

平成 20 年度（2008 年度）

卒業論文要旨集

Abstracts of the 2008 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

顔の微細動作に基づく感情推定手法

An Emotion Estimation Method Based on Precise Analysis of Facial Expressions

インテリジェントソフトウェアシステム学講座 0312005008 荒川祐子

指導教員：羽倉淳 横松理樹 藤田ハミド

1.はじめに

人と人とのコミュニケーションにおいて、顔の表情は非言語情報を伝達することのできる重要な役割を果たしている。人は顔の微細な動きを無意識に見極め、表情を理解していると考えられる。

岩手県立大学ソフトウェア情報学部藤田研究室では「宮澤賢治プロジェクト」と題し、人間同士のコミュニケーションに近いインターフェースを持つ「仮想宮澤賢治」(賢治システム)の実現を試みている¹⁾。そこで重要な課題の一つとして、利用者の表情から感情を推定する手法の構築が挙げられる。

近年、表情認識・生成をはじめとする顔表情に対する研究が活発に行われている。これらの研究では、表情記述形式として、Ekmanらが開発したFACS(Facial Action Coding System)を利用したものが主である²⁾。FACSはAU(Action Unit)という視覚的に識別可能な表情動作の最小単位として設定されており、これらAUの組み合わせで表情を記述する手法である。

また、従来の研究ではFACSで定義されている基本6感情+1(喜び・驚き・恐怖・怒り・嫌悪・悲しみ・中立)に直接関連付けることで感情推定を行っている。認識できる表情は定義された感情に限定されてしまう。すなわち、実際の表情は意図的に作られたものもあれば、自発的に表出されたものもあり、多様である。一方で、人間は社会的な表情を表出することであり、従って感情とは異なる表情を表出している場合も多いと考えられる。人が、コミュニケーションを行う際には、より細かな観察・分類を行い、多様な感情を読み取っていることが予想される。そこで、本研究ではより詳細に表情の動きを分析することで表情を隋意・不隋意表情を含めて認識し、感情を推定する手法の構築を試みる。

2. 表情による感情推定

2.1 従来法

従来の表情認識システムの研究には、表情の動的侧面を十分に用いていない、表情の分類がFACSの基本カテゴリーに留まっている等の問題点が挙げられる。

そこで、本研究では表情変化を顔の微細動作に着目することによって認識できると考える。そして、顔の微細動作から得られる情報を用いて表情および感情をより詳細に理解する手法を提案する。

ここでは、表情を端的に表す特徴点を定義し、その特徴点の移動量から顔のバージョンの動きを抽出する。カメラから人の顔を検出し、特徴点のx座標データ、

y座標データを得る。その座標データから特徴点の動いた角度と距離を計算によって求める。その得られた移動角度を4パターン、移動距離を3パターンの全部で12パターンの分類に分ける。その分類された動きのパターンを分析することによって動きの特徴を抽出する。その抽出された動きのパターンを定義することにより、表情を認識することが可能となる。

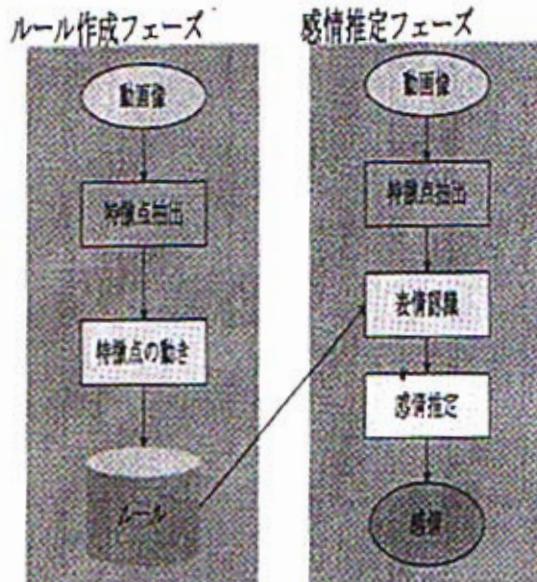


図1：システムの全体像

2.2 表情時系列の詳細化によるアプローチ

特徴点とは、表情における顔の動作を特徴的に表す位置を定義した点である。特徴点の検出にはAAM(Active Appearance Model)を用いる。特徴点の数は、左右の眉に14点、左右の目に17点、口に8点とした。AAMは特徴点を設けた顔画像から統計モデルを作成し、それを用いて画像からモデルと同じと思われる箇所に特徴点を定めることができる。

2.3 微細動作に着目したアプローチ

AAM³⁾から得られた特徴点の座標データについて移動量を計算する。まず、すべての座標データを一つ先のフレームデータから引き算をし、移動量を

だす。そのデータを利用し、計算を進めていく。移動距離については $\sqrt{x^2 + y^2}$ で求めることができる。次に移動した角度を求めるために以下の計算を行う。

$$\theta = \tan^{-1}(y/x)$$

これにより、移動距離と移動した角度が求められる。次に、得られたデータをもとに分類を行う。分類については、経験上、距離を2種類、角度については4種類に分類しており、全部で8種類に分類ができる。距離については、0.002以下と0.002より大きいものに分類する。角度については $0^\circ < \theta \leq 90^\circ$, $90^\circ < \theta \leq 180^\circ$, $-180^\circ < \theta \leq -90^\circ$, $-90^\circ < \theta \leq 0^\circ$ の4種類に分類する。

移動距離と移動角度により8種類に分類したデータをSOM(4)を用いてさらに分類を行う。SOMとは自己組織化マップ(Self-Organizing Map)の略である。T.Kohonenにより1981年頃に発表された教師なし学習ニューラルネットワークで、入力パターン群をその類似度に応じて分類する能力を自律的に獲得していくニューラルネットワークである。これを利用し、フレーム数1~Nのデータを100種類に類似性に基づいて分類を行う。

3. 特徴ルール抽出

SOMで分類したデータを確率推論を利用して、ある事象 X_1 が起きたとき、次に X_2 が起きる確立を求める。それにより X_1 が起きたときに、ある表情が認識される確率を求めることが可能である。図2のモデルの各ノードは、顔全体の動作パターンを示し、アーケは統計的に獲得された遷移確率を包含する。

図2のグラフにおいて、ノード61, 79のように親が存在しないノードには、それらの確立度で定まる確率が付随する。これを事前確率という。事前確率は、統計的に算出する。

これに対して、他のノードは親ノードを持つので、条件付確率が付随する。条件付確率とは、ある事象 X_1 が起きた時、次に X_2 が起きる確率のことである。条件付確率を出すことで、動きの予測をパーセンテージで出すことが可能となる。

このモデルを基本6感情の随意・不随意感情=合計12個作ることにより表情が認識可能である。

このモデルを利用することにより、一瞬の表情からある程度の確信を得て、その表情の意味を理解することが可能となる。ある表情が表出し終わった時点で確率と共に出力を行う。パーセンテージで表示することによって、その後の感情推定フェーズにより詳細に感情を推定することができると思われる。

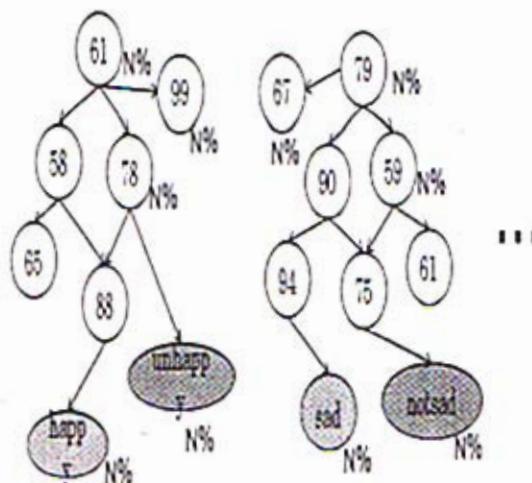


図2：統計的表現解釈ルール

4. 評価方法

本手法によって得られた表情ルールが認識の際にどのくらい適合するかを測るために、以下のような実験を計画している。被験者の表情データについて表情ルールの適合率を求める。一方で被験者に表情データを提示し、感じた感情を時間的要因を踏まえた上で書き記させる。その両者の整合性を検証する。

5. おわりに

本稿では、表情を理解する際に重要な顔の微細動作に着目することで、表情をより詳細に理解する枠組みを提案した。この手法により、演技による作られた表情と自発的に表出された表情の認識が可能になると想われる。

今後は評価、感情推定部のシステム構築を行い、問題点を解決していく。

参考文献

- 1) Hamido Fujita, et al: Virtual Cognitive Model for Miyazawa Kenji Based on Speech and Facial Images Recognition, WSEAS Transactions on Circuits and Systems, Issue 10, Vol.5, pp.1536~1543(2006)
- 2) Paul Ekman, Wallace V. Friesen, 工藤力 訳: 表情分析入門, 誠信書房 (1987).
- 3) Cootes, T. F. et al: Active Appearance Model, European Conference on Computer Vision 1998, Vol.2, pp.489~498, 1998
- 4) T. Kohonen. Self-Organization and Associative Memory, volume 8 of Information Sciences. Springer Verlag, 1984.

音節単位の音声特徴変化に基づく合成音声による感情表出の研究

Research of Emotion Expression Using Synthetic Speech Syllables

インテリジェントソフトウェアシステム学講座 0312005026 大橋真里菜

指導教員：藤田ハミド 横松理樹 羽倉淳

1. はじめに

現在、我々は「仮想宮澤賢治」(賢治システム)の実現¹⁾を試みている。賢治システムとは感情を伴った応答をユーザと行う疑似人格システムであり、その機能の1つに朗読がある。本システムが目指す朗読機能は、表現する感情が与えられた文章を音声合成によりその感情を表現しつつ読み上げるものである。朗読の場合、読み上げる文章が固定されているため、感情を表現するために音のピッチ・パワー・時間といった音声特徴を変化させる必要がある。

この音声特徴を表出したい感情に従い、変化させる方法として、音声特徴と感情との対応付けを利用する方法²⁾がある。しかし、明確な対応付けの定義ではなく、既存手法の精度にはまだ改良の余地がある。

以上の背景から、本研究では感情表現を目的とした新しい音声合成手法を提案する。従来の音声合成手法は発話全体の音声特徴を変化させている。しかし、音声学などの分野において発話内にある音節のピッチを変化することで、主題の提示、対比や感情的意味を伝えることが可能であること³⁾、感情により音節ごとの音声特徴が変化すること⁴⁾が示されている。本研究ではこの点に着目し、人の感情を込めた発話から獲得した音節単位の音声特徴変化を再現することで感情表出を行う音声合成手法を提案する。

2. 音節に基づく合成音声による感情表出

本システムの概観を図1に示す。本手法は、意図的発話データの音節毎の音声特徴と感情とを対応付けることで音声合成ルールを構築する。また、感情に対応したルールを用いて音声合成を行うことで、合成音声による感情表出を試みる。

2.1. 感情と意図的発話データ

意図的発話データは、ある感情を伝えることを意図して人が発話した音声データである。表出を試みる感情としてEkman⁵⁾らが提唱した基本6感情「怒り・喜び・嫌悪・恐怖・悲しみ・驚き」と、これらに当てはまらない感情として「平常」を加える。発話者としては、音声による感情表出経験が豊富で、変

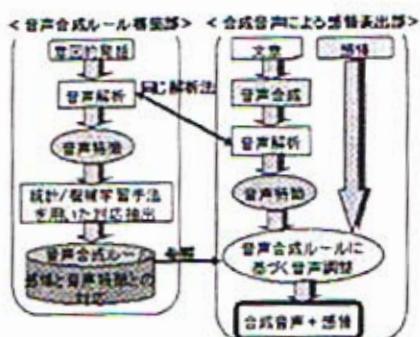


図1：音節に基づく合成音声による感情表出の概観

化を明瞭にできる人物が望ましい。

2.2. 音声合成ルール構築部

音声合成ルール構築部では、次の手順にそって意図的発話データからルールを構築する。

(1) 発話毎の音節の構築

はじめに、意図的発話データを音素に分割する。次に、音節の定義に基づき、音素列を音節へと変換する。音節の定義には聞こえ度⁶⁾を用いる。聞こえ度とは、聴き手側からみた各音素の聞こえ方の度合いの違いによるもので、音声と同じ大きさ・高さ・長さで発した場合、遠くに届くものほど聞こえ度が大きいという考え方である。

(2) 音節単位の音声特徴抽出

音節単位でピッチ・パワー・音節持続時間を音声特徴として捉える。ピッチ・パワーについては最大値・最小値・範囲・平均値・分散・標準偏差・四分位点・第1四分位点と第3四分位点の差を用いる。

(3) 発話の変化の抽出

発話はいくつかの音節で成り立ち、音節単位で音声特徴が変化するという点に着目し、(2)で求めた値の回帰直線を求める。この回帰直線の傾きが、音節単位の音声特徴変化を端的に示していると考える。

(4) クラスター分析を用いた対応抽出

回帰直線の傾きを角度に変換し、その値に対してクラスター分析を行う。その結果得られた感情毎のクラスター中心値を感情に対応する音声特徴と捉え、

音声合成ルールを構築する。

2.3. 合成音声による感情表出部

合成音声による感情表出部では、音声合成ルールに基づき、感情に合わせた合成音声を生成する。詳細な手順は次の通りである。

(1) 音声合成と合成音声の音声特徴抽出

文章に対し、音声合成を行う。この時、表現したい感情にあったパラメータに従い音声合成を行う。パラメータとは音声合成時に設定できる項目を指す。作成された合成音声を、ルール構築時と同じ手法で音声解析を行う。

(2) 音声合成ルールの参照と合成音声の出力

ユークリッド距離を用い、合成音声の音声特徴解析結果が最も近いクラスター中心値を持つルールを求める。これにより、求めたクラスター中心値に対応づけられた感情が、表現したい感情と同じであれば、合成音声を出力する。異なる場合はパラメータを変更し、再度音声合成を行う。

3. 評価実験

提案手法の妥当性を評価するために合成音声に対する聴取実験を行った。実験では、比較1として、合成音声が表出を試みた感情と合成音声に対して被験者が付加した感情との比較を、比較2として、発話者が意図した感情と被験者が意図的発話データを聞いた際に推定した感情との比較を、比較3として、被験者が合成音声と意図的発話データに対して付加した2つの感情の比較をそれぞれ行う。

3.1. 音声合成ルール構築実験

意図的発話データとして、1名の発話者が10種類の文章を、7感情を意図して発話したもの用いる。この意図的発話データから2.2で示した手法により音声合成ルールを構築した。今回は各クラスター数を3とし、各感情に対し3つのルールを求めた。また、音声特徴抽出にはWaveSurfer⁷⁾を、クラスター分析には統計処理環境R⁸⁾を用いた。

3.2. 合成音声による感情表出実験（聴取実験）

聴取実験として2種類の実験を行った。実験1は合成音声に対し、感情を付ける実験である。実験には音量・速度・抑揚・ピッチが異なる81個の合成音声を用いる。実験2では意図的発話データ、70個に対し感情を付ける。発話者本人を含む被験者24名は音声をランダムで聞き、推定した感情を7感情から1つだけ回答する。この結果、2つの感情が一致すれば、意図した感情が伝達できているといえる。

4. 結果と考察

図2に比較3の結果（人の声と合成音声の伝達率の差）を示す。実験の結果、喜びでは合成音声の伝達率が高い結果となった。理由として、文書の影響や合成音声の方が人の演技よりも大きいことが考えられる。喜び以外の感情では意図的発話データの伝達率が高い結果となった。理由として、音声合成ルール構築時にクラスター数を固定していることが考えられる。また、合成音声と最も近いクラスター中心値の持つ感情を付加していることや、今回作成した合成音声では割り当てられない感情があったことなどが挙げられる。

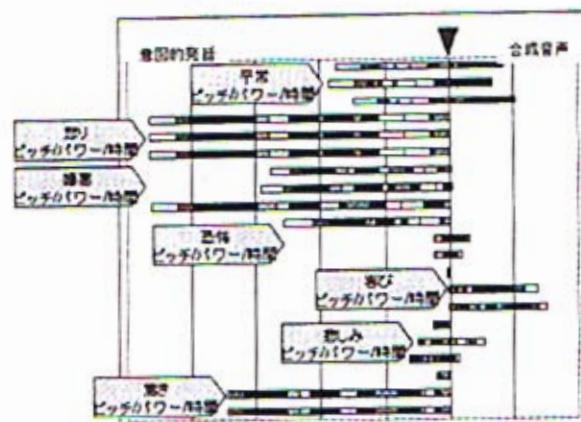


図2：比較3における実験結果（感情単位）

5. おわりに

研究では、合成音声による感情表出を行うために音節単位の音声特徴変化に着目した手法を提案した。今後の課題としてはクラスター・実験の詳細な解析、従来手法との比較・統合の検討、クラスターと合成音声の再構築を行う。

参考文献

- 1) 富澤賀治プロジェクト：<http://www.fujita.soft.iwate-pu.ac.jp/KENJI/index.html>.
- 2) Pierre-Yves Oudeyer：“The production and recognition of emotion in speech: features and algorithms”，技術評論社（2003）。
- 3) 小泉保：“音声学入門”，東京大学書林（1996）。
- 4) 広瀬啓吉：“韻律と音声言語情報処理-アクセント・イントネーション・リズムの科学”，丸善株式会社（2006）。
- 5) 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）。
- 6) 斎藤純男：“日本語音声学入門”，株式会社三小堂（2007）。
- 7) WaveSurfer：<http://www.speech.kth.se/wavesurfer/>.
- 8) R-Project：<http://www.r-project.org/>.

音声合成を用いた音声からの感情推定

A Study of Estimation of Emotion in Speech Using Speech Synthesis

インテリジェントソフトウェアシステム学講座 0312005046 木下 おりみ

指導教員：藤田ハミド 横松理樹 羽倉淳

1. はじめに

現在、我々は感情を伴った応答をユーザと行う擬似人格システム「仮想宮沢賢治」¹⁾の実現を試みている。本システム実現のためには、ユーザが発話した音声から、ユーザが持っている感情の推定を行う必要がある。従来手法では感情と音声特徴との対応づけは「人の意図的発話」から構築されている。²⁾しかし、大量のデータを集めることが困難である、音声収集時にノイズが混入する、訓練していない一般人が意図的発話をを行うのは困難であるといった問題を抱えている。結果、感情と音声特徴との十分な対応づけを構築できず、認識精度が低いというのが現状である。

以上のような背景から、本研究では感情と音声特徴の対応づけを「合成音声」から構築することを試みる。本手法は異なる音声特徴を持つ合成音声とその音声から聴衆が推定した感情との対応づけに基づき、人の音声からの感情推定を行うものである。合成音声は直接PCに録音できることから、ノイズ混入の軽減が期待できる。また、合成音声を聞き感情を付与することは、意図的発話より負荷が少ない。さらに聴衆の立場に立った感情推定ルールの構築を試みることから、システムの目的により適合していると考えられる。以上のことから、本手法は前述した問題の解決や推定精度の高い手法の実現が期待できる。

2. 音声合成を用いた音声からの感情推定

図1に本手法の概要を示す。本手法は、感情と音声特徴との対応づけである感情推定ルールを抽出する部分(感情推定ルール抽出部)と、作成した感情推定ルールに基づき音声に込められる感情を決定する推定部から構成される。

2.1. 感情推定ルール抽出部

(1) 合成音声収集

音声パラメータを操作し、様々な合成音声を作成する。また、作成した合成音声を人に聞かせることにより感情付与を行う。感情には、Ekmanらの提唱する6感情「怒り・嫌悪・恐怖・幸福・悲しみ・驚

き」³⁾と「平常」の合計7感情を使用する。

合成音声を生成する文章には、多様な感情を表現できるものを用いる。また、感情付与の際に、より人間同士の対話に聞こえるようにシチュエーションとして人の声に対し合成音声が回答するという対話形式を取り入れる。

(2) 音声解析

合成音声に対し音声解析を行い、音声特徴を抽出する。音声特徴としては、パワーとピッチについて、それぞれ発話単位の最大値、最小値、範囲、平均、分散を求める。

(3) 特徴からのルール作成

(2)で求めた音声特徴を標準化し、この値に対し、クラスター分析を行う。クラスター分析によって得られた各クラスター中心値とそのクラスタに割り当てられた感情との対応づけを感情推定ルールとする。

2.2. 推定部

与えられた人の音声を感情推定ルール抽出時と同様の手法で解析し、音声特徴を求める。次に、得られた音声特徴とともに近い音声特徴の集合を持つルールの感情を求め、これを推定された感情とする。

この流れを通じ、音声からの感情推定を実現する。この点については、従来手法と同様である。

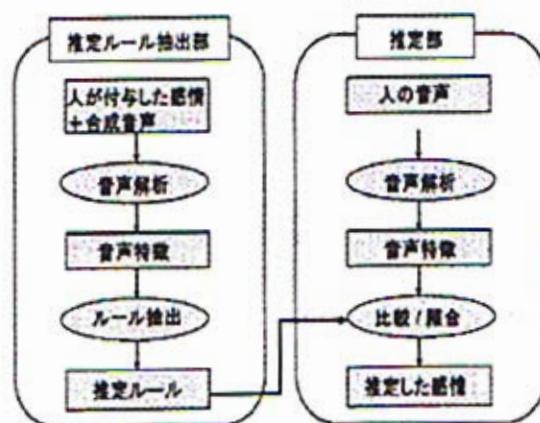


図1 音声合成を用いた音声からの感情推定

3. 評価実験

本手法の有用性を評価するために実験を行う。実験の内容は、合成音声への感情付与、感情推定ルールの構築および評価実験である。

3.1. 合成音声への感情付与

最初に、合成音声への感情付与を行う。合成音声は、音量・読みの速さ・抑揚の大きさ・声の高低の4つの音声パラメータを操作し、81パターン用意する。感情を付与する際は、より感情をつけやすいよう人の声に対し合成音声が答える対話を被験者が聞き、合成音声の応答に感情の付与をする。また合成音声データを流す際、音声特徴の変化がわからぬようランダム再生し、直後に7感情中1感情を回答する。今回は20代女性1名が感情付与を間を置いて5回実施し、それぞれの結果に対し評価した。

3.2. 感情推定ルールの構築

3.1で得られた感情が付与された合成音声に対し、それぞれ2.1で述べた手法で感情推定ルールを抽出する。クラスター分析においては、クラスター数を3とし、各感情に対し3つの推定ルールを構築した。

3.3. 評価実験

構築した感情推定ルールを評価するため、以下の3つの比較評価を行った。

(比較1)

意図的発話データを被験者が聞き、発話内から推定した感情を回答する。意図的発話データとしては、音声による演技経験者が10個の文章に7感情を意図して発話したデータを利用した。これらの被験者の回答結果と発話者の感情を比較する。

(比較2)

発話者の感情と、比較1と同じ意図的発話データに対し、感情推定ルールを用いてシステムが推定した感情とを比較する。

(比較3)

比較1で得た発話に対する聴者の感情と、比較2で得た感情ルールに基づいてシステムが推定した感情とを比較する。

本実験の被験者は、24名である。また、この中に発話者も含まれる。

3.4. 実験結果

図2に比較実験の結果を示す。図の左側が被験者、右側がシステムによる正答率を示している。

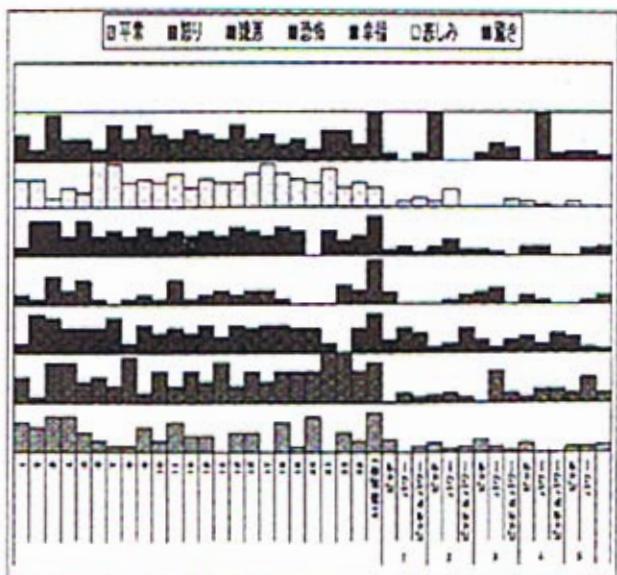


図2 比較実験の結果

4. 考察

今回の手法による結果は全体的に低いが特徴によっては高い。これは判別に必要な音声特徴の選択や抽出が十分ではないと考えられる。また、発話者と比べると幾つかは同等の結果が得られる。これは発話者に合ったルールであるといえる。

5. おわりに

本研究では、感情推定を行うための感情と音声特徴との対応づけを、従来の人の発話ではなく、人が合成音声を聞き感情を付与した音声データから構築する手法を提案した。また、従来の手法における音声データの収集の負荷の問題点の解決を試みている。

結果、システムによる感情推定は人によるものよりも全体として劣っていた。しかし、一部では高い結果も得られた。

今後は、さらに実験結果を解析し、それらを反映することでシステムによる感情推定の向上を図る。また、従来手法での感情推定ルールとの比較・統合も検討し、認識精度を高めることがあげられる。

参考文献

- 1) 宮沢賢治プロジェクト:
<http://www.fujita.soft.iwate-pu.ac.jp/KENJI/index.html>
- 2) Pierre-Yves Oudeyer: "The production and recognition of emotions in speech: features and algorithms", International Journal of Human Computer Interaction, 59(1-2):157-183, (2003).
- 3) 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).

表情に基づくユーザ Profile を用いた感情推定

Emotion Estimation from Facial Expression based on Profile Information of User

インテリジェントシステム学講座 0312005112 千田調和
指導教員：藤田ハミド 羽倉淳 横松理樹

1.はじめに

本研究は、藤田研究室で取り組んでいる「宮澤賢治プロジェクト」における表情からの感情推定部を担うものである。本プロジェクトの目的は、不特定のユーザがコンピュータを意識することなくコンピュータとの相互作用を可能にすることである。そのために人間の感情をコンピュータに認識させる手法の構築が必要である。感情推定手法の一つとして2)表情を用いた感情推定手法の構築を試みてきた。しかし、現手法における問題点として、不特定の人物に対しての認識率が低いことや、恐怖や悲しみなどの認識が困難である点が挙げられる。これは、表情には個人差が存在し、それが表情認識精度に影響与えているためと考えられる。ここでは、表情の個人差を反映した表情認識手法を取り入れることで感情推定精度の向上を試みる。

2.ユーザ Profile

表情の個人差を反映するためにユーザ Profile (性別、年齢) を用いる。これは、性別によって骨格の作りなどにより表情に違いがあることや、年齢を重ねることによる表情筋の衰えなど、表情の動作の違いに直接関係があると考えるからである。これにより、ユーザ Profile が近い人は表情も似ているという仮定を立て、表情を分類して認識していく手法を提案していく。この手法により表情の個人差による表情認識への影響は軽減され、不特定の人物への対応が可能となり、感情推定の精度向上に繋がると予測する。使用する Profile データは、性別 (男性、女性) と年齢 (10代、20代、30代、40代、50代、60代以降) である。

3.ユーザ Profile を用いた感情推定手法

表情を線形システムとして捉え、特定の感情を表す感情をシステムモードとして同定する。ここで同定結果は特徴点を用いる遷移行列として表現し、後述のテンプレートとして用いる。このテンプレートは人間の各感情の表情を表しており、これを使って人間の感情を推定する。そして、ユーザの表情に最も近い動作のテンプレートをユーザ Profile によって選択する手法を用いたシステムを構築する。システムの全体像を図1に示す。以下、感情推定手法の詳細を述べる。

3.1. 特徴点抽出

特徴点の検出には AAM (Active Appearance Model) 3) を用いる。この手法は特徴点を設けた顔画像から作成した顔のモデルを利用し、画像からモデルと似ていると思われる顔を見つけだし特徴点を定めることの出来る手法である。特徴点は表情を追うために顔の3パート (眉、目、口) に定めたもので、左右の眉に14点、左右の目に16点、口に8点とした。

3.2. テンプレート作成法

テンプレートは、感情ラベルと各パートの動きを予測する線形システムで構成される。線形システムは遷移行列を

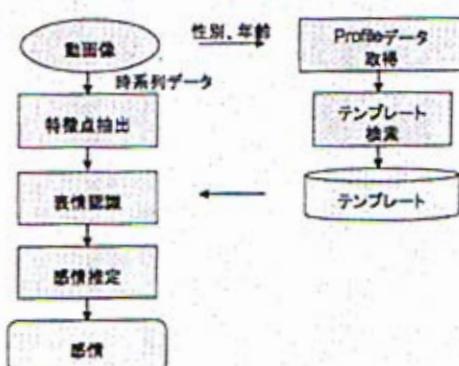


図1：システム全体像

持ち、遷移行列により現時点の各特徴点の座標からの1フレーム後の座標を予測する。

作成手法としては、表情の時系列データから各感情が現れていると思われる区間のフレームを指定し、その時系列データから遷移行列を作成する。そして各区間の遷移行列の類似度を基にクラスタリングを行い、新たに遷移行列を作成しテンプレートとして用いる。

3.3. ユーザ Profile によるテンプレート検索

ユーザ Profile データによって使用するテンプレートを検索する。これは、ユーザの表情の動きにより近いテンプレートを使用して表情認識の精度を向上させるためである。

テンプレートはあらかじめ性別と年齢の組み合わせによる12のカテゴリに分けDB上に管理されており、ユーザから得られたProfileのカテゴリを検索する。一つのカテゴリ上には複数のテンプレートが存在するため、さらにテンプレートを絞り込み最終的に使用するテンプレート (ユニバーサルテンプレート) を決定する。これは、表情をさらに特定し表情を一般化するためである。ユニバーサルテンプレートはユーザの顔のバーツ間の距離を用いて選択する。使用するバーツ間の距離は、左右の目の間の距離 α 、眉の長さ β 、眉毛の両端の距離と鼻とを結んだ合計の距離 γ とする。これらのバーツを使用する理由として、表情から感情を認識するのに、目や眉が大きく関係しているため、顔のバーツ間の距離が似ている人の表情に類似性があると仮定したことである。

4. 表情認識と感情推定

本手法では、カメラから得られた各特徴点の時系列データと用意されたテンプレートとをマッチングし、どの感情を示す表情なのかを認識し感情を推定する。マッチング手法はまず、1フレーム後の各特徴点の座標をテンプレートの持つ線形システムによって予測する。次に、カメラから得られる特徴点の時系列データの値と予測された値を比

較し誤差を抽出する（以下予測誤差）。

表情認識時には予め無表情時の時系列データを用いて、各テンプレートに予測誤差範囲を定めておく。この予測誤差範囲は、各テンプレートに対して無表情時の時系列データを用いて抽出した最大値(MAX)と最小値(MIN)を基に定められるものである。予測誤差が最大値を超えた場合はHigh、最大値と最小値の範囲内ならMiddle、最小値を下回ればLowとシステムが判断する。基本的に予測誤差がLowの状態の時その感情を認識する。Lowの状態が存在しない場合はMiddleとHighの状態を考慮に入れて感情推定を行う。

5. 実験

ユーザProfileによるテンプレートの分類によって表情の個人差が与える影響を軽減し、感情推定の向上に影響を与えることと、ユニバーサルテンプレートを特定することによる感情推定精度の向上への影響を検証するために以下の2つの実験を行った。

5.1. 実験1

同一カテゴリ（男性、20代）に属する被験者3名に対し感情推定を行う。使用するテンプレートは被験者と同一カテゴリ（男性、20代）のテンプレートと他のカテゴリ（女性、20代）のテンプレートである。

5.2. 実験2

実験1と同じ被験者3名（男性、20代）に対し感情推定を行う。使用するテンプレートはどちらも被験者と同一のカテゴリのものだが、一つはユニバーサルテンプレートとして近いものであり、もう一つは選択されなかったテンプレートである。

尚、認識率は各感情が表れている3人分の区間の内、どのくらいの割合でシステムが認識したのかを示している。認識率は、正確に表情が示す感情だけを認識した結果と誤認識を含む結果（他の感情も認識）、誤認識のみの結果（表情が示さない感情のみ）を求めた。

5.3. 実験結果

実験1では、幸福に関しては、誤認識はあったものの、どちらのテンプレートでも表情が示す感情の認識率が高かった。怒りに関しては、誤認識はあったが同じカテゴリのテンプレートを使用した際の認識率が高い結果となった。しかし、驚きと嫌悪に関しては異なるカテゴリ（女性20代）のテンプレートを使用した際の認識率の方が誤認識の確率の低く、高い結果となった。恐怖と悲しみについてはどちらのカテゴリでも正確な感情の認識率は低く、誤認識率が高い結果となった。

実験2では、喜びと恐怖に関しては、大きな認識率の違いは見られなかった。怒り、嫌悪に関しては、ユニバーサルテンプレートでの認識率の方が高い結果が出た。しかし、驚きに関しては、ユニバーサルテンプレートによって認識率が下がる結果となった。悲しみに関しては、ユニバーサルテンプレート以外では全く認識出来なかつたのに対し、ユニバーサルテンプレートを使用した際には、誤認識はあるものの認識率が上がる結果となった。

6. 考察

今回実験1においてユーザProfileに基づき、同一カテゴリのテンプレートを分類することによって一部で表情が示す感情の認識率が高くなることが確認出来た。しかし、

表1：同一カテゴリ（男性、10代）の認識結果

%	正	正／誤	誤
幸福	0.5	84.2	15.3
怒り	1.4	67.0	31.6
驚き	55.2	13.7	31.1
嫌悪	21.2	39.7	39.1
恐怖	0.0	39.3	60.7
悲しみ	0.0	0.0	100.0

表2：異なるカテゴリ（女性、20代）の認識結果

%	正	正／誤	誤
幸福	0.0	83.6	16.4
怒り	0.5	20.5	79.1
驚き	79.8	18.6	1.6
嫌悪	48.4	44.6	7.1
恐怖	0.0	55.8	44.2
悲しみ	0.0	79.3	20.7

表3：ユニバーサルテンプレートでの結果

%	正	正／誤	誤
幸福	3.3	96.7	0.0
怒り	18.1	7.9	74.0
驚き	34.4	43.7	21.9
嫌悪	40.2	41.3	18.5
恐怖	0.0	66.3	33.7
悲しみ	0.0	59.2	40.8

驚きや嫌悪のようにカテゴリ以外のテンプレートを使用した際の認識率の方が高い結果が出てしまった。また、実験2により、怒りや嫌悪、悲しみなどユニバーサルテンプレートを使用することによって認識率が向上したと思われる結果を確認することが出来た。これらから、使用するテンプレートによって表情からの感情認識率が変化することがわかる。今後は、検証するテンプレートと被験者を増やし実験を行い、認識率を上げるためにテンプレートの特徴や分類方法を分析してすることにより、感情推定精度の向上に繋がると考える。

7. おわりに

本研究では、表情からの感情推定の精度向上のため、個人の表情の違いに着目したユーザProfileを用いた表情からの感情推定手法を提案した。

しかし、今回の実験ではユーザProfileによる感情推定精度の大幅な向上を確認するまでには至らなかった。今後は、考察で挙げた点を改善し、個人の表情の違いについての特徴を分析し、感情推定精度の向上を目指す。

8. 参考文献

1) Hamido Fujita, et al: Virtual Cognitive Model for Miyazawa Kenji Based on Speech and Facial Images Recognition, WSEAS Transactions on Circuits and Systems, Issue 10, Vol.5, pp.(2006)

2) 佐藤徹、他3名：線形システム同定に基づく自然な表情からの感情認識手法、岩手県立大学ソフトウェア情報学部（2007）

3) Cootes, T. F. et al: Active Appearance Model, European Conference on Computer Vision 1998, Vol.2, pp.489-498, 1998

擬人化システムの利用者に応じた表情合成手法

A Facial Expression Synthesis suitable for user profiles

インテリジェントソフトウェアシステム学講座 0312005121 中坪祐介
指導教員：羽倉淳 森松理樹 藤田ハミド

1.はじめに

岩手県立大学ソフトウェア情報学部藤田研究室では、「宮澤賢治プロジェクト」が進行中である。これは、コンピュータ上で宮澤賢治の人格を仮想的に再現し、三次元顔モデル上での表情合成に取り組んでいる¹⁾。そこでは Ekman の提唱した表情記述法 FACS(Facial Action Coding System)²⁾を基にした表情合成手法が実現されている。「賢治の再現」という目標における人間らしさを追及すると、相手に応じて有機的に表情を表出することが理想である。本システムのような対話型の擬人化システムの研究では、発話内容からユーザの感情を推定し、表出する表情を選択するシステム³⁾など、ユーザの動的な情報に応じた表情合成がほとんどである。しかし、人間のコミュニケーションにおいては相手の表情や音声等の動的な情報以外にも、相手と自身との関係や、相手がどのような人物であるか等の静的な情報が影響を与えることが考えられる。そこで、相手の静的な情報から、相手を識別することでその相手に応じた表情合成手法を提案する。

2. ユーザ profile

ユーザを識別する情報には、人間の属性として一般的な年齢や性別、職業などがあげられる。それらのユーザ固有の性質を「ユーザ profile」として扱い、システムの入力とし、ユーザを識別する材料とする。Ekman らに示されている「表示規則」⁴⁾によれば、対話の相手の社会環境・立場の違いにおいて、人間はその場にふさわしい表情を表出すると言わわれている。さらに発達心理学においては、年齢による他者感情の認知の成熟の違いが明らかにされている⁵⁾。これらの考え方から、本研究では「年齢」をユーザ profile とする。社会環境や、他者感情の認知の成熟の違いが顕著に現れるのは「子供」と「大人」であると推察し、それぞれに対する表情表出の違いに注目する。対象とする子供の年齢は、他者感情の認知の発達に関する心理学的知見⁶⁾をもとに 3~10 歳までと決定した。

3. 表情合成の概要

賢治システムの表情の合成は、表情動作の最小単位 AU (Action Unit) や、皮膚の皺を三次元の顔モデル上

で作成し、動作バーツとして繰り返し使用し実現している⁷⁾。システムが表現する感情は Ekman の基本六感情（怒り・嫌悪・恐怖・幸福・驚き・悲しみ）である。感情と関連付けた AU と皺の組み合わせを、予め「表情合成スクリプト」に記述する。

3.1.1. 感情パラメータ入力・上位感情抽出

感情はシステムにおいて、それぞれの感情名と強度(0~100)で構成される。受け取った感情は強度順に整列され、出力対象となる上位感情を決定する。本研究では、ユーザが子供である場合には表出する感情を单一とする。

3.1.2. 感情出力部

出力する感情と、その強度をシステムが受け取ることで、表情合成スクリプトが読み込まれ、表情が合成される。各動作バーツには、動作量を決定する数値があり、その大小を変えて呼び出すことが可能である。動作量は、感情の強度により計算される。

3.1.3. 発声出力部

賢治システムの合成音声と同期させた発話動作を行う。発話する文章のデータを受け取り、表情合成スクリプトが読み込まれ、発話の動作が合成される。データは、音素列と、出力時間長から構成されている。受け取った順に音素の口の形を出力時間長動かすことで発声動作が行われる。

4. ユーザ profile に応じた表情合成手法

ユーザによってシステムが違う表情を表出するために、ユーザ profile に対応する表情合成スクリプトが必要である。システムがユーザ profile を認識することで、各 profile 専用の表情合成スクリプトを呼び出し、表情合成を行う。次に、表情合成スクリプト作成の基となる表情の抽出過程を示す。

5. 子供に対する表情の抽出

表情表出の研究分野では、表情の分析や、表情の認知心理の研究が行われているが、子供に対しどのように表情表出が適しているかという知見は無い。そこで、子供に関する分野、職業に携わる人の知識を参考に考えていく必要がある。今回、本研究では 2 つのアプローチから、子供に対する表情を抽出する。

一つは、子供の関心を引く表情を保育士の経験的知識や考えを基に「変化の大きさ」「示す感情の明確さ」

と仮定し、Ekman の表情分析を参考に抽出する方法である。これにより抽出した表情は、表現する動作バーツの一部を、他の動作バーツより動作量を大きく出力することで変化の大きさを表現する。さらに、使用する AU を増やすことで、変化に富んだ表情の表現を試みる。

もう一つは、保育士の表情表出を基に抽出した表情である。保育士が子供に絵本を読み聞かせている表情を撮影し、各感情の表情を抽出したものである。撮影した映像を観察し、実際の表情の印象を伝えることを重視しながら表情を抽出したものである。ここでは、子供が親しみやすい表情を抽出することが目的である。

6. 作成、及び合成実験

実装環境は Autodesk 社の 3D レンダリングソフト「Maya8.5」上で、AU や皺の作成は同ソフトの「BlendShape」機能を、各モジュールはスクリプト「MEL」を利用した。各 AU は、顔モデル上にある格子状の頂点を、エリア毎に三次元方向で制御・動作させる方式で作成する。以下に、それぞれのアプローチから抽出した表情の合成結果を図 1、図 2 に示す。また、それぞれの動作バーツの組み合わせを表 1 に示す。



図 1 顔モデル 1：従来の恐怖
顔モデル 2：保育士の考え方を基に抽出した恐怖



図 2 顔モデル 1：従来の恐怖
顔モデル 3：保育士の表情から抽出した恐怖

表 1: 図 1 図 2 の表情の動作バーツ

	動作 (AU-AU の呼称)
顔モデル 1	AU20: 眼を横に引く / AU5: 上瞼を上げる AU7: 瞼を緊張 / AU1: 内眉を上げる / 額の皺
顔モデル 2	AU1 / AU5 / AU7 の動作量を 1.5 倍 発声の動作を 1.5 倍 AU6: 眼を上げる / 額の皺 / AU20: 眼を横に引く
顔モデル 3	AU4: 眉全体を下げる 外眉の角度を下げる / 上瞼が押し下がる

7. 評価

恐怖を例に三種類の表情を示したが、いずれも子供に対する効果は明らかでない。システムの表情が子供に与える印象や効果を検証するために、一般ユーザ 10 人を対象に実験を行った。表情の抽出過程を考慮し、「子供が親しみやすい」「子供の関心を引く」「子供への感情の伝わりやすさ」の三つを評価基準とした。感情毎に三種類の表情を動画で提示し、評価基準に基づき四段階評価を付けた。評価は低いほうから順に 1 から 4 までの点数とし、その平均値を求めた。例として恐怖の結果を表 2 に示す。

表 2 恐怖の実験結果

評価基準	顔モデル 1	顔モデル 2	顔モデル 3
関心を引く	1.1	2.2	2.3
親しみやすい	1.1	2	2
感情が伝わる	1.3	2.3	2

顔モデル 2 と 3 は、動作の組み合わせが大きく異なる。これは、恐怖を表す動作の違いであると考えられる。モデル 3 が自ら恐怖する表情であることに対し、モデル 2 は子供に対し恐怖を与える意味の表情であると考えることができる。意味の異なる恐怖を示したことで評価に差が表れにくかったと推察される。

8. おわりに

本研究では、ユーザに応じた表情合成を実現するため、扱う profile を年齢に限定し、子供に対する表情の抽出・合成を行った。調査や実験を通して、同じ感情で複数の表現をすることが人間らしい表情の実現に向けての課題であることが表面化した。今後、同じ感情の中でも複数の表現を子供に対し表すルールの構築が課題となる。

参考文献

- [1] 檜山洋一、羽倉淳、博松理樹、藤田ハミド、『FACS を用いた上位感情抽出式表情合成手法』、岩手県立大学学士論文(2008)
- [2] Paul Ekman, Wallace V. Friesen, 工藤力(訳)：表情分析入門、誠信書房(1987)
- [3] 目良和也、黒澤義明、市村匠、『話し言葉における感情を考慮した知的インターラクションシステムの構築』、広島市立大学
- [4] 新井邦二郎：図でわかる発達心理学、福村出版(1997)
- [5] 無藤隆：保育・看護・福祉ブリマーズ 発達心理学、ミネルヴァ書房(2001)