

平成 22 年度 (2010 年度)

# 卒業論文要旨集

Abstracts of the 2010 Graduation Theses

岩手県立大学ソフトウェア情報学部

Faculty of Software and Information Science  
Iwate Prefectural University

# 顔画像からのエゴグラム推定手法の構築

An Ego-gram Estimation Method from Facial Image

インテリジェントソフトウェアシステム学講座 0312005043 菊地昭文  
指導教員：羽倉淳 樽松理樹 藤田ハミド

## 1. はじめに

本研究では、対話型システムにおいて、交流分析<sup>1)</sup>に基づくユーザとの相互作用を実現するために、エゴグラム推定手法の構築を行う。エゴグラムの推定手法には、東大式エゴグラム<sup>2)</sup>がある。しかし、その推定のためには60問程度の設問に答える必要があり、対話システム利用時にこの方法でエゴグラムを推定することは現実的ではない。一方で、人間は他者の顔を見るとその人の性格をある程度言い当てることのできる場合がある。実際、人相学や相貌心理学等では同様のことを体系付ける試みがなされている<sup>3)</sup>。そこで本研究では、顔画像から抽出する顔特徴とそれらの一般的な印象を関連付けることで、エゴグラムを推定する手法を提案する。

## 2. 顔特徴に基づくエゴグラム推定

### 2.1 エゴグラム推定の目的と意義

対話型システムにおいて、個々のユーザに合わせた応答を行うには、ユーザ情報の取得が重要となる。ここでは、ユーザ情報として性格を推定することを、例えば、内向的な人間に対してはシステム側が積極的に働きかけを行うといったユーザの性格に合わせた行動選択が実現するための第一歩とする。

本研究の性格推定では、エゴグラムを利用する。エゴグラムは、人間の性格特性を要素として5つの自我状態（CP：批評的な親、NP：養育的な親、A：大人らしさ、FC：自由な子供、AC：顧慮した子供）により表し、要素間の強弱によって人間の性格を表現する事ができる<sup>4)</sup>。

### 2.2 エゴグラム推定のアプローチ

個人のエゴグラムを取得するためには、東大式エゴグラムで行われるように、多くの設問への回答が必要である。対話型システムにおいて、ユーザにこれらの設問への回答を求めることは、システム利用に対する意欲を減衰させる恐れがある。すなわち、設問への回答を省略し、エゴグラムを取得するための手法が必要となる。

そこで本研究では、ユーザのエゴグラムを推定するために、5つの自我状態と、顔から受ける事が可能な印象、そしてその印象を想起させる顔特徴を関連付けることによって、エゴグラムの推定を行う。ここでは、顔特徴から推定されたエゴグラムが東大式エゴグラムにより近づく事を目標とする。システムの全体図を図1に示す。

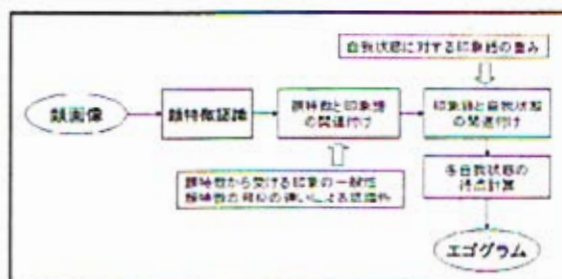


図1—システム全体図

## 3. エゴグラム推定手法

### 3.1 顔特徴認識

システムはユーザの正面顔の情報が取得できる状況を想定し、正面顔の特徴認識によって得られた顔特徴を入力情報とする。顔特徴は、他者に印象を与える要因となる部分特徴とし、顔特徴を表すものとして、特徴語を定義する。特徴語は、顔特徴と印象との関連について記述あった3冊の文献<sup>5)6)7)</sup>を参考に、22個の特徴語を定義した。定義した特徴語は、表1に示されるものである。顔特徴は、得られた顔画像に対して指定した特徴点の抽出を行い、各特徴点座標の相互距離を基に定量的に表現する。例えば、眉間の広さは、眉に設定した特徴点のうち最内部の2点間の距離を測定する。この測定値について、事前に多数人から測定した同様の距離のデータから最大値と最小値を利用して設定した「広い—狭い」の特徴範囲と比較し、特徴を決定する。特徴点の抽出には FSE (Face Sensing Engine)<sup>8)</sup>及びOpenCV<sup>9)</sup>による画像処理を用いる。抽出する特徴点は、顔の広さを認識するために顔肌色領域における最上部に一点、眉間の距離および曲がり、眉と目の距離、目の大きさ、丸み、目角、両目間距離を認識するために、左右の眉に6点と左右の目に8点、口の大きさ、口角、唇の厚みを認識するために口に6点、顔型を認識するために顔輪郭上に11点とした。

### 3.2 顔特徴と印象語の関連付け

顔特徴と関連がある印象として印象語を定義する。上述した22個の特徴語から得られる印象として17個の印象語を定義した。印象語は表2において示されるものである。ここで定義した印象語と特徴語は、文献によって関連付けられているものであり、本研究でもその関係性を利用する。17個の印象語と、特徴語との関係性を表1に示す。

### 3.3 印象語と自我状態の関連付け

印象語と自我状態の関係性を示すために、エゴグラム、及び交流分析について書かれた文献<sup>1)2)3)10)11)</sup>において自我状



態の説明から特性の調査を行い、印象語の意味合いとを比較する事により、関連性を考慮し定義した。各印象語と自我状態の関係性を表2に示す。

### 3.4 エゴグラム作成のための得点設定

東大式エゴグラムと同様に、本研究でも自我状態を20点満点で得点を算出する<sup>19)</sup>。これは、認識された各顔特徴について関連する印象語に得点付けを行い、各自我状態に計算していくことで求める。基本得点として1つの印象語に対し、認識された場合に2点、認識されなかった場合は1点、対立する印象語が認識された場合は0点として、関連する自我状態に得点を与える。この時、各特徴語、及び各印象語がすべて等価なものであるとは考えにくい。そこで、それぞれの影響力には差異があるものと仮定し、以下を重みとして差異を反映した得点設定を行う。

- ・ 各特徴語に対し、文献の記述による重複から一般性の違いによる重みを考慮し、一般性の強い特徴語から得られる得点が強くなるように設定
- ・ 各特徴語に対し、顔部位の認識性の違いが、印象の与え方に影響すると考え重みを考慮し、認識性の強い特徴語から得られる得点が強くなるように設定
- ・ 各印象語に対し、自我状態の特性の説明の記述から、自我状態に対する重要性の違いによる重みを考慮し、重要性の強い印象語の得点が強くなるように設定

最後に、各自我状態に計算された得点を最大で20点になるように正規化し、エゴグラムを作成する。

## 4. 実験

本手法を用いたエゴグラムの推定を65名に対して行った。本手法を用いたエゴグラムの認識性を確認するため、同被験者に東大式エゴグラムのための60問からなる質問に回答してもらい、本手法から成るエゴグラムと東大式エゴグラムの比較を行った。比較は、得られた得点を1~5の5段階にコード化して行う。本手法で求めた得点については最大得点である20点を5分割してコード化した。本手法によるエゴグラムと東大式エゴグラムの結果を比較し、各自我状態でコード値が一致した割合を表3に示す。

## 5. 考察

今回の実験では、顔特徴から推定したエゴグラムと東大式エゴグラムの比較をした際、5段階に分けたコード値が一致ものは、全体の33.5%となった。このうち、CPとFCは4割以上の一致が見られ、現手法においてある程度の評価ができていたものと考えられる。対して、ACに関しては15.4%と低い値となった。これはACの自我状態に関わる定義がうまくいっていなかった可能性が示唆され、顔特徴との関連性と、得点の重み付けについて見直しが必要であると考えられる。

## 6. おわりに

本研究では、交流分析に基づくユーザとの相互作用の実現を目的として、顔特徴に着目したエゴグラム推定手法を提案した。現状ではCPとFCに関して東大式エゴグラムと同様の

表1：特徴語と印象語の関係表

特徴語	特徴から受ける印象				
	親しみやすい	寛容的	親切	外向的	情緒的
太顔	親しみやすい	寛容的	親切	外向的	情緒的
端正な顔	親しみやすい	情緒的	知的		
目が広い	親しみやすい	寛容的	外向的	情緒的	知的
目が細い	寛容的				
目が丸い	寛容的				
目が細い	内向的	情緒的			
目が大きい	親しみやすい	寛容的	外向的	情緒的	
目が細い	内向的	情緒的			
目が細い	親しみやすい	寛容的	親切	外向的	情緒的
目が広い	親しみやすい	寛容的			
目が細い	親しみやすい	寛容的			
目が細い	寛容的	知的			
目が大きい	寛容的	外向的	情緒的		
目が細い	内向的				
目が細い	寛容的	親切			
目が上向き	寛容的				
目が下向き	内向的				

表2：自我状態と印象語の関係表

自我状態	関連する印象語				
	親しみやすい	寛容的	情緒的	外向的	知的
CP	親しみやすい	寛容的	情緒的	外向的	知的
NP	親しみやすい	寛容的	寛容的	知的	
A	知的				
FC	外向的	情緒的	情緒的	情緒的	
AC	外向的	外向的	情緒的	情緒的	

表3：コード値が一致した割合 (%)

全体	CP	NP	A	FC	AC
33.5	46.2	32.3	32.3	41.5	15.4

コード値を得ることが40%以上で可能であるが、ACに関しては低い割合となった。今後は一致度の低かったACに関わる手法の見直しを中心に改善を行う。

## 7. 謝辞

本研究にあたり、顔画像の提供および東大式エゴグラムの調査に快く協力して頂いた皆様へ心から感謝します。

## 参考文献

- 1) 桂枝作, 杉田峰康, 白井幸子: 交流分析入門, チーム医療 (1984)
- 2) 杉田峰康: 交流分析, 日本文化科学社 (1985)
- 3) 東京大学医学部心療内科: エゴグラム・パターン-TEG 東大式エゴグラム第二版による性格分析, 金子書房 (1995)
- 4) L・コルマン, 須賀哲夫 訳, 福田忠郎 訳: 相貌心理学序説, 北大路書房 (2005)
- 5) Robert L. Whiteside: Face Language, LIFETIME BOOKS (1974)
- 6) Leslie A. Zebrowitz, 羽田節子 訳, 中尾ゆかり 訳: 顔を読む, 大修館書店 (1999)
- 7) 渋谷昌三: 第一印象で相手の心をつかむ本, PHP 文庫 (2006)
- 8) 沖電気: Face Sensing Engine, 沖電気 (2008)
- 9) <http://openconv.jp/> (2011/1/13 アクセス)
- 10) ジョン・M・デュセイ, 新里里春 訳: エゴグラム, 創元社 (1980)
- 11) イアン・スチュアート, ヴァン・ジョインズ, 深沢道子 監訳: TA TODAY, 実務教育出版 (1991)



# はずれ値除去による音声からの感情推定向上の試み

## An Idea of Improvement of Emotion Estimation from Speech by Removing Outliers

インテリジェントソフトウェアシステム学講座 0312007100 竹松 佑士

指導教員：藤田ハミド 樽松 理樹 羽倉 淳

### 1. はじめに

私は音声からの感情推定に関する研究に取り組んでいる。感情推定の研究としては、表情から推定を試みる研究が主流<sup>1)</sup>となっている。これに対し、音声に着目した理由は、電話など相手の表情やしぐさを得ることができない場合でも利用可能であり、その活用範囲が広いと考えたためである。

音声からの感情推定方法としては、音声から抽出した音声特徴と感情との関連ルールを構築し、これを用いて別の音声から感情を推定する手法<sup>2)</sup>が主流である。しかし、過去の研究では、無声音、ノイズ、振幅の急激な変化などによって生じる本来のデータと比べてかけ離れた値であるはずれ値<sup>3)</sup>を除去していない。私は、はずれ値によって誤った音声特徴を抽出し、それが感情の推定精度を低下させる要因になっていると考えた。本研究では、このはずれ値を除去することにより、感情推定の精度向上を試みる。なお、本研究では、はずれ値を、“前後の値と比べて、統計的に異常に大きいあるいは小さい値”と定義した。本研究でははずれ値の定義は、前述の定義<sup>3)</sup>とは異なる。これは、音声学におけるはずれ値は音声データ列の一定の流れに沿っていないデータであると考えたからである。

### 2. はずれ値除去による音声からの感情推定の向上

本提案手法の概要を図1に示す。本手法は、感情とはずれ値を除去した音声特徴とを対応付けてルールを構築する学習部と、提案手法の有効性を評価する推定部から成り立つ。

#### 2.1 発話データ

本研究で用いる感情は、ポール・エクマンらの述べている怒り・悲しみ・驚き・喜び・恐れ・嫌悪の6感情<sup>4)</sup>に平常を加えた計7感情である。本手法では「話し手」ではなく「聞き手」の感情推定を行う。そのため発話データとしては、ある人の発話に対し、「聞き手」が前述の感情の一つを付与したものをを用いる。具体的には、N人の「話し手」が感情を込めて発話したM個の発話データに1人の「聞き手」が感情を付与する。

#### 2.2 学習部

##### (1) 全体の流れ

M個の発話データに対して以下の処理を行う。

各発話データから単位時間内での振幅と基本周波数を抽出する。これにより1発話から振幅データ列と基本周波数データ列が求まる。次に2つのデータ列のはずれ値を除去する。はずれ値を除去した振幅データ列と基本周波数データ列から、平均値、中央値、標準偏差、範囲(最大値-最小値)を算出し、これを音声特徴として用いる。

最後に上記の工程で求めた全発話データの音声特徴と感情との対応付けを行い、得られた結果を感情推定のルールとする。

##### (2) はずれ値の判定と除去

はずれ値の判定と除去は1発話から一定数の区間単位に順次区切り、その区間毎に行う。

はずれ値の判定方法を次に示す。初めに切り出された各区間の中心のデータをはずれ値の候補とする。はずれ値かの判定については、そのデータが区間内の他のデータと比べてかけ離れた値であるかを検定する統計学的手法を用いる。かけ離れた値と検定されたら、その値をはずれ値と判定する。

具体的に用いる検定手法はスミルノフ・グラブス検定<sup>5)</sup>である。スミルノフ・グラブス棄却検定とは、データの中に、1つだけ他のデータとかけ離れている値を棄却するかどうかを検定する方法である。本検定手法における帰無仮説は、“他のデータとかけ離れた値ははずれ値ではない。”であり、対立仮説は、“他のデータとかけ離れた値ははずれ値である。”となる。具体的には、式(1)で示す検定統計量Tを求め、この値が棄却域内であれば、帰無仮説を棄却し、かけ離れた値をはずれ値とみなして棄却する。

$$T = |X - \mu| / s \quad (1)$$

式(1)において、Xは検定対象の値、 $\mu$ とsはそれぞれの区間内の平均値と標準偏差を示す。

はずれ値と判定された値は、MovingMedian<sup>6)</sup>を用いて除去する。MovingMedianとは、対象データを区間内のデータの中央値で置き換える方法である。

##### (3) ルール構築方法

感情推定ルールを構築するために決定木を用いて、感情と音声特徴との対応づけを行う。決定木の生成手法はCART法<sup>7)</sup>を用い、説明変数としては音声特徴の値を、目的変数としては感情名を用いる。



### 3. 評価実験(推定部)

提案手法の有効性を評価実験により検証した。ルール構築時と同じように1感情を付加した推定用のK個の発話データから音声特徴を抽出する。音声特徴を基にルールで推定を行い、推定用データに付加された感情と同じ感情の予測値を精度とする。

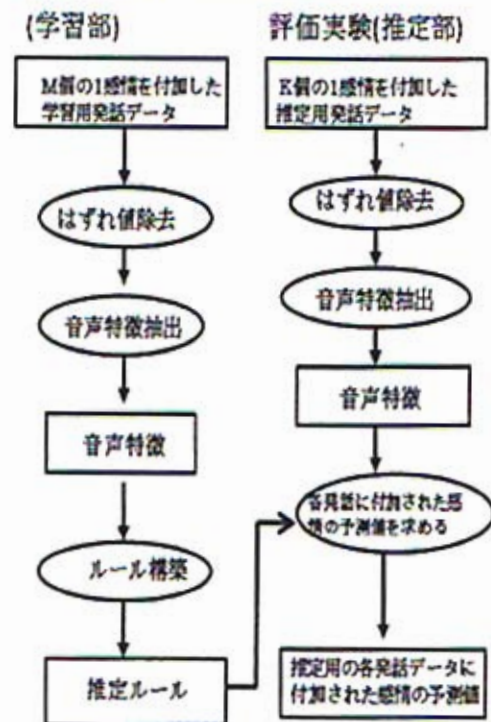


図1.はずれ値除去による音声からの感情推定

#### 3.1 発話データ

発話データとして、20代女性1名(話し手)が2.1で示した7感情に対し、各感情10発話、計70発話を用意したものに対して、20代男性1名(聞き手)が感情の再ラベル付けを行ったものを用いる。各発話データのサンプリング周波数は64kHz、量子化は16bit、フレーム長は10msecとした。

#### 3.2 実験方法

はじめに70発話のデータを11個のデータセットに分け、各データセットを検証データと訓練データに分ける。訓練データから構築した決定木決定木を用いて検証データで推定を行い、検証データに付与された感情の予測値を求める。また、はずれ値を除去した場合としない場合の予測値を比較することによって、はずれ値除去の有効性を評価する。

#### 3.3 実験結果

表1は、データ列全体の個数に対するデータ列におけるはずれ値比率を、表2は、精度の平均値と、はずれ値を除去した後に精度が向上あるいは低下した検証データの個数を示す。

表1: はずれ値比率

振幅データ列	基本周波数データ列
4.5%	0.5%

表2: はずれ値の有無による実験結果の比較

音声特徴	はずれ値除去		予測値の変化	
	する	しない	向上	低下
振幅と基本周波数	0.333	0.268	12	22
振幅のみ	0.300	0.218	16	34
基本周波数のみ	0.178	0.205	22	11

#### 4. 考察

表2に示すように振幅と基本周波数、振幅のみの実験では精度が低下した。理由として、表1で示すように振幅では、はずれ値ではないデータまで除去された可能性がある。一方で、基本周波数のみの実験では若干精度が向上した。理由として適切なはずれ値除去に成功したからだと考えられる。しかし、表2で示すように予測値の低下する検証データも多数存在している。

#### 5. 終わりに

本研究では、はずれ値除去による感情認識の精度向上を試みた。有用なはずれ値除去の手法としてはまだ不十分であり、はずれ値の判定など、より適した手法の考案が必要である。

また、今回1人分の音声でしか実験を行わなかったが、より多くの人の音声を集めて本提案手法の有効性を評価する必要がある。

#### 参考文献

- 1) 國弘威, 吉川榮和: "動的表情認識による感情推定手法に関する基礎研究", 京都大学大学院エネルギー科学研究科エネルギー社会・環境科学専攻(1999)
- 2) Pierre-Yves Oudeyer: "The production and recognition of emotions in speech features and algorithms", International Journal of Human-Computer studies, vol. 59(1-2), pp. 157-183(2003)
- 3) 平坂文男: "実験音声学のための音声分析", 丸善株式会社(2009)
- 4) P. Ekman, W. V. Friesen, 工藤力訳: "表情分析入門-表情に隠された意味をさぐる", 誠信書房(1987)
- 5) 宗森信, 佐藤寿邦: "データのとり方とまとめ方", 共立出版株式会社(2009)
- 6) 金明哲: "Rによるデータサイエンス データ解析の基礎から最新手法までの詳細", 森北出版(2007)



# 音楽理論に基づく音声合成による感情表出手法の研究

Research on Emotional Speech Synthesis Using Musical Theories

インテリジェントソフトウェアシステム学講座 0312007115 千葉央輝

指導教員：藤田ハミド 樽松理樹 羽倉淳

## 1. はじめに

音声は、大きく言語情報と非言語情報から構成されている。従来の音声合成は、言語情報を中心に行われてきたが、音声合成の利用範囲の拡大とともに非言語情報にも着目されるようになった。結果、感情表現を試みる音声合成手法の研究が活発化している<sup>1)</sup>。しかし、評価実験などにおける人の認識精度は、実験参加者や発話内容などの実験条件に依存する点が多い。よりの確かな合成音声による感情表出を行うためには、更なる改善が必要である。

一方、音楽と感情の関係に関しては、いくつかの知見が得られている。それらを利用したシステム開発<sup>2)</sup>も進められている。

以上のような背景から、本稿では、音楽と感情の関係を音声合成に利用することを考え、音楽理論に基づいた新たな音声合成手法を提案する。本手法は従来手法と異なり、理論を活用することから、学習データを必要とする従来手法よりも開発効率が向上し、より実用的な手法となることが期待できる。

## 2. 音楽理論に基づく音声合成手法

本章では、音楽理論に基づく音声合成手法について説明する。

### 2. 1. 入力

入力としては、表出を試みる感情を付与した文章を与える。ここで表出を試みる感情としては、Ekmanらが提唱した怒り、喜び、悲しみ、驚き、恐れ、嫌悪の基本6感情<sup>3)</sup>を用いる。

### 2. 2. 基本旋律

始めに、読み上げる文に対し、日本語の基本的なアクセントやイントネーション<sup>4)</sup>に基づき、各音の高さを付与する。これを基本旋律と呼ぶ。本手法は、この基本旋律を音楽理論に基づき変化させることで感情表出を試みる。

### 2. 3. パラメータ設定

EMOTIONAL EXPRESSION IN MUSIC<sup>5)</sup>を参考に、基本旋律に対し、①テンポ、②ボリューム、③音高、④音程、⑤音程差、⑥コード・スケール、の6つのパラメータを、この順番で変更していく。以下、各パラメータについて説明する。

①テンポ：“速く”と“遅く”の2通りがある。前者は読み上げ速度を10%増にし、後者は10%減する。

②ボリューム：“変化あり”と“変化なし”の2通りがある。前者は一定の周期でボリュームが大きくなったり小さくなったりする変化を加える。後者は、音節一つ一つが持つごく小さなボリュームの変化のみとし、特に変化は加えない。

③音高：“高い”と“低い”と“変化なし”の3通りがあり、“高い”は文全体の音高を半音上げ、“低い”は半音下げる。

④音程：“上昇”と“下降”の2通りがある。1文節において音節数÷2(切捨て)の数だけの音節を、前者は1全音上げ、後者は1全音下げる。

⑤音程差：“大きい変化”と“小さい変化”の2通りがある。前者は次の音節との音程差が1音以下なら、1音より大きく、後者は次の音節との音程差が3全音以上なら、3全音より小さくなるように変更する。

⑥コード・スケール：Marc-Antoine Charpentierが1642年に”Rules of Composition”<sup>6)</sup>として提案したコード・スケールと感情との関係を基に設定する。

⑤までで変更した基本旋律の音のうち、⑥で設定したコード・スケールに含まれない音については、最近傍の高さの音に変更する。最近傍の高さの音が複数あった場合、高い音を選択する。

以上の考えに基づき、各感情に対して付与したパラメータ設定を表1に示す。なお、コード・スケールが1感情に対し複数あるため、感情毎にパラメータの組み合わせ数は異なっている。

表1:意図する感情と各パラメータの組み合わせ

No	表出感情	テンポ	ボリューム	音高	音程	音程差	コード・スケール
1	怒り	速く	変化なし	高い	上昇	小	F
E							
3	嫌悪	遅く	変化なし	変化なし	変化なし	小	Fm
4	恐れ	速く	変化あり	高い	上昇	小	B♭m
5							Em
6	喜び	速く	変化なし	高い	上昇	大	D
7							G
8							A
9	悲しみ	遅く	変化なし	低い	下降	小	Cm
10							Bm
11							Fm
12	驚き	遅い	変化なし	高い	上昇	大	B♭m
13							Gm
14							G



### 3. 評価実験

#### 3. 1. 概要

本手法の妥当性を評価するために次に示す聴取実験を行った。実験参加者は、基本旋律の合成音声とパラメータ設定後を加えた合成音声を聴き、各感情に対し、「弱くなった」「少し弱くなった」「同じ」「少し強くなった」「強くなった」の5段階で回答する。これらの回答結果を比較することで、表出される感情に対する各パラメータ設定の妥当性を評価する。

今回の実験には、10代から40代の男15名、女26名、合計41名が参加した。合成音声としては、内容から感情が推定しにくいニュース記事を、クリプトン社の歌声合成ソフト“巡音ルカ”<sup>7)</sup>を用いて作成した。

#### 3. 2. 実験結果と考察

実験結果の概要を表2に示す。ここでNoは表1のパターンNo、表出感情は音楽理論を基にしたとき表出される感情を示す。各感情の列の値は、それぞれのパターンに対し、「弱くなった」を-2、「少し弱くなった」を-1、「同じ」を0、「少し強くなった」を1、「強くなった」を2とし、全回答の合計を求めたものである。値が大きいくほどその感情を強く感じた人が多かったと考えられる。▲が最大値、▼が最小値、黒抜きが表出を意図した感情である。

表2: 実験結果

No	表出感情	喜び	驚き	嫌悪	恐れ	悲しみ	怒り
1	怒り	37	▲38	-15	-14	▼-18	-12
2	怒り	23	▲40	-9	-6	▼-15	6
3	嫌悪	-1	1	▼-4	-3	▲2	▼-8
4	恐れ	12	▲38	4	9	-5	10
5	恐れ	22	▲42	-4	7	-12	6
6	喜び	▲56	44	▼-25	-13	-19	-7
7	喜び	44	▲47	-17	-4	-17	-1
8	喜び	33	▲42	-18	-22	▼-25	1
9	悲しみ	▼-40	-37	41	30	▲42	18
10	悲しみ	-30	-23	20	19	▲39	14
11	悲しみ	▼-46	-36	38	30	▲55	10
12	驚き	35	▲39	-20	-8	▼-21	-4
13	驚き	48	▲49	-15	-13	▼-20	-10
14	驚き	44	▲47	▼-17	-4	▼-17	-1

「喜び」と「驚き」は類似した結果となった。強くなったと感じた上位は#6、#13、#7、#14、#2などのパターンとなっているが、同時に悲しみ以外のパターンは割合にほとんど高い結果となっている。悲しみの全パターンは、強くなったと感じた回答がなく、弱くなったと感じた回答が多い結果となった。これは、喜びと驚きを意図した設定では、表出される感情に大きな違いがなく、また、悲しみを感ずることはなくなると思われる。

「嫌悪」と「悲しみ」も類似した結果となった。強くなったと感じた上位は、どちらも悲しみのパターンとなっている。これは、嫌悪と悲しみを意図した設

定では、表出される感情に大きな違いがなく、どちらも悲しみが表出されると言える。

「恐れ」に関する評価回答は、#9、#11のパターンで20人以上が強くなったと感じた回答した。

「怒り」に関する評価回答も、強くなったと感じた上位に悲しみのパターンとなっている。

この実験結果から、音楽理論に基づく設定は、音声合成の場合には当てはまらない場合があることがわかった。この点を考慮し、再設定したパラメータを表3に示す。ここで黒抜きになっている部分の変更された点である。

表3: 実験結果に基づく感情と対応する各パラメータ

表出感情	テンポ	ボリューム	音高	音程	音程差	コードスケール
怒り	ゆっくり	変化あり	低い	下降	小	Fm, Em
嫌悪	ゆっくり	変化なし	低い	下降	小	Fm
恐れ	ゆっくり	変化あり	低い	下降	小	B♭m, Em
喜び	速く	変化あり	高い	上昇	大	D, G, A
悲しみ	ゆっくり	変化なし	高い	変化なし	小	Cm, Bm, Fm
驚き	速く	変化あり	高い	上昇	大	G

#### 4. おわりに

本研究では、合成音声による感情表出を行うために音楽理論に基づいた新たな手法を提案した。今後の課題としては、表3で示したパラメータ設定の有用性の検証、さらなる精度向上のためのパラメータの修正と追加を行うことである。また、現在合成音声生成は、手動で行っている。手法を幅広く利用するためには、本手法の自動化も必要となる。

#### 参考文献

- 1) M. Schröder: "Emotional Speech Synthesis: A Review", Proc. Eurospeech 2001, Vol. 1, pp. 561-564, 2001.
- 2) 三浦: "「怒り、喜び、悲しみ…」MIDI 楽曲に感情表現を付加するシステム「MOR2ART」", DTM MAGAZINE, Vol. 178, pp. 98-99, 2009.
- 3) P. Ekman, W.V. Friesen 著, 工藤訳: "表情分析入門", 誠信書房, 1987.
- 4) 松崎, 河野: "よくわかる音声", アルク, 1998.
- 5) A. Gabrielsson and P.N. Juslin, "EMOTIONAL EXPRESSION IN MUSIC", HANDBOOK OF AFFECTIVE SCIENCE, Oxford University Press, pp. 503-534, 2009
- 6) H. Fujita, J. Hakura and M. Kurematu: "Cognitive Modeling in Software and Relation to Human Emotional Reasoning", New Trends in Software Methodologies, tools and Techniques, pp. 145-165, 2007
- 7) クリプトン社: "巡音ルカ", <http://www.crypton.co.jp/mp/pages/prod/vocaloid/cv03.jsp>, (2011/01/07 アクセス)



# 訓練データの細分化を用いた音声からの感情推定 An Estimation of Emotion in Speech Using Subdivided Training Data

インテリジェントソフトウェアシステム学講座 0312009301 天沼 沙織

指導教員：藤田ハミド 樽松理樹 羽倉淳

## 1. 背景

私は、音声からの感情推定の研究に取り組んでいる。この分野の代表的な手法は次のようなものである。はじめに、感情を含んだサンプル音声から音声特徴を抽出し、表現を試みた感情と音声特徴との間の関係を抽出する。この関係を音声からの感情推定規則として利用し、新たに入力された音声の音声特徴から感情推定を試みる。しかし、この手法では推定精度が低いという問題が残っている。この要因の一つとして、同じ感情、同じ人であっても語彙や状況などによって複数の表現が存在しているのを考慮していないことがあげられる。この問題を解決するために、従来手法に訓練データの細分化を加えた手法を提案する。本手法により前述した問題の解決や推定精度の向上が期待できる。

## 2. 提案手法

本手法の概要について図1に示す。本手法は、①感情を含んだサンプル音声から音声特徴の抽出、②音声特徴による階層的クラスタ分析、③クラスタ分割指標によるクラスタ分割の最適化、④表現を試みた感情と音声特徴と関係抽出、⑤推定規則(④で求めた関係)を用いた音声からの感情推定、の5つの工程から構成される。①から④において推定規則を得ることから、これらを学習部、⑤は推定規則から感情を推定することから推定部と呼ぶ。以後、各部分について説明する。

### 2.1. 学習部

#### (1) サンプル音声

サンプル音声に付与する感情としては、Ekman<sup>①</sup>らの提唱する6感情「恐怖・怒り・嫌悪・幸福・悲しみ・驚き」に「平常」を加えた7感情を使用する。また、サンプル音声収集時においては、音素による影響をさけるため、同一発話に対し、異なる感情を付与したものをを用いる。

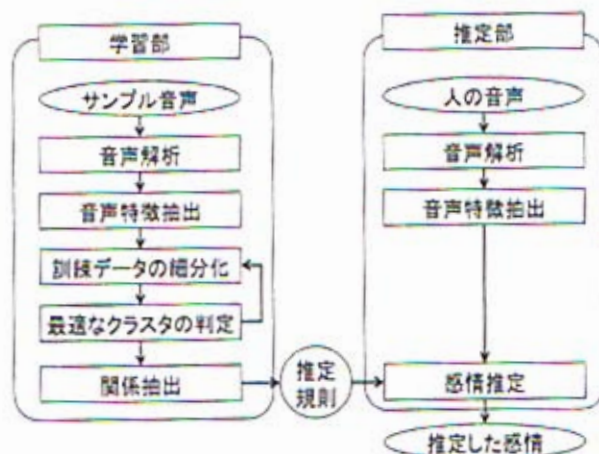


図1：提案手法の概要

#### (2) サンプル音声の解析・音声特徴抽出

各サンプル音声より、ピッチ(基本周波数)、パワーを抽出、さらに最大値、最小値、平均値、範囲、分散、四分位値、四分位値範囲を求める。これらの値を音声特徴として用いる。

#### (3) 訓練データの細分化

サンプル音声の細分化に対しては、感情ごとに音声特徴による階層的クラスタ分析を用いる。

#### (4) 最適なクラスタの判定

細分化された訓練データに対し、クラスタ集合評価基準により、クラスタ集合を評価する。一定の基準を満たしていない場合は、(3)に戻り、分割基準を変更し、再クラスタリング(細分化)を行う。

現在、評価基準としては、次のものを考えている。得られたクラスタ集合に対し、クラスタ内指標として、クラス内のデータ数、クラス内分散、クラスタ間指標として、クラス間分散、データが異なるクラスに含まれる割合を端的に示すクラスの重なり率、クラス数を求める。これらをクラスタ集合の属性値と見て、主成分分析を行う。また、同一データから構築した分類器によりサンプル音声に付与された感情が推定される割合(以後、推定値)を求める。これらの値の関係に基づき、基準を設定する。



#### (5)関係抽出

(4)のクラスタに基づいて分類器の生成を行う。分類器として、構築される関係が人にとってわかりやすいこと、従来研究で利用されることが多いことに着目し、回帰木を用いる。

### 2.2. 推定部

新たな音声から感情を推定する。感情推定手法としては、新たな音声に対し、(2)で用いた手法で音声特徴を抽出し、(5)で求めた関係と適合し、各感情に対する予測値を求める。最も高い予測値を得た感情を推定した感情とする。また、最大予測値を持つ感情が複数あった場合は、一定の規則によって感情を選択する。

## 3. 評価実験

訓練データの細分化の有用性およびクラスタ集合評価基準の妥当性を評価するために、以下の実験を行った。

### 3.1. 実験内容

サンプル音声として、20代女性1名が、「おはよう」「ありがとう」「すみません」などの10種類の短い文を指定した7感情でそれぞれ読上げた70個の発話を用いた。また実験においては、これらのサンプル音声を、発話内容によって訓練データ(関係構築用データ)63発話とテスト用データ7発話に分けた10種類のデータセットを構築した。これらのデータセットにおける関係抽出・推定に対し、交差検査法を行った。

#### (実験1:細分化の有用性の評価)

訓練データの細分化が有用であるか評価するために実験を行った。クラスタの分割基準としては、全クラスタの距離の平均を用いる。

#### (実験2:クラスタ集合評価基準の妥当性)

提案したクラスタ集合評価基準の妥当性について、回帰木の推定値とクラス内分散・クラス間分散・クラスの重なり率・クラス数・クラス内のデータ数からえた主成分値との関係から、評価した。

### 3.2. 実験結果

図2に実験1の結果の一例として、パワーを用いた場合の正解数の比較を、図3に実験2の結果として、回帰木の推定値と第一主成分値と関係を示す。

## 4. 評価・考察

実験結果から、訓練データの細分化を行うことで、全体的に従来手法よりも高い感情推定精度を得ることができた。しかし、テストセットの組み合わせにより感情推定精度が低い場合も存在した。

また、回帰木の推定値と指標から得られた第一主成分値との間に関係が見られ、クラスタ集合評価基準として有用性があると考えられる。

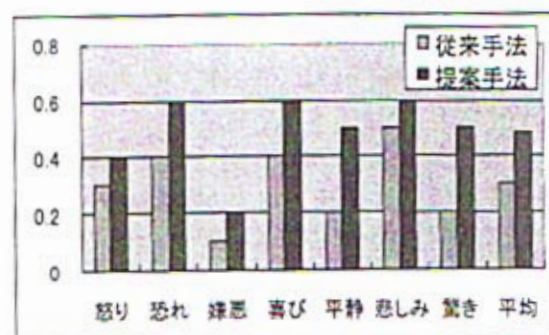


図2: 実験結果(パワー, 正解数)

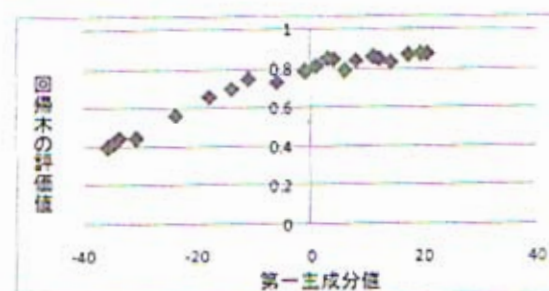


図3: 回帰木の推定値と第一主成分値

## 5. おわりに

本研究では、音声からの感情推定精度の向上を目的とし、訓練データの細分化における音声からの感情推定手法の提案をした。

## 参考文献

- 1) R.W.Picard: Affective Computing, The MIT Press, (1997).
- 2) P.Ekman, W.V.Friesen 著, 工藤力訳: 表情分析入門-表情に隠された意味をさぐる-, 誠信書房, (1987).