

平成 19 年度（2007 年度）

卒業論文要旨集

Abstracts of the 2007 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

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

An Emotional Recognition Method for Voluntary Facial Expression
Based on Linear System Identification

インテリジェントソフトウェアシステム学講座 0312004076 佐藤徹

指導教員：藤田ハミド 羽倉淳 榎松理樹

1. はじめに

近年、人とコンピュータがインタラクティブな対応をすることが重要となっており、その分野にHCI(Human Computer Interaction)がある。

そのHCIへの応用を目的として、表情からの感情を認識するシステムの研究が行われている。これらの代表的な手法としてFACS(Facial Action Coding System)¹⁾という表情記述手法で用いられているAU(Action Unit)を用いたパターンマッチング手法や、ニューラルネットワークを用いて認識する手法がある²⁾。これらの手法はFACSで定義されている基本6表情(喜び、驚き、恐怖、怒り、嫌悪、悲しみ)を被験者に練習させ、演じられた表情の認識を行っている。しかし、FACSを使用すると認識できる表情は定義された44種類に限定されてしまうため、これらのシステムをHCIシステムに応用した際に、Userの自然な表情を認識できるのかは疑問である。また近年、表情解析により、自然な表情と演技された表情では顔パーツ(眉、目、口など)の動きが違ってくる³⁾。

そこで本研究は、数多く存在する自然な表情の動作を線形システムとして扱いテンプレートを構築する。線形システムを認識に用いることで、AUに依存せず、時系列データからボトムアップに表情を分析することが可能である。これを実現することで人とコンピュータのコミュニケーションの向上を図ることができると考える。

2. 感情認識手法

本研究ではリアルタイムで表情を認識し、さらに表情認識結果から感情を認識するシステムの構築を目指す。カメラから人の顔を検出し、設けられた特徴点をキャプチャすることによって得られる時系列データを用いて表情を認識し、感情を認識する。システムの全体像を図1に示す。本手法は大きくテンプレート作成フェーズと感情認識フェーズから成る。感情認識フェーズでは、テンプレート作成フェーズで予め作成しておいたテンプレートを用いて感情を認識する。以下、それぞれのパートについて述べる。

2.1 特徴点抽出

特徴点の抽出にはAAM(Active Appearance Model)⁴⁾を用いる。特徴点は各パーツの動きを捉えるために定めたものであり特徴点の数は左右の眉に14点、左右の目に16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の特徴点を定めることでできる。あらゆる顔の画像を用いてモデルを作成することにより、人種、性別などによる顔の形を問わず様々な人の顔から特徴点抽出が可能である。

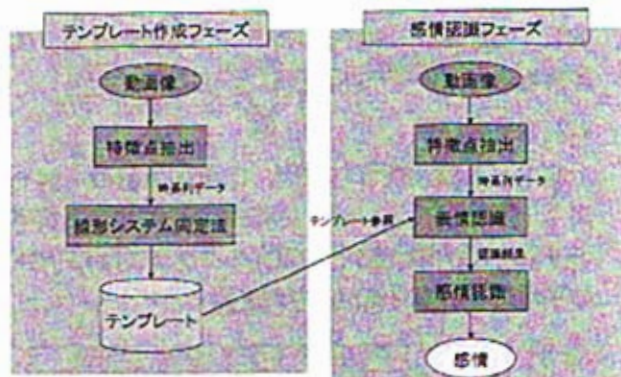


図1:システムの全体像

2.2 テンプレート作成手法

テンプレートは感情ラベルとパーツの動き方を予測する線形システムから成る。本研究では表情譜⁵⁾のようにパーツ毎のある動き方を1つの線形システムとみなし、その線形システムによって感情を表すパーツの動きを定義する。線形システムは遷移行列を持ち、遷移行列により顔の特徴点の座標から1フレーム後の座標を予測できる。

テンプレートの作成手法はまず、カメラを用いて表情の時系列データを取得する。その時系列データを取得したときの動画画像をテンプレート作成者が観察し、感情が現れていると思われる区間のフレームを指定し、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に基づき、クラスタリングを行い、それにより作られた遷移行列をテンプレートに用いる。

本研究では、感情を示す表情はあくまで基本6表情のいずれかに分類でき、各表情の下層に演技の表情や自然な表情が存在するという考え方で行う。テンプレートは基本6表情のいずれかにカテゴライズされており、その中で複数の表情を示すテンプレートを持つ。

2.3 表情認識

表情認識は各テンプレートを用いて時系列データを予測し、実際の値との予測誤差を求めることによって行う。予測誤差の求め方はまずテンプレートが持つ線形システムにより時系列データの1フレーム後の特徴点の座標を予測する。次に予測した値と実際の値との誤差を予測誤差としている。

表情認識には予め無表情時の予測誤差範囲を定めておく必要がある。無表情は人が顔の筋肉を動かしていない自然な状態とする。ある予測誤差が最大値を上回ったらそのテンプレートとは表情が異なり、最小値を下回ればそのテンプレートと表情が似ているということである。この最大値、最小値はテンプレート毎に求めておく。

さらに各テンプレートを用いて求めた、Userの表情に対する予測誤差の3つの状態(予測誤差が最大値を上回

たら High, 最大値と最小値の範囲内なら Middle, 最小値を下回ったら Low とする)をシステムが判断する。

2.4 感情認識

感情認識は各テンプレートを用いて求めた表情状態から行う。認識する感情は基本的に Low 状態の感情である。しかし、Low の状態が存在しない場合でも表情が変化していないわけではない。よって、Low が存在しない場合でも各表情状態の関係性から感情を認識する。

3. 実験

本手法を用いて感情認識を行うため予備実験として、表情認識を試みた。

3.1 データ取得環境

この実験では、被験者が他者と会話をしているところを 2 分間撮影し、自然な表情の認識の可否の確認を行った。

撮影対象は被験者 A の正面の顔である。また、室内で撮影し、外部からの光は遮断しており室内灯のみとなっている。カメラは IEEE-1394b 高速デジタルカメラ Dragonfly Express[®]を使用した。

3.2 テンプレート作成

実験時と同じく被験者の自然な表情を撮影するために、被験者と他者が会話をしているところを数分間撮影した。被験者はそのときの動画から表情変化を観察し、表情が変化したと思った区間を指定してもらった。さらにその指定された区間から遷移行列を作成し、その表情に基本 6 表情のいずれかのラベルを付けてもらった。作成したテンプレートは喜び 1 つ、怒り 1 つ、驚き 1 つ、嫌悪 1 つである。

3.3 結果

被験者が認識した表情の区間ではシステムはほぼ 100% 被験者が認識した表情を認識することができた。しかし被験者が認識した表情以外の表情もシステムは認識していることが多い。被験者の表情認識区間内のシステムの表情認識結果の割合を表 1 に示す。システムは、被験者の認識表情とは別の表情も同時に認識していることが多いが、被験者が表情認識した区間でのシステムの表情認識結果の割合は、被験者とシステムとで表情が一致しているものがすべて高くなっている。また喜び、嫌悪、怒りの表情はシステムが同時期に認識することが多かった。しかし同時期でもシステムは若干表情を認識するタイミングが異なっている。

3.4 考察

被験者が表情認識した区間を一瞬一瞬で見れば、システムは被験者が認識した表情とは別の表情も認識しており、一致率が低いように思われる。しかし、被験者が表情認識した一定の区間で見れば、システムは被験者の認識した表情を一番高い割合で認識できていることがわかる。

また被験者の表情変化に対して、システムは同区間で異なる表情を認識したものが多くみられた。これは表情変化時のある段階での表情は、異なる表情を示すものでも類似していることを示している。

以上のことから、表情を認識するには静的要素だけでは難しく、動的要素を視野に入れてある程度のスパンで認識を行う必要があることが考えられる。よって現在のシステムの表情認識結果から、さらに表情を特定するために表情認識を行うタイミングを検討する必要がある。

被験者/システム	フレーム	被験者の表情認識区間内のシステムの表情認識結果の割合			
		喜び	怒り	驚き	嫌悪
喜び	60-158	1	0	0	0.825
喜び	259-288	1	0.068966	0	0.068966
嫌悪	486-495	1	1	0	1
怒り	498-540	0.071429	1	0	0.071429
嫌悪	598-650	0.057692	0.057692	0	1
喜び	695-749	1	0.09434	0	0.09434
喜び	787-863	0.973884	0	0.302632	0
驚き	975-1008	1	0	0	0
嫌悪	1120-1140	0.038492	0	1	1

表 1: 被験者の表情認識区間内のシステムの表情認識結果の割合

4. まとめ

本研究では、人が他者とコミュニケーションを行う際に表出するごく自然な表情から感情を認識することを目的として、表情動作に着目した感情認識手法を提案した。

本研究では、表情動作が線形システムで表現できると仮定した上で、線形システム同定法に基づき、感情を表出したときの表情動作をテンプレートとして遷移行列で定義した。さらにテンプレートを感情毎に用意し、それらのテンプレートを用いた表情予測を行うことで表情動作から感情を認識することを試みている。

今回は、提案した手法により自然な表情からの感情認識を行う前段階として自然な表情の認識を試みた。結果として表情認識に関しては高い精度が得られた。

今後は表情認識の次のステップである感情認識を実現するために、感情を含む表情動作と感情を含まない表情動作とで線形システム同定法に基づいた遷移行列同士の比較を行い、感情を含む表情動作特有の動作の抽出を行う必要がある。

そしてここでは自然な表情の認識を目的として表情認識実験を行ったが、本論文では自然な表情からの感情認識を試みることで本手法による感情認識の有用性を示す。

参考文献

- (1) Paul Ekman, Wallace V. Friesen, 工藤力 訳: 表情分析入門, 誠信書房(1987)
- (2) 下田 隆, 他 2 名: 動的顔画像からのリアルタイム感情認識システムの試作, ヒューマンインタフェース学会論文誌 Vol.1, No.2, pp.25-32 (1999)
- (3) 叶冠峰, 他 2 名: 顔動画像からの特徴点抽出を用いた表情認識, 岐阜大学大学院工学研究科 (2005)
- (4) 山平慎吾, 他 2 名: 表情譜を用いた自然な表情の認識手法, 岩手県立大学ソフトウェア情報学部(2006)
- (5) 内田英子, 他 2 名: 高速カメラを用いた顔面表情の動的変化に関する分析, (株)ATR 人間情報通信研究会, (株)ATR 知能映像通信研究会(2000)
- (6) 西山正統, 他 2 名: 表情譜・顔パーツのタイミング構造の記述とその自動獲得, 京都大学大学院情報科学研究科 (2005)
- (7) Coates, T. F. et al: Active Appearance Model, European Conference on Computer Vision 1998, Vol.2, pp.489-498, 1998
- (8) 高速デジタルカメラ Dragonfly Express:
www.j-imaging.com/product.php?product_id20283

感情表現を意図した音声合成システムの提案

An Idea of a Speech Synthesis System with Emotion Expression

インテリジェントソフトウェアシステム学講座 0312004115 中里央
指導教員：橋松理樹 羽倉淳 藤田ハミド

1. はじめに

人との親和性が高いユーザインタフェースを構築する一つの方法として感情情報処理を組み込むことが試みられている。そのような感情情報処理の一つとして、音声による感情表現がある。これは表現したい感情によって音声を変化させることで、より人に近いイメージをユーザに喚起させることが目的である。しかし、音声による感情表現手法が確立していない点、音声と感情との関係が不明瞭である点などから、従来の音声合成研究¹⁾においては感情表現まで意図しているものは少ない。より質の高い音声合成を実現するためにも、この問題の解決が求められている。

以上のような問題に対し、本稿では、人の感情毎の音声特徴と合成音声の音声特徴とを対応づけ、それを用いることで感情表現を意図した音声合成を行うシステムを提案する。従来手法²⁾では、事前に感情と音声特徴との対応づけを行うものが多い。しかし音声特徴は複雑であり、また適切な理論もないため、事前に適切な対応づけを構築することは困難である。これに対し、本手法は合成する音声の特徴を人の発話の特徴に近づけることでその労力の軽減を図る。

2. 感情表現を意図した音声合成システム

2.1. 概要

本稿で提案する音声合成システムは、①感情・音声合成パラメータ対応付け作成フェーズ、②音声合成フェーズから構成される。感情・音声合成パラメータ対応付け作成フェーズにおいては、人の発話の解析結果に基づき、感情毎の音声合成パラメータを決定する。ここで音声合成パラメータとは、音声合成を行う際に設定する値であり、音量や速度、音程などが該当する。音声合成フェーズにおいては、表現したい感情に対し、①のフェーズで作成した対応付けで決定した音声合成パラメータを用いて音声合成を行う。これにより、人の音声による感情表現の再現を試みる。

2.2. 感情・音声合成パラメータ対応付け作成フェーズ

本フェーズでは、初めに、人の感情毎の発話データから音声特徴を抽出する。発話データは、指定した感情を人が意図的に表現した発話音声であり、それらを音声解析した結果得る特徴量が音声特徴となる。音声特徴としては、ピッチの平均や分散、パワーの平均や分散などが挙げられる。次に抽出した各特徴に対し、各値を最大値で割る

ことで正規化 (0 から 1 の値に変換) する。正規化を行うのは、人と合成音声との特徴の値域をそろえるためである。次に特徴量を 20 個の領域に均等分割し、度数分布を求め、感情毎に最頻度数領域を特徴量に対する各感情の特徴領域とする。なお該当する領域が複数ある場合は、すべて特徴領域とする。また複数の感情が同時に表現される場合 (以後、複合感情と表記) に対しては、対象となる感情すべての度数を用いて特徴領域を決定する。次に音声合成に対し、可能なパラメータの組合せを用いた合成音声に対し、発話データと同じ方法で音声特徴の抽出・正規化を行う。最後に、正規化して得られた合成音声の各特徴と先に作成した感情毎の特徴領域を比較する。結果、もっとも多くの特徴領域が重なった感情を、その音声合成パラメータに対応づける。該当する感情が複数あった場合は、すべて対応づける。

2.3. 音声合成フェーズ

音声合成フェーズにおいては、対応付け作成フェーズにおいて作成した対応付けに従い、表現したい感情に対応付けられた音声合成パラメータを選択する。このとき、表現したい感情に対して、複数の音声合成パラメータが存在する場合は、ランダムで一つを選択する。また表現したい感情に対する音声合成パラメータが無い場合は、デフォルトのパラメータを用いる。最後に選択したパラメータに基づき音声合成を行う。

3. 音声合成システムの実装

2章で述べた提案手法に対し、実装を行った。本章では、実装方法について説明する。

3.1. 感情・音声合成パラメータ対応付け作成フェーズ

発話データとしては、人が文章を読み上げた音声で 0.3 秒以上の無音区間で区切ったものを用いる。また発話に与える感情については、Ekman³⁾が提言している 6 表情をベースとした 6 感情 (幸福、怒り、悲しみ、嫌悪、恐れ、驚き) および中立 (前述のもの以外) を用い、発話者自身が読み上げた文章に対し、これらの感情を付与している。このとき、複数の感情をつけることを許すほか、つける範囲は任意の区間とする。各発話に対し、発話内容と感情を付与した文章と比較することで、その発話の感情を決定する。発話内容中に異なるパターンの感情が付与されている場合は、もっとも長い区間に適用するパターンの感情を発話の感情とする。

対応付け生成に用いるために音声特徴としては、ピッチ・パワーそれぞれの最大値、平均値、分散および発話速度（発話内のモーラ数を発話時間で割った値、ただし促音の無音区間は発話時間とみなす）を用いる。音声特徴抽出には京都大学音声メディア研究室の開発したピッチ抽出プログラム⁴⁾を用いた。

音声合成に対しては、沖電気工業株式会社開発の SMARTTALK⁵⁾（以後 ST と略記）を用いた。今回利用するパラメータとしては、今回着目している音声特徴に影響を与えるものとして、①抑揚、②高さ、③音量、④声質、⑤速さ、⑥話者の6つを選択した。

3.2. 音声合成フェーズ

音声合成フェーズでは、3.1 章で述べた ST を用いて音声合成を行う。読み上げる文章には、表現したい感情が事前に与えられている。表現する感情の組合せに対し、対応付けられたパラメータ設定を取り出す。なお複数候補がある場合は、ランダムで選択する。

4. 評価実験

本手法の有用性を評価するため、構築したシステムを用いた実験を行った。実験では、合成音声にシステムが表現した感情と、被験者である聴者が推定した感情との比較を行う。感情としては、前述の6感情の組合せを用い、また被験者も推定したすべての感情を答えさせた。結果に対し、完全一致および部分一致（一つでも一致した感情がある）の二つの観点から比較し、本手法の有用性を検証する。

4.1. 実験概要

対応付け作成の発話データとしては、読み聞かせの経験が豊富な男性1名が、児童文学一作品を読み聞かせた音声を0.3秒以上の無音区間で区切ったものを用い、また感情については、発話者本人が付与し、3.1章の方法で決定した。

評価実験においては、20代男性8名を被験者とした。合成した音声データは、次の通りである。感情としては、①嫌悪、②恐れ、③幸福、④驚き、⑤悲しみ、⑥怒り、⑦嫌悪と怒り、⑧幸福と驚き、⑨中立の9パターンを対象とした。実験では、最初に中立の読み上げを聞かせた後、前述の順番で3回読み上げを行った。この時、言語によるバイアスを考慮し、日本語として意味を成さない文を読み上げた。なお文はすべて異なっている。またパラメータセットについても同一感情であっても毎回異なっている。

4.2. 実験結果

実験結果の一部として、部分一致率の結果を表1に示す。において、表現する感情の①から⑨は4.1章で示した感情である。全体として、一致率は

低いものであった。その中で④驚きは1回目が高さ0.63(8名中5名)、3回目が1と高い結果となっている。また感情によって一致率にばらつきが生じている。

表1: 部分一致率 (0.6 以上は網掛け)

回	表現する感情								
	①	②	③	④	⑤	⑥	⑦	⑧	⑨
1	0.25	0.53	0.5	0.63	0	0.63	0.25	0.38	0
2	0.25	0	0	0.13	0	0	0.25	0.38	0.13
3	0.13	0.25	0	1	0	0.13	0.63	0.13	0.38

4.3. 考察

全体として一致率が低い、また同じ感情でも一致率が異なることから、感情とパラメータセットの対応付けが不十分であると考えられる。この要因としては、感情とパラメータセットとの対応付けの際に、音声特徴間の関係性を十分に考慮していない点、強制的に対応づけている点が挙げられる。例えば、比較的一致率の高い「驚き」における回答とパラメータセットの関係に着目した結果、直前の発話からスピードが上がった場合、または抑揚が大きくなった場合、前者は10回中7回、後者は10回中5回、両方では10回中2回、回答している。このことから被験者は直前の発話の変化に着目していると考えられる。この点について、現在の手法は考慮できていない。この結果からも考えられるように、感情とパラメータセットとの対応付けについて、着目する音声特徴の検証も加えた改善を行う必要がある。

5. おわりに

本稿では、人の感情毎の音声特徴と合成音声の音声特徴とを対応づけ、それを用いることで感情表現を意図した音声合成を行うシステムを提案した。評価実験の結果、感情ごとに一致率は異なり、全体として一致率は低いものであった。今後は実験結果の解析を進め、そこで得られた知見を元に、人の発話の音声特徴と合成音声の音声特徴との高い対応付けを実現するように、手法を改善する必要がある。

参考文献

- 1) D.Gibson, I.Mertins and R.K.Moore(Eds), "HANDBOOK OF MULTIMODAL AND SPOKEN DIALOGUE SYSTEMS", Kluwer, USA, 2000
- 2) P. Y. Oudeyer, "The production and recognition of emotions in speech: features and algorithms", International Journal of Human Computer Interaction, Vol.19(1-2), pp.157-183, 2003
- 3) P. Ekman, W.V.Friesen, "表情分析入門—表情に隠された意味をさぐる", 工藤力(訳・編), 誠信書房, 東京, 1987
- 4) 京都大学音声メディア研究室, ピッチ抽出プログラム, <http://vision.kuee.kyoto-u.ac.jp/lecture/dsp>
- 5) SMARTTALK, <http://www.oki.com/jp/Cng/Softsw/11S/sm.html>

FACS を用いた上位感情抽出式表情合成手法

A Facial Expression Synthesis for Selected Emotions Based on Facial Action Coding System

インテリジェントソフトウェアシステム学講座 0312004125 檜山洋一
指導教員：羽倉淳 博松理樹 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、現在「宮沢賢治プロジェクト」¹⁾が進行中である。これはコンピュータ上で宮沢賢治の人格・感性を仮想的に再現し、ユーザとの実時間でのコミュニケーションを図る試みである。これを実現すべく、システム側の感情表現手段の一つとして、三次元顔モデル上での高度な表情出力が必要となる。本研究は、同システムの表情出力手法を可るものである。

表情合成の代表的手法としては、Ekmanの提唱した表情記述法FACS (Facial Action Coding System) ²⁾で定義されている、表情動作の最小単位AU (Action Unit) を顔モデル上で組み合わせることで表情を作り出す手法³⁾が挙げられる。これは予め決められた動作のパーツを繰り返し使用する方式であるため、計算量が少なく、誰の顔であっても同等の変化を得られるという利点がある。その一方で、AUの組み合わせによってはどんな感情を表しているのか読み取りにくくなる、解剖学的に取り得ない動きの組み合わせとなる、といった複数感情の合成に対する障害も存在する。

そこで本研究では、従来手法をベースとしながらも、これらの問題点を解決するための新手法を提案する。まず、どの感情を表しているのかを明確なものとするため、各 AU を使用段階ごとに分類し、感情と関連付け、出力に用いる AU を決定する。さらに、解剖学的に起こり得ない組み合わせを極力排除するため、複数の感情を表現する際に、強く表現したい感情を複数の AU (顔全体) で表現し、その中に弱く表現されるべき感情として単一の AU (一部分) を加える方式を取る。これらの工夫により、従来法の利点を活かしつつ、複数の感情をより自然に表現する事を試みる。

2. 上位感情抽出式表情合成手法

提案手法の外観を図 1 に示す。ここでは、入力された感情パラメータ内で一番強い感情をメイン感情、2 番目に強い感情をサブ感情と呼び、表情出力の対象とする。この二つを抽出した後、顔全体でメイン感情を、一部分でサブ感情を出力する。以下に本手法の詳細を述べる。

2.1. 感情パラメータ入力

ここでは、システムが表現すべき感情が

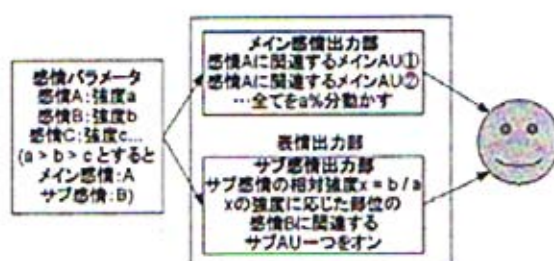


図 1 表情出力の全体像

Ekman の基本 6 感情 (怒り・嫌悪・恐怖・幸福・驚き・悲しみ)、それぞれの強度で構成される感情パラメータで与えられると仮定する。このパラメータを受け取った後、強度順で上位 2 つを抽出し、メイン感情・サブ感情として表出対象とする。このように 2 つに限定する理由は、表現する感情が増えるほど使用する AU も増加し、従来手法のデメリットが発生しやすくなるからである。

2.2. メイン感情出力部

表現する 2 つの感情を選択した後、まずメイン感情に対応する表情を出力する。顔全体で表現するのがメイン感情の特徴であるが、6 感情の単一表情を作成するのであれば、AU 全 60 個のうち、顔への影響が特に強い 17 個のみで再現可能⁴⁾と言う研究結果が出ている。よってメイン感情ではこの 17 個のみを用いるものとし、ここでは、これらの AU をメイン AU と呼ぶ。メイン AU はメイン感情の強度値に合わせて動作量を調節できるものとする。これにより、顔全体での表現されるべきメイン感情の強度を表現する事ができる。

メイン感情と関連付けられた全メイン AU を、強度値の割合だけ動作させ、メイン感情の出力を終える。

2.3. サブ感情出力部

次にサブ感情の出力である。残り 43 個の AU はメイン AU よりも表情に与える影響が小さいと仮定し、これらをサブ AU と呼ぶ。サブ AU では基本的に動作量の調節は行わず、オン・オフの切り替えのみを行う。これは、サブ AU に動作量を設定した場合、メイン感情が極端に低い値の際に、サブ感情はさらに弱くなり、サブ AU の動作量もわずかとなって両感情が明確に表現されない可能性があるためである。

実際の出力の際は、出力部位決定のため、サ

ブ感情（部分）のメイン感情（全体）に対する相対強度を計算する。出力部位は顔の三領域（眉・目・口）に分けられ、影響力が強い順に眉、口、目である。ただし眉は影響力が特に強い上、眉の AU が全てメイン AU に登録されているため、目と口のみとする。よって相対強度が強い場合は口の、弱い場合は目のサブ AU でサブ感情と関連付けられているものを一つだけ出力する。

3. 作成及び合成実験

上記のアルゴリズムをシステムとして実装し、本手法の有用性を検証した。実装環境は Autodesk 社の 3D レンダリングソフト「Maya 8.5」上であり、AU は同ソフトの「Blend Shape」機能を、アルゴリズムの各モジュールは同ソフトのスクリプト「MEL」を利用した。

各 AU は図 2 に示されるように、顔モデル上に存在する格子状の頂点をエリア毎に 3 次元方向で制御・動作させる方式で作成する。

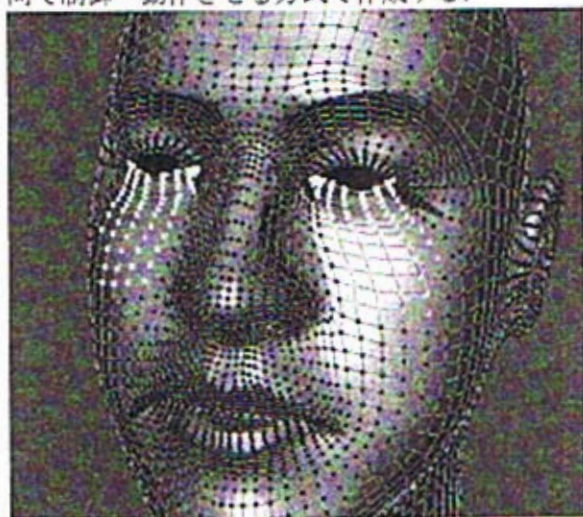


図 2 AU6（頬を上げる）の動作頂点エリア

各 AU と感情の関連付けは、Ekman の表情分析入門などを基に行った。図 3、図 4 に実装したシステムが合成した各種の表情を示す。



図 3 左：嫌悪 50% 右：嫌悪 100%

図 3 はメイン感情に強度差がある場合の比較

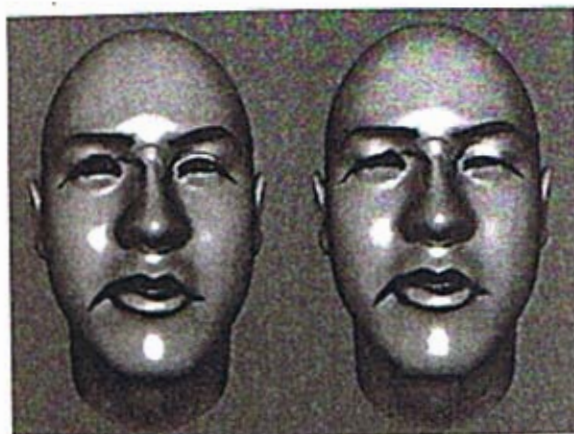


図 4 左：嫌悪 右：嫌悪+悲しみ

である。当初の想定通り、メイン感情しか存在しない場合においても、強度によって各部の動作量が変化できる事を確認した。

対して図 4 は嫌悪の表情に悲しみの AU41「顔を力なく下げる」を重ねた場合との比較である。全体で嫌悪を表現しながら、同時に顔を下げて悲しみを表現している。AU 同士の干渉も無く、人間が十分とり得る表情を合成できる事が確認された。

4. おわりに

本研究では複数感情表現手法として、表現すべき感情のうち、強度の大きな上位 2 感情を抽出し、それぞれに対応する顔の部位を全体変化・部分変化させる表情合成手法を提案し、その成果を示した。

賢治プロジェクトの一部として稼働させる際、システムの反応を自然なものとするためにも、表情変化はアニメーションで行うのが理想的である。実時間で変化する感情パラメータに逐次対応し、なおかつ処理量の大きいアニメーションを実現する上で、計算量を抑えつつ複数感情の表現にも対応でき、より自然な表情を合成できる本手法は極めて有用であるといえる。

今回はメイン感情とサブ感情で使う AU を厳格に分けたが、さらなるバリエーションの増加のため、今後は両者の AU 区分が薄れる可能性もある。

参考文献

- 1) H. Fujita, J. Hakura, M. Kuromatsu, 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) 竹田智史, 岡田守, “表情豊かな手話アニメーションの生成”, 高知工科大学学士論文 (2001).
- 4) 坂口竜己, 山田寛, 森島繁生, “顔画像モデルを基にした 3 次元感情モデルの構築とその評価”, 電子情報通信学会論文誌 A, vol.J80-A, no.8, pp1279-1284 (1997).

表情MAPによる感情推定手法 An Emotion Recognition Method with Facial Expression MAP

インテリジェントソフトウェアシステム学講座 0312004130 藤田祈
指導教員：羽倉洋 樽松理樹 藤田ハミド

1. はじめに

近年のコンピュータシステムの急速な普及により老若男女、多様なユーザ層がコンピュータを利用している。一方で、そのインターフェイスに着目すると、ユーザがその意図をコンピュータに的確に伝えるためには、多くのスキルが要求される。

この問題に対して、利用者の感情を推定し、利用者とコンピュータの間で、人間同士のコミュニケーションに近いインターフェイスを実現しようとする試みが行われてきている(例^{1),2)}。著者らのグループでも、このようなインターフェイスを持つシステムの究極の実現例の一つとして「仮想宮澤賢治」(賢治システム)の実現を試みてきている³⁾。そこでの重要な課題のひとつとして、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。賢治システムは、博物館といった不特定多数の利用者が来訪する場所に設置することを目的としている。従って、感情推定ではこれらの不特定の利用者がシステムに対面した際に示す自然な表情から感情をシステムが自動的に推定する必要がある。さらに、感情の大きさを推定し、複数の感情が入り交じった表情からそれらの感情を推定することで、利用者の反応に対してよりの確に反応できることが求められる。

そこで、本研究ではこれらの要求を満たすため感情推定手法の構築を試みる。表情から感情を自動推定する手法は多く提案されてきている⁴⁾が、第2節で概述するように、これらの多くはFacial Action Coding System (FACS)に基づいており、上記の要求の全てを満たすには至っていない。そこでここでは、表情を端的に表わす特徴点を定義し、これらの特徴点の位置と表出された表情とを多数の表情データから確率的に関係付けた表情MAPの構築と、表情MAPに基づく感情推定手法を提案する。以下、第2節では、従来のコンピュータによる表情分析・認識、感情推定手法を概観し、そこでの問題点を議論する。第3節では、表情MAPに基づく感情推定手法を詳述する。第4節では、評価方法について示し、最後に今後の可能性に言及する。

2. 従来の感情推定手法

表情から感情を自動推定する手法の代表的な研究として、FACSに基づく手法⁵⁾がある。FACSは表情をAU (Action Unit) と呼ばれる、表情を構成する顔の部位の最小パーツによってコーディングする手法である。これにより、人間の表情が記号で表現できることから、表情から感情を自動推定する多く試みにおいて利用されてきている。

これらの研究では、特定の演技された表情の認識においては優れた結果を示しているが、感情の強弱が認識で

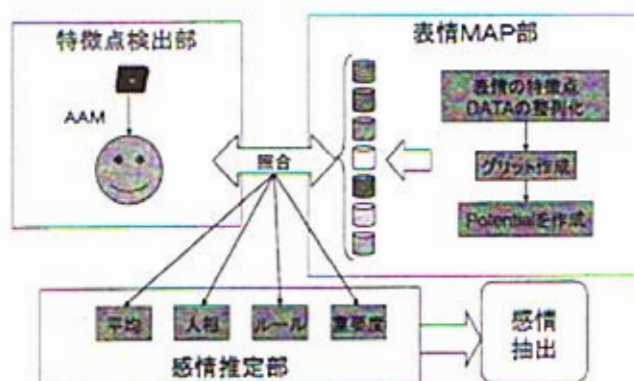


図1 システム概観

きない場合が多く、また、自然な表情から感情を推定することは困難である。さらに、利用者の中立の表情が既知であることを前提としているといった問題もある。

そこで、ここでは特定の感情を表す表情においては、大部分の人間が同様の位置に、目、口といった部位を移動させるというFACSと同様の考え方をを用いた上で、それらの部位の位置から確率的に感情を推定する手法を提案する。以下、その詳細を示す。

3. 感情推定手法

ここでは、表情を端的に表わす特徴点を定義し、これらの特徴点の位置と表出された表情とを多数の表情データから確率的に関係付けた表情MAPを作成し感情推定を行う手法を提案する。

図1に示すように、カメラから人の顔を検出し、構築した表情MAPに検出された特徴点を照合する。表情MAPには、MAP上の座標に対して、その位置に特徴点があった場合の特定の感情である確率が示されている。以下、この確率を感情Potentialと呼ぶ。したがって、ユーザの全ての特徴点を表情MAPと照合させ、その結果を感情推定部で提案する4手法のいずれかに出力することで表情に表れた6感情の強度を推定する。以下、各部について詳述する。

3.1 特徴点検出部

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。ここでは、特徴点の数は経験的に、左右の眉12点、左右の目が12点、口が7点の計31点とした。各特徴点は、IDを持っており、特徴点ごとに表情MAPが構築される。MAP適用時には、このIDを基に、各特徴点に対するMAPと照合することになる。

3.2 表情MAP部

特定の感情を表す表情に対して、上述の3.1の特徴点ごとの位置にあることが多いかを多数の被験者による表情データから確率的に表現し、表情Potentialとして表現する。

本小節では、これを特徴点ごとに表現した表情 MAP に関して、特にその作成方法について詳述する。

3.2.1 表情 MAP 作成データの特徴点の整列化

表情 MAP は、基本 6 感情 + 中立に対してそれぞれの特徴点ごとに構築される。この構築には、多数の被験者の該当する表情に対して、表情表出時の各特徴点の位置が同じ顔の位置に存在するように整列する必要がある。すなわち、人、または、撮影条件によって異なる特徴点の位置を全ての人に対して概ね同じ位置になるよう再配置する必要がある。ここでは、このための手法として Procrustes Analysis⁵⁾を用いる。その概要を図2に示す。

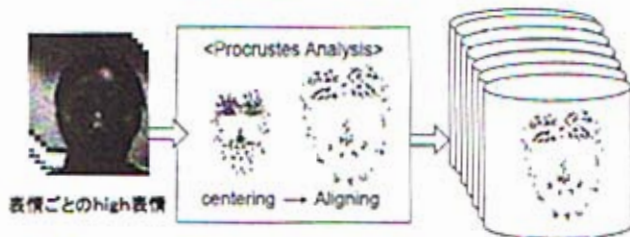


図2

Procrustes Analysis による特徴点の整列化と重ね合わせ

3.2.2. グリッド作成手法

整列化された特徴点データは表情 MAP 上の一点である。この点に対して、感情 Potential を張ることも可能であるが、照合の高速化や MAP のデータ量の軽減を考慮し、ここでは、各特徴点の MAP を 2 次元のグリッド空間として表現する。各 MAP 作成に用いられる特徴点は一般に特定の領域内を構成する。この領域内を各グリッド大きさが同一になるよう分割するものとする。したがって、グリッド空間のグリッド数も大きさも特徴点、感情ごとに異なることとなる。ここでは、ID を k とした場合、以下のように X 方向のグリッド数 M と Y 方向のグリッド数 N を決定する：

$$M = (\max(x_k) - \min(x_k)) / \text{width}$$

$$N = (\max(y_k) - \min(y_k)) / \text{width}$$

ここで、 x_k 、 y_k は ID が k の特徴点の MAP 作成時のデータ集合の要素である。

3.2.3. 感情 potential 作成手法

上記のグリッド空間の各グリッドに対して、感情 potential を生成する。ここでは、各特徴点・感情に対応する MAP 上に MAP 作成用のデータを配置した際の特徴点の密度に着目して感情ポテンシャルを生成するものとする。すなわち、各グリッドの中心から一定の半径内に存在する MAP 作成用のデータ数を全データ数で除した値をそのグリッドのポテンシャルとする。したがって、MAP 作成時に多くの特徴点が重複した位置における感情 potential は高く、まばらな位置では低くなる。

3.3. 感情推定部手法

上述の表情 MAP を用いた感情推定では、まず、カメラから得られた画像に対して、Active Appearance Model⁶⁾を適用し、顔と特徴点の自動抽出を行う。抽出された各特徴点を ID 毎に各感情の MAP と照合し、そこでの感情 potential を読み取ることで、特徴点ごとにその位置が特

定の感情を表す可能性を数値として獲得する。従って、特徴点ごとに各感情を表す値が得られることとなる。得られた特徴点数分の数値を用いて、顔全体が特定の感情をどれ位の可能性で示しているかを決定するための手法として、ここでは、全ての特徴点の平均を用いる手法、人相学に基づく分類による手法、感情ごとに各 ID に重み付けしその線形和を用いる手法、そして、経験的ルールに基づく手法の4つの手法を検討中である。

4. 評価方法

本手法によって、感情の強度、複合感情、及び、不特定多数の人物の感情が推定精度を測るために、以下のような実験を計画している。すなわち、MAP 作成時にデータ提供を受けていない被験者に対して、映画などの映像を提示し、シーンごとに感じた感情を書き記させる。一方で、その際の表情から提案手法を用いて感情推定をおこない、その両者の整合性を検証する。

5. おわりに

本稿では、不特定の利用者がシステムに対面した際に示す自然な表情から感情を推定でき、感情の大きさを推定し、複数の感情が入り交じった表情からそれらの感情を推定する表情からの感情推定手法の構築を目指し、表情 MAP の導入と、それに基づく感情推定手法について述べた。

今後は、これらの実験を行い表情から表出された感情ごとのデータから感情推定部を行い、その後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) 岡田謙一他, ヒューマンコンピュータインタラクション, オーム社, (2002).
- 3) B. Fasel, and J. Luetten, Automatic Facial Expression Analysis: A Survey, Pattern Recognition, Vol. 36, No.1, pp.259-275, (2003).
- 4) Y. Tian, T. Kanade, and J. F. Cohn, Recognizing Action Units for Facial Expression Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, (2001).
- 5) Any Ross, et al : Procrustes Analysis: Department of Computer Science and Engineering University of South Carolina, SC 29208
- 6) Cootes, T. F. et al : Active Appearance Model: European Conference on Computer Vision 1998, Vol. 2, pp. 489-498, 1998