博士学位論文

ソフトコンピューティングによる
マルチモーダル感情判別に関する研究

平成 23 年 3 月

九州工業大学大学院生命体工学研究科

佐藤　芳紀

目次

図目次

表目次

第１章　序論

近年の機械技術の発展により、ユビキタスをキーワードに日常生活における機械の利用の場はますます拡大している。その結果、″機械が環境に融合し、人間の変化に適応して必要な情報を必要なときに提供し、快適かつ安全な環境を実現する″という、″Ambient Intelligence″が注目されるようになってきた[1-4]。すなわち、従来の機械は与えられた一定のタスクを忠実に実行することが重要な課題であったが、これからは機械が人間と共存し、環境の変化に柔軟に対応することが望まれている。機械と人間が共存するには、機械が人間の内部状態を的確に判断し、自身が取るべきタスクを自律的に導き出し人間に働きかける必要がある。さらに、選択したタスクが状況に適していたのかを機械自身が判断するには、選択したタスクに対する人間のリアクションを取得する機能が不可欠となる。これらの必要要素を満足する方法の１つとして、機械が人間と同等の感性を備え持つことが考えられる。つまり、機械が人間との心理的なコミュニケーションを通じて自身の行動を決定することである。機械が人間との心理的なコミュニケーションを実現する知的インタフェースとして、機械による感情の判別が多く研究されている[5-6]。感情を取得するためのモダリティとして、表情[7-9]、ジェスチャー[10]、音声[11-13]、さらには脳波や心拍などの生体信号[14]が用いられている。しかし、単一のモダリティのみでは人間の表出する様々な感情を正確に捉えるには限界がある[15]。そこで、人間が五感を駆使しているように、複数のモダリティを同時に扱うマルチモーダル感情判別が注目されつつある[5-6, 16-19]。

人間の感情を分類する方法として、ニューラルネットワーク[20]やサポートベクタマシン[21]、隠れマルコフモデル[22]などの機械学習による方法が多く提案されており、70% を超える精度が報告されている[5]。しかし、機械学習による感情の分類には多量の感性データが必要であり、また結果として得られた知識は人間には理解し難く、追加学習も難しいという問題がある。一方、機械学習における問題を解消する方法の１つとして、ルールベースによる感情の分類がある。すなわち、人間が理解可能な″言葉″を用いるファジィ理論[24]によって感情を定義する。実環境における感情判別において、ファジィ理論のも

つ〝あいまいさ〟は実環境中の変動に対して高いロバスト性を示す[25]。また、処理時間の観点から、ファジィ推論による感情判別がリアルタイム処理に適している[26]。さらに、ファジィ推論は追加学習にも適している。そこで、本研究ではファジィ推論ベースのマルチモーダル感情判別について、各モダリティにおけるルールの構築手法および実用的な実装手法について提案する。第一の方法として、表情による感情判別について述べる。各感情によってバラエティに富む表情を分類するために、SOM を用いた表情筋の変化の量子化を提案する。さらに、SOM によって形成されたマップに対し分離度を定義し、感情の分類に必要な表情筋を評価し、言葉ベースでルール化する手法を提案する。第二の方法として、音声による感情判別について述べる。音声による感情判別では、音声中に含まれる韻律情報のみを抽出し、表出感情を判別する。韻律情報をコミュニケーションにおける１つのメッセージとしてとらえることで、言語に依存しないコミュニケーションを可能にする。本論文では、韻律情報による感情の分類のために、感情表出による韻律情報の変動分布を推定する方法として推計統計量を用いる手法を提案する。さらに、表情による感情判別結果と音声による感情判別結果を統合したマルチモーダル感情判別システムについて、拡張性およびリアルタイム性を考慮した実装について述べる。

本論文は 6 章からなり、構成は図 1.1 に示す通りである。第 1 章は序論であり、本研究の背景および位置づけについて述べている。第 2 章では提案手法の基本となる感情について、従来研究によってヒューマン・ユニバーサルであると認められている基本 6 感情について述べ、ルールの学習および評価のために収集した基本感情データについて述べる。第 3 章では、表情による感情判別について述べる。具体的には、SOM を用いて各表情に対する表情筋の変化をベクトル量子化し、分離度を定義して定量的にルールを構築する。第 4 章では、音声による感情判別について述べる。この手法では、感情音声中の韻律情報を推計統計学に基づいて分類する方法でルールを構築する。第 5 章では、表情による感情判別と音声による感情判別を統合したマルチモーダル感情判別について、効果的な統合手法について述べ、その実装方法を示す。少ないオーバヘッドでシングルモダリティの誤判別を互いに抑制し合う方法を提案し、その有効性について議論する。第 6 章は、本論文の結論であり、本提案手法の特長、有効性および今後の展望についてまとめている。

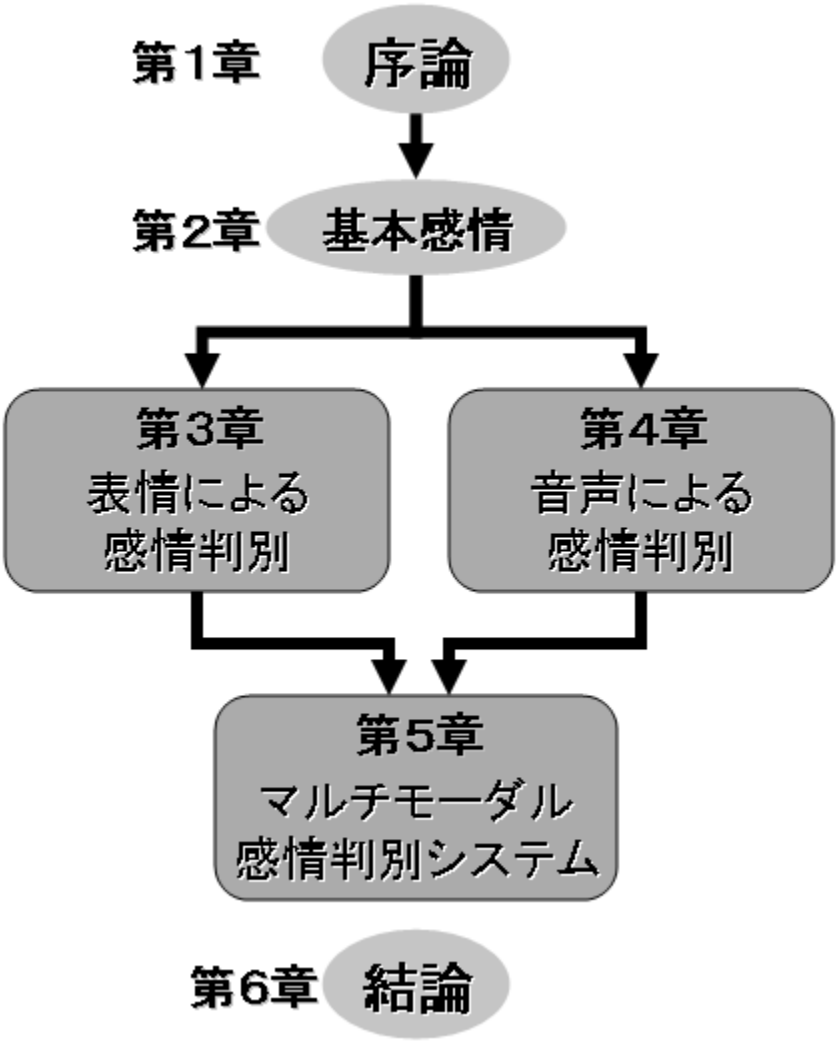

第1章
序論
第2章
基本感情
第3章
表情による
感情判別
第4章
音声による
感情判別
第5章
マルチモーダル
感情判別システム
第6章
結論

図 1.1. 本論文の構成

第 2 章　基本感情

2.1 緒言

生体は種の保存のために様々な進化を遂げてきた[27]。感情も生体が進化の過程で獲得してきたものであり、生体の生存のために適切な行動を促進する機能を担う[28]。本章では、生体の行動に基づく感情のモデルについて、これまでの知見について述べる。また、ヒューマン・ユニバーサルであるとされる普遍的感情について述べる。最後に、近年最も多く支持されている Ekman と Friesen の基本感情[29]に基づいて感情データを収集した手順について述べる。

2.2 感情のモデル

2.2.1 感情の次元

生体の持つ感情は多種多様であるが、根本は〝快―不快〟の 1 次元上に還元できるという考えが広く持たれている[30]。生体の生存にとって有益であるものに対しては接近行動を促進する〝快〟感情、生体の生存を脅かすものに対しては回避行動を促す〝不快〟感情が発生する。人間については、〝快―不快〟に加えてさらなる感情次元が存在するという考えがある。Wundt は心理学の研究法は自己観察にあるとし、図 2.1 に示すように内観法に基づいて〝快―不快〟に加えて〝緊張―弛緩〟および〝興奮―沈静〟の次元を追加した感情の 3 次元モデルを主張した[31]。また、Schlosberg は表情の分類実験を基に図 2.2 に示すように〝快―不快〟と〝注意―拒否〟を直行軸として感情カテゴリーを円環的に配列し、さらに感情の活性の次元として〝緊張―弛緩〟を加えた 3 次元モデルを提案した[32-33]。Schlosberg のモデルは色相関のアナロジーであり、隣り合った感情は混同されやすいが、対極の感情は補色の関係に似て誤判別が少ない。

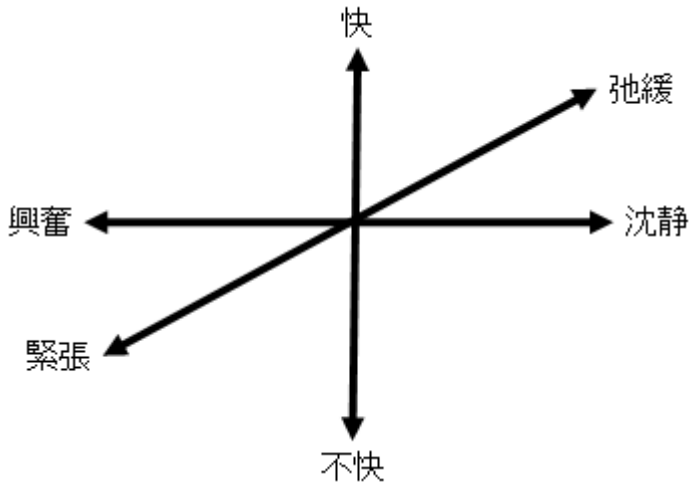


図 2.1. Wundt の感情の 3 次元モデル[31]

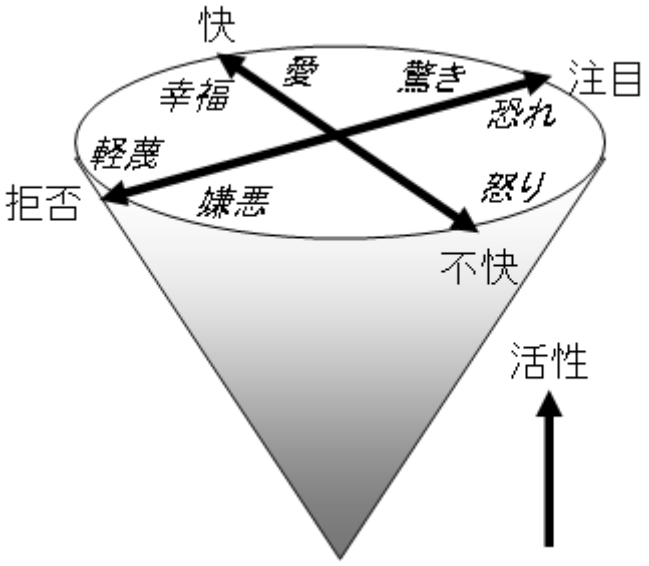


図 2.2. Schlosberg の感情モデル[32, 33]

2.2.2 普遍的感情の定義

異なる文化を持つ人間同士のコミュニケーションにおいて、共通の感情による非言語コミュニケーションは重要な役割を持つ[34]。ヒューマン・ユニバーサルな感情は異なる文化をもつ人間同士においても創造的協調を促進させる。ダーウィンは〝悲しみ、幸福、怒り、軽蔑、嫌悪、恐怖、驚き〟の基本的感情は文化によらず普遍的に同じ方法で表現されると示唆した[35]。また、Plutchikは表 2.1 に示す 8 基本行動が 8 つの基本的感情に対応すると主張した[28]。

表 2.1. Plutchik の 8 基本行動と対応する基本的感情

基本行動	対応する基本的感情
攻撃する	怒り
食べる	受容
所有する	喜び
停止する	驚き
逃げる	恐れ
排出する	嫌悪
失う	悲しみ
探索する	期待

その後、Ekman と Friesen は従来の感情モデルについてまとめ、表情認知に基づく比較文化的研究の結果から、〝怒り、喜び、驚き、悲しみ、嫌悪、恐怖〟の基本 6 感情がヒューマン･ユニバーサルであると結論した[29]。現在では、Ekman と Friesen の結論が最も多く支持されている。ただし、人間同士の感情の表出と感受の過程において、Shigeno は表出者の〝恐怖〟は〝驚き〟もしくは〝悲しみ〟として認識されることを確認している[15]。

2.3 基本感情データの収集

第 3 章以降で提案する感情判別システムの構築のために、ルールの学習および評価用として Ekman と Frisen の基本感情を基に感情データを収集した。ただし、Shigeno の実験結果より、本論文では ″恐怖″ は ″驚き″ もしくは ″悲しみ″ に含まれる感情として扱う。結果として、本論文では ″怒り(*Ang*)、喜び(*Hap*)、驚き(*Sur*)、悲しみ(*Sad*)、嫌悪(*Dis*)″ に加え、特に感情を含まない ″平静(*Nat*)″ の表情および音声を扱う。20 歳代の男性被験者 10 名に対し、表 3.2 に示した内容の感情データを収集した。日常での感情表現を想定しているので、被験者には感情表出について特別に訓練を受けていない人物を選んだ。収集したデータは各感情に対し 25 サンプルである。無発話状態の感情データを得るために、25 サンプル中 5 サンプルは発声せずに表情のみで感情を表出する。また、10 サンプルは各感情に共通の発話語として ″おはよう″、″こんにちは″、″こんばんは″ を発話する。4 サンプルは規定語の発話であり、各感情を表現し易いと思われる言葉を予め設定した。残りの 6 サンプルは自由語の発話であり、各被験者が最も感情を表現し易い発話語を用いて感情を表出する。

表 2.2. 収集した感情データの内容

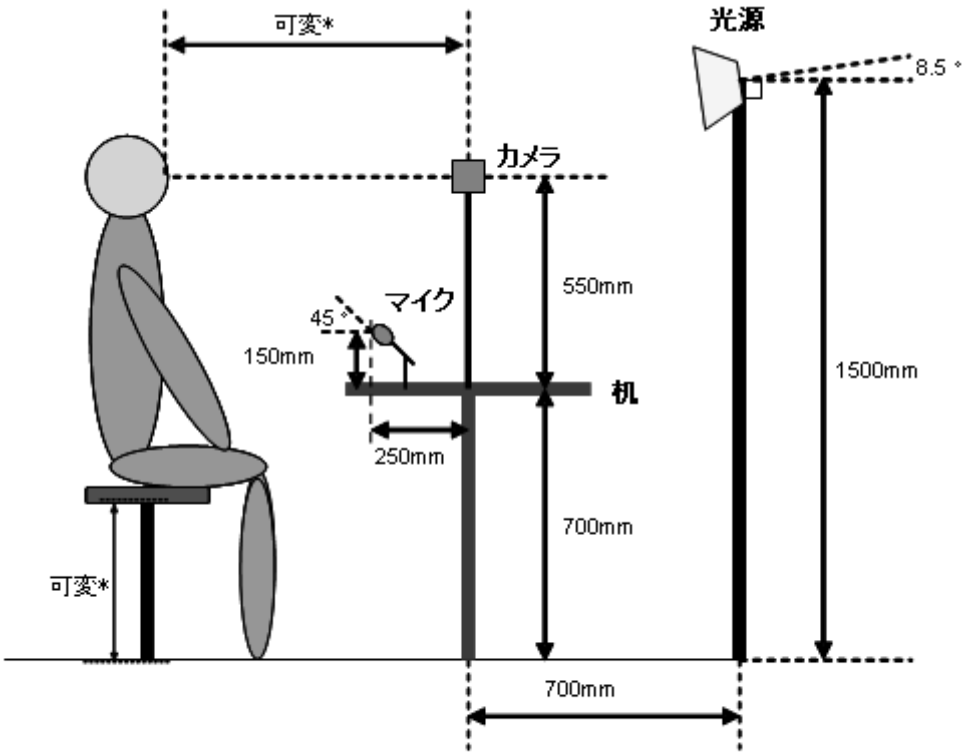
(a)″*平静*″、″*怒り*″, ″*喜び*″

No.	*平静*	*怒り*	*喜び*
0~4	(無発話)	(無発話)	(無発話)
5~7	"おはよう"	"おはよう"	"おはよう"
8~10	"こんにちは"	"こんにちは"	"こんにちは"
11~14	"こんばんは"	"こんばんは"	"こんばんは"
15~16	"神酒研究室"	"ふざけるなよ"	"やったあ"
17~18	"基本情報"	"何考えてるの"	"修論終わった"
19~24	(自由語)	(自由語)	(自由語)

(b)″*驚き*″、″*悲しみ*″、″*嫌悪*″

No.	*驚き*	*悲しみ*	*嫌悪*
0~4	(無発話)	(無発話)	(無発話)
5~7	"おはよう"	"おはよう"	"おはよう"
8~10	"こんにちは"	"こんにちは"	"こんにちは"
11~14	"こんばんは"	"こんばんは"	"こんばんは"
15~16	"ええっ"	"申し訳ありません"	"気持ち悪い"
17~18	"そうなんだ"	"つらそうに見えたよ"	"ありえない"
19~24	(自由語)	(自由語)	(自由語)

図 2.3 に感情データ取得環境を示す。本環境は、表情を取得するための Web カメラおよび感情音声を取得するためのマイク、安定した光量を確保するための光源から構成される。Web カメラは USB Video Device Class 規格に準拠しており、640x480 ピクセル 24 ビットカラー画像を 15fps で取得する。マイクは量子化ビット数 16 ビット、サンプリングレート 44.1KHz のモノラルマイクである。被験者はカメラおよびマイクの正面に置かれた椅子に座る。椅子からカメラまでの距離および椅子の高さは調整可能であり、顔全体がカメラに収まるように予め調整する。図 2.4 に感情データの収集手順を示す。最初に、椅子に座った状態の被験者の顔全体がカメラに収められるよう、椅子の位置および高さを調整する。また、録音レベルについて、平静時の長母音の音量が 20 dB となるように調整する。続いて、記録者は被験者に対して表出感情と表出音声を指示し、録音、録画を開始する。被験者は無表情で表出内容の指示を受け、約 3 秒の間隔を空けて指示された感情を表出する。感情の表出後、被験者は無表情状態に戻る。25 サンプルのデータを収集する毎に 5 分程度の休憩を設けた。



＊椅子からカメラまでの距離および椅子の高さは顔全体がカメラに収まるように調整する

図 2.3. 感情データ収集環境

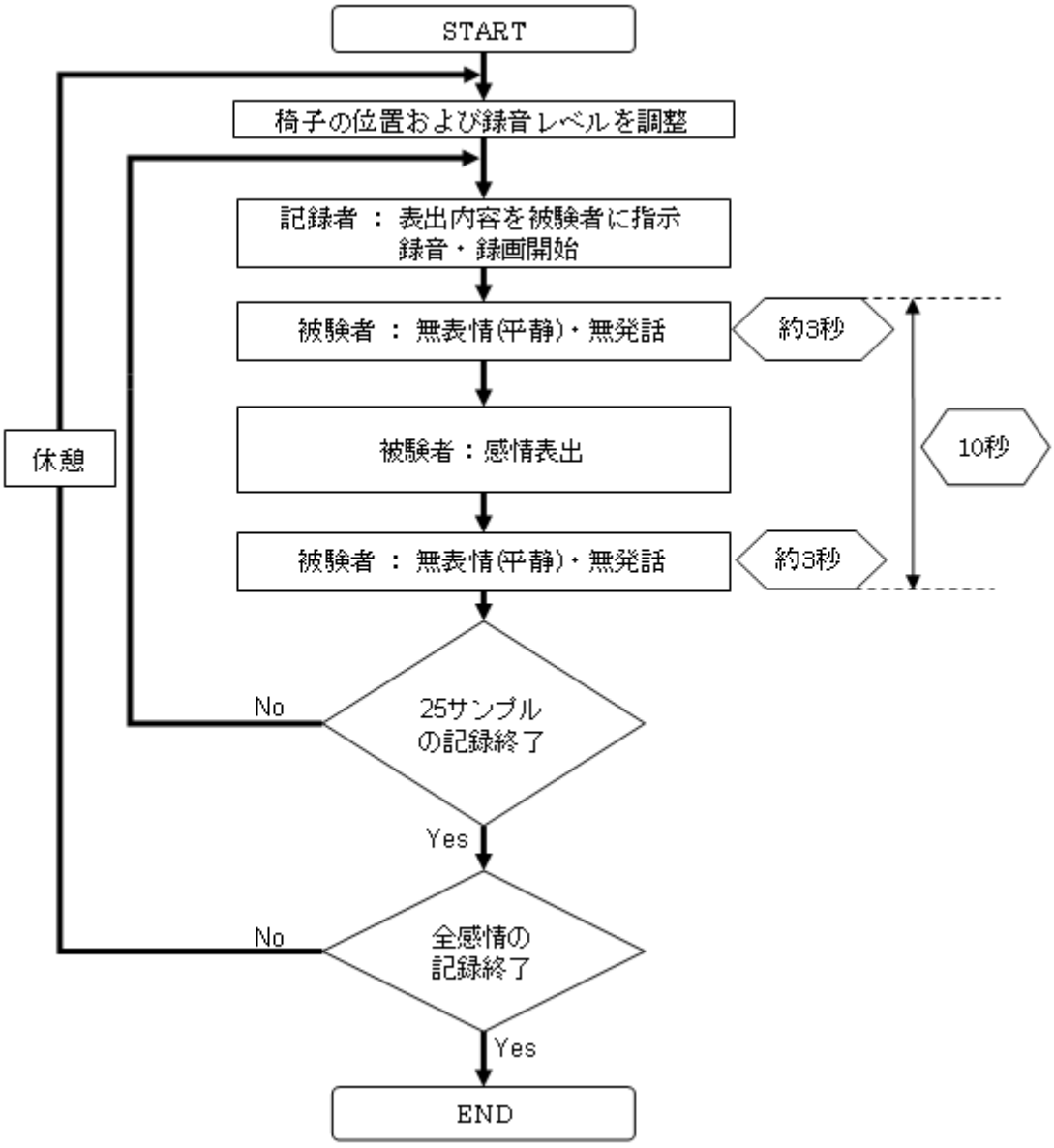


図 2.4. 感情データ収集手順

図 2.4 に示す手順で取得した感情データから、特に感情がよく表れているデータを選別するために、10 名の被験者による図 2.5 に示す感情判別実験を行った。記録者は記録済みの感情データから、表情のみのデータ、音声のみのデータ、表情と音声両方を含むデータをランダムに提示する。被験者は提示されたデータに含まれている感情をアンケート形式で回答する。10 名中 9 名以上の回答が一致したデータを感情がよく表れているデータとした。結果として表 2.3 に示す数のデータが感情がよく表れているデータとして選別された。感情判別システムのルールの構築および評価は、表 2.3 で示したデータを用いる。

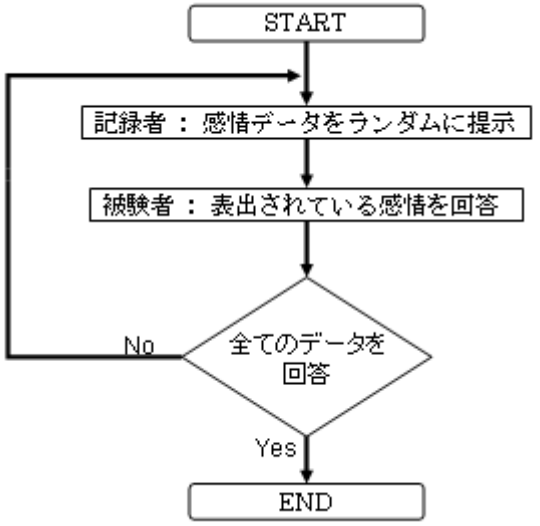


図 2.5. 感情データ選別手順

表 2.3. 感情がよく表れているデータのサンプル数

	平静	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
表情のみ	221	97	169	210	56	41
音声のみ	121	66	128	120	134	44
表情+音声	170	139	174	187	129	74

2.4 結言

本章では、感情のモデルについてのこれまでの知見を述べた。普遍的感情は非言語コミュニケーションにおいて重要な役割を持っており、Ekman と Friesen の基本 6 感情が現在最も多く支持されている。そこで、Ekman と Friesen の基本 6 感情を基に感情データを収集し、その中から特によく感情が表出されているデータを選別した。**第 3 章**以降では、選別したデータを用いて感情判別ルールを構築、評価する。

第 3 章　表情による感情判別

3.1 緒言

近年の機械技術の発展により、ユビキタスをキーワードに日常生活における機械の利用の場はますます拡大している[36-39]。人と機械が共存する Ambient Inteligence の時代では、機械自身が周囲環境を理解し、人と機械が協働してコミュニケーションをとりつつ、問題を解決しなければならない[1-4]。そのためには、周囲環境の認識とともに人間の置かれている状況を認知する機能が必要であり、感情の判別は重要な役割を持つ[40]。人間同士の対面的コミュニケーションにおいて、表情は感情を豊かに伝える[41-42]。感情を判別する手法として、ニューラルネットワークによる方法[43-46]、サポートベクタマシンによる方法[9]、隠れマルコフモデルによる方法[47-48]のような機械学習ベースのものが挙げられる。しかし、機械学習による感情の分類には一般的に以下の問題がある：

1) 機械学習によって得られたルールは人間には理解し難い
2) 学習には特徴がよく表れている多量のデータが必要である
3) 追加学習が難しい

一方、感情を言葉で分類する手法として、ルールベースによる手法がある。Ekman と Friesen は表情筋の変化によって形成されるあらゆる表情を言葉で分類するために FACS(Facial Action Coding System)を開発した[49]。FACS は顔の動作を 44 種類の AU(Action Unit)にコード化したもので、表情の分類の他に精神医学の分野でも幅広く利用されている。Mufti と Khanman はルールを言葉で定義可能なファジィ理論の持つ〝あいまいさ〟が実環境中の変動に対してロバストであると主張している[24]。また、Seyedarabi らは処理時間の観点から、ファジィ推論による感情判別がリアルタイム処理に適していることを示唆している[25]。さらに、ファジィ推論は追加学習にも適していると一般的に言われ

ているが、Razak らによると、ファジィルールの構築について、入力する特徴の増加は無駄なルールセットの増加を招き、その結果判別率の低下につながるという意見もある[26]。

本章では、無駄なルールセットの増加を抑制したファジィ推論ベースの表情による感情判別システムを提案する。ファジィルールの構築においては、表情筋は各表情に対して多様な変化を見せる[50]ので、その分布を SOM によってベクトル量子化する。さらに分離度を定義して定量的に感情を分離し、感情の判別に必要なルールのみを抽出する手法を提案する。また、感情判別実験の結果から提案システムの妥当性について議論する。

3.2 システム構成

表情による感情判別システムを図 3.1 に示す。提案システムは〝*平静*〟状態からの表情筋の変化を基に感情を判別する。本システムは表情を取得するためのカメラ、表情筋に従う特徴点抽出部、筋肉長算出部、ファジィ推論による感情判別部で構成される。

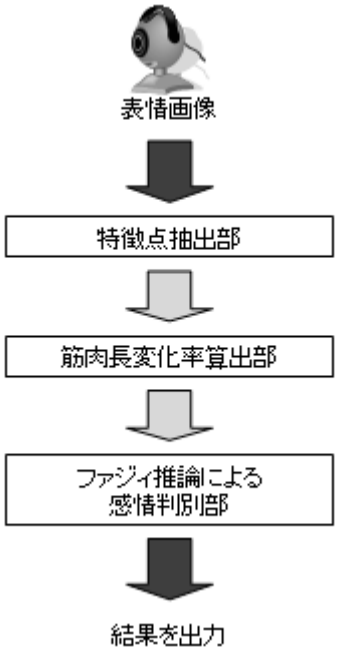


図 3.1. 表情による感情判別システム

特徴点抽出部では、まず haar-like 特徴を用いた分類器[51-52]を用いてカメラ画像中から顔領域を検出する。続いて、検出した顔領域の中心を起点にして、ラスタスキャンによって表情筋の位置の推定に必要な特徴点を探索する。筋肉長算出部では得られた特徴点の座標を用いて各表情筋と口の縦、横方向の開きの〝*平静*〟からの変化率 ΔM_i を(3-1)式に従って算出する。ここで、口の縦方向の開きは上唇の上端から下唇の下端までの長さとした。

$$\Delta M_i = \frac{(M_i - \overline{M}_i)}{\overline{M}_i} \tag{3-1}$$

ここで M_i は i 番目の表情筋または口の長さ、$\overline{M}_i$は "*平静*" 状態の i 番目の表情筋または口の長さの平均をそれぞれ表す。

感情判別部では、得られた表情筋変化率から表出された感情を判別する。本システムでは、感情を言葉で表現可能なファジィ推論を採用した。筋肉長変化率 ΔM_i に対するファジィセットは{*Negative*(*N*), *Zero*(*Z*), *Positive*(*P*)}の 3 種類を用いた。例えば、"*平静*" 状態と比較して目が見開いており("眼輪筋" is *Positive*)、眉が引き上げられ("前頭筋" is *Negative*)、頬が延びていれば("大頬筋" is *Positive*)、表出された感情は "*驚き*" である。

3.3 SOM による感情分類

3.3.1 SOM

SOM は Kohonen により高次元データの特徴をベクトル量子化によって視覚化するために提案された、大脳皮質の視覚野をモデル化したニューラルネットワークである[53]。SOM は高次元のデータ間に存在する非線形な統計学的関係を自己組織的に分類し、低次元のノードの格子上に写像する。SOM による高次元データの分類の有用性から、プロセス解析、機械の知覚機能、さらには生物学、医学、経済学の分野での様々な応用がなされている[54-55]。

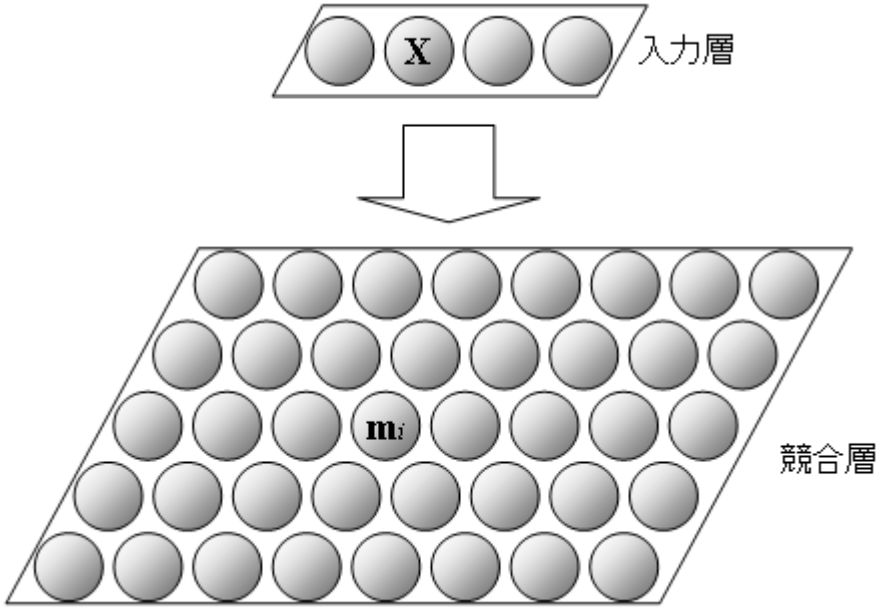


図 3.2. 2 次元 SOM の構造

SOM は図 3.2 に示すような入力層と競合層の 2 層からなるネットワークであり、反復的に学習する。入力変数を $\mathbf{X}=[x_1,x_2,...,x_n]^T \in \mathbf{R}^n$ として定義すると、各競合層ユニットそれぞれにモデルと呼ばれる重みベクトル $\mathbf{m}_i=[\mu_{i1},\mu_{i2},...,\mu_{in}]^T \in \mathbf{R}^n$ を結びつける。

$d(\mathbf{X}, \mathbf{m}_i)$ で示される $\mathbf{X}$ と $\mathbf{m}_i$ 間の距離を導入すると、入力ベクトル $\mathbf{X}$ に最も近い重みベクトルのニューロンは Best-Matching Unit あるいは勝者ユニット c と呼ばれ、(3-2)式のように定義される。

$$d(\mathbf{X},\mathbf{m}_c)=\min_i\{d(\mathbf{X},\mathbf{m}_i)\} \qquad (3\text{-}2)$$

ここで、一般的な距離としてユークリッド距離が多く用いられる。勝者ユニットが発見されると、SOM の重みベクトルは勝者ユニットおよびその近傍が入力ベクトルにより近づくように更新される。ユニット i に対する重みベクトルの更新式は

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \alpha(t)h_{ci}(t)[\mathbf{X}(t) - \mathbf{m}_i(t)] \tag{3-3}$$

で与えられる。ここで、t は離散時間である。$\mathbf{X}(t)$ は t においてランダムに選ばれた入力データであり、$h_{ci}(t)$ は勝者ユニット c の近傍カーネル、$\alpha(t)$ は t における学習係数をそれぞれ示す。近傍関数 $h_{ci}(t)$ は学習回数と勝者ユニットからの距離に従って減少する関数であり、収束するためには $t \to \infty$ のとき $h_{ci}(t) \to 0$ であることが必要である。学習係数 $\alpha(t)$ は $(0 < \alpha(t) < 1)$ であり、時間に関して単調減少する関数である。

SOM によって量子化されたモデルのコードブック・ベクトルの境界を示す方法として、U-matrix が提案されている[56-57]。U-matrix は近接したコードブック・ベクトル間の平均距離を諧調度の濃淡によって表す。近接した $\mathbf{m}_i$ 間の平均距離が小さいならば薄い色合いが用いられ、逆に濃い色合いは距離が大きいことを示す。ただし、一般に U-matrix の解釈については使用者に委ねられる。

3.3.2 Fukui の境界線抽出手法の SOM への適用

SOM によって量子化された結果に対して定量的に境界を求めるため、U-matrix に代わる境界抽出手法として、画像処理における境界抽出手法を SOM へ適用する方法を提案する。つまり、SOM の 2 次元マップを 2 次元の画像データに見立てる。画像処理における境界抽出手法として、1 次微分フィルタによる方法[58-59]、テンプレートマッチングによる方法[60]、Fukui の統計的方法[61]などがある。Fukui の方法は、大津メソッドとしてよく知られている大津の判別分析法[62]をベースとした領域抽出手法である。大津メソッドでは、ある輝度値を閾値としてクラスを 2 分した場合の、各クラス内の分散 σ_w^2 とクラス間の分散 σ_b^2 を定義し、クラス内分散とクラス間分散の比 σ_w^2/σ_b^2 が最小となる輝度値を探索する。大津メソッドは境界を 2 つの局所領域の統計的性質によって決定するため、輝度勾配に基づく方法と比較して高いロバスト性を持つ。Fukui の方法では、(3-4)式に示す画像全体を 2 つの領域に分割した際の分

離度 η を算出する。分離度は 0 から 1 までの値をとり、2 つの領域が完全分離可能ならば分離度は 1 となり、2 つの領域が分離できないならば分離度は 0 に近づく。

$$\eta = \frac{\sigma_b^2}{\sigma_T^2} \tag{3-4}$$

$$\sigma_b^2 = n_1(\overline{P_1} - \overline{P})^2 + n_2(\overline{P_2} - \overline{P})^2 \tag{3-5}$$

$$\sigma_T^2 = \sum_{i=1}^{n1+n2}(P_i - \overline{P})^2 \tag{3-6}$$

ここで、n1、n2 はそれぞれ探索領域 1、探索領域 2 内の画素数を表す。P_i は位置 i の特徴量を表す。特徴量として、各画素中の輝度、色相、彩度などが利用可能である。$\overline{P_1}$、$\overline{P_2}$、$\overline{P}$はそれぞれ領域 1、領域 2、領域全体の P_iの平均を表す。

ここで Fukui の方法を SOM に適用するために、SOM の実行結果に対して閾値を設ける。設定した閾値を超える値を持つユニットの集合を領域 h、閾値を超えないユニットの集合を領域 l とする。Fukui の方法を基に、SOM における分離度ξを次のように定義する。

$$\xi = \sqrt{\frac{\sigma_b^2}{\sigma_T^2}} \qquad \begin{cases} \xi > 0.5 : \text{分離可能} \\ \xi \le 0.5 : \text{分離不可能} \end{cases} \tag{3-7}$$

$$\sigma_b^2 = n_h(\overline{P_h} - \overline{P})^2 + n_l(\overline{P_l} - \overline{P})^2 \tag{3-8}$$

$$\sigma_T^2 = \sum_{i=1}^{n_h+n_l}(P_i - \overline{P})^2 \tag{3-9}$$

ここで、n_hは領域 h の要素数、n_lは領域 l の要素数をそれぞれ表す。P_iは位置 i に対象感情が存在するかを表しており、存在するならば 1、存在しないならば 0 とする。$\overline{P_h}$、$\overline{P_l}$、$\overline{P}$はそれぞれ領域 h、領域 l、領域全体の P_iの平均を表す。分離度ξは偏差に対して線形に変化させるために分散比の平方とする。

また、分離度ξは 0 から 1 までの値をとり得るので、中央の 0.5 を超えると分離可能であるとする。

3.3.3 SOM によるファジィルールの構築

SOM を用いて表情に関するファジィルールを構築するために、学習用データとして **2.3** 節で収集した感情がよく表出されているデータ群から、5 名分の成人男性被験者のデータを半数選択した。学習用データから図 3.3 において点線で示した 12 本の表情筋と口の縦方向、横方向の開きについて、(3-1)式に従ってそれぞれの *"平静"* 状態からの筋肉長変化率ΔM_iを算出し、14 次元のベクトルデータとした。SOM の実行には MATLAB 上で利用可能な SOM Toolbox 2.0[64]を利用した。近傍カーネルは Gaussian を用いた。その他パラメータについては、SOM Toolbox の自動選択機能を用いている。SOM Toolbox 2.0 によって得られた結果を図 3.4 に示す。

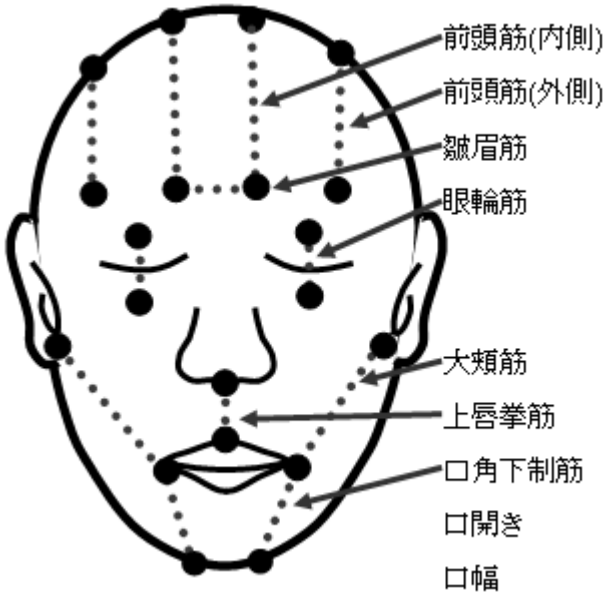
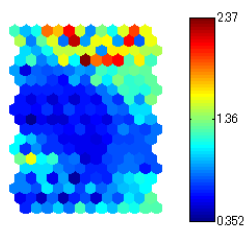
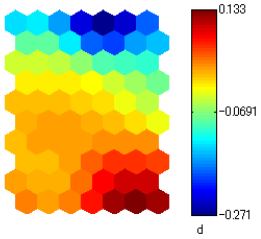


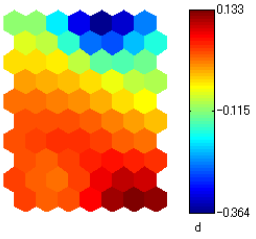
図 3.3. 表情形成時に作用する主要表情筋[63]



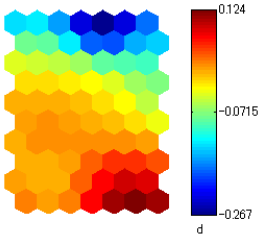
(a)U-matrix



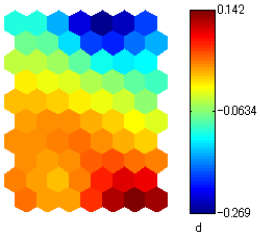
(b)左前頭筋(内側)



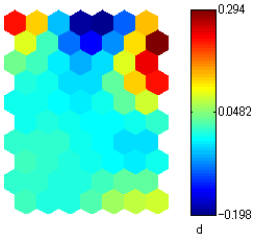
(c)右前頭筋(内側)



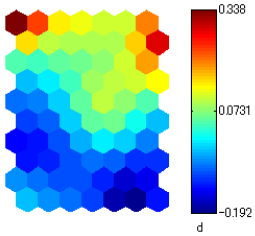
(d)左前頭筋(外側)



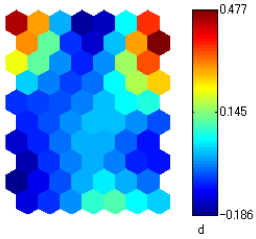
(e)右前頭筋(外側)



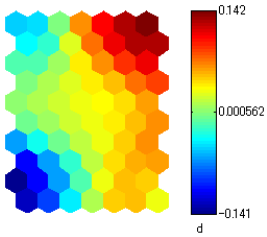
(f)皺眉筋



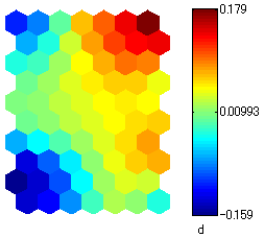
(g)左眼輪筋



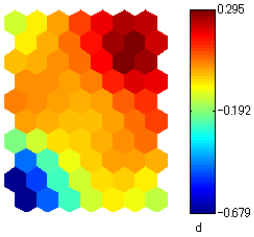
(h)右眼輪筋



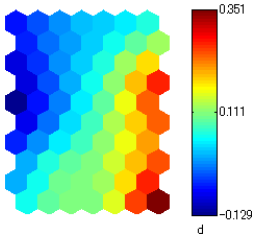
(i)左大頬筋



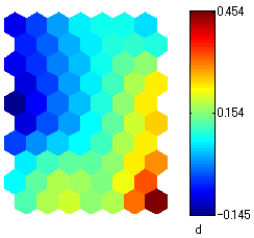
(j)右大頬筋



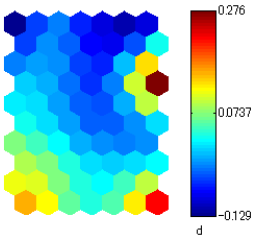
(k)上唇挙筋



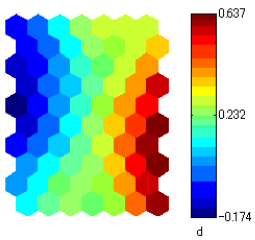
(l)左口角下制筋



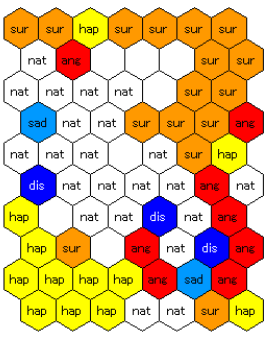
(m)右口角下制筋



(n)口幅



(o)口開き



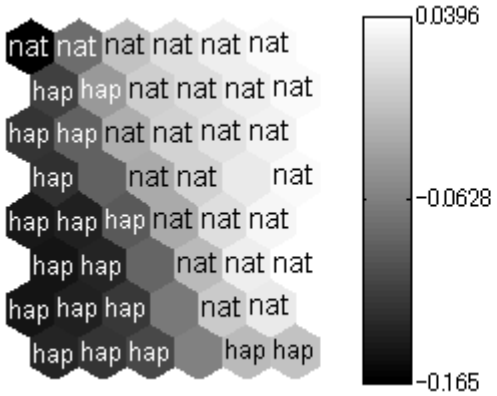
(p)ラベル

図 3.4. SOM の実行結果

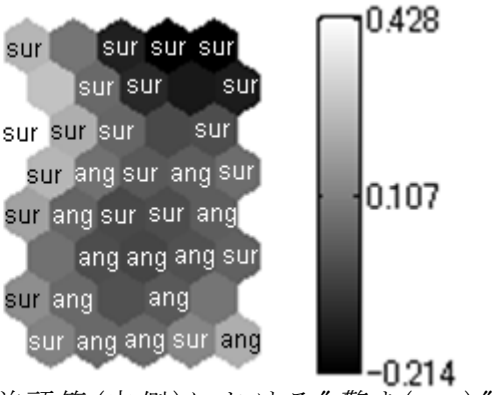
図 3.4(p)より〝*平静*(nat)〟、〝*喜び*(hap)〟、〝*驚き*(sur)〟、〝*怒り*(ang)〟については分布の集中が確認できた。前頭筋は〝*驚き*(sur)〟に対して比較的小さな値を取り、〝*怒り*(ang)〟に対しては比較的大きな値をとることが確認できた。一方、大頬筋と上唇挙筋は〝*驚き*(sur)〟に対して大きな値をとり、〝*喜び*(hap)〟に対して小さな値をとる。

感情の違いによるそれぞれの表情筋の違いを詳細に調べるため、6 感情から 2 感情のペアのデータセットを作成した。組合せは ${}_6C_2 = 15$ 通りが考えられる。感情ペアのデータセットを入力として SOM によるベクトル量子化を実行した例を図 3.5 に示す。図 3.5(a)は〝*平静*(nat)〟と〝*喜び*(hap)〟のペアにおける右側の大頬筋の SOM の実行結果である。〝*平静*(nat)〟は表情筋変化率が 0 付近の値をとり、〝*喜び*(hap)〟は〝*平静*(nat)〟よりも小さな値の分布となった。これは、〝*喜び*(hap)〟の表情は〝*平静*(nat)〟の表情と比較して頬が縮むことを表す。一方、図 3.5(b)に示す左内側の前頭筋では〝*驚き*(sur)〟は一様に分布しており〝*怒り*(ang)〟の分布と区別できない。これは、左内側の前頭筋では〝*怒り*(ang)〟と〝*驚き*(sur)〟を区別できないことを表す。

図 3.5 の SOM の実行結果に対し分離度 ξ を求めた結果を図 3.6 に示す。図 3.6(a)では、右側の大頬筋は閾値を-0.086 としたときに最も高い分離度が得られ、その値は 0.818 で 0.5 を超えた。すなわち、分離可能である。一方、図 3.6(b)では、左側の内側前頭筋は閾値を 0.067 としたときに分離度が最大値 0.338 をとったが、分離の基準とした 0.5 を超えなかったので、分離不可能と判断した。

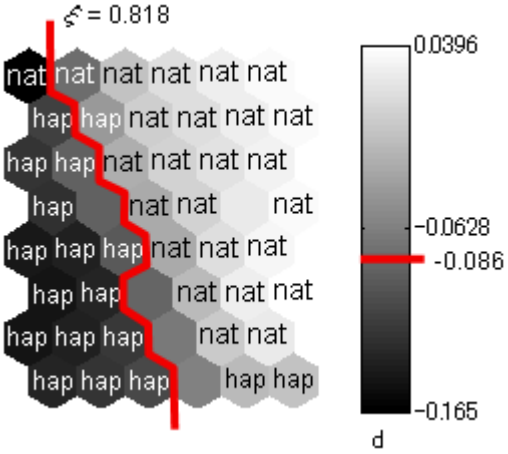


(a)右大頬筋における″*平静*(nat)″と″*喜び*(hap)″の分布

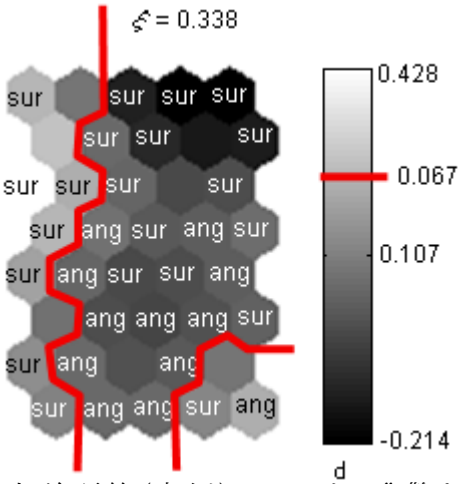


(b)左前頭筋(内側)における″*驚き*(sur)″と″*怒り*(ang)″の分布

図 3.5. 感情ペアに対する SOM の実行結果の例



(a)右大頬筋における″*平静*(nat)″と″*喜び*(hap)″の分布



(b)左前頭筋(内側) における″*驚き*(sur)″と″*怒り*(ang)″の分布

図 3.6. 分離度による境界抽出結果の例

分離度ξに従い、ファジィルールを作成した。まず、本システムは〝*平静*〟状態からの筋肉長の変化を利用しているので〝*平静*〟における各筋肉のラベルを全て *Zero*(*Z*) と割り当てた。次に、〝*平静*〟から分離可能な筋肉に対し、〝*平静*〟から伸びた筋肉を *Positive*(*P*)、〝*平静*〟から縮んだ筋肉を *Negative*(*N*) にそれぞれ割り当てた。結果として得られたラベルを表 3.1 に示す。続いて、任意の 2 感情に対し〝*Z*〟と〝*P*〟のどちらとも分離不可能であった筋肉は〝*Z or P*〟、〝*Z*〟と〝*N*〟のどちらとも分離不可能であった筋肉は〝*Z or N*〟にそれぞれ割り当てることで結果として表 3.2 に示すラベルが得られた。

表 3.1. "*平静*" 状態との比較によるラベルの割り当て

表情筋	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
左前頭筋(内側)	*Z*	*Z*	*Z*	*N*	*P*	*Z*
右前頭筋(内側)	*Z*	*Z*	*Z*	*N*	*P*	*Z*
左前頭筋(外側)	*Z*	*Z*	*Z*	*N*	*P*	*Z*
右前頭筋(外側)	*Z*	*Z*	*Z*	*N*	*P*	*Z*
皺眉筋	*Z*	*Z*	*Z*	*P*	*P*	*N*
左眼輪筋	*Z*	*Z*	*Z*	*P*	*P*	*N*
右眼輪筋	*Z*	*Z*	*Z*	*P*	*N*	*N*
左大頬筋	*Z*	*P*	*N*	*P*	*Z*	*Z*
右大頬筋	*Z*	*P*	*N*	*P*	*Z*	*Z*
上唇挙筋	*Z*	*Z*	*N*	*P*	*P*	*Z*
左口角下制筋	*Z*	*Z*	*Z*	*Z*	*Z*	*P*
右口角下制筋	*Z*	*Z*	*Z*	*Z*	*Z*	*P*
口開き	*Z*	*Z*	*Z*	*Z*	*Z*	*Z*
口幅	*Z*	*Z*	*P*	*Z*	*P*	*Z*

表 3.2. 感情ペアの比較によるラベルの割り当て

表情筋	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
左前頭筋(内側)	*Z*	*Z or P*	*Z*	*N*	*P*	*Z or P*
右前頭筋(内側)	*Z*	*Z or P*	*Z*	*N*	*P*	*Z or P*
左前頭筋(外側)	*Z*	*Z or P*	*Z*	*N*	*P*	*Z or P*
右前頭筋(外側)	*Z*	*Z or P*	*Z*	*N*	*P*	*Z or P*
皺眉筋	*Z*	*N or Z*	*Z*	*P*	*P*	*N*
左眼輪筋	*Z*	*Z*	*N or Z*	*P*	*P*	*N*
右眼輪筋	*Z*	*Z*	*N or Z*	*P*	*N*	*N*
左大頬筋	*Z*	*P*	*N*	*P*	*Z or P*	*Z*
右大頬筋	*Z*	*P*	*N*	*P*	*Z or P*	*Z*
上唇挙筋	*Z*	*Z*	*N*	*P*	*P*	*Z*
左口角下制筋	*Z*	*Z or P*	*Z*	*Z*	*Z*	*P*
右口角下制筋	*Z*	*Z or P*	*Z*	*Z*	*Z*	*P*
口開き	*Z*	*Z*	*Z*	*Z*	*Z*	*Z*
口幅	*Z*	*Z*	*P*	*Z*	*P*	*Z*

表 3.2 に示したルールには冗長な要素が含まれており、前頭筋に関わる 4 つの筋肉ついては、各感情に対して同一のルールが得られた。また、口の縦方向の開きは全ての感情において *″平静″* と区別できなかった。そこで、前頭筋については外側に従う特徴点は前髪に隠れて取得困難な場合があるため内側のみを採用した。また、口の縦方向の開きは除外した。

結果として得られたファジィルールを表 3.3、メンバーシップ関数を図 3.7、扱う特徴点と表情筋を図 3.8 にそれぞれ示す。本システムでは、図 3.8 に示す 17 個の特徴点に従って 10 本の表情筋と口の横開きを基に感情を判別する。図 3.7 のグラフは、横軸が筋肉長変化率、縦軸がメンバーシップ関数のグレード μ をそれぞれ示す。メンバーシップ関数については、Fukui の方法で取得した境界を相補型メンバーシップ関数の交点に置いて作成した。表 3.3 における *″Z or P″* と *″Z or N″* のメンバーシップ関数のグレード μ はそれぞれ(3-10)式および(3-10)式で算出する。

$$\mu(Z\ or\ P) = 1.0 - \mu(N) \tag{3-10}$$

$$\mu(N \text{ or } Z) = 1.0 - \mu(P) \tag{3-11}$$

表 3.3. 表情による感情判別システムで用いるファジィルール

表情筋	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
M_1, M_2 : 前頭筋	*Z*	*Z or P*	*Z*	*N*	*P*	*Z or P*
M_3 : 皺眉筋	*Z*	*N or Z*	*Z*	*P*	*P*	*N*
M_4, M_5 : 眼輪筋	*Z*	*Z*	*N or Z*	*P*	*P*	*N*
M_6, M_7 : 大頬筋	*Z*	*P*	*N*	*P*	*Z or P*	*Z*
M_8: 上唇挙筋	*Z*	*Z*	*N*	*P*	*P*	*Z*
M_9, M_{10} : 口角下制筋	*Z*	*Z or P*	*Z*	*Z*	*Z*	*P*
M_{11}: 口幅	*Z*	*Z*	*P*	*Z*	*P*	*Z*

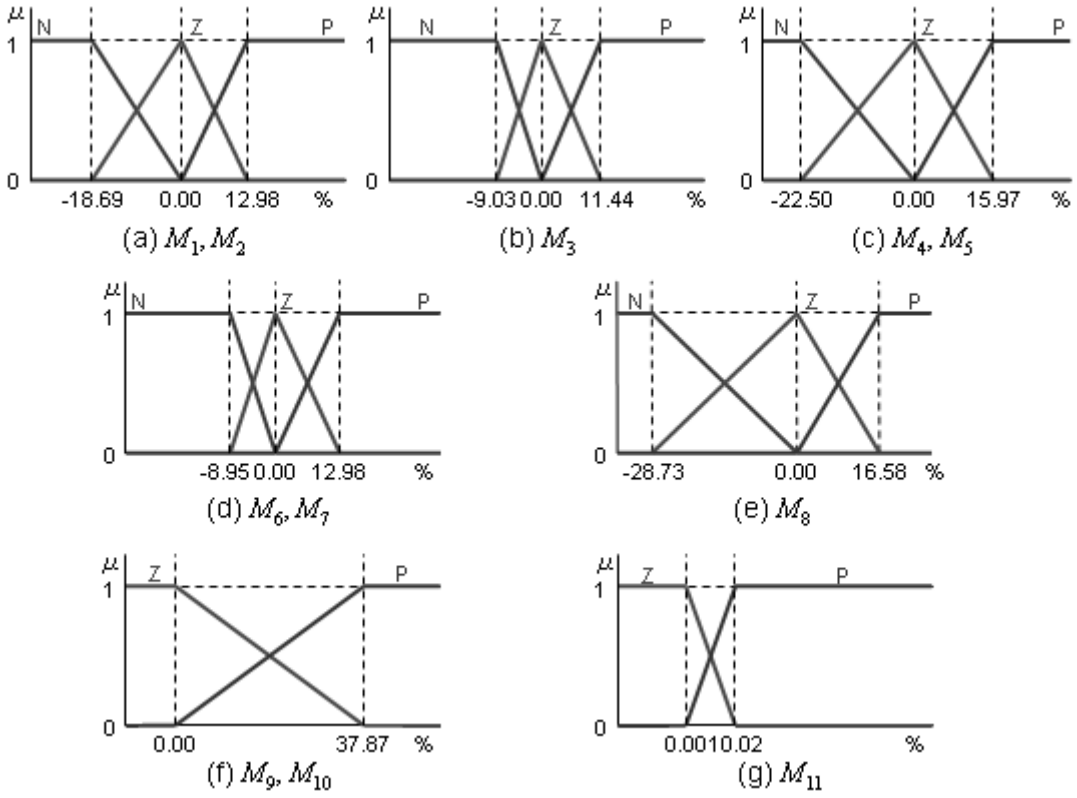


図 3.7. 表情による感情判別システムで用いるメンバーシップ関数

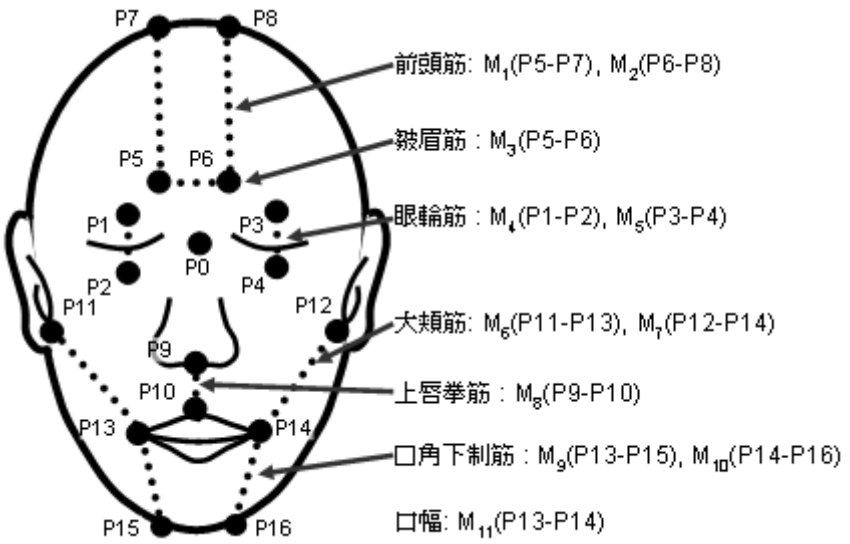


図 3.8. 表情による感情判別システムで用いる特徴点と表情筋

3.4 表情による感情判別実験

提案システムの妥当性を検証するために行った感情判別実験結果を表3.4に示す。実験用データとして、2.3節で収集した感情がよく表出されているデータ群から、学習用に用いたデータを除外したものを入力データとして用いた。表3.4より、〝*怒り*〟と〝*驚き*〟については未知の被験者を用いた場合の正答率がルール作成に関わった人物の正答率を上回った。これは、未知の人物の中に大頬筋を大きく変化させて感情を表現した人物が含まれており、大頬筋に関する〝*怒り*〟と〝*驚き*〟の適合度がより大きな値を示したことが要因である。〝*平静*〟、〝*喜び*〟、〝*嫌悪*〟についてはルール作成に関わっていない人物に対しても 75%を超える判別結果が得られた。一方、〝*怒り*〟と〝*悲しみ*〟は60% 未満の比較的低い判別率となった。〝*怒り*〟の表情は〝*平静*〟や〝*嫌悪*〟に誤判別しやすい傾向が見られた。表情のみでは判別が難しい感情の判別率の向上には、他のモダリティを用いたマルチモーダルなシステムへの拡張が必要と考えられる。

表 3.4. 表情による感情判別システム 感情判別結果

(a) ルール作成に関わった被験者に対する感情判別結果

出力 / 入力	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*	正答数 / 総数	正答率 [%]
平静	**58**	2	1	0	0	0	58 / 61	95.1
怒り	4	**10**	0	2	2	2	10 / 20	50.0
喜び	1	1	**38**	1	2	0	38 / 43	88.4
驚き	4	4	2	**27**	1	1	27 / 39	69.2
悲しみ	0	1	4	3	**10**	3	10 / 21	47.6
嫌悪	0	0	0	0	1	**3**	3 / 4	75.0
計							146 / 188	77.7

(b) 未知の被験者に対する感情判別結果

出力 / 入力	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*	正答数 / 総数	正答率 [%]
平静	**48**	3	2	0	0	0	48 / 53	90.6
怒り	3	**15**	0	0	1	7	15 / 26	57.7
喜び	1	0	**28**	0	4	0	28 / 33	84.8
驚き	7	7	2	**46**	3	0	46 / 65	70.8
悲しみ	5	0	0	0	**5**	1	5 / 11	45.5
嫌悪	0	0	0	0	2	**11**	11 / 13	84.6
計							153 / 201	76.1

3.5 結言

本章では、ファジィ推論による表情による感情判別システムを提案した。提案システムは、"*平静*" 状態からの表情筋の変化を基に感情を判別する。ファジィルールの構築には、SOM を用いた定量的な境界抽出手法として、Fukui の境界線抽出手法を SOM へ適用する方法を提案した。提案手法は感情の判別に対して重要でない筋肉の特定にも貢献している。

感情判別実験の結果、″*平静*″、″*喜び*″、″*嫌悪*″ に関しては 75% を超える判別率が得られた。一方、″*怒り*″ の表情は ″*平静*″ や ″*嫌悪*″ に誤判別しやすい傾向が見られた。

第 4 章　音声による感情判別

4.1 緒言

人間同士のコミュニケーションの場では、音声は感情を伝達するために表情に次いで重要な役割を持っていると思われる。音声から感情を判別する手法としては、人間の肉声をコーパスとして予め記録し、パターンマッチングを行う手法[65-66]があり、パターンの学習には表情による感情判別システムと同様にニューラルネットワークによる方法[26]、サポートベクタマシンによる方法[65, 67-68]、HMM による方法[69]等の機械学習ベースのものがある。しかし、コーパスにより感情を判別する場合、膨大な肉声データが必要であり、あらゆる言語を網羅したコーパスの作成は一般に難しい。一方、音声中に含まれる韻律情報を基にして感情を判別する方法がある[70]。一般に、共通の言語を持たない環境においても、人間は言語に依存しない声の高さや大きさなどを利用して相手に感情を伝達することが可能である[71-72]。Zeng らは音声信号中のピッチとエネルギーが感情判別に最も貢献すると報告している[6]。

本章では韻律情報として声の「大きさ」、「抑揚強度」および「高さ」を利用したファジィ推論ベースの感情判別システムを提案する。一般に、感情音声は〝*平静*〟状態からの韻律情報の変動によって形成される[71]。そこで、各感情における韻律情報の変動分布を推定するために推計統計学を用いた手法を提案する。本手法では、韻律情報から統計量を算出し、仮説検定による手法でファジィルールを構築する。

4.2 システム構成

本論文で提案する音声による感情判別システムを図 4.1 に示す。本システムは音声信号を取得するためのマイクロフォン、韻律パラメータ算出部、ファジィ推論による感情判別部で構成される。本システムでは、図 4.1 の韻律パラメータ算出部において、声の大きさおよび抑揚強度を得るための短時間平均パワーと声の高さを得るためのピッチの平均値を算出する。

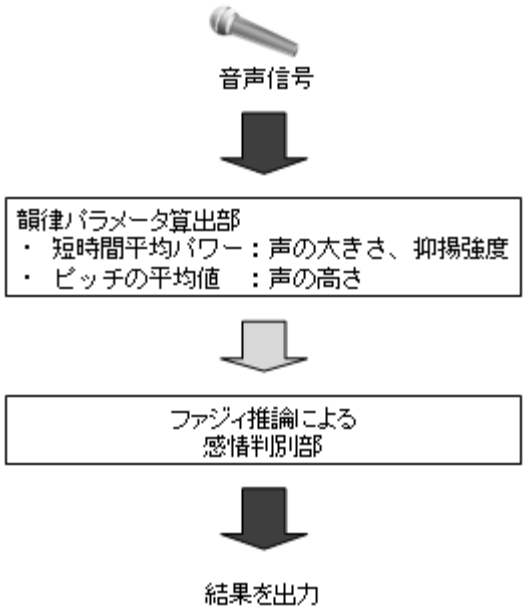


図 4.1. 音声による感情判別システム

i 番目のフレームにおける短時間平均パワー$POW(i)$は常用対数を用いて(4-1)式に従って算出される。

$$POW(i) = 10 * \log_{10} \frac{1}{N} \sum_{n=0}^{N-1} x_n^2 \qquad [dB] \tag{4-1}$$

ここで、N は 1 フレームに含まれる入力サンプルデータ数、x_n は [0, 1] に正規化された入力サンプルデータをそれぞれ表す。本システムでは、1024 入力サンプルを 1 フレームとし、フレームシフトを 512 サンプルとした。本システムでは、声の大きさ L および抑揚強度 IS を短時間平均パワーを用いてそれぞれ(4-2)式および(4-3)式に示すように定義した。

$$\text{声の大きさ} L := \frac{\text{短時間平均パワーの最大値}[dB]}{\text{初期有声フレームの短時間平均パワー}[dB]} \tag{4-2}$$

$$\text{抑揚強度} IS := \frac{\text{短時間平均パワーの最大値}[dB]}{\text{短時間平均パワーの平均値}[dB]} \tag{4-3}$$

(4-2)式において、初期有声フレームの短時間平均パワーは腹筋に力が入りきれていない状態の声の大きさに関与している。短時間平均パワーの最大値は腹筋に最も力が入った状態における声の大きさであり、これらの比を声の大きさと定義した。(4-3)式においては、短時間平均パワーの平均と最大値との比の大きさを抑揚強度と定義した。

ピッチの平均値は声の高さを表す。本システムでは、1024 入力サンプルを 1 フレームとして(4-4)式に示す自己相関関数 $R_{xx}(k)$ からピッチを算出し、その平均値 $\overline{P}$ を声の高さと定義した。

$$R_{xx}(k) = \frac{1}{N}\sum_{n=0}^{N-1} x_n * x_{n+k} \tag{4-4}$$

ここで、N は 1 フレームに含まれる入力サンプルデータ数、x_n は [0, 1] に正規化された入力サンプルデータ、k は入力サンプル x_n とのラグをそれぞれ表す。

感情判別部では、得られた韻律パラメータから表出された感情を判別する。本システムでは、各感情に対する韻律情報のルールを言葉で表現可能なファジィ推論を採用した。例えば、声の大きさが小さく、抑揚強度も小さく、声の高さも低い値を示すならば、表出された感情は "*悲しみ*" である。

4.3 推計統計学的手法による感情分類

4.3.1 仮説検定

母集団の母数に関する主張を仮説といい、母集団からとられた無作為標本の値によって母数に関する仮説の棄却もしくは採択を決定することを仮説検定という[73]。仮説は棄却されることを期待して設定されるので、「帰無仮説」と呼ばれ、H_0で表される。一方、「帰無仮説」が棄却されると受け入れられることになる仮説を「対立仮説」と呼び、H_1で表す。「帰無仮説」を誤って棄却してしまう確率は有意水準αと呼ばれる。検定には、与えられた仮説に対して適当な統計量Tを選び、「帰無仮説」が真のときの統計量Tの標本分布を用いる。このとき、Tを検定統計量といい、「帰無仮説」が棄却されることになるTの実現値の範囲を棄却域という。標準的な検定と選択する検定統計量の関係を表4.1に示す。表4.1において、n、μ、σ^2、はそれぞれ標本数、標本平均、標本分散を表す。検定統計量Tの分布の左右の両裾を棄却域に選ぶ検定を両側検定、片裾を棄却域に選ぶ検定を片側検定と呼び、以下の手順で行う：

1) 「帰無仮説」H_0と「対立仮説」H_1を定める
2) 有意水準αの値を決める
3) 検定統計量Tおよび棄却域を選ぶ
4) 与えられたデータからのTの実現値を求める
5) Tの実現値が棄却域に含まれるならば「帰無仮説」H_0を棄却する。Tの実現値が棄却域に含まれないならば「帰無仮説」H_0を採択する

表 4.1. 標準的な検定と検定統計量の関係

仮説	条件	検定統計量	検定統計量の分布
平均の検定 H_0： $\mu = \mu_0$ H_1： $\mu \neq \mu_0$ H_1： $\mu > \mu_0$ H_1： $\mu < \mu_0$	母集団の分布は正規分布 σ^2 は既知	$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ 大標本の場合: $z = \frac{\bar{x} - \mu_0}{\frac{u}{\sqrt{n}}}$	標準正規分布
小標本の 平均の検定 H_0： $\mu = \mu_0$ H_1： $\mu \neq \mu_0$ H_1： $\mu > \mu_0$ H_1： $\mu < \mu_0$	母集団の分布は正規分布 σ^2 は未知	$t = \frac{\bar{x} - \mu_0}{\frac{u}{\sqrt{n}}}$	自由度n-1の t 分布
分散の検定 H_0： $\sigma^2 = \sigma_0^2$ H_1： $\sigma^2 \neq \sigma_0^2$ H_1： $\sigma^2 > \sigma_0^2$	母集団の分布は正規分布	$\chi^2 = \frac{(n-1)\mu^2}{\sigma^2}$	自由度n-1の χ^2 分布
平均の差の検定 H_0： $\mu_1 = \mu_2$ H_1： $\mu_1 \neq \mu_2$ H_1： $\mu_1 > \mu_2$ H_1： $\mu_1 < \mu_2$	母集団の分布は正規分布 σ_1^2、σ_2^2 は既知	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	標準正規分布
平均の差の検定 H_0： $\mu_1 = \mu_2$ H_1： $\mu_1 \neq \mu_2$ H_1： $\mu_1 > \mu_2$ H_1： $\mu_1 < \mu_2$	母集団の分布は正規分布 $\sigma^2 = \sigma_1^2 = \sigma_2^2$	$t = \frac{\bar{x}_1 - \bar{x}_2}{u\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $u = \frac{(n_1 - 1)\mu_1^2}{n_1 + n_2 - 2} + \frac{(n_2 - 1)\mu_2^2}{n_1 + n_2 - 2}$	自由度n_1 - n_2 − 2 の t 分布

4.3.2 仮説検定によるファジィルールの構築

ファジィルールの作成において、表情による感情判別と同様に、2.3 節で収集した感情がよく表出されているデータ群から 5 名分の成人男性被験者のデータの半数を用いて韻律パラメータを算出し、表 4.2 に示す統計量を算出した。表 4.2 において、歪度(skewness)および尖度(kurtosis)は分布の非対称性および尖り具合を示す指標であり、(4-5)式および(4-6)式でそれぞれ定義される[74]。歪度および尖度の絶対値は、それぞれ 10 以上になると母集団が正規分布から外れることが経験的に知られている。

母集団が正規分布に従うとき、歪度および尖度はそれぞれ 0 に近づく。ここでは、表 4.2 において、歪度および尖度の絶対値は全て 10 未満であったので、各母集団は正規分布に従うと仮定した。

$$skewness = \frac{n}{(n-1)(n-2)} \sum_i \left(\frac{x_i - \bar{x}}{s} \right)^3 \tag{4-5}$$

$$kurtosis = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_i \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{4-6}$$

表 4.2. 韻律パラメータの統計量

(a) 声の大きさ L

	平静	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
標本数	27	15	25	28	35	7
標本平均	2.344	3.438	3.166	2.858	2.000	2.093
標本分散	0.755	1.534	1.032	0.564	0.442	0.049
歪度	1.694	-0.029	0.719	0.046	2.469	-1.297
尖度	3.530	-1.011	0.297	-0.626	9.606	1.249

(b) 抑揚強度 IS

	平静	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
標本数	27	15	25	28	35	7
標本平均	1.754	2.242	2.187	1.881	1.515	1.656
標本分散	0.111	0.345	0.191	0.191	0.099	0.048
歪度	1.622	1.926	0.863	0.947	2.187	-0.715
尖度	3.546	5.017	0.865	1.325	7.110	-0.974

(c) 声の高さ $\overline{P}$ [Hz]

	平静	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
標本数	27	15	25	28	35	7
標本平均	133.87	138.53	169.40	165.80	119.99	115.91
標本分散	139.05	485.34	517.50	1239.54	412.76	85.36
歪度	0.616	-0.422	0.949	0.470	0.389	1.577
尖度	0.837	1.412	2.538	-0.333	2.017	2.512

韻律パラメータの大小関係を定量的に評価し、ファジィルールを作成するために推計統計学を用いた。ここでは、5 名の成人男性被験者から得られた韻律パラメータの母集団は歪度および尖度を根拠に正規分布に従うと仮定し、仮説検定を実施した。検定の種類は平均の差の検定であり、両側検定を選択した。有意水準αは 0.2 とした。帰無仮説 H_0 は〝2 感情に対するパラメータの平均は等しい〟とし、対立仮説 H_1 は〝2 感情に対するパラメータの平均は等しくない〟とした。すなわち、

$$H_0 : \mu_{pi} = \mu_{pj} \tag{4-7}$$

$$H_1 : \mu_{pi} \neq \mu_{pj} \tag{4-8}$$

(p: L, IS, $\overline{P}$, i, j : *平静*, *怒り*, *喜び*, *驚き*, *悲しみ*, *嫌悪*)

である。帰無仮説が棄却できなければ、2 感情に対する韻律パラメータは同一の母集団から発生したとみなす。

表 4.2 から得られた検定統計量を表 4.3 に示す。表 4.3 において、帰無仮説を棄却できなかった要素を太字で示す。表 4.3 の結果より、声の大きさに関しては〝*怒り*〟と〝*喜び*〟、〝*喜び*〟と〝*驚き*〟、〝*悲しみ*〟と〝*嫌悪*〟はそれぞれ帰無仮説を棄却できなかった。従って、(4-9)式および(4-10)式にそれぞれ示すように〝*怒り*〟と〝*喜び*〟、〝*喜び*〟と〝*驚き*〟、〝*悲しみ*〟と〝*嫌悪*〟はそれぞれ同一の母集団から発生したとみなした。

$$\mu_{L怒り} = \mu_{L喜び} = \mu_{L驚き} \tag{4-9}$$

$$\mu_{L悲しみ} = \mu_{L嫌悪} \tag{4-10}$$

同様に、抑揚強度については〝*平静*〟と〝*驚き*〟、〝*平静*〟と〝*嫌悪*〟、〝*怒り*〟と〝*喜び*〟は帰無仮説を棄却できなかったので、それぞれ(4-11)式および(4-12)式に示すように同一の母集団から発生したとみなした。

$$\mu_{IS\,平静} = \mu_{IS\,驚き} = \mu_{IS\,嫌悪} \tag{4-11}$$

$$\mu_{IS\,怒り} = \mu_{IS\,喜び} \tag{4-12}$$

声の高さでは〝*平静*〟と〝*怒り*〟、〝*喜び*〟と〝*驚き*〟、〝*悲しみ*〟と〝*嫌悪*〟は帰無仮説を棄却できなかったため、それぞれ(4-13)式、(4-14)式および(4-15)式に示すように同一の母集団から発生したとみなした。

$$\mu_{\overline{P}\,平静} = \mu_{\overline{P}\,怒り} \tag{4-13}$$

$$\mu_{\overline{P}\,喜び} = \mu_{\overline{P}\,驚き} \tag{4-14}$$

$$\mu_{\overline{P}\,悲しみ} = \mu_{\overline{P}\,嫌悪} \tag{4-15}$$

表 4.3. 検定統計量の算出結果

(a) 声の大きさ L

μ_{Lj} \ μ_{Li}	平静	怒り	喜び	驚き	悲しみ	嫌悪
平静		3.030	3.122	2.343	1.706	1.343
怒り	3.030		**0.717**	1.656	4.239	4.067
喜び	3.122	**0.717**		**1.241**	5.018	4.881
驚き	2.343	1.656	**1.241**		4.737	4.642
悲しみ	1.706	4.239	5.018	4.737		**0.660**
嫌悪	1.343	4.067	4.881	4.602	**0.660**	

(b) 抑揚強度 IS

μ_{ISj} \ μ_{ISi}	平静	怒り	喜び	驚き	悲しみ	嫌悪
平静		2.963	3.997	**1.219**	2.868	**0.936**
怒り	2.963		**0.312**	2.088	4.524	3.394
喜び	3.997	**0.312**		2.546	6.575	4.422
驚き	**1.219**	2.088	2.546		3.734	1.931
悲しみ	2.868	4.524	6.575	3.734		1.440
嫌悪	**0.936**	3.394	4.422	1.931	1.440	

(c) 声の高さ $\overline{P}$

$\mu_{\overline{P}j}$ \ $\mu_{\overline{P}i}$	平静	怒り	喜び	驚き	悲しみ	嫌悪
平静		**0.761**	6.989	4.542	3.371	4.311
怒り	**0.761**		4.239	3.115	2.790	3.388
喜び	6.989	4.239		**0.447**	8.668	9.326
驚き	4.542	3.115	**0.447**		8.668	9.326
悲しみ	3.371	2.790	8.668	6.118		**0.832**
嫌悪	4.311	3.388	9.326	6.639	**0.832**	

同一の母集団から発生したデータを統合し、統計量を再び算出した結果を表4.4に示す。表4.4の統計量に従って、図4.2に示すメンバーシップ関数および表4.5に示すファジィルールを作成した。図4.2のグラフは、横軸が韻律パラメータ、縦軸がメンバーシップ関数のグレードμをそれぞれ示す。得られたデータの母集団は正規分布に従うと仮定しているのでメンバーシップ関数はガウス曲線を採用するべきであるが、計算の簡略化のために、三角型メンバーシップ関数で近似した。ここでは、それぞれのパラメータの標本平均値μを三角型メンバーシップ関数の頂点とし、標本標準偏差σを用いて$\mu \pm 2\sigma$を三角型メンバーシップ関数の幅とした。韻律パラメータに対するファジィセットとして、ファジィ集合に{*Low*(*L*), *Middle*(*M*), *High*(*H*)}の3種類のラベルをそれぞれ割り当てた。

表 4.4. 統合後の韻律パラメータの統計量

(a) 声の大きさ L

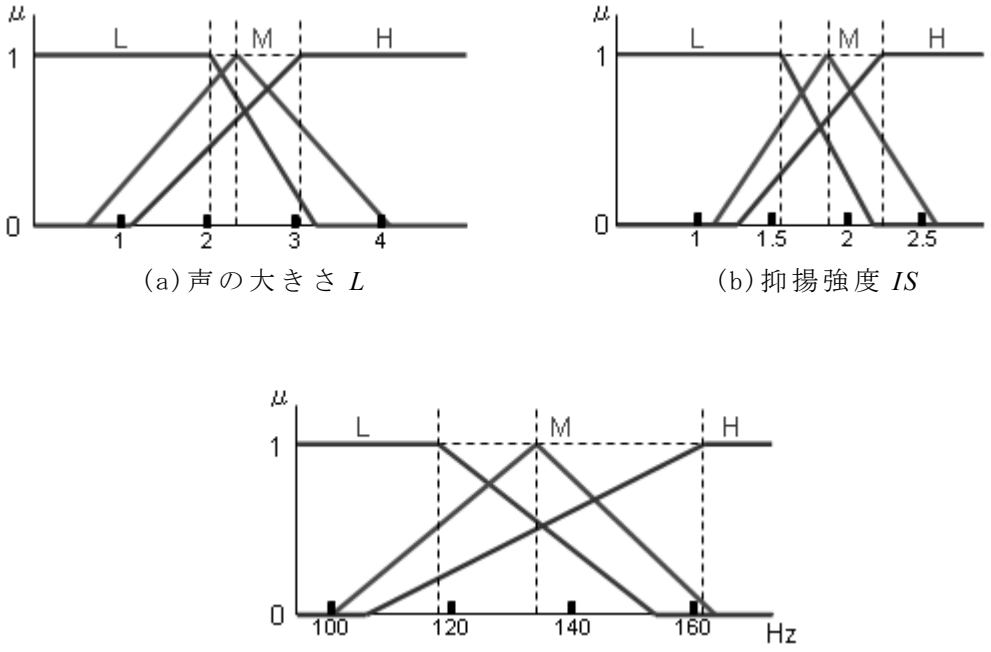
	平静	*怒り+喜び+驚き*	*悲しみ+嫌悪*
標本数	27	68	42
標本平均(μ)	2.344	3.084	2.016
標本分散	0.755	0.966	0.375
標本標準偏差σ	0.869	0.983	0.613

(b) 抑揚強度 IS

	平静+驚き+嫌悪	*怒り+喜び*	*悲しみ*
標本数	62	40	35
標本平均(μ)	1.817	2.208	1.515
標本分散	0.137	0.242	0.099
標本標準偏差σ	0.371	0.492	0.314

(c) 声の高さ $\overline{P}$ [Hz]

	平静+怒り	*驚き+喜び*	*悲しみ+嫌悪*
標本数	42	53	42
標本平均μ	135.53	167.50	119.31
標本分散	259.01	885.75	357.14
標本標準偏差σ	16.09	29.76	18.90



(a)声の大きさ L　　(b)抑揚強度 IS

(c)声の高さ $\overline{P}$

図 4.2. 音声からの感情判別システムで用いるメンバーシップ関数

表 4.5. 音声による感情判別システムで用いるファジィルール

韻律パラメータ	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
声の大きさ	*M*	*H*	*H*	*H*	*L*	*L*
抑揚強度	*M*	*H*	*H*	*M*	*L*	*M*
声の高さ	*M*	*M*	*H*	*H*	*L*	*L*

4.4 音声による感情判別実験

音声による本システムを用いた感情判別実験を行った結果を表4.6に示す。実験用データとして、表情による感情判別と同様に2.3節で収集した感情がよく表出されているデータ群から、学習用に用いたデータを除外したデータを入力として用いた。表4.6より、表3.4より、*"喜び"*については未知の被験者を用いた場合の正答率がルール作成に関わった人物の正答率を上回った。これは、未知の人物の中に声の高さを大きく変化させて*"喜び"*を表現した人物が含まれていたため、声の高さに関する*"喜び"*の適合度がより大きな値を示したことによる。また、表情による感情判別システムが不得意としていた *"悲しみ"* についてはルール作成に関わっていない人物に対しても70%を超える判別率が得られた。一方、*"喜び"* と *"驚き"* はルール作成に関わった人物に対しても 60%未満の低い判別率となった。しかし、Shigenoは人間同士の音声のみでのコミュニケーションにおいても*"喜び"* と *"驚き"* は比較的誤判別され易い傾向にあることを示している[15]。

表 4.6. 音声による感情判別システム 感情判別結果

(a) ルール作成に関わった被験者に対する感情判別結果

出力 / 入力	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*	正答数 / 総数	正答率 [%]
平静	**21**	0	1	0	4	1	21 / 27	77.8
怒り	1	**9**	1	3	0	0	9 / 14	64.3
喜び	5	8	**17**	2	1	0	17 / 33	51.5
驚き	2	2	9	**20**	3	0	20 / 36	55.6
悲しみ	0	0	0	2	**40**	0	40 / 42	95.2
嫌悪	0	0	0	0	1	**8**	8 / 9	88.9
計							115 / 161	71.4

(b) 未知の被験者に対する感情判別結果

出力 / 入力	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*	正答数 / 総数	正答率 [%]
平静	**18**	5	0	0	0	1	18 / 24	75.0
怒り	2	**12**	1	1	0	1	12 / 17	70.6
喜び	0	2	**16**	6	0	0	16 / 24	66.7
驚き	2	2	7	**12**	0	0	12 / 23	52.2
悲しみ	2	1	4	1	**22**	0	22 / 30	73.3
嫌悪	1	1	0	0	2	**6**	6 / 10	60.0
計							86 / 128	67.2

4.5 結言

本章では、ファジィ推論を用いた音声による感情判別システムを提案した。音声による感情判別システムは、韻律として声の大きさ、抑揚強度および高さを用いて感情を判別する。ファジィルールの構築では、ルール作成用に収集したデータの歪度および尖度を根拠に、韻律パラメータの母集団は正規分布に従うと仮定して仮説検定による手法を提案した。

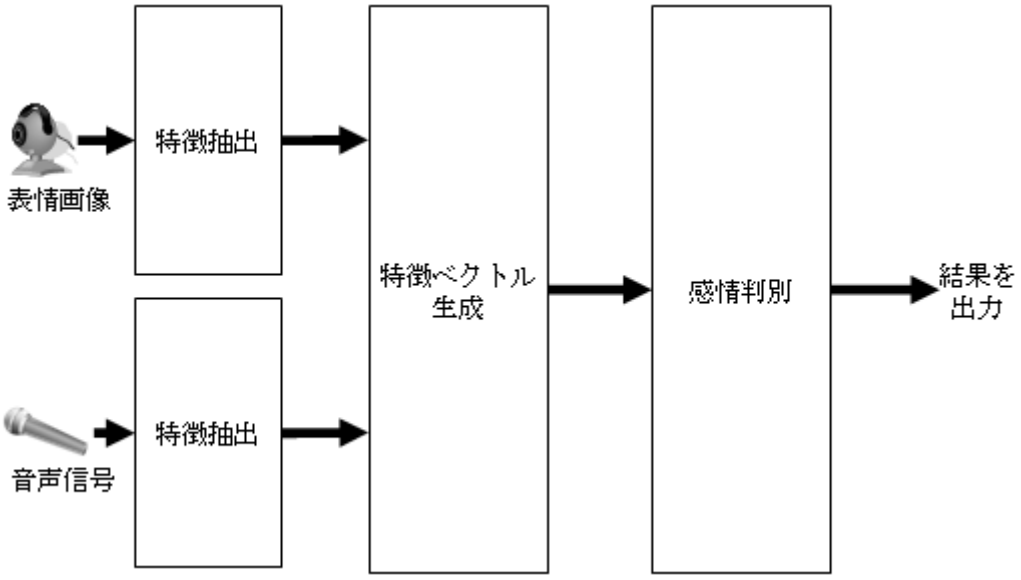
感情判別実験の結果、*″悲しみ″*に対しては 70% を超える良好な判別率が得られた。一方、*″喜び″*と *″驚き″*に関しては 60% 未満の低い判別率となったが、人間同士のコミュニケーションの場においても、音声のみの場合は *″喜び″*と *″驚き″*は比較的誤判別されやすい傾向が報告されており、人間同士のコミュニケーションにおける感情の誤認識と同様の傾向が得られたと考えられる。

第 5 章　マルチモーダル感情判別システム

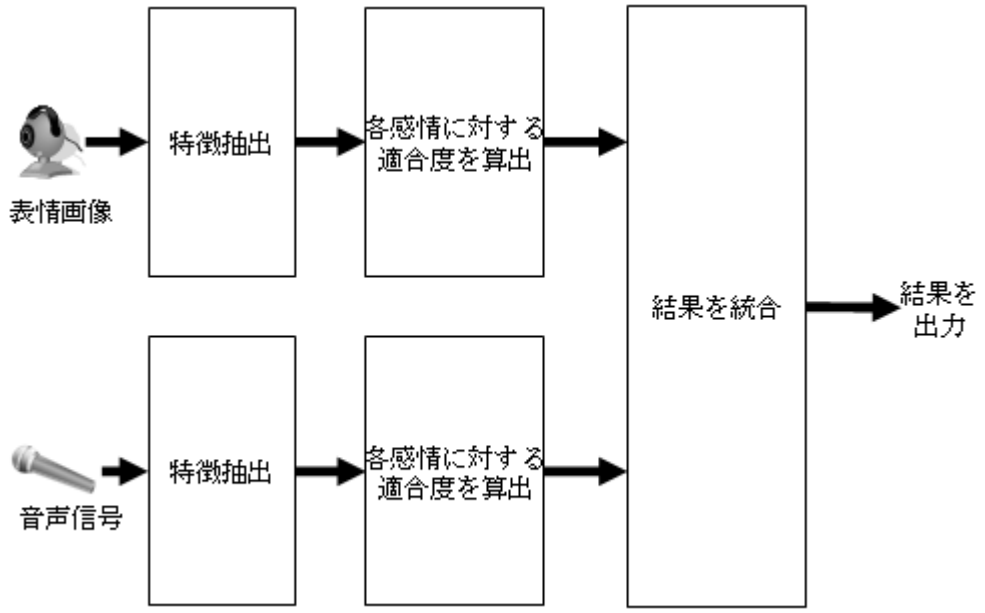
5.1 緒言

人間同士のコミュニケーションでは、互いの感情を読み取るために複数の感覚情報を同時に利用している。感情判別システムにおいても、複数のモダリティによる感情判別を同時に実行するマルチモーダル感情判別システムが注目されつつある[6, 16-19]。モダリティの統合においては、図 5.1(a)に示す特徴レベルの統合[17, 75-76]と図 5.1(b)に示す決定レベルでの統合[17, 77-78]が考えられる。特徴レベルの統合は、各モダリティから得られる特徴をベクトルデータとして統合する。各モダリティの特徴間に存在する相互関係がベクトルデータ中に現れるという利点があるが、各モダリティの特徴が異なる時間間隔で得られる場合､特徴ベクトル作成のタイミング調整が難しいという問題がある。一方、決定レベルの統合では、各モダリティが出力した各感情に対する適合度を統合する。各モダリティでの適合度が更新されるタイミングで統合処理を実行するので、特徴ベクトル作成のタイミング同期は必要ないが、次の判別結果が到達する前に統合処理を終了させる必要がある。

本章では、表情による感情判別および音声による感情判別を統合したマルチモーダル感情判別システム[79]とその実装方法について提案する。各シングルモダリティにおいて、リアルタイム性を考慮し到達した入力データから直ちに特徴を抽出するために、特徴抽出時のタイミング調整による待ち合わせが不要な決定レベルでの統合を採用する。各シングルモダリティでの誤判別を互いに抑制し合う方法を提案し、ヘテロジニアスマルチコアプロセッサ上でリアルタイムかつ高精度な感情判別システムを構築する。



(a) 特徴レベルでの統合



(b) 決定レベルでの統合

図 5.1. モダリティの統合レベル

5.2　システム構成

本マルチモーダル感情判別システムの構成を図5.2に示す。各モダリティに対する処理の並列実行性と拡張性を考慮し、階層モジュール型モデルを採用した。下位モジュールはシングルモダリティでの感情判別を個別に実行し、上位モジュールは各下位モジュール群からの感情判別結果が更新される毎に結果を統合し、最終的な感情判別結果を出力する。提案モデルでは、下位モジュールの追加が容易に行える構造をしており、更なるモダリティの追加が可能である。

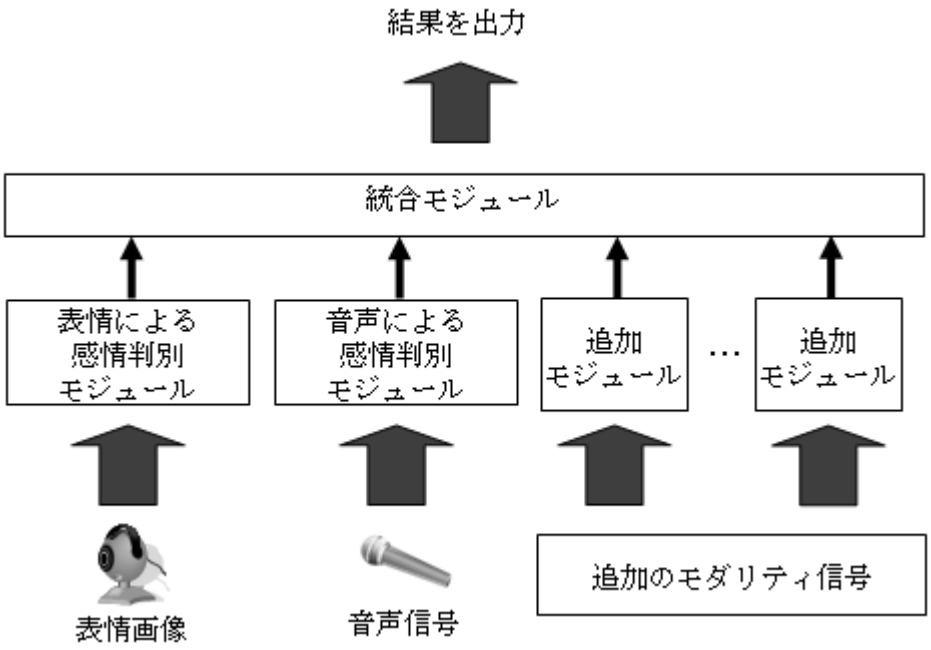


図 5.2. 階層モジュール型マルチモーダル感情判別システム

判別結果の統合については、次に示す方法が主に考えられている[17]:

1)Maximum法　：　全ての適合度中最も高い適合度を示した感情を出力する

2)Average法　：　各モダリティの同一感情同士の適合度の平均を算出し、最も高い平均値を示した感情を出力する

3)乗算法　　：　各モダリティの同一感情同士の適合度を乗算し、最も高い乗算値を示した感情を出力する

本システムでは、3.4節および4.4節で述べたように、表情と音声単独の感情判別システムにおいて、各モダリティには判別が容易な感情と不得意な感情に違いが見られた。各モダリティにおける判別結果を表5.1に示す。表5.1より、音声の "*喜び*" の判別は困難であるが表情による感情判別システムでは "*喜び*" は高い判別率を示す。また、"*悲しみ*" の表情は判別が難しいが、"*悲しみ*" の音声は判別率が73%以上であり容易に判別可能であることがわかる。そこで、上位モジュールでの感情判別結果の統合には互いのモダリティの判別ミスを抑制し合う方法が効果的であると考え、また統合に要するオーバヘッドを小さくするために図5.3に示すmin-MAX法を採用した。図5.3において、^はmin演算を示す。まず、各シングルモダリティから得られる感情に対する適合度に対し感情毎にmin演算を行い、小さい方を統合値として採用する。各感情に対する統合値の中から最も大きな値を持つ感情をMAX演算で選択し、最終的な感情判別結果として出力する。

表 5.1 シングルモダリティでの感情判別結果

(a) 表情による感情判別率[%]

感情 / ルール作成に関与	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
○	95.1	50.0	88.4	69.2	47.6	75.7
×	90.6	57.7	84.8	70.8	45.5	84.6

(b) 音声による感情判別率[%]

感情 / ルール作成に関与	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*
○	77.8	64.3	51.5	55.6	95.2	88.9
×	75.0	70.6	66.7	52.2	73.3	60.0

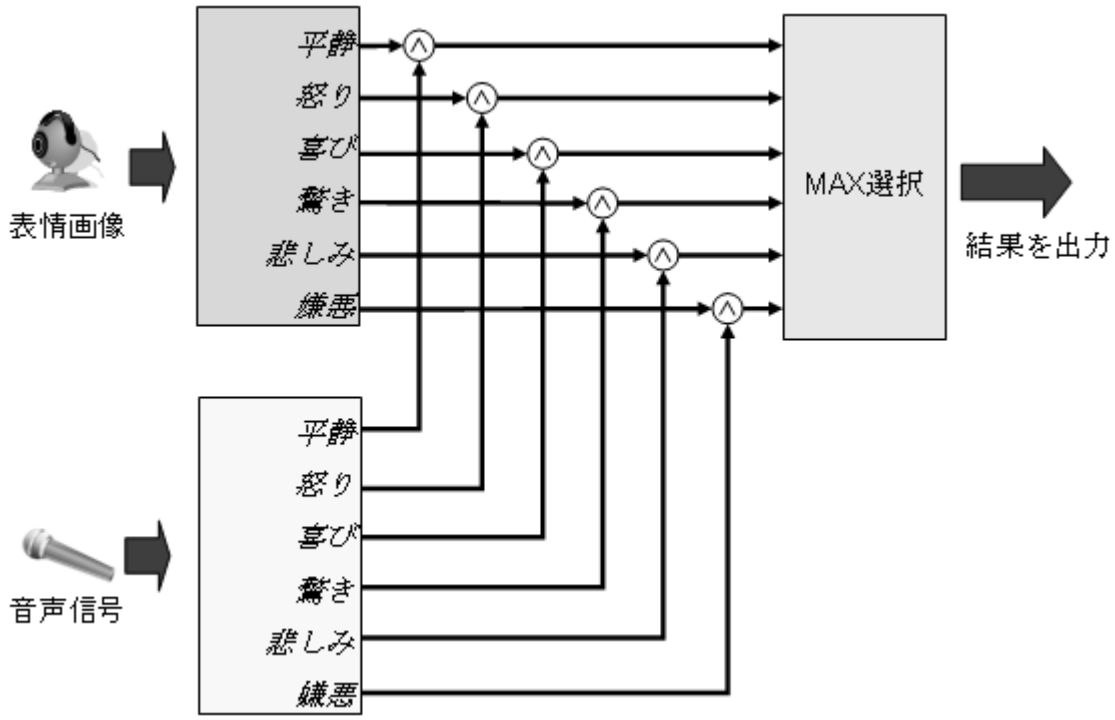


図 5.3. min-MAX 法による適合度の統合

5.3 マルチコアプロセッサへのシステム実装

提案したマルチモーダル感情判別システムでは、各モジュールを同時に実行可能な並列処理と、高速なストリームデータ処理および統合処理が要求される。そこで、図5.4に示すマルチコアプロセッサ Cell Broadband Engine に着目した[80-81]。Cell Broadband Engine は1個の PPE(Power Processor Element) と8個の SPE(Synergistic Processor Element) の2種の異なるアーキテクチャを採用したヘテロジニアスマルチコアプロセッサである。それぞれのプロセッサコアと主メモリおよび外部 I/O は相互接続バス Element Interconnect Bus によって接続されている。PPE は 64bit Power Architecture に準拠した汎用演算コアである。2way Multi-threading 機構を備えており、複数の処理をスレッドと呼ばれる単位で管理し、同時実行可能する。PPEはOSや汎用プログラムの実行に加え、SPEリソースの管理を担当する。SPE は 128bit SIMD(Single Instruction Multiple Data) 型アーキテクチャを採用したプロセッサであり、ストリームデータ演算に特化している。

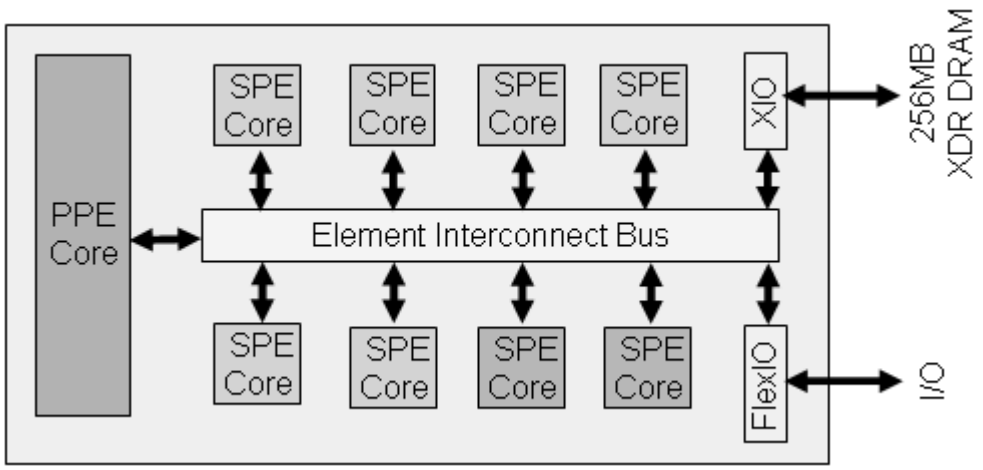


図 5.4. Cell Broadband Engine ブロック図

提案システムの実装概念を図5.5に示す。ここでは提案システムを6SPEが使用可能な PLAYSTATION®3 上に実装した。各モジュールを並列プログラムの実装単位であるスレッドとして各プロセッサコアに割当てる。画像処理は音声信号処理と比較して演算コストが大きいので、5個の SPE を表情スレッドに、1個の SPE を音声スレッドにそれぞれ割当てた。カメラやマイクロフォンから得られるデータは DMA転送によってSPEコアへ直接送られる。各 SPE コアは表情筋の座標パラメータや韻律パラメータを抽出し、PPE コアへ転送する。PPE コアは SPE コアから得られたパラメータからファジィ推論によって感情を判別し、さらに親スレッドにおいて判別結果を統合し、最終的な判別結果を得る。

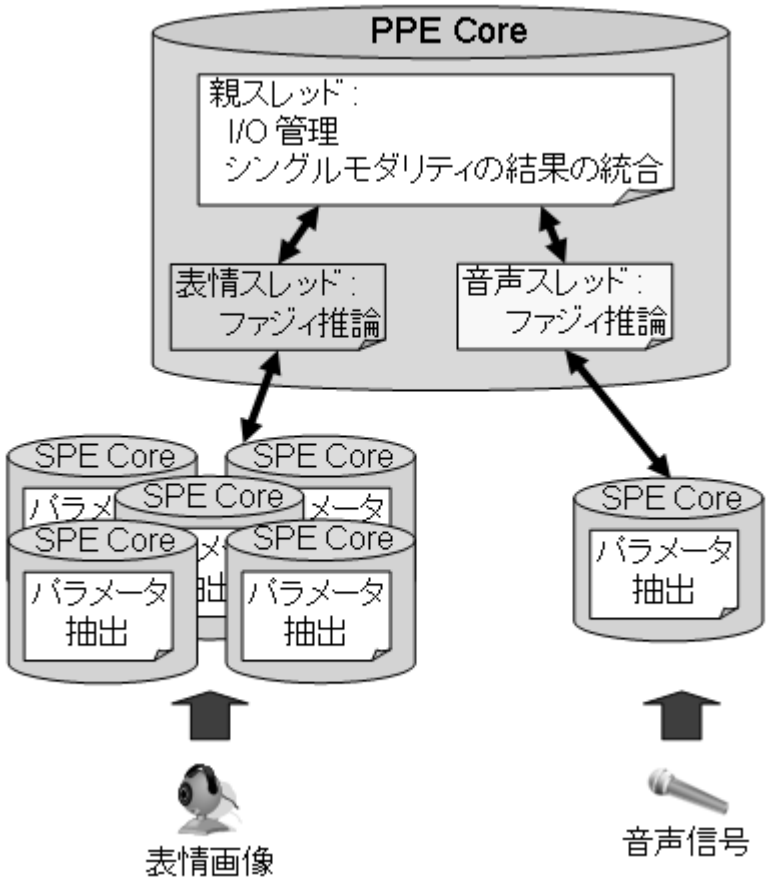


図 5.5. Cell Broadband Engine への提案システム実装概念

提案システムの概観を図5.6に示す。表情画像を取得するためのWebカメラおよび音声信号取得用のモノラルマイクが PLAYSTATION3 に USB 経由で接続されている。Webカメラは USB Video Device Class 規格に準拠しており、640x480ピクセル、24ビットカラー画像が取得可能である。安定した光量を得るために、Webカメラの背後に光源を設置した。音声信号は量子化ビット数16ビット、サンプリングレートは44.1KHz である。

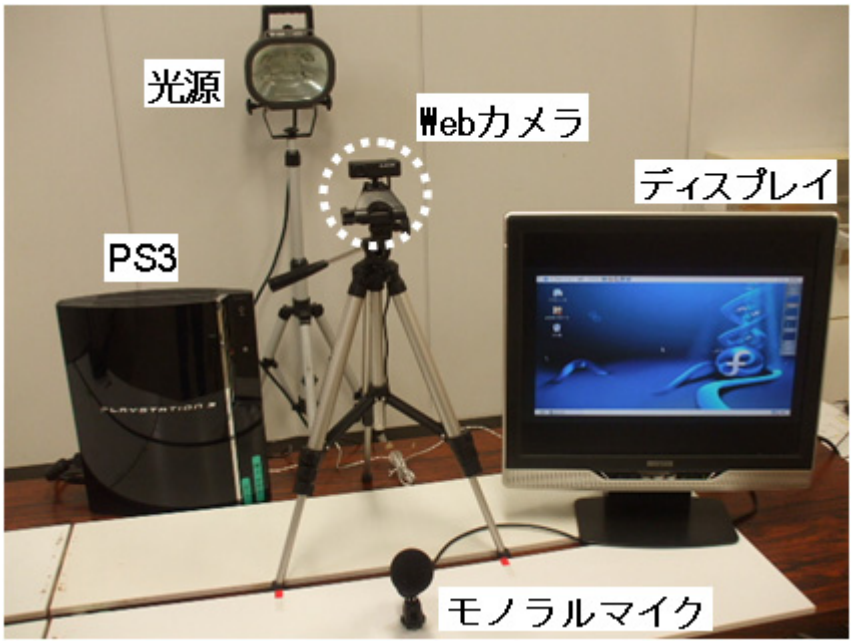


図 5.6. 提案システム概観

感情判別結果は図5.7に示すように、Webカメラで取得した表情画像と適合度グラフとともにディスプレイに表示される。適合度グラフは表情(face)と音声(speech)に対する6感情の適合度を棒グラフで表示する。感情判別結果は表情画像の左上部に表示される。Face および Speech が表情および音声の判別結果を、Multi が統合後の最終判別結果を表す。

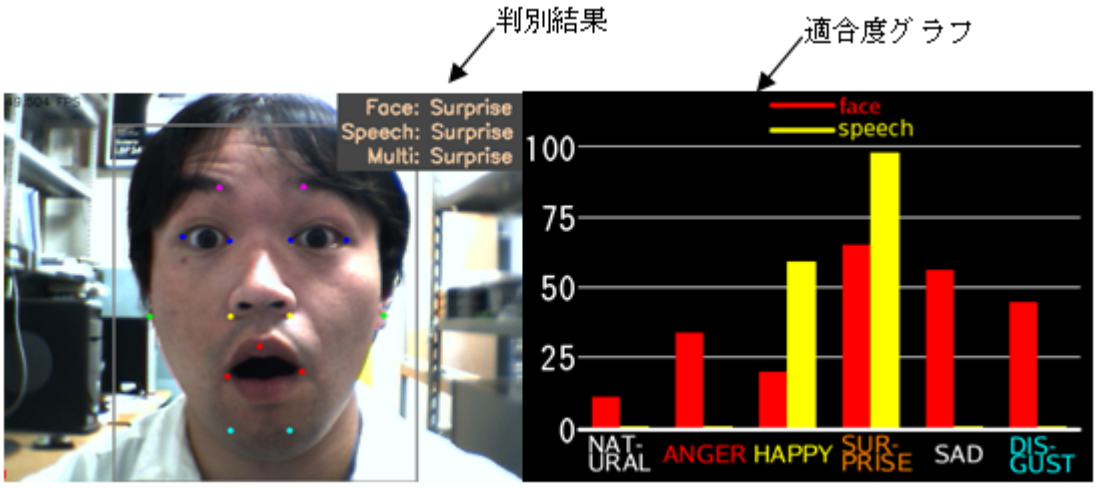


図 5.7. 感情判別結果の例

5.4 マルチモーダルシステムによる感情判別実験

提案システムの妥当性を検証するために感情判別実験を行った。実験用データとして、シングルモダリティにおける感情判別実験と同様に 2.3 節で収集した感情がよく表出されているデータ群から、学習用に用いたデータを除外したものを入力データとして用いた。

5.4.1 判別精度実験結果

マルチモーダル感情判別実験の結果を表 5.2 に、シングルモダリティでの感情判別結果との比較を表 5.3 にそれぞれ示す。表 5.2 より、全体として 80% を超える判別率が達成できている。表 5.3 より、min-MAX 法による統合で互いのモダリティの誤判別を抑制し合った結果、*"怒り"* の判別率は両シングルモダリティでの感情判別率を上回っていることがわかる。また、表情による感情判別システムでは不得意としていた *"悲しみ"* についても音声による感情判別結果と統合して 70% を超える判別率を達成している。

表 5.2. マルチモーダル感情判別システム 感情判別結果

(a) ルール作成に関わった被験者に対する感情判別結果

出力 / 入力	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*	正答数 / 総数	正答率 [%]
平静	**66**	0	0	0	6	1	66 / 73	90.4
怒り	2	**50**	0	3	0	1	50 / 56	89.3
喜び	4	5	**72**	3	3	0	72 / 87	82.8
驚き	5	3	8	**56**	6	0	56 / 78	71.8
悲しみ	0	0	0	6	**59**	0	59 / 65	90.8
嫌悪	0	1	0	0	4	**25**	25 / 30	83.3
計							328 / 389	84.3

(b) 未知の被験者に対する感情判別結果

出力 / 入力	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*	正答数 / 総数	正答率 [%]
平静	**79**	4	0	0	2	0	79 / 85	92.9
怒り	1	**49**	4	0	1	3	49 / 59	83.1
喜び	0	0	**68**	11	0	0	68 / 79	86.1
驚き	6	3	21	**65**	0	0	65 / 95	68.4
悲しみ	2	6	2	2	**44**	1	44 / 57	77.2
嫌悪	0	1	1	2	7	**32**	32 / 43	74.4
計							337 / 418	80.6

表 5.3.感情判別システム 感情判別率の比較

(a) ルール作成に関わった被験者に対する感情判別率 [%]

モダリティ	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*	Total
表情	95.1	50.0	88.4	69.2	47.6	75.7	77.7
音声	77.8	64.3	51.5	55.6	95.2	88.9	71.4
マルチモーダル	90.4	89.3	82.8	71.8	90.8	83.3	84.3

(b) 未知の被験者に対する感情判別率 [%]

モダリティ	*平静*	*怒り*	*喜び*	*驚き*	*悲しみ*	*嫌悪*	Total
表情	90.6	57.7	84.8	70.8	45.5	84.6	76.1
音声	75.0	70.6	66.7	52.2	73.3	60.0	67.2
マルチモーダル	92.9	83.1	86.1	68.4	77.2	74.4	80.6

5.4.2 実行速度に対する考察

マルチモーダル感情判別システムの1フレーム当りの処理時間を図5.8に示す。比較として、プロセッサコアを4個搭載した Intel Core2 Quad Q6600 2.4GHz (C2Q) および Cell Broadband Engine と同一クロック周波数の、ハイパースレッディングテクノロジによって複数スレッドを同時実行可能な Intel Pentium4 2.4GHz (P4) の1フレームの処理時間を同時に示している。比較用の Intel プロセッサでは、コンパイラは gcc-4.1.1 および icc-10.1 を、最大最適化オプションを付与して使用した。図5.8より、Cell Broadband Engine は1フレーム当り 21.4 ms で処理を行う。すなわち、約 46.7 fps で動作可能であり、30 fps を超えるリアルタイム処理を実現している。一方、比較用プロセッサではいずれの場合も30fpsに達しておらず、リアルタイム処理が達成できたとは言えない。さらに、比較用プロセッサでは、判別結果の統合処理やファジィ推論を担当したプロセッサコアにおいて、特徴抽出を担当したプロセッサコアの処理の終了を待つためのブロッキングが常に発生し、結果としてプロセッサ全体の性能を最大に使い切ることができなかった。つまり、比較用プロ

セッサは全てのコアが同等の性能を持つホモジニアスマルチコアプロセッサであるので、本システムのように処理コストに偏りのある階層型モデルには不適であると考えられる。

min-MAX法による判別結果の統合には 55 クロックを要した。 Cell Broadband Engine のクロック周波数は 3.2GHz であるので、モダリティの統合によるオーバヘッドは約 17.2 ns となり、極めて小さなオーバヘッドで統合処理が実現できた。以上より、Cell Broadband Engine がリアルタイムなマルチモーダル感情判別システムに適していることを確認した。

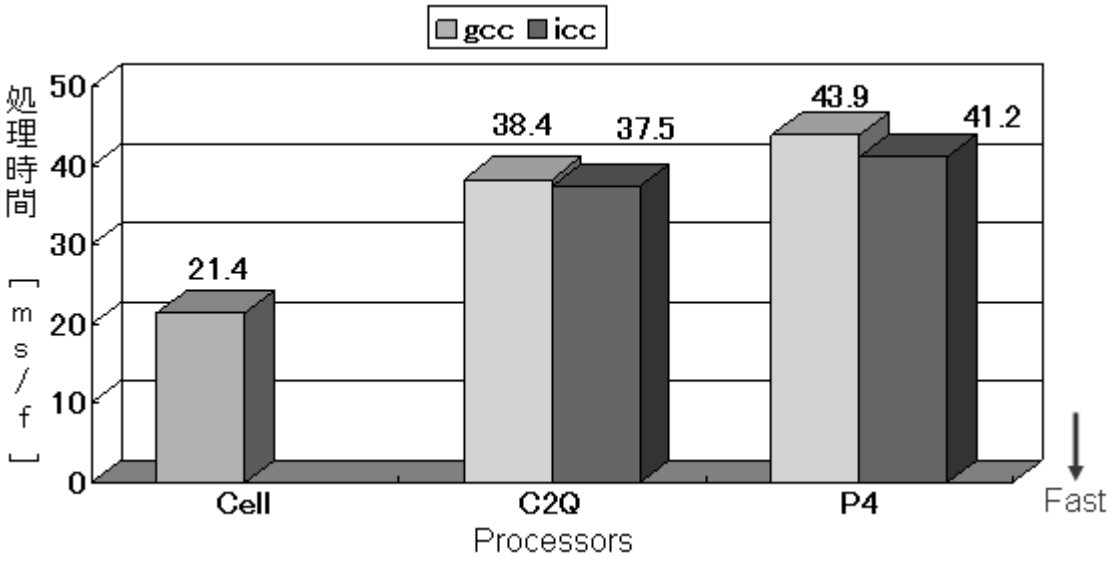


図 5.8. 処理時間の比較

5.5 結言

本章では、高精度な感情判別の実現のために表情による感情判別および音声による感情判別を統合したマルチモーダル感情判別システムを提案した。本システムは、各モダリティ間の同期処理を簡単化するために、特徴ベクトル作成時のタイミング調整が不要な決定レベルでの統合を採用、拡張性に優れた階層構造アーキテクチャを有する。min-MAX 法によるモダリティの統合は、互いの

モダリティの判別ミスを抑制し合うように働き、平均して 80% を超える判別率が得られた。また、階層構造はヘテロジニアスマルチコアプロセッサ Cell Broadband Engine のプログラムモデルに合っており、45fps を超えるリアルタイム実行が実現できた。決定レベルでの統合においては、各モダリティにおける判別結果の統合に要するオーバヘッドが大きくなると感情判別処理が間に合わなくなるという問題点があるが、min-MAX 法によるモダリティの統合に要したコストは 55 クロックであり、極めて小さなオーバヘッドで統合が実現できた。つまり、3.2GHz で動作する Cell Broadband Engine の場合では約 17.2ns のオーバヘッドでモダリティを統合可能であり、次の統合処理に十分間に合うので、リアルタイム実行の観点からも min-MAX 法は有効であると考えられる。

第 6 章 結論

本論文では、機械と人間が共存する社会において重要となる知的インタフェースの１つとして感情判別システムを提案した。人間の感情を分類する方法として、機械学習による方法が多く提案されており、70% を超える判別率が報告されている。しかし、機械学習による感情の分類には多量の感性データが必要であり、また結果として得られた知識は一般に人間には理解し難く、追加学習も難しいという問題がある。一方、機械学習における問題を解消する方法の 1 つとして、人間が理解可能な〝言葉〟を用いるルールベースで感情を定義する方法がある。そこで、本論文ではルールベースの感情判別に関する下記システムを提案、その効果について明らかにした。

- 表情による感情判別

ファジィ推論を用いた表情による感情判別システムを提案した。各感情における表情筋の変化を SOM によってベクトル量子化し、さらに画像処理における分離度を SOM に適用する方法を提案して、定量的に表情筋の変化を分類した。本分類手法は感情の判別に対して重要でない筋肉の特定にも貢献することを明らかにした。感情判別実験の結果、〝*平静*〟、〝*喜び*〟、〝*嫌悪*〟に関しては 75% を超える判別率が得られた。一方、〝*怒り*〟の表情は〝*平静*〟や〝*嫌悪*〟に誤判別しやすい傾向が見られることを示した。

- 音声による感情判別

ファジィ推論を用いた音声による感情判別システムを提案した。音声による感情判別システムは、韻律として声の大きさ、抑揚強度および高さを用いて感情を判別する。ファジィルールの構築では、ルール作成用に収集したデータの歪度および尖度を根拠に、韻律パラメータの母集団は正規分布に従うと仮定して仮説検定による手法を提案した。感情判別実験の結果、〝*悲しみ*〟に対しては 70% を超える良好な判別率が得らることを示した。一方、〝*喜び*〟と〝*驚き*〟に関しては 60% 未満の低い判別率となったが、人間同士のコミュニケーションの場においても、音声のみの場合は〝*喜び*〟と〝*驚き*〟は比較的誤判別されやす

い傾向が報告されているため、人間同士のコミュニケーションにおける感情の誤認識と同様の傾向が得られたと考えられる。

- マルチモーダル感情判別

高精度な感情判別の実現のために、表情による感情判別および音声による感情判別を統合したマルチモーダル感情判別システムを提案した。min-MAX 法によるモダリティの統合は、互いのモダリティの判別ミスを抑制し合うように働き、平均して 80% を超える判別率が得られた。提案システムは階層モジュール型モデルであり、モダリティの拡張性に優れている。また、階層構造はヘテロジニアスマルチコアプロセッサ Cell Broadband Engine のプログラムモデルに合っており、45fps を超えるリアルタイム実行が実現できた。min-MAX 法によるモダリティの統合に要したオーバヘッドは 55 クロックであり、リアルタイム実行の観点からも min-MAX 法は有効であることを示した。

今後は、人間の疲労、ストレス検知システムや、人間の感情に合わせて自身の行動を選択するシステムなどへの応用に取り組んでいく。さらに、将来的な展望として人間との感性コミュニケーションによる自律型行動生成システムの実現へ本システムの寄与が大いに期待できる。

謝辞

本研究を進めるにあたり、熱心な御指導と有益な御助言を賜わりますとともに、終始身をもって研究者としての心得を御教示下さいました、本学生命体工学研究科、神酒勤教授に慎んで感謝の意を表します。また、本論文を執筆するにあたり、多くの有益なご助言をいただいた山川烈特任教授、粟生修司教授、堀尾恵一准教授に感謝の意を表します。共に学び、励ましあってきた神酒研究室の皆さんに感謝いたします。

なお、この学位論文の研究の一部は、21 世紀 COE プログラム「生物とロボットが織りなす脳情報工学の世界」(拠点番号 J19) の推進事業として実施いたしました。関係各位ならびに関係部署に深く感謝いたします。

参考文献

[1] N. Streitz, ″Ambient Intelligence Research Landscapes: Introduction and Overview,″ Lecture Notes in Computer Science, Vol.6439, pp.300-303 (2010)

[2] G. Acampora, M. Gaeta, V. Loia and A. V. Vasilakos, ″Interoperable and adaptive fuzzy services for ambient intelligence applications,″ ACM Transactions on Autonomous and Adaptive Systems, Vol.5, Issue.2, pp.1-26 (2010)

[3] D. J. Cook, J. C. Augusto and V. R. Jakkula, ″Ambient intelligence: Technologies, applications, and opportunities,″ Pervasive and Mobile COmputing, Vol.5, Issue.4, pp.277-298 (2009)

[4] E. Aarts and B. Ruyter, ″New research perspectives on Ambient Intelligence,″ Journal of Ambient Intelligence and Smart Environments, Vol.1, No.1, pp.5-14 (2009)

[5] N. Sebe, I. Cohen and T. S. Huang, ″Multimodal approaches for emotion recognition: a survey,″ Proceedings of the SPIE - The International Society for Optical Engineering, Vol.5670, pp.56-67 (2005)

[6] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.31, No.1, pp.39-58 (2009)

[7] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey, Pattern Recognition," Vol.36, pp.259-275 (2003)

[8] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Vol.22, No.12, pp.1424-1445 (2000)

[9] T. Wu, M. S. Bartlett and J. R. Movellan, "Facial Expression Recognition Using Gabor Motion Energy Filters," IEEE CVPR workshop on Computer Vision and Pattern Recognition for Human Communicative Behavior Analysis, pp.42-47 (2010)

[10] D. Heylen, "Head gestures, gaze and the principle of conversational structure," International Journal of Humanoid Robotics, Vol.3, No.3, pp.1-27 (2006)

[11] P. Y. Oudeyer, "The production and recognition of emotions in speech: Features and algorithms," Int. J. Human-Computer Studies, Vol.59, pp.157-183 (2003)

[12] S. Mitsuyoshi et al., "Non-verbal Voice Emotion Analysis System," International Journal of Innovative Computing, Information and Control, Vol.2, No.4, pp.819-830 (2006)

[13]T. Pao, Y. Chen and J. Yeh, "Emotion Recognition and Evaluation from Mandarin Speech Signals," International Journal of Innovative Computing, Information and Control, Vol.4, No.7, pp.1695-1709 (2008)

[14]J. Kim and E. Andr, "Multi-Channel Biosignal Analysis for Automatic Emotion Recognition," BIOSIGNALS, pp.124-131 (2008)

[15]S. Shigeno, "Cultural similarities and differences in the recognition of audio-visual speech stimuli," International Conference on Spoken Language Processing-1998, paper 1057, pp.281-284 (1998)

[16]A. Jaimes and N. Sebe, "Multimodal Human Computer Interaction: A Survey", COMPUTER VISION AND IMAGE UNDERSTANDING, vol.108, pp.116-134 (2007)

[17]C. Busso et al., "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information," Proceedings of the ICMI'04, pp.205-211 (2004)

[18]E. Schapira and R. Sharma, "Experimental evaluation of vision and speech based multimodal interfaces," Workshop on Perceptive User Interfaces, pp.1-9 (2001)

[19]S. Emerich, E. Lupu and A. Apatean, "Emotions Recognition by Speech and Facial Expressions Analysis," 17th European Signal Processing Conference, EUSIPCO'09, pp.24-28 (2009)

[20]甘利俊一, "神経回路網の数理 ―脳の情報処理様式―," 産業図書 (1978)

[21]V. Vapnik, ″The Nature of Statistical Learning Theory,″ NY Springer (1995)

[22]L. E. Baum and T. Petrie, ″Statistical Inference for Probablistic Functions of Finite State Markov Chains,″ The Annals of Mathematical Statistics, Vol.37, No.6, pp.1554-1563 (1966)

[23]L. A. Zadeh, ″Fuzzy Algorithms,″ *Information and Control*, Vol.12, pp.94-102 (1968)

[24]M. Mufti and A. Khanam, ″Fuzzy Rule Based Facial Expression Recognition,″ IEEE Intelligent Agents, Web Technologies and Internet Commerce'06, pp.57-61 (2006)

[25]H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi, ″Recognition of Six Basic Facial Expressions by Feature-Points Tracking using RBF Neural Network and Fuzzy Inference System,″ IEEE International Conference on Multimedia and Expo 2004, Vol.2, pp.1219-1222 (2004)

[26]A. A. Razak, R. Komiya and M. I. Z. Abidin, ″Comparison Between Fuzzy and NN Method for Speech Emotion Recognition,″ Third International Conference on Information Technology and Applications, Vol.1, pp.297-302 (2005)

[27]R. Dawkins, 垂水 雄二訳, ″進化の存在証明,″ 早川書房 (2009)

[28]R. Plutchik, ″The emotions: Facts, theories and a new model,″ Randam House (1962)

[29]P. Ekman and W. V. Friesen, ″Constants across cultures in the face and emotion,″ Journal of Personality and Social Psychology, Vol.17, No.2, pp.124-129 (1971)

[30]A. M. Isen, K. A. Daubman and G. P. Nowicki, ″Positive affect facilitates creative problem solving,″ Journal of Personality and Social Psychology, Vol.52, No.6, pp.1122-1131 (1987)

[31]岡田顕宏, 阿部純一, ″心理学における感情研究の歴史と動向,″ 日本ファジィ学会誌, Vol.12, No.6, pp.730-740 (2000)

[32]H. Schlosberg, ″A scale for judgment of facial expressions,″ Journal of Experimental Psychology, No.29, pp.497-510 (1941)

[33]H. Schlosberg, ″The description of facial expressions in terms of two dimensions,″ Journal of Experimental Psychology, Vol.44, No.4, pp.229-237 (1952)

[34]J. L. Tsai, B. Knutson and H. H. Fung, ″Cultural Variation in Affect Valuation,″ Journal of Personality and Social Psychology, Vol.90, No.2, pp.288-307 (2006)

[35]C. Darwin, ″The expression of the emotions in man and animals,″ University of Chicago Press (1872)

[36]M. Weiser, R. Gold and J. S. Brown, ″The origins of ubiquitous computing research at PARC in the late 1980s,″ IBM Systems Journal, Vol.38, pp.693-696 (2010)

[37]M. Weiser, ″Ubiquitous Computing,″ Computer, Vol.26, No.10, pp.71-72 (1993)

[38]G. D. Adowd, E. D. Mynatt, ″Charting past, present, and future research in ubiquitous computing,″ ACM Transactions on Computer-Human Interaction, Vol.7, pp.29-58 (2000)

[39]K. Kakousis, N. Paspallis, G. A. Papadopoulos, ″A survey of software adaptation in mobile and ubiquitous computing,″ Enterprise Inormation Systems, Vol.4, Issue.4, pp.355-389 (2010)

[40]福田収一, 綿貫敬一責任編集, ″感覚・感情とロボット 人と機械のインタラクションへの挑戦,″ 工業調査会 (2008)

[41]山下利之, ″心のインタラクション,″ 社団法人 日本機械学会編, 福田収一責任編集, HCD ハンドブック―人間中心設計, pp.57-75 (2006)

[42]山下利之, ″心理学から見たコミュニケーション―ヒューマンコンピュータインタラクションとの関連―,″ 機械の研究, 第 59 巻, 第 1 号, pp.198-203 (2007)

[43]M. S. Bartlett, P. A. Viola, T. J. Sejnowski, B. A. Golomb, J. Larsen, J. C. Hager and P. Ekman, ″Classifying Facial Action,″ Advances in Neural Information Processing Systems, Vol.8, pp.823-829 (1996)

[44]Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu, ″Comparison between Geometry-based and Gabor-Wavelets-based Facial Expression Recognition using Multi-Layer Perceptron,″ IEEE Proceedings of the Second International Conference on Automatic Face and Gesture

Recognition, pp.454-459 (1998)

[45]C. L. Lisetti and D. E. Rumelhart, ″Facial Expression Recognition using a Neural Network,″ Proceedings of the 11th International Flairs Conference, AAAI Press, pp.328-332 (1998)

[46]H. Kobayashi and F.Hara, ″Facial Expression Recognition and its Degree Estimation,″ IEEE Conference on Computer Vision and Pattern Recognition, pp.295-300 (1993)

[47]J. Cohn, A. Zlochower, J. J. Lien, Y. T. Wu and T. Kanade, ″Automated Face Coding: A Computer-Vision based Method of Facial Expression Analysis,″ 7th European Conference on Facial Expression Mearurement and Meaning, pp.329-333 (1997)

[48]Z. Zeng, J. Tu, B. M. Pianfetti and T. S. Huang, ″Audio-Visual Affective Expression Recognition Through Multistream Fused HMM,″ IEEE Transactions on Multimedia, Vol.10, pp.570-577 (2008)

[49]P. Ekman and W. V. Friesen, ″Facial Action Coding System: A Techniquie for the Measurement of Facial Movement,″ Consulting Psychologists Press, Palo Alto (1978)

[50]P. Ekman, W. V. Friesen, ″UNMASKING THE FACE,″ MALOR BOOKS, Cambridge (1993)

[51]P. Viola and M. J. Jones, ″Rapid Object Detection using a Boosted Cascade of Simple Features,″ IEEE Computer Vision and Pattern

Recognition, Vol.1, pp.511-518 (2001)

[52]R. Lienhart and J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection," IEEE ICIP 2002, pp.900-903 (2002)

[53]T.Kohonen, "Self-organizing maps," Springer series in information sciences (2001)

[54]S. Kaski, J. Kangas, and T. Kohonen, "Bibliography of self-organizing map (SOM) Papers: 1981-1997," Neural Computing Surveys, Vol.1, pp.102-350 (1998)

[55]M. Oja, S. Kaski, and T. Kohonen, "Bibliography of self-organizing map (SOM) Papers: 1998-2001," Neural Computing Surveys, Vol.3, pp.1-56 (2003)

[56]A. Ultsch, H. Siemon, "Technical Report 329," University of Dortmund, Dortmund, Germany, (1989)

[57]M. A. Kraaijveld, J. Mao, A. K. Jain, "A Non-Linear Projection Method Based on Kohonen's Topology Preserving Maps," IEEE Trans. Neural Networks, Vol.6, pp.548-559 (1995)

[58]J. Canny, "A computational Approach To Edge Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.8, pp.679-714 (1986)

[59]O. R. Vincent, O. Folorunso, "A Descriptive Algorithm for Sobel Image Edge Detection," Proceedings of Informing Science & IT Education

Conference 2009, pp.97-107 (2009)

[60]R. Maini and J. S. Sohal, "Performance Evaluation of Prewitt Edge Detector for Noisy Images," GVIP Journal, Vol.6, Issue 3, pp.39-46 (2006)

[61]K. Fukui, Edge Extraction Method Based on Separability of Image Features," IEICE TRANS. INF. & SYST., Vol.E78-D, No.12, pp.1533-1538 (1995)

[62]N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Trans. Sys., Man, and Cybernetics, SMC-9, No.1, pp.62-66 (1979)

[63]岡本道雄監訳, R. V. Putz, R. Pabst, "Sobotta 図説人体解剖学 第４版," 医学書院 (1996)

[64]J. Vesanto, J. Himberg, E. Alhoniemi and J. Parhankangas, "SOM toolbox for Matlab 5," In Technical Report A57 (2000)

[65]L. Devillers and L. Vidrascu, "Real-life emotions detection with lexical and paralinguistic cues on Human-Human call cener dialogs," International Conference on Speech and Language Processing, pp.801-804 (2006)

[66]M. Shami and W. Verhelst, "An evaluation of the robustness of existing supervised machine learning approaches to the classification of emotions in speech," Speech Communication, Vol.49, Issue.3, pp.201-212 (2007)

[67]K. P. Truong and D. A. Leeuwen, ″Automatic Discrimination between Laughter and Speech,″ Speech Communication, Vol.49, Issue.2, pp.144-158 (2007)

[68]I. Vasilescu and L. Devillers, ″Detection of Real-Life Emotions in Call Centers,″ Proceedings of The 18th Inter. Conf. on Spoken Language Processing, pp.1841-1844 (2005)

[69]S. Matos, S. S. Birring, I. D. Pavord and D. H. Evans, ″Detection of Cough Signals in Continuous Audio Recordings Using HMM,″ IEEE Trans. Biomedical Eng., Vol.53, No.6, pp.1078-1083 (2006)

[70]福田収一, 松浦慶総, ″音による感情理解,″ 日本機械学会論文集(C 編), Vol.62, No.598, pp.2293-2298 (1996)

[71]森重実, ″感情の判別分析からみた感情音声の特性,″ 電子情報通信学会論文誌, Vol.J83-A, No.6, pp.726-735 (2000)

[72]森山剛, 森真也, 小沢慎治, ″韻律の部分空間を用いた感情音声合成,″ 情報処理学会論文誌, Vol.50, No.3, pp.1181-1191 (2009)

[73]村上正康, 安田正實, ″統計学演習,″ 培風館 (1989)

[74]C. I. Bliss, ″Statistics in biology statistical methods for research in the natural sciences,″ Vol.1, McGraw-Hill (1967)

[75]B. Schuller, R. Muller, B. Hornler, A. Hothker, H. Konosu and G. Rigoll,″ Audiovisual Recognition of Spontaneous Interest within Conversations,″ Proceedings of the Ninth ACM Int'l Conf. Multimodal

Interfaces, pp.30-37 (2007)

[76]Z. Zeng, Z. Zhang, B. Pianfetti, J. Tu and T. S. Huang, "Audio-Visual Affect Recognition in Activation-Evaluation Space," Proceedings of the 13th ACM Int'l Conf. Multimedia, pp.828-831 (2005)

[77]H. J. Go, K. C. Kwak, D. J. Lee and M. G. Chun, "Emotion Recognition from Facial Image and Speech Signal," Proceedings of the Int'l Conf. Soc. of Instrument and Control Engineers, pp.2890-2895 (2003)

[78]S. Hoch, F. Althoff, G. McGlaun and G. Rigoll, "Bimodal Fusion of Emotional Data in an Automotive Environment," Proceedings of the 30th Int'l Conf. Acoustics, Speech, and Signal Processing, Vol.2, pp.1085-1088 (2005)

[79]Y. Sato, T. Miki and K. Honda, "Multimodal Emotion Extraction from Facial Expressions and Voice, and Its Multi-core Processor Implementation", Proceedings of the SCIS & ISIS 2008, pp.1932-1937 (2008)

[80]J. A. Kahle et.al., "Introduction to the Cell Multiprocessor," IGM J. Res. & Dev.49, No.4, pp.589-604 (2005)

[81]T. Chen et.al., "Cell Broadband Engine Architecture and its first implementation," IBM developerWorks (2005).