ファジー制御を用いた音声における情緒性評価法

森山　剛[†]　　小沢　慎治[†]

A Measurement of Human Vocal Emotion Using Fuzzy Control

Tsuyoshi MORIYAMA[†] and Shinji OZAWA[†]

あらまし　本論文では，音声の韻律成分を計測し，含まれている情緒性が既知である教師音声からあらかじめ学習しておいた，音声の物理パラメータと含まれる情緒性との関係を用いて，音声に含まれる情緒性を評価するシステムを提案する．ここでは『感情』と『情緒性』について区別した上で，音声コミュニケーションにおける情緒性の伝達についてのモデルを示している．更に，コミュニケーションの当事者である話し手及び聞き手の他に観測者という立場を考え，観測者の計算モデルを実装する形で情緒性評価システムを構築する．教師音声の学習には，言葉によってシステムを記述することのできる利点をもつファジー制御を用いた．構築したシステムによる情緒性評価と被験者による主観評価とを比較する実験を行ったところ，人並の情緒性評価が行えることが示された．

キーワード　感情情報，ファジー制御，音声処理，韻律，感性情報処理

1. まえがき

近年，豊かさに対する認識が変化しつつある中で，豊かさを提供するものに求められている役割もまた，変化しつつあるといわれている．すなわち，これまで人はめんどうな仕事の代行という役割をものに求めてきたが，そのことは結果として人間同士のコミュニケーションの機会を奪い，その中で達成されていた情緒的な代謝 [1] を不足させてしまったというのである．そのため，今度は「人にやさしい機械」というキャッチフレーズなどに代表される，情緒的代謝のパートナーとしての役割を，ものに期待するようになってきたのである．ものを作る立場からそのような期待にこたえるためには，このような情緒的な役割を担うことのできるものをつくる必要がある．実際には，人間同士で日常的に行われている情緒のコミュニケーションを分析し，ものと人間との間のコミュニケーションにおいて備えるべき能力，すなわち人間の表出した情緒を認識したり，情緒表現を行ったりすることをものに身につけさせることが必要となる．

本研究では,『情緒性』とは感情表現を客観的に観察する側に共有されるステレオタイプを指していると考え，いわゆる『感情』が話し手などの主観的な心理状態を指しているのに対して区別している．例えば，もともと怒っているような話し方の話し手では，本人が「怒り」の『感情』を抱いていないにもかかわらず,「怒り」の『情緒性』が含まれている場合が存在する．また演劇などの場合でも，申し合わされた「怒り」の感情表現を演者と観客の間で共有しており，演者が本当に怒っているか否かにかかわらず「怒り」の情緒性が共有される．コミュニケーションにおける文脈的状況や話し手の感情に比して，このステレオタイプが音声の物理的側面から観測可能である，という点については，Streeter ら [2] も指摘している．

従来，情緒性に関する研究では，感情の構造や性質に対する心理学的研究 [3], [4] や，顔画像における表情の認識合成の研究 [5] などは盛んに行われているが，音声に関する研究は，いまだ十分な成果を上げているとはいえない．しかし聴覚メディアの重要性や音声インタフェースの有用性などから，音声からの情緒性情報の抽出 [6], [10] や平静音声への情緒性付与 [11], [17] を目的として，音声に含まれる情緒性と物理的なパラメータの関係を明らかにしようという試みが行われている．

従来，情緒性の伝達に関して，音声の韻律成分が大

[†] 慶應義塾大学理工学部情報工学科，横浜市
Department of Information and Computer Science, Keio Univ., 3–14–1 Hiyoshi Kouhoku-ku, Yokohama-shi, 223-8522 Japan

きく寄与しているという報告があるが，この韻律成分については，更にパラメータ分析の容易さも指摘されていることから，本研究では音声の韻律成分（ピッチやパワーの時間構造，発話長）を計測することによって情緒性の評価を行う手法を提案する．このためには，ある情緒性が伝搬される際の韻律パラメータ間の関係について考慮する必要があるが，筆者らが行った統計的実験の結果，韻律パラメータ間には高い相関があり [18]，更に感情語についても，これを評価語として種々の音声についての主観評価実験を行った結果，感情語間に高い相関が認められた．したがって，情緒性や韻律パラメータ間に存在する非線形関係を考慮し，これらの間の関係付けを行う必要がある．関係付けを行う方法としては，線形モデルやニューラルネットワークによる方法などが考えられるが，システムのチューニングに関して専門的な知識を必要としたり，また一度構成されたシステムを更新するのに多大な労力を要したりという問題がある．これに対して，人間の判断など，あいまいさを含む複雑かつ非線形な制御アルゴリズムを実装することができ，既に工学的に実績を上げているファジー制御の方法は，入力されるパラメータに関する重み付けや制御規則が言語で記述され，感覚的な把握が容易であるため，比較的容易にシステムの構築及び更新を行うことができる．したがって，ファジー制御を感情音声の合成に用いた例もある [28]．本論文では，ファジー制御を新たに音声に含まれる情緒性の評価に適用する方法を提案する．

2. 情緒性評価のモデル

2.1 情緒性のコミュニケーション

音声による情緒性のコミュニケーションの模式図を図 1 に示す．ここでは，コミュニケーションの参加者としての話し手と聞き手が，更にコミュニケーションには参加していない第三者としての観測者が考えられる．話し手が音声に込めようと意図した感情は，話し手を取り巻く環境などの外的要因，直前の感情状態及び生理的状態などの内的要因などの影響を受けて音声に含まれると考えられるが，これらの要因が観測されたとしても情報の部分性 [20] を考慮すると，意図した感情そのものをこれらの要因から間接的に決定するのは困難であると考えられる．また，コミュニケーションの当事者である聞き手が話し手の音声から知覚する感情も，聞き手自身の外的要因，直前の感情状態及び内的要因などの影響を受けていると考えられ，話し手

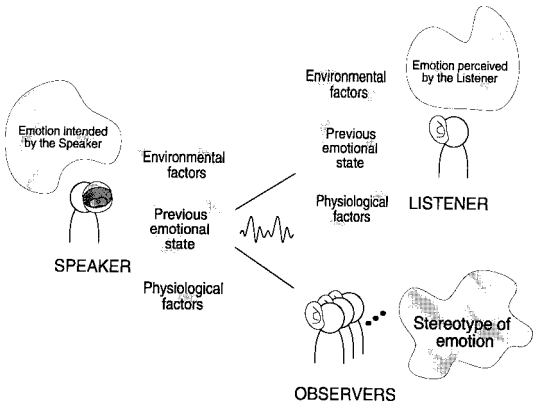


図 1　情緒性のコミュニケーション
Fig. 1　Emotion communication by speech.

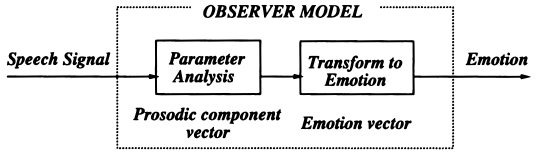


図 2　情緒性評価の計算モデル
Fig. 2　Computational model of measurement of emotion stereotypes.

同様，これらの要因から決定することは困難である．これに対して，話し手から発せられる音声を客観的にとらえ，また前述の情緒性（感情のステレオタイプ）に基づいて音声に含まれる感情の評価を行う観測者においては，知覚される情緒性を統計的手法などにより，客観的に観測することができると考えられる．音声の物理的様相と音声に含まれる情緒性とを関係づけるためには，音声の物理的特徴とともに，音声に含まれる情緒性を観測する必要がある．本論文では，この観測者の観点から情緒性を観測することとする．更に，情緒性が音声の物理的様相によって規定されるととらえると，物理的特徴の間の関係は，いわゆる Zadeh（1973）の「不適合の原理」[21] が成立するため，特徴パラメータ間の制約条件を記述し，それらの条件に対する解を観測者に知覚される情緒性として求める，といった枠組みが有効であると考えられる．

2.2 情緒性評価のモデル

図 2 に観測者による情緒性評価の計算モデルを示す．観測者は音声信号を聞いて情緒性のステレオタイプに基づいた評価を行うが，ここでは計算機上で情緒性評価を実現するために，音声信号を分析して得られ

る物理パラメータと，統計的調査によって得られる情緒性との関係をあらかじめ学習しておき，未知音声のパラメータから情緒性にマッピングすることで情緒性の評価を行う．具体的には，ファジー推論器における if – then 形式のルールでこの関係を記述するとともに，音声から分析されるパラメータに対する感度をメンバシップ関数の形状に反映させることによってシステムを構築する．

3. ファジー制御

ファジー制御システムは，ファジー制御ルールとこれに基づく推論の形式である構造，及び推論の計算法である演算とによって決定される [22]．本研究では，Mamdani が最初に応用に用いたファジー関係の合成則に基づく推論法 [25] を実装する．

N 個のルールの i 番目は式 (1) のような if – then 形式で書かれ，ルール間は or 結合されている．

$$R^i\text{: if } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \text{ then } y_j \text{ is } B_i \quad (1)$$

本研究では，x は韻律パラメータの平静時からの変化分であり，y_j は j 番目の情緒性，A_i は「大きい」や「高い」などのファジー値をとるファジー変数，B_i は「やや含まれている」などのファジー値をとるファジー変数である．すなわち，N 個の if – then ルールによって x，y のファジー関係（ある韻律成分のある変動によってある情緒性がある度合含まれる）を記述する．この関係を獲得する方法に関しては後述する．

ここで，x_1，x_2 に対する入力値を x_{10}，x_{20} としたとき，i 番目のルールの前件部（条件部）命題の適合度 w_i を，

$$w_i = A_{i1}(x_{10}) \wedge A_{i2}(x_{20}) \quad (2)$$

により計算する．ただし，ここではファジー変数 A_{i1}，A_{i2} に相当するメンバシップ関数をそのまま A_{i1}，A_{i2} と書くこととする．

次に，前件部の適合度 w_i が求まると，後半部命題のファジー変数 B_i に対するメンバシップ関数 B_i の現在の度合と w_i との min 演算を行い，次の関数を得る．

$$\tilde{B}_i(u) = w_i \wedge B_i(u) \quad (3)$$

$\tilde{B}_i$ は B_i を度合 w_i のところでカットしたものである．したがって，出力に関するメンバシップ関数 $h(u)$ は式 (3) の最大値をとる次の式 (4) の演算によって得られ，その操作出力値 u_0（ここではその情緒性が含まれる度合）は関数 $h(u)$ の面積の重心として求められる（M は出力メンバシップ関数の峰数）．

$$h(u) = \max_i \tilde{B}_i(u) \quad (4)$$

$$u_0 = \frac{\int h(u)udu}{\int h(u)du} = \frac{\sum_m^M h(u_m)u_m}{\sum_m^M h(u_m)} \quad (5)$$

4. 情緒性評価システム構築

図 3 に構築する情緒性評価システムを示す．入力は平静音声（あらかじめ得ておく）と評価対象となる感情音声とする．それぞれについてパラメータ分析を行い，韻律成分を算出する．感情音声のパラメータを平静音声のそれで正規化したもの（以下変動率）を，ファジー制御システムへ入力し，出力として情緒性の含まれる度合を得る．

情緒性評価システムの本体であるファジー推論器は，

- 入力メンバシップ関数の設計
- ファジー制御ルールの設計

の二つに分けて設計される．これらの設計は，ともに 1）初期設定，2）教師による学習（修正），の二つのステップからなっている．ここではまず，音声に含まれる情緒性と物理パラメータの関係を明らかにするために，含まれる情緒性が既知の音声（教師音声）を収録する必要がある．次にこの教師音声の収録の方法と，システム構築の詳細について説明する．

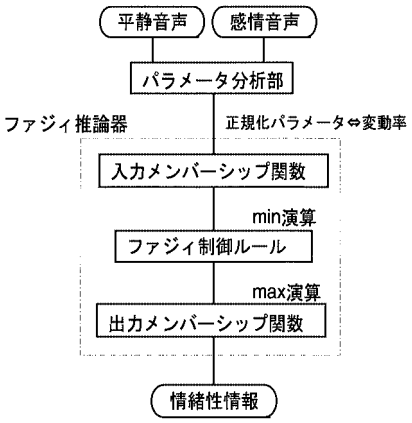


図 3　提案する情緒性評価システム
Fig. 3　Proposed emotion measurement system.

表 1　データ収録に用いた言葉
Table 1　Words used for recording speech data.

#	言葉	モーラ数	アクセント型
1.	えー	2	起伏式頭高型
2.	おい	2	〃
3.	なぜ	2	〃
4.	なんだ	3	起伏式頭高型
5.	みろよ	3	〃
6.	まだか	3	〃
7.	そんな	3	平板式
8.	いやだ	3	起伏式中高型
9.	おまえ	3	平板式

4.1　教師音声の収録

教師音声は「怒り」「悲しみ」「喜び」「恐れ」そして「平静」の五つの情緒性について，表 1 に示す九つの言葉を用い，計 45 感情音声を収録した．例えば「怒り」の音声を収録する場合，話し手がある言葉を用いて行った一発話に対して主観評価実験を行い，すべての被験者が「怒り」のみが含まれていると判定した場合のみ，その音声を「怒り」音声として収録し，それ以外の場合には再び話し手（男性声優 1 名）に発話してもらい，収録されるまでこれを繰り返した．ここで被験者はあらかじめ，言葉の意味に影響されずに情緒性の評価を行えるよう訓練されており，更に平静音声からの相対評価を行うことによって情緒性以外の要因を正規化が行われていると考えられる．ここで被験者は大学生 8 名（男子 6 名，女子 2 名）である．

ここで用いた情緒性は，Plutchik によって提案された円環モデルの八つの基本感情 [23] のうち，予備実験により聞き手にとって認識しやすいと判断され，かつこの円環上で等距離にある四つを選択したものである．表 1 に挙げた言葉は，日常生活において様々な情緒性を含んで発せられるような言葉から無作為に選択した．このような言葉を用いることによって話者が自然に感情表現を行えると考えられる．またこれらの言葉を用いることで，言葉による物理パラメータのばらつきをファジー推論器の設計に反映させる．単語レベルの言葉を用いたのは，コミュニケーションにおける状況や文脈の影響を排除するためである．また，ここでは話し手を 1 名としているが，これはファジー推論器の設計において，同じ情緒性を含む音声の間のばらつきを観測する際に，仮に話し手を複数にすると，音声の物理的特徴（モーラ数など）の言葉によるばらつきだけでなく，話し手の間の個性のばらつきをも同時に観測してしまうため，結果として得られる推論器が情緒性を評価しているのか，個性を評価しているのかがふめいりょうになってしまうのを避けるためである．また話者を 1 名とすることにより，話者を特定することになるが，**8.**で述べるような応用が考えられる．

表 2　韻律パラメータ
Table 2　Selected and calculated prosodic parameters.

a.	パワー平均値
b.	パワーダイナミックレンジ
c.	発話長
d.	ピッチ平均値
e.	ピッチダイナミックレンジ

収録した音声は標本化周波数 44.1 kHz，16 bit 直線量子化で A–D 変換し，ファイルに保存した．

4.2　物理パラメータの選定と算出

従来の研究では，情緒性を含んだ音声の韻律的特徴が調べられてきた [9], [15]〜[17], [26]．韻律的特徴とは声の高・低，強・弱，リズム・テンポを指し，発話意図などのパラ言語情報や感情情報などの非言語情報を伝達する担い手といわれている [26]．本研究でも，この韻律的特徴に注目し，表 2 に挙げる 5 パラメータをすべての音声について算出する．なお，ピッチ及びパワーに関しては，フレーム長約 20 ms の分析フレーム内において平均値を算出するが，フレーム周期約 10 ms でこれをシフトしながら，音声全体に対してその時間的軌跡を求め，音声全体における平均値及びダイナミックレンジをパラメータとして用いる．また発話長とは，まず音声区間以外での音圧レベルの平均を求め，音声区間の境界においてこれを超えた点を音声の始端及び終端として決定し，始端から終端までの時間長を指す．

ピッチ抽出には，FFT ケプストラム分析（リフタリングしきい値 2.8 ms，20.0 ms）を用い，パワーは時間波形における短時間平均パワーを算出した．ピッチは，時間波形の零交差数及び短時間平均パワーにしきい値を設けることにより，有声/無声の自動判別を行い [9]，有声部についてのみ算出した．パワー及び発話長に関しては，無声/有声の別なく算出した．

4.3　メンバシップ関数初期設計

メンバシップ関数の初期値形状は連続直角三角型の形状とし，入力メンバシップ関数は五つの韻律パラメータそれぞれに対して，出力メンバシップ関数は四つの情緒性それぞれに対して設定する．

入力メンバシップ関数は二つの峰を有し，それぞれ

がとるファジー値は，ピッチ平均値では「低い」「高い」，発話長では「短い」「長い」，その他では「小さい」「大きい」である．出力メンバシップ関数は五つの峰を有し，それぞれ 1〜5 までがその情緒性が含まれる度合に対応している．最左峰はその情緒性が含まれる度合が 5 段階のうち 1 で,「全く含まれていない」を意味している．

入力メンバシップ関数については，後述する最適化の方法によって形状を修正し，最終形状を得る．

4.4 ファジー制御ルール初期設定

ファジー制御ルールにおいて記述すべきファジー関係は，五つの韻律パラメータそれぞれが二つのファジー値を有することから $2^5 = 32$ 通りの組合せが存在する．ここで，各入力メンバシップ関数の左の峰に 0，右の峰に 1 の符号をそれぞれ割り当てることによって，この 32 通りの組合せを 5 bit の 2 進数で番地表現し，それぞれの番地に，四つの情緒性それぞれの評価値（含まれる度合）からなる 4 次元ベクトルを割り当てることでルールを作成する．

制御ルールの初期値には，それぞれの情緒性に典型的な韻律パラメータの変動の組合せ（例えば「怒り」に関しては 11001）に対して評価値 5（「全く含まれている」）を，その他の組合せについては，典型的な組合せからのマンハッタン距離に応じて順次低い評価値を与える（例えば，10001 には 4，00001 には 3 を初期設定する）．なお，それぞれの情緒性に典型的なパラメータ変動の組合せは，次の韻律パラメータ変動分布より目視により求める．

5. システム最適化

5.1 韻律パラメータ変動分布

入力メンバシップ関数及びファジー制御ルールを設計するために，各情緒性に典型的な韻律パラメータの挙動を知る必要がある．そこで，教師音声すべてについて韻律パラメータを計算し，感情音声のパラメータを平静音声のそれで正規化したものを変動率 c_{ijk} とし，次式で定義する．

$$c_{ijk} = \frac{p^e_{ijk}}{p^n_{ik}} \times 100\ [\%] \tag{6}$$

p^e_{ijk}，p^n_{ik} は感情音声及び平静音声のパラメータである．i は韻律パラメータを示し（$i = 1, \cdots, 5$），j は情緒性（$j = 1, \cdots, 4$），k は表 1 に挙げた言葉を指している（$k = 1, \cdots, 9$）．これを韻律パラメータ別に，情緒性ごとの分布にまとめたものを韻律パラメータ変動分布と呼び，図 4 に示す．

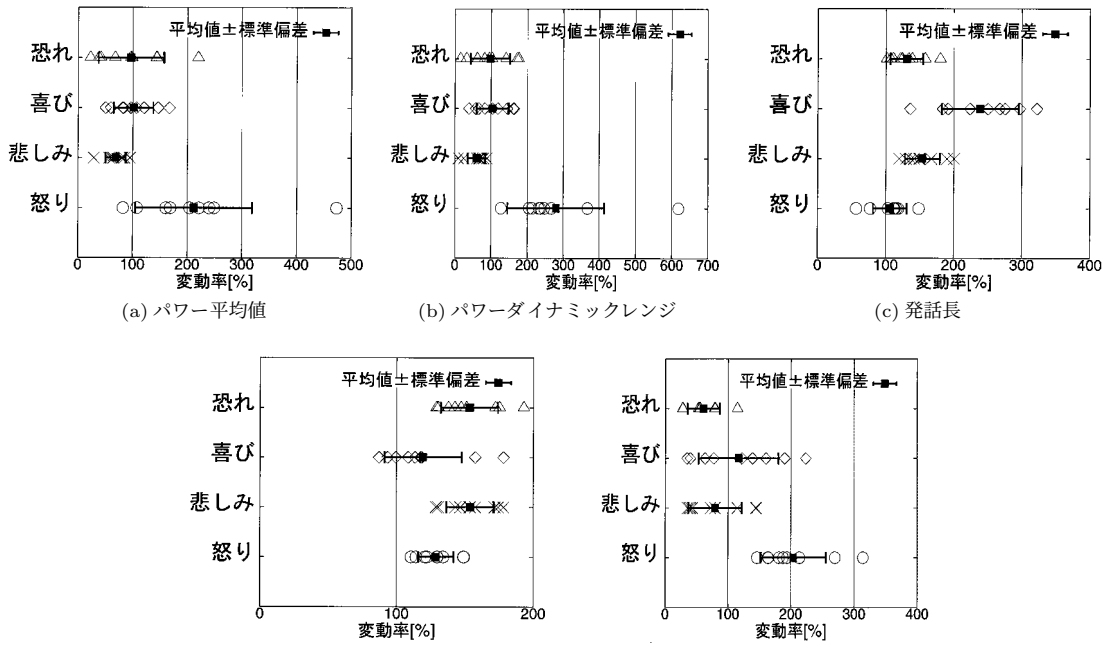
図 4 から,「怒り」はピッチやパワーの平均値，ダイナミックレンジともに増大し，発話長は他の情緒性に比べて短くなり，また「悲しみ」は，ピッチ，パワーともにダイナミックレンジが小さくなり，発話長は長くなることがわかった.「喜び」は，パワーの平均値やダイナミックレンジにおいて平静からあまり変化が見られないが，発話長の増大が特徴的である.「恐れ」は「悲しみ」と非常に近い変動分布を示した（相関係数 0.9755，有意水準 5%）．これらの知見は従来報告されている知見とほぼ一致している [9], [15]〜[17]．しかし「喜び」の発話長に関しては，従来の平静からあまり変化がないとの報告 [15], [17] とは異なり，長くなる傾向にあることがわかった．この違いは感情表現の個人性に起因していると考えられるが,「喜び」においては時間構造の変化があまり重要ではないという報告 [17] もあることから，主観評価で「喜び」の知覚に影響があまりなかったと考えられる．

本研究では，この韻律パラメータ変動分布から入力メンバシップ関数の形状を決定し，その形状をもとに，教師音声の韻律的特徴から求められるシステムの評価値と教師との誤差を最小化するように，ファジー制御ルールの変更を行う．この際，一度決定された入力メンバシップ関数形状も同時にチューニングする．

5.2 入力メンバシップ関数最適化

入力メンバシップ関数は，システムへ入力される韻律パラメータに対して，情緒性間の相違に敏感に反応するような特性を有し，本研究では二峰のメンバシップ関数であるから，物理空間をなるべく 2 分するように設計しなければならない．ここでは，前節で算出した韻律パラメータ変動分布において，各情緒性のセントロイド間の距離を最も強調するようなメンバシップ関数の形状となるよう最適化した．例えばある韻律パラメータにおいて，二つの情緒性セントロイド間で変動分布の平均値が近く，標準偏差が小さいというような場合には，ファジー変数（峰）の交点を二つの平均値の中間にくるようにし，こう配を急しゅんにすることによって，入力される変動率の微小な差異にも大きな出力差を生むようにした．

このように最適化を行い，更に後述するファジー制御ルールの最適化プロセスにおいて変形した結果得られた最終形状を図 5 に示す．



(a) パワー平均値　(b) パワーダイナミックレンジ　(c) 発話長

(d) ピッチ平均値　(e) ピッチダイナミックレンジ

図 4　韻律パラメータ変動分布　（△ 恐れ，◇ 喜び，× 悲しみ，○ 怒り）
Fig. 4　Dynamic distribution of the prosodic parameters.

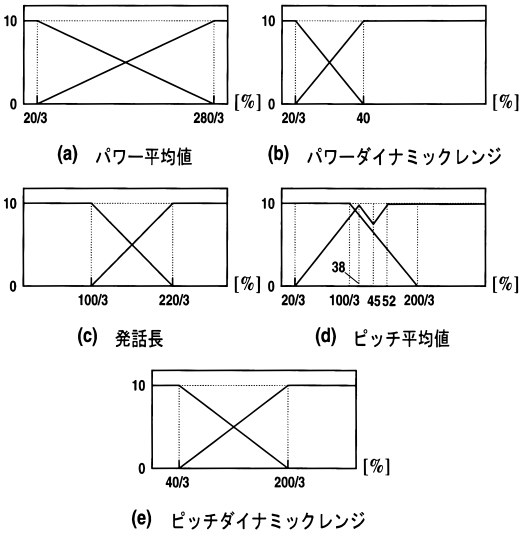


図 5　入力メンバシップ関数
Fig. 5　Input membership functions.

5.3　ファジー制御ルール最適化

ここでは，前節で与えられた入力メンバシップ関数を前提にして，**4.4** においてマンハッタン距離で初期設定された箇所について変更を行う．すなわち，**4.4** ではすべての箇所について韻律パラメータの役割は等しいという仮説に基づいて決めていたが，例えば「怒り」に特徴的なパラメータの変動のうち，ピッチ平均値の増大が生じていても，他のともに起こるべきパラメータの変動がなければ，必ずしも「怒り」は知覚されない．このように韻律パラメータの間には非線形性が存在し，ファジー制御ルールにこれを反映させる必要がある．

例として {1,1,1,1,1} の「怒り」について見ると，初期値として 3 が与えられている．ここでパワー平均値や発話長は「怒り」の伝達に特徴的なパラメータであり，図 4 を見ると，それぞれ「大きい」及び「短い」というファジー値をとる組合せが「怒り」の評価を高くし，それ以外は低くする，という基準が考えられる．そこで，{1,1,1,1,1} は発話長ビットが 1 であることから「怒り」の評価値を下げ，3 に代えて 1 を与えるように変更する．本研究では，ファジー制御を用いた手法の提案が主たる目的であるので，その他，変動分布から最適化しづらい箇所については筆者の主観で変更を加えた．

表 3 に最適化されたファジー制御ルールを示す．表 3 中の a〜e は，表 2 の韻律パラメータである．

表 3　ファジー制御ルール
Table 3　Fuzzy control rules.

入力メンバシップ関数					出力メンバシップ関数			
					怒り	悲しみ	喜び	恐れ
a	b	c	d	e	A	B	C	D
0	0	0	0	0	1	1	1	1
0	0	0	0	1	1	3	1	1
0	0	0	1	0	1	4	1	2
0	0	0	1	1	4	1	2	1
0	0	1	0	0	1	3	2	1
0	0	1	0	1	1	2	5	1
0	0	1	1	0	1	4	2	4
0	0	1	1	1	2	1	4	2
0	1	0	0	0	1	2	1	5
0	1	0	0	1	4	1	2	2
0	1	0	1	0	1	2	1	5
0	1	0	1	1	5	1	2	2
0	1	1	0	0	1	5	2	1
0	1	1	0	1	1	1	5	1
0	1	1	1	0	1	5	2	4
0	1	1	1	1	3	1	5	1
1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	1	2	5
1	0	0	1	0	1	1	1	5
1	0	0	1	1	5	1	3	1
1	0	1	0	0	1	5	2	2
1	0	1	0	1	1	1	5	3
1	0	1	1	0	1	4	2	4
1	0	1	1	1	2	1	5	1
1	1	0	0	0	1	5	2	2
1	1	0	0	1	5	2	4	1
1	1	0	1	0	1	2	1	5
1	1	0	1	1	4	1	1	1
1	1	1	0	0	1	2	4	4
1	1	1	0	1	3	3	4	1
1	1	1	1	0	1	2	4	1
1	1	1	1	1	1	1	4	1

6. 実　　験

構築した情緒性評価システムの性能を評価するために，含まれる情緒性が未知のサンプル音声を用い，主観評価値とシステムによる評価値とを比較する実験を行った．本研究では教師音声の収録に表 1 に示す言葉を用いている．これらの言葉は日常生活において様々な情緒性が含まれるようなものをアクセント型やモーラ数などの音声条件に関して無作為に選択したものであるが，頭高型のアクセント型を有するものが大半を占めた．そこでサンプル音声には，“いいよ”，“よせよ” 及び “やめろよ” の三つの言葉を用いた．“いいよ” 及び “よせよ” は 3 モーラ頭高型の言葉であり，“よせよ” は言葉の意味が情緒性評価に影響を与える可能性があると考えられるものである．また，“やめろよ” は

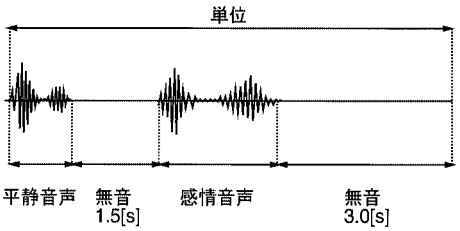


図 6　音声データの構造
Fig. 6　Structure of the speech data.

音声条件が教師音声とは異なっており，更に “よせよ” 同様言葉の意味の影響が考えられる．それぞれの言葉を教師音声と同一話者に 3 回発話してもらい，そのすべてを実験で用いた．それぞれ音声 1〜3，音声 4〜6 及び音声 7〜9 と呼ぶこととする．発話の際に込める感情に関しては，多種多様となるよう話し手に指示し，具体的には話し手に任せた．またサンプル音声の評価の基準となり，かつ変動率の算出に必要な平静音声も併せて収録した．

6.1　主観評価実験

被験者は教師音声収録時と同じ大学生男子 6 名女子 2 名の計 8 名で，全員が四つの情緒性すべてに評価を終えるまで繰り返し聞かせた．その際，音声は図 6 のように平静なものと情緒性を含んだものを組にし，3 秒程度の無音区間を挟んで再生した．評価は 5 段階評定尺度を用い，「前半の音声に比べ，後半の音声にそれぞれの情緒性がどの程度含まれているかを評価して下さい」とあらかじめ指示してから聞かせた．システムによる評価値と比較を行う際には，5 段階評定尺度に対応する点数を付し，それぞれの音声について 8 名の間の平均値と標準偏差を算出し，これを誤差範囲を有する主観評価値とする．結果は次節のシステムによる評価実験の結果とともに示す．

6.2　システムによる評価実験

6 サンプル音声それぞれについて，式 (6) に基づいて各韻律パラメータの変動率を算出し，構築した情緒性評価システムに入力することにより，情緒性の評価を行った．

6.3　比 較 結 果

図 7 に主観評価実験及びシステムによる評価実験の結果を併せて示す．ここでは，×印は主観評価値の被験者間平均値を，□印はシステムによる評価値を表している．主観評価値については更に，被験者間の標準偏差を誤差範囲として線分で示している．出力される

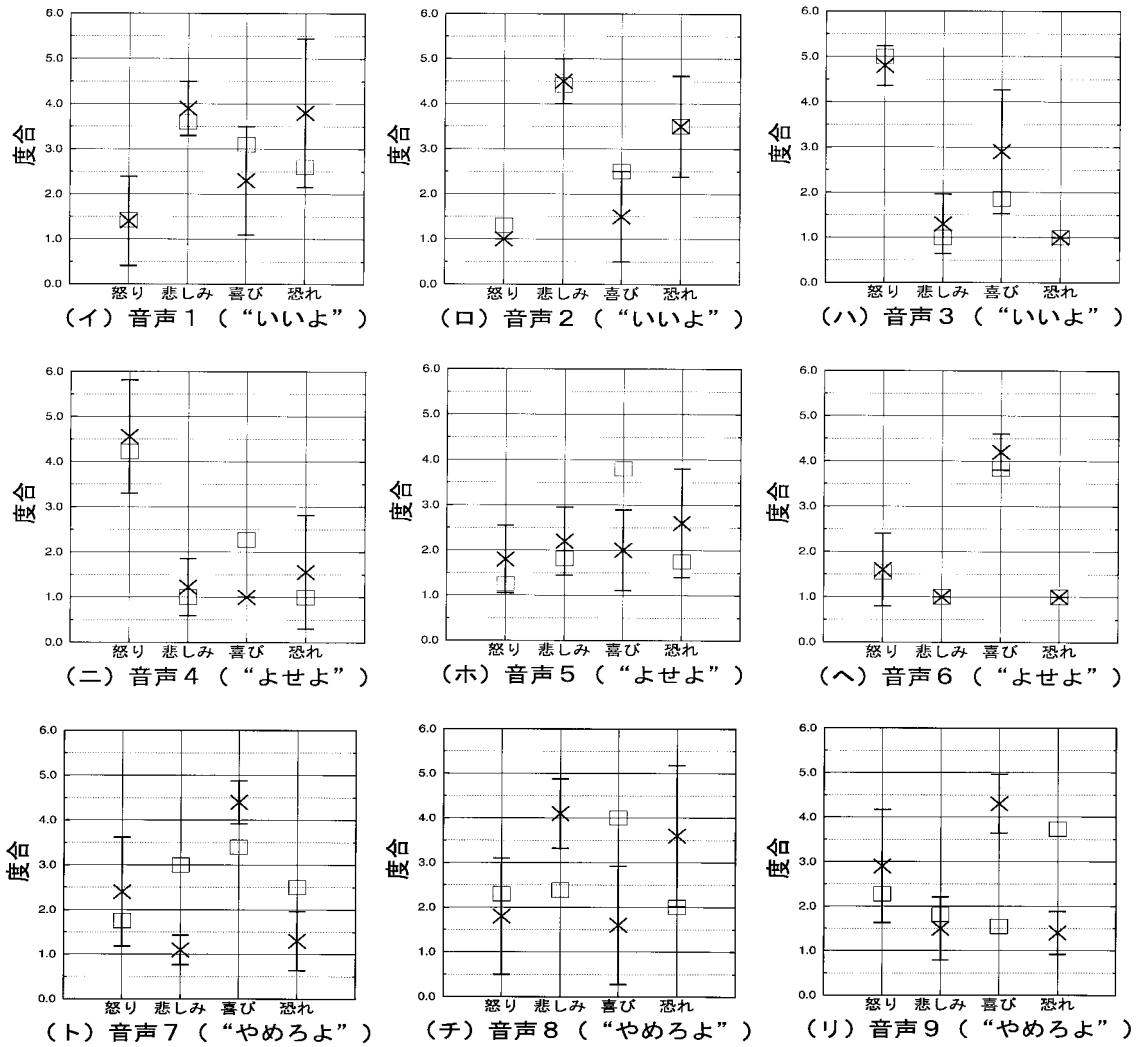


図 7　含まれている情緒性が未知の音声に対する情緒性評価
Fig. 7　Emotion measurement for unknown speeches.

評価値のダイナミックレンジは 1.0〜5.0 であり，1.0 以下の評価値は 1.0，5.0 以上は 5.0 に丸めた．

図 7 において主観評価値と本システムの評価値とを比較すると，情緒性「喜び」と言葉（ト）〜（リ）については全体的に誤差が大きいものの,「怒り」については誤差が小さく，特に（イ）〜（ハ）及び（ニ）〜（ヘ）では,「喜び」以外のすべての情緒性についてほぼ主観評価に一致する評価を行うことができた（相関係数 0.8 以上，有意水準 5%）．

またシステムの評価値の定量的な評価を行うために，主観評価値とシステムの評価値のユークリッド距離を平均誤差と定義する．図 7 の結果から，音声データ別に情緒性についての平均誤差を算出したものを表 4 に，情緒性別に音声データについての平均誤差を算出したものを表 5 に示す．表 5 では，更に用いた言葉別に平均を求めている．

7.　検討及び考察

言葉の意味の情緒性評価への影響について見ると，音声 4〜6 では，“よせよ” という比較的ネガティブな情緒性評価を誘発する可能性のある言葉を用いているが，図 7（へ）からもわかるように，韻律的特徴をも

表 4　システム評価値の音声データ別平均誤差
（情緒性についての平均）
Table 4　Average error of system outputs by speech data (average through emotions).

音声		平均誤差
“いいよ”	音声 1	0.58
	音声 2	0.35
	音声 3	0.39
“よせよ”	音声 4	0.59
	音声 5	0.90
	音声 6	0.11
“やめろよ”	音声 7	1.19
	音声 8	1.55
	音声 9	1.51

表 5　システム評価値の情緒性別平均誤差
（音声データについての平均）
Table 5　Average error of system outputs by emotion (average through speech data).

情緒性	平均誤差			
	“いいよ”	“よせよ”	“やめろよ”	合計平均
怒り	0.18	0.31	0.59	0.36
悲しみ	0.23	0.20	1.31	0.58
喜び	0.95	0.14	2.05	1.38
恐れ	0.40	0.47	1.71	0.86

とに情緒性評価を行っている本システムの結果において「喜び」が多く含まれるとされた音声について，主観評価値は「悲しみ」や「恐れ」に引っ張られることなく,「喜び」の評価となっている．したがって，従来指摘されているように [9]，言葉の意味とはある程度独立に情緒性評価が行えることが本実験の結果からも示された．

ここで本手法を線形モデルを当てはめる情緒性評価手法と比較すると，例えば音声 6 の場合，本システムでは韻律パラメータ変動率 187.80, 195.77, 190.10,139.22, 174.16 が図 5 の (a)〜(e) の横軸へそれぞれ入力される．線形モデルではこれらの韻律パラメータに線形結合係数である重みを決定し，入力されたパラメータとの線形演算によって情緒性を算出するが，この場合，パワー平均値やダイナミックレンジ，ピッチダイナミックレンジの増大が「怒り」に寄与するために,「怒り」が多く含まれると評価されることが予想される．これに対して本システムでは,「怒り」が含まれる際にはパワー平均値の増大と発話長の増大はともに起こらないというような非線形関係をファジー制御ルールに反映させていることによって，結果として「怒り」の度合は小さく，逆に発話長の増大の寄与から「喜び」が多く含まれると評価されている．この結果は，主観評価の結果と整合しており，本手法が有効であることを示唆している．

また音声条件からの影響について明らかにするために，すべてのサンプル音声間の韻律的特徴の類似度を検定した結果，音声 1 と音声 7，音声 1 と音声 9，音声 5 と音声 7 などで類似度が高いことがわかった（相関係数 0.9 以上，有意水準 5%）．これらは，主観評価の結果から含まれる情緒性が異なると評価されているにもかかわらず，韻律的特徴は類似している．一方音声 3 と音声 4 の間でも相関は大きくなり（相関係数 0.79, 有意水準 5%），音声条件の同じ音声については，含まれる情緒性から受ける韻律的特徴の変化も類似しているということがわかった．例えば表 4 を見ると，音声 1〜3 及び 4〜6 では，言葉の音声条件が教師音声と一致していることから，音声の韻律的特徴と情緒性の評価の対応関係が学習したものと一致しており，良好な結果となったと考えられ，一方音声 7〜9 では，音声条件の相違から「怒り」を除いて全体的に誤差が大きくなってしまったと考えられる．本システムは，教師音声の韻律的特徴とその情緒性との対応関係を学習しているために，教師音声で用いた言葉と音声条件の等しいものに関しては，その韻律的特徴から情緒性評価を行うことができる．したがって，言葉のアクセント型とモーラ数の組合せにいくつかの類型を見出し，本システムの前処理としてその類型を判別するような機構を設けることによって，任意の言葉について本システムを適用することができると考えられる．

また表 5 を見ると,「喜び」や「恐れ」において誤差が大きくなっている．Davitz らは，情緒性判断の難易度に関する調査 [24] の中で,「怒り」や「悲しみ」に比べて「恐怖」や「愛」が判断しにくいことを報告している．したがって「喜び」や「恐れ」の主観評価値の信頼度が低い可能性もあるが，筆者らが行った実験で，韻律成分のみで情緒性を評価あるいは合成することが困難であることがわかった [19]．すなわち,「喜び」や「恐れ」は韻律成分以外に，声色の変化などに特徴が現れると考えられ [27]，今後はこのようなパラメータについて検討する必要があると考えられる．

8. むすび

本研究では，音声から観測される物理量からそこに含まれる情緒性を計測する，という問題に対して，ファジー制御の機構を用いた手法を提案した．特に情緒性のコミュニケーションモデルを示し，観測者の視点に着目し，観測者の計算モデルを実装する形で情緒性評

価システムを構築した点は，従来の問題をより明確に整理し，解を与えたという点で有効であるといえる．

本システムの性能を評価するために行った実験では，韻律パラメータの間などに存在する非線形関係をファジー制御ルールに反映させることによって，従来線形モデルなどでは困難であった場合についても情緒性の評価が行えており，ファジー制御を用いた本手法の有効性が示された．中でも「怒り」のように韻律に特徴が現れると報告されている情緒性については非常に人に近い情緒性評価を行うことができた．

本研究で構築した情緒性評価システムは，声優の感情表現の訓練の補助や，主人のその日の話し方に従って（そこから情緒性を抽出して）応答を変化させるペットロボットなどのように，特定のユーザの平静音声を登録して用いるようなアプリケーションに有効であると考えられる．今後，アクセント型とモーラ数の組合せに有限個の類型を見出すなどして，言葉についての制約なく情緒性の評価を行えるよう検討を行うことが課題である．また，韻律以外の声色に関係するような物理量についても検討していく．

文　献

[1] 大橋　力，小田　晋，日高敏隆，村上陽一郎，情緒ロボットの世界，講談社，東京，1985.

[2] L.A. Streeter, N.H. Macdonald, R.M. Krauss, W. Apple, and K.M. Galotti, “Acoustic and perceptual indicators of emotional stress,” J. Acoust. Soc. Am., vol.73, no.4, pp.1354–1360, April 1983.

[3] 福井康之，感情の心理学，川島書店，東京，1990.

[4] I.R. Murray and J.L. Arnott, “Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion,” J. Acoust. Soc. Am., vol.93, no.2, pp.1097–1108, Feb. 1993.

[5] 佐藤　順，川上文雄，森島繁生，“感情音声による感情空間の構築,” 信学総全大 A-399, p.400, 1996.

[6] 水木久美子，“感情を含む音声に関する基礎研究—単音節の定常的解析,” 人間工学，vol.20, no.4, pp.225–230, April 1984.

[7] 伊東久美子，“感情を含む音声に関する基礎研究（II）—合成単母音 [え] による音響パラメータの評価,” 人間工学，vol.21, no.2, pp.81–87, Feb. 1985.

[8] 伊東久美子，“感情を含む音声に関する基礎研究（III）—単母音 [え] の非定常的解析,” 人間工学，vol.22, no.4, pp.211–217, April 1986.

[9] 平賀　裕，斉藤善行，森島繁生，原島　博，“音声に含まれる感情情報抽出の一検討,” 信学技報，HC93–66, Jan. 1994.

[10] 藤崎博也，大野澄雄，冨田　修，丸山英晃，“話者の感情が音声の韻律的特徴に及ぼす影響,” 音響講論 2–4–3, pp.269–270, March 1995.

[11] 小林　豊，新美康永，“音声の感情を反映する韻律情報制御方式について,” 音響講論 2–8–7, pp.233–234, Oct. 1993.

[12] 新村貴彦，村上憲也，“平静音声と感情音声の音韻パラメタの比較,” 音響講論 3–P–19, pp.319–320, Oct. 1993.

[13] 羽尻公一郎，山下崇晴，小川　均，天白成一，“日本語母音単音節における感情表現と韻律情報との関係,” 音響講論，2–4–1, pp.265, March 1995.

[14] 今井憲一，都木　徹，宮坂栄一，“声質変換における感情付与を目的としたニューラルネットによるピッチパターン制御,” 音響講論 1–8–23, pp.189–190, March 1993.

[15] 上床弘幸，小林　豊，新美康永，“音声の感情表現の分析とモデル化,” 信学技報，SP92–131, Jan. 1993.

[16] 重永　実，小川　孝，中尾光志，“単語音声による感情表現について,” 信学技報，SP95–15, May 1995.

[17] 北原義典，東倉洋一，“音声の韻律情報と感情表現,” 信学技報，SP88–158, March 1989.

[18] 森山　剛，斎藤英雄，小沢慎治，“音声における感情表現語と感情表現パラメータの対応付け,” 信学技報，SP95–67, Oct. 1995.

[19] 森山　剛，細田康弘，小沢慎治，“音声における感情情報の認識・合成システム,” 音響講論，1–2–10，pp.193–194, Sept. 1998.

[20] 松原　仁，橋田浩一，“情報の部分性とフレーム問題の解決不能性,” 人工知能，vol.4, no.6, pp.695–703, 1989.

[21] L.A. Zadeh, “Outline of a new approach to the analysis of complex systems and decision processes,” IEEE Trans. Systems, Man and Cybernetics, vol.SMC-3, no.1, pp.28–44, 1973.

[22] 菅野道夫，ファジィ制御，日刊工業新聞社，東京，1988.

[23] R. Plutchik, Emotion, A psychoevolutionary synthesis, Harper & Row, 1980.

[24] J.R. Davitz and J.R. Davitz, “Correlates of accuracy in the communication of feelings,” J. Commun., vol.9, pp.110–117, 1959.

[25] E.H. Mamdani, “Application of fuzzy algorithms for control of simple dynamic plant,” Proc. IEE, vol.121, no.12, pp.1585–1588, 1974.

[26] 藤崎博也，“音声の韻律的特徴における言語的・パラ言語的・非言語的情報の表出,” 信学技報，HC94–09, Sept. 1994.

[27] 重永　実，“感情の判別分析からみた感情音声の特性について,” 音響講論 1–P–3, pp.287–288, Sept. 1996.

[28] 金澤博史，クリスマエダ，竹林洋一，“計算機との対話のための非言語情報の認識と合成,” 信学論（D–II），vol.J77–D–II, no.8, pp.1512–1521, Aug. 1994.

（平成 11 年 2 月 25 日受付，6 月 3 日再受付）

森山　剛　（正員）

平 6 慶大・理工・電気卒．現在，同大大学院博士課程在学中．音声情報処理，感情情報処理に関する研究に従事．平 10 電子情報通信学会学術奨励賞受賞，日本音響学会，言語処理学会各会員．

小沢　慎治　（正員）

昭 42 慶大・工・電気卒．昭 47 同大大学院博士課程了．昭 45 慶大・工・電気助手．同大専任講師・助教授を経て現在，同大電気工学科教授．その間，昭 59 米メリーランド大訪問助教授．ディジタル通信及び画像・音声のディジタル信号の処理の研究に従事．著書「ディジタル信号処理」,「基礎通信工学」（実教出版）工博．電気学会，情報処理学会，IEEE 各会員．