

音声に基づく感情推定は基本周波数に基づいて行う。マイク入力された音声データに対し、自己相関関数を用いて基本周波数を推定する。また、現在は未実装だが、基本周波数値の範囲による感情推定を行い、推定した感情に認知科学に基づいた重みづけを行ったものをモーダル値とする予定である。

### 2.3 会話文に基づく感情推定手法

会話文に基づく感情推定手法は松本らの手法を用いた [3]。マイクから入力される音声信号に対し、音声認識することで会話文章を取得する。得られた会話文章に対し、MeCab [6] を用いて形態素解析を行う。形態素解析によって得られた各要素に対して、感情語辞書を参照し感情属

性を付与する。これら全ての要素の感情属性の値を足し合わせた合計に認知科学に基づいた重みづけを行ったものをモーダル値とする。

## 3. システムの実験結果

構築したシステムの評価を行うため、顔表情の特徴点抽出と会話文に基づく感情推定の結果を示す。音声について、自己相関分析を用いた基本周波数推定は汎用的な手法であるので評価は紙面の都合上、省略する。

### 3.1 顔表情の特徴点抽出結果

構築したシステムで行った顔表情の特徴点抽出の結果を図 2 に示す。本手法では目、鼻、口、眉の動きを表現する 38 個の特徴点を抽出する。図 2 より、各特徴点はそれぞれの顔の部位を表現していることが分かる。

### 3.2 会話文に基づく感情推定

本実験で用いた会話文 (表 1) に基づく感情推定の結果の一例を表 2 に示す。表 1 は複数の文から成る会話文を感情推定し、そのモーダル値を出力したものである。

表 2 の各モーダル値は文が持つ感情の値を表しており、最も大きいものがその文を最もよく表す感情である。本システムは全体的に“Hope”の値が大きくなる傾向があるため、“Hope”を除いた各文のモーダル値を見ると、文 1、文 2、文 t、それぞれに対して、“Sorrow”、“Love”と“Suprise”、“Hate”がそれぞれ大きな値を持っていることが分かる。これらの感情は各文の感情を表現するものであり、手法の有効性が確認できる。

表 1: 実験で用いた会話文

文 1	鍵をなくしてしまうなんて、困ったなあ。
文 2	あの鈴木君が結婚すると知らされたときはすごく驚いたよ。
文 t	僕のプリンを勝手に食べるなんて、信じられないよ。

表 2: 各会話文のモーダル値

	文 1	文 2	...	文 t
Hate	-0.146	-0.697		2.562
Pleasure	-0.187	0.708		0.068
Hope	1.668	-0.754		1.689
Love	0.278	1.874		-1.410
Sorrow	0.839	-0.877		0.812
Anxiety	0.169	0.914		-0.487
Surprise	-0.118	1.456		0.528

## 4. 今後の課題

我々はこれまで、複数のモダリティを統合した感情推定手法を開発することを目指し、これまでは要素技術の開発の統合の準備を行ってきた。今後の課題としては以下が挙げられる。

Multi-modal Emotion Estimation Using Facial Expression, Voice, and Conversational Sentence and its Application to Actroid Koki Maeda (Tokushima Univ.), Fuji Ren (Tokushima Univ.), and Shun Nishide (Tokushima Univ.)

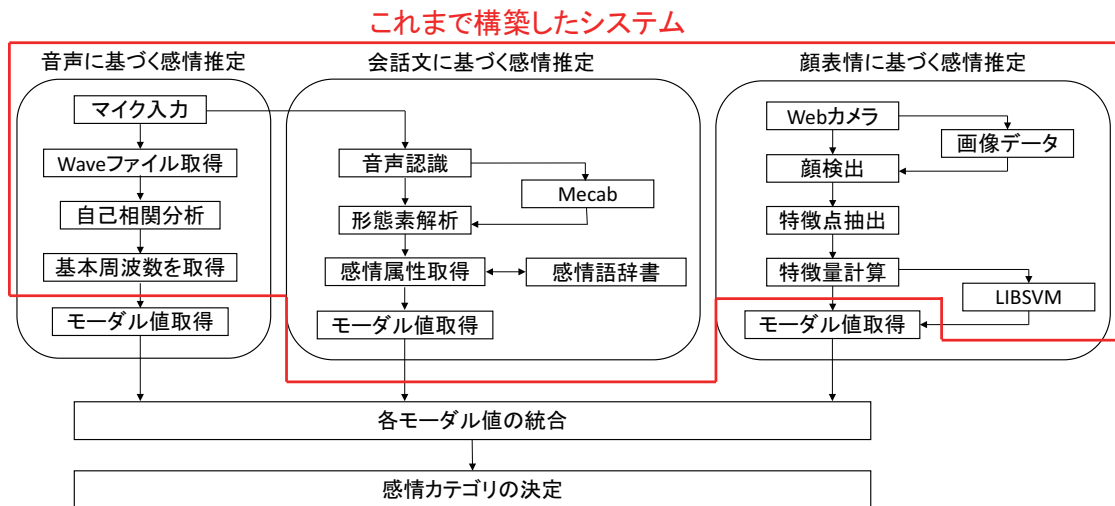


図 1: 本システムの構成図

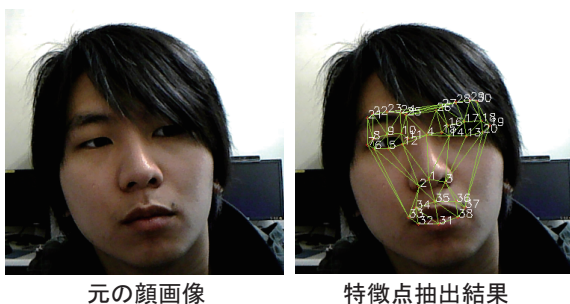


図 2: 顔の特徴点抽出結果

アクトロイド自身に搭載されたマイク・カメラを用い、複数のモダリティを統合して対話者の感情を推定できるシステムを構築する。今後は顔表情、音声、会話文の三つの情報を用いて感情推定した結果をアクトロイドで出力するインターフェイスの作成を行う予定である。例えば、感情推定の結果「喜び」の感情カテゴリが出力されると、アクトロイドの表情で「喜び」を表現する。これにより、アクトロイドも人間と同様に感情表現をすることが可能になると考えられる。

## 5. まとめ

本稿では、顔表情、音声、会話文の三つの情報を用いて感情推定を行うシステムを提案した。本システムでは、顔表情、音声、会話文の各々に基づく感情推定を別々に行い、出力されるモーダル値を統合することで感情推定を行う。本稿ではこれまで構築した要素技術について報告し、今後の方針としてモーダル値の統合とロボットへの応用について述べた。今後は各課題に取り組み、システム全体の完成を目指す予定である。

### 謝辞

本研究は JSPS 科研費 15H017012 の支援を受けた。

### 参考文献

- [1] 光吉 俊二, 任 福継: “人間の感情を測定する,” 電気学会誌, Vol. 125, No. 10, pp.641-644, 2005.
- [2] 谷 卓哉, 長谷川 浩司, 坂本 博康, 坂田 年男, 藤田 浩, 福島 重廣: “正準相関分析と注視特性による顔表情画像からの感情の測定法,” 知能と情報: 日本知能情報フアジイ学会誌, Vol. 22, No. 1, pp. 52-64, 2010.
- [3] 松本 和幸, 三品 賢一, 任 福継, 黒岩 真吾: “感情生起事象文型パターンに基づいた会話文からの感情推定手法,” 自然言語処理, Vol. 14, No. 3, pp. 239-272, 2007.
- [4] P. Ekman and W. V. Friesen, “Facial Action Coding System: Investigator’s Guide,” Consulting Psychologists Press, Palo Alto, CA, 1978.
- [5] D. Feng, F. Ren, and S. Nishide, “Automatic Facial Feature Points Extraction and Expression Recognition Based on Video Database,” Proceedings of International Conference of Emerging Applications of Information Technology, WIT529, 2015.
- [6] MeCab (和布蕪): <http://taku910.github.io/mecab/>
- [7] A. Mehrabian, “Communication without words,” Psychology Today, Vol. 2, No. 9, pp. 52-55, 1968.

1. モーダル値の算出とその統合による感情カテゴリの決定
2. ロボットへの応用

#### 4.1 モーダル値の統合による感情カテゴリの決定

複数のモダリティを統合した感情推定を実現する上で、各モダリティの出力するモーダル値の重み付けが大きな課題である。一般的に人間の感情について、顔表情は 55%、音声情報は 38%、言語情報は 7%の重みがあるといわれており [7]、本研究でもこれをもとにシステムを構築する。

モーダル値の統合においては二つの方法が考えられる。

1. 時系列に取得される三つのモーダル値を数ステップ分、感情カテゴリごとに足し合わせ、合計が最も大きい感情カテゴリを出力する。
2. 各モーダル値に対して数ステップ分の値を感情カテゴリごとに足し合わせ、各モーダル値ごとに全感情カテゴリの値の合計に対する各感情カテゴリの割合を求める。最後の各モーダル値の感情カテゴリの値の割合を用いて尤度を計算し、最も尤度が高い感情カテゴリを出力する。

今後はこれらの二つの方法を実験的に検証していき、システムに組み込む予定である。

#### 4.2 ロボットへの応用

本研究の応用として、人間に近い容姿を持つアクトロイドへの実装を予定している。人間との対話において、