ニューラルネットワークによる感情音声認識に関する一考察*

高橋和彦^{*1}, 中津良平^{*2}

Remarks on Emotional Speech Recognition using Neural Networks

Kazuhiko TAKAHASHI^{*3} and Ryohei NAKATSU^{*3} Department of Mechanical Engineering, Yamaguchi University,
2-16-1 Tokiwadai, Ube-shi, Yamaguchi, 755-8611 Japan

This paper investigates the characteristics of recognizing emotions contained in human speech. The concept of artificial neural networks (ANNs) is adopted for a recognition algorithm. An approach based on the Hidden Markov model (HMM) is also investigated as an alternative recognition method. Using a large database of phoneme-balanced Japanese words, both recognition systems are trained and tested. To evaluate the emotion recognition results obtained by using ANNs or HMMs, emotion recognition testing is carried out with human subjects. The obtained average emotion recognition rates are 51% using ANNs, 32% using HMMs, and 55% with humans. Experimental results confirm that the emotion recognition rate achieved by using the ANNs in the speaker- and context-independent mode is feasible and that ANNs are well suited to this task.

Key Words: Emotion, Speech, Neural Network, Hidden Markov Model, Human

1. 緒言

通常, 人間どうしのコミュニケーションは音声のやり取りによって行なわれているが, 音声には言語情報の様なバーバル情報の他, 話し手の意志, 年齢や性別の様な個人の特性を表す情報, その時の感情やムード等様々なノンバーバル情報が含まれている。例えばテキストを発話する場合, それをどのように表現するかにより, 幾つかの異なった意味を表現することが可能である。一例として, 英語における“really”は, 発話の仕方によって質問, 感嘆, 疑念, 明確等の意味を表す事ができる。従って, テキストのみの理解では話者の発言の意味を解釈することができない事も多く, 感情の様なノンバーバル情報が人間どうしのコミュニケーションにおいて重要な役割を果たしていることは明かである^[1,2]。

音声における感情認識は, 多くの潜在的なアプリケーションが期待される。一つは, 音声認識の補助機能としての役割である。音声認識では伝統的に感情をノイズとみなしているが, 音声中の感情を認識することにより, 音声から感情成分を取り除くことができ, 音声認識システムにおける認識性能を改善することができる可能性がある。他の可能性として, 「感情翻訳機」の様なものが考えられる。言語や文化が異なっている場合, 感情はしばしば異なる意味が反映されている。例えば, ある種の日本語で感嘆

を示すイントネーションは, 英語で疑惑を表す事ができる。従って, 言語間で感情を翻訳する手法は, 国際的なコミュニケーションを改善するのに役立つ可能性がある。この他, 感情認識機能を機械(コンピュータ)のインタフェース(例えば, 画面等に登場するエージェント)に持たせることで, 人間と機械のコミュニケーションをより普通にできる可能性が考えられる。従って, 人間と機械がより自然にインタラクションするためにも, 人間と機械がバーバル情報とノンバーバル情報を使用してコミュニケーションできる能力は, 知的マン-マシンインタフェースにおいて不可欠である^[3]。

この様にノンバーバル情報の重要性が認識されてはいたが, 従来の研究は表情認識^[4]やジャスチャー認識^[5]等画像に含まれるノンバーバル情報を取り扱ったものが多く, 音声に含まれるノンバーバル情報に関する研究^[6]は少ない。この点と感情認識の応用のポテンシャルに鑑み, 我々は既報^[7]においてモジュール構造のニューラルネットワーク(以下, NN)を用いた感情認識システムを構築し, 大量の感情音声データベースを用いた学習により, 話者独立・コンテキスト独立の条件下で平均 50-55%程度の認識結果が得られる事を報告したが, 提案する感情認識システムの特徴や認識結果の妥当性については必ずしも明らかではなかった。

そこで本研究では, NN による感情認識システムの性能と有効性を明らかにすることを目的に, 隠れマルコフモデル(以下, HMM)により構築した感情認識システムによる認識実験とヒトによる感情の主

* 原稿受付 2001年10月9日。

^{*1} 正員, 山口大学工学部(〒755-8611 宇部市常盤台 2-16-1)。^{*2} ATR(〒619-0288 京都府相楽郡精華町光台 2-2)。

E-mail: kylyn@yamaguchi-u.ac.jp

観評価実験を行い、NN、HMM 及びヒトによる認識について検討する。第 2 章では、音声に含まれる感情について述べる。第 3 章で感情認識システムを構築する手法について説明する。HMM は与えられた入力に対して予め設定された遷移確率に従って状態が遷移する非決定的な状態機械であり、音声認識の研究で幅広く用いられているが^[11]、最近では表情やジェスチャ等の時変パターン^[12, 13]の認識^[12, 13]へも適用されている。HMM は扱う対象の構造が比較的明らかな場合は非常に有効であり、例えば、意味内容の認識では音素連鎖という音声構造が事前にわかっているため HMM は音声認識に適している。音声に含まれる感情の構造は未だ明かではないが、本研究では NN との比較対象として HMM を使用して感情認識システムの構築を試みる。第 4 章では、NN、HMM、ヒトによる感情認識実験結果について述べ比較検討を行う。

2. 音声に含まれる感情

普通、人間は日常の会話、特にビジネス上の会話では、主に音声の中に含まれる意味内容の理解に基づいている。そのため、音声認識の研究では、音声に含まれる感情の情報を単なる変動やノイズとして扱ってきた。また、人間が発話する音声では、意識的な場合も無意識的な場合も意味内容と感情が絡み合っている点が更に問題を複雑にしており、特に無意識な状態では、コンテキストが感情的な特徴よりもより重要な役割を果たしている。これは、音声に含まれる感情表現の強度が、状況にきわめて依存している事を示している。本研究では、最終的には無意識の感情表現と意味内容表現が混合した場合においても感情認識ができることを目的としているが、現状では無意識な感情表現の認識は上述の理由により困難であるため、ここでは意識的な感情表現を伴って発話された音声を対象とする。

本研究では、以下の条件下で構築された音声データベースを用いている。

- (1) 意識的に表現された感情を対象とする。これは、人間にとって認識し易く^[14]、データ収集が容易であるためである。
- (2) 8 種類の感情(喜び、からかい、恐れ、哀しみ、嫌悪、怒り、驚き、通常)を対象とする。これは、感情表現を取扱った研究^[15, 16]を参考に選択した。
- (3) 話者独立、コンテキスト独立のため、多くの異なる話者により発話された大量の音韻バランスの取れた単語を使用する。具体的には、音声サンプルは

日本人の成人男性 50 名、成人女性 50 名から収集され、100 個の音韻バランスの取れた日本語(例: 「台所」、「勢い」、「十一月」)を使用した。

(4) 音韻特徴と韻律特徴の二種類の特徴を音声特徴量とする。これらの特徴量は言葉を発話する際に密接に結びついており、韻律特徴のみの制御では感情表現が困難であると考えられるためである。

尚、データベース構築用に選択した単語の殆どは、日常生活において特別な感情無しで発話しているものであり、普通の人がある感情で意識的に発話することは難しい。そのため、プロフェッショナルの声優(男性及び女性)にこれら 100 語について各単語当たり 8 種類の感情で発話してもらったものを録音し、被験者にその録音した音声を聴いてもらい、それを真似て発話してもらうことでデータベースを構築した。従って、全体で各話者毎に 800 個の音声サンプルが得られる。

3. 感情認識システム

図 1 は処理の流れを示したもので、音声特徴抽出部と感情認識部とからなっている。音声特徴抽出部では、入力音声から音声特徴パラメータが抽出された後、音声パワーを用いて音声区間の抽出が行なわれ抽出された音声区間を用いて入力音声の各々に応じた音声特徴量が決定される。この音声特徴量は、感情認識部への入力として用いられる。感情認識部では、2 段階の感情認識が行なわれる。第 1 段階は、8 つの感情を学習した NN 或いは HMM に音声特徴量が入力される。第 2 段階として論理部があり、NN 或いは HMM の出力を処理して感情認識を行なう。以下、各部分での処理について説明する。

Fig. 1 Processing flow of emotion recognition in speech.

3.1. 音声特徴抽出部

音韻特徴を表すパラメータとして、LPC (Linear Predictive Coding) パラメータ^[9]を用いる。LPC パラメータ ($c_i, i = 1, 2, \dots$) は、音声特徴を表現するための代表的なパラメータであり、音声認識においてよく用いられている。韻律特徴を表すパラメータは、エネルギー、ピッチ、及び音韻特徴の時間変化である。エネルギー及びピッチに対応するパラメータは LPC パラメータから求められる音声パワー (P) とピッチ情報 (p)、時間変化 (δ) は LPC パラメータの時間変化を表すパラメータが対応する。

次に、入力音声から音声特徴パラメータの計算について示す。アナログ音声は、ローパスフィルタ通過後に AD コンバータにより 11kHz, 16bit でサンプリングされデジタル音声に変換される。デジタル音声は、256 のサンプリング点を含むフレームの連続として表され、各フレーム毎に LPC 分析が行われて音声特徴パラメータが求められる。第 k フレームに関する音声特徴パラメータを次式で表す。

$$f_k^T = [P_k \ p_k \ \delta_k \ c_{1k} \ c_{2k} \ \dots \ c_{12k}] \quad (1)$$

但し、LPC パラメータは 12 個とした。このパラメータの時系列が音声区間抽出部に送られる。

図 2 に音声区間抽出と音声特徴抽出の概要を示す。音声区間の抽出は、音声パワーの情報をもとに行う。入力された音声パワーは、予め設定した閾値 P_{th} と比較され、 P_{th} よりも音声パワーが大きいフレームが連続してある程度続いた場合に音声が存在すると判断される。またその後、音声パワーが P_{th} よりも小さいフレームがある程度続くと音声終了したと判断される。背景雑音が音声抽出に影響を与えるので、システム動作開始時に背景雑音レベルを計測し、これに基づいて P_{th} を定めている。

抽出された音声区間の全体から等間隔になるように配置された 20 フレームを取り出し、これを $f_1, f_2, \dots, f_{20}$ とする。これら 20 フレームの音声特徴パ

Fig. 2 Speech period extraction and speech feature extraction.

ラメータをまとめて感情認識部への入力となる 300 次元 (15×20) の特徴ベクトル F_k を定義する。

$$F_k^T = [f_1^T \ f_2^T \ \dots \ f_{20}^T] \quad (2)$$

3.2 感情認識部

3.2.1 NN による認識^[10]

図 3 に NN による感情認識部の構成を示す。認識部は、8 種類の感情に対応する 8 つの NN と各 NN からの出力 $v_i (i = 1, 2, \dots, 8)$ を統合する論理部で構成されている。本構成は、各感情に対応する NN の学習を個別に行うことができるため、教師データの追加・変更や NN の構造変更に伴う再学習や認識部の再設計が容易である。各 NN は 4 層構造のシグモイドネットであり、入力層ユニット数は特徴ベクトル F_k と同次元の 300、出力層ユニット数は 1 である (中間層ユニット数は個々の感情によって異なる)。各 NN は、対応する感情の場合に 1、それ以外は 0 を出力するように誤差逆伝播法によって学習を行なうが、学習速度の改善のためモーメント法と適応学習率を使用している。各 NN からの出力は、論理部において統合されて最終的な認識結果が与えられるが、簡単のため、最大値 $\max(v_i)$ を与える NN が対応する感情を認識結果としている。

Fig. 3 Emotion recognition using NNs.

3.2.2 HMM による認識

HMM で認識系を構成する場合、特徴ベクトルのシーケンス $F = f_1^T, f_2^T, \dots, f_{20}^T$ を時変パターンとみなして認識部を設計する。図 4 に HMM による感情認識部の構成を示す。NN の場合と同様に、認識部は、8 種類の感情に対応する 8 つの HMM と、各 HMM からの出力を統合する論理部から構成されている。簡単のため、ベクトル量子化 (VQ) と離散

型 HMM の組み合わせを使用し、状態遷移は全ての状態間での遷移を許す ergodic モデルを使用した。特徴ベクトルのシーケンス F は、HMM へ入力する前に VQ を行ってシンボルシーケンス $O = O_1, O_2, \dots, O_N$ へ変換される。ここで、VQ としてカテゴリ別 VQ^[12]を用いた。これは、各カテゴリ C_i ($i = 1, 2, \dots, 8$) の HMM 毎に VQ コードブック B_i を用意し、シンボルシーケンスの生成をカテゴリ別に行なうものである。コードブック B_i はカテゴリ C_i の特徴ベクトルから LBG アルゴリズムによって生成する。HMM の学習には Baum-Welch アルゴリズム、認識には Viterbi アルゴリズムを使用する。認識時の尤度 $P(O_i)$ の計算においては、カテゴリ別 VQ を用いているので、確率計算において VQ のコードワードと特徴ベクトルシーケンス間の距離を考慮したペナルティ関数^[12]を導入している。論理部では、各 HMM の尤度 $P(O_i)$ が入力され、最大値 $\max P(O_i)$ を与える HMM に対応する感情のカテゴリを最終的な認識結果とする。

Fig. 4 Emotion recognition using HMMs.

4. 感情認識実験

感情認識システムの性能を評価するため、感情音声データベースを用いて NN または HMM の学習と感情認識実験を行なう。学習は男性データと女性データに対し個別に行なうため、男性の感情音声認識システム、女性の感情音声認識システムが各々構築される。学習には、男女各々話者 30 人分の発話データを使用し、残りのデータをテスト用とした。

NN の学習では、はじめに話者 10 人分のデータで構成した小規模データベースを用いて、各 NN のネットワーク構造の最適化を試みた。その結果、喜び、からかい、恐れ、通常の感情に対しては、入力

層・中間層 1-中間層 2-出力層のユニット数が 300-32-8-1 の構造が、哀しみ、嫌悪、怒り、驚きの感情に対しては 300-16-4-1 の構造が良好な学習・認識結果を与えたため、これらの構造に固定して以降の実験を行なった。話者 30 人分のデータを用いた学習の終了条件は、許容最大誤差が 0.001 以下とし、学習回数が 16 万回を越えても収束しない場合は、学習を打ち切った。NN の学習終了後の認識率は、学習データについて 8 感情平均で男性データに対して 57.4%、女性データに対して 56.9%であり、テストデータについて男性データに対して 49.5%、女性データに対して 52.8%の結果が得られた。また、テストデータに対する識別結果の詳細を表 1, 2 に示す。ここで認識率とは、入力データの属する感情カテゴリと論理部からの出力が一致、即ち認識が成功した割合の事であり、平均とはこの認識率（表中の対角成分）の単純平均である。比較のため、ラベル付き自己組織化マップ (SOM)^[10]による認識を試みた。120×120 サイズの 2 次元 SOM により男性、女性いずれの場合も 8 感情平均で学習データについて 70% 程度、テストデータについては 25%以下の認識率であった。これより、階層型 NN による非線形写像が感情認識に有効であることがわかる。

HMM の学習においても、はじめに小規模データベースを用いてコードブックサイズと HMM の状態数の最適化を試みた。コードブックサイズを 16~

Table 1 Classification rate (NN, Male) [%].

Out \ In	J.	T.	F.	S.	D.	A.	Su.	N.
Joy	31.0	10.4	5.9	2.7	4.6	5.4	10.3	12.6
Teasing	13.6	42.0	8.6	2.2	6.2	2.9	10.6	8.8
Fear	6.2	8.1	40.2	10.9	8.9	4.1	4.0	9.2
Sadness	5.6	5.0	11.6	66.1	3.8	3.4	2.9	16.8
Disgust	7.7	11.5	13.1	4.5	63.3	3.0	5.1	9.1
Anger	7.4	5.5	6.1	2.7	3.1	70.1	9.1	10.9
Surprise	11.9	8.7	5.3	1.1	2.3	4.1	55.1	3.3
Neutral	16.6	8.8	9.2	9.8	7.8	7.0	2.9	29.3

Table 2 Classification rate (NN, Female) [%].

Out \ In	J.	T.	F.	S.	D.	A.	Su.	N.
Joy	30.5	9.9	2.1	0.4	1.9	6.0	9.8	13.9
Teasing	14.7	45.2	4.2	1.2	6.0	3.6	14.7	7.0
Fear	4.8	14.2	47.8	17.0	12.5	4.4	3.7	17.4
Sadness	0.3	1.3	29.3	75.4	4.7	2.2	0.0	16.5
Disgust	2.8	8.2	6.7	2.8	71.6	2.4	3.0	5.9
Anger	24.4	7.5	2.8	1.1	1.2	72.3	14.6	11.6
Surprise	12.8	7.6	0.4	0.1	0.6	2.7	52.6	0.8
Neutral	9.7	6.1	6.7	2.0	1.5	6.4	1.6	26.9

128 と変化させ、各々の場合に対し状態数を 8~48 と変化させて実験を行なった結果、コードブックサイズ 48、状態数の 16 の時に最大認識率が得られ、この組み合わせにより以降の実験を行なった。話者 30 人分のデータを用いた学習終了後の認識率は、学習データについて 8 感情平均で男性データに対して 49.5%、女性データに対して 44.5%の結果が得られた。テストデータに対する識別結果を表 3、4 に示す。8 感情平均で男性データに対して 32.2%、女性データに対して 32.6%の認識率であった。NN による結果と比較すると、HMM による識別の傾向は NN のそれと類似しているが認識率は低くなっている。HMM は対象の情報を用いて構造決定ができる場合は比較的高い認識率を得ることが可能だが、本研究では感情の構造が未知であるため事前情報を用いることができず学習結果によってのみ HMM の構造やパラメータを決定をせざる得なかった事、また学習によるコードブックの量子化精度の向上にも限界がある事等が認識率の低さに影響を与えている可能性が高いと考えられる。

NN 或いは HMM によって得られた感情認識結果を評価するための比較対象として、ヒトにより感情音声进行分类する主観評価実験を行った。被験者は日本人成人男性 22 名、日本人成人女性 6 名である。実験は被験者にヘッドフォンを通じて感情音声を聴いてもらい、その音声はどの感情のものであるかを

分類してもらうものである。はじめに、練習として声優によって発話された音声为正解カテゴリとともに 2 回提示され、これを各感情について 4 回行なった後に、本実験を行なう。本実験では、各感情について 4 サンプルを感情音声データベースからランダムに選択して被験者に 2 回提示し、そのサンプル音声に属すると判断した感情カテゴリに分類してもらった。これを男性音声、女性音声について各 4 回行なった。このヒトによる識別結果を表 5、6 に示す。8 感情平均で男性データに対して 58.2%、女性データに対して 51.7%の認識率であった。

以上の実験結果より、以下のことがいえる。

- ・何れの認識結果においても、男性データと女性データの認識結果は類似している。
 - ・HMM による認識結果は、NN のそれよりも悪くなっているが、NN、HMM とともに次の点で類似した特徴がある。即ち、負の感情（怒り、悲しみ等）は認識しやすいのに対し、正の感情（喜び等）の認識率が低い。
 - ・ヒトは負の感情を認識しやすいが、幾つかの正の感情は認識し難い。
 - ・通常については、NN、HMM とともに認識率が低い、ヒトはこの感情についての認識率が高い。
 - ・NN による感情認識の可能性は、ヒトによる感情認識の可能性とほぼ同程度である。
- 従って、話者独立、コンテキスト独立の条件下で

Table 3 Classification rate (HMM, Male) [%].

Out \ In	J.	T.	F.	S.	D.	A.	Su.	N.
Joy	27.1	18.0	3.3	2.8	6.9	14.0	27.4	12.4
Teasing	12.4	25.1	10.8	3.3	9.8	5.8	13.5	13.7
Fear	7.0	10.4	37.6	19.4	22.3	8.9	4.6	15.9
Sadness	2.0	3.7	14.1	51.4	14.8	2.5	1.2	8.4
Disgust	4.2	8.9	15.7	8.9	21.8	6.8	2.6	8.9
Anger	17.8	8.5	4.8	6.9	4.7	37.5	8.7	15.0
Surprise	10.0	5.7	1.0	0.4	1.3	2.6	32.5	1.1
Neutral	19.5	19.7	12.7	6.9	18.4	21.9	9.5	24.6

Table 4 Classification rate (HMM, Female) [%].

Out \ In	J.	T.	F.	S.	D.	A.	Su.	N.
Joy	21.3	15.9	1.3	1.3	9.4	7.1	14.8	9.3
Teasing	20.7	18.6	1.6	2.3	7.3	10.6	16.8	12.3
Fear	4.8	11.3	44.3	26.0	18.3	5.5	4.0	13.2
Sadness	2.2	4.4	37.5	55.7	18.9	2.0	1.6	16.7
Disgust	7.6	16.0	7.8	6.5	28.7	7.0	6.7	17.0
Anger	16.2	9.6	1.2	0.8	3.4	46.1	20.7	6.2
Surprise	10.2	7.0	0.2	2.0	1.4	10.3	23.0	2.2
Neutral	17.0	17.2	6.1	5.4	12.6	11.4	12.4	23.1

Table 5 Classification rate (Human, Male) [%].

Out \ In	J.	T.	F.	S.	D.	A.	Su.	N.
Joy	18.6	7.7	0.0	0.0	0.0	1.3	13.1	1.6
Teasing	3.5	61.6	4.6	1.5	7.9	1.8	8.2	1.4
Fear	0.9	2.4	58.7	18.0	1.5	2.9	1.9	1.4
Sadness	1.5	4.4	30.3	70.6	1.9	7.7	1.9	15.5
Disgust	1.0	4.9	3.5	4.3	86.7	6.2	0.7	3.1
Anger	6.0	4.7	1.3	0.9	0.9	53.8	19.3	6.0
Surprise	7.4	5.5	1.0	0.3	1.0	2.4	45.4	0.6
Neutral	61.1	8.8	0.6	4.4	0.1	23.9	9.5	70.4

Table 6 Classification rate (Human, Female) [%].

Out \ In	J.	T.	F.	S.	D.	A.	Su.	N.
Joy	26.5	20.0	0.0	2.0	0.1	6.7	25.0	3.2
Teasing	4.1	39.9	1.5	1.1	7.8	4.8	6.1	2.7
Fear	0.6	1.8	42.4	10.0	0.1	0.7	2.2	1.7
Sadness	0.9	4.8	48.9	71.9	0.6	3.4	5.7	8.9
Disgust	0.4	6.0	5.1	8.2	90.1	7.3	3.4	3.1
Anger	6.6	4.9	0.8	3.2	0.6	37.8	9.5	6.0
Surprise	3.7	9.2	0.2	0.7	0.4	4.2	31.0	0.7
Neutral	57.2	13.4	1.1	2.9	0.1	35.1	17.1	73.7

NN による感情認識システムで得られた平均 51% の感情認識率は有効であり、ノンバーバル情報を利用した機械と人間とのコミュニケーションへの利用の可能性が示された。この感情認識システムの応用として、コンピュータ上のエージェントの感情認識エンジンに組み込みリアルタイムシステムとして動作させて、感情情報により人間とインタラクションでできるエージェント²⁾の実現性も示されているが、現在の認識率では主にインタラクティブムービーシステム等のような認識精度がそれほど深刻な影響を与えないようなエンタテインメント分野での利用が中心になると考えられる。更に認識率の向上ができれば、コンピュータや機械の複雑な操作をユーザが快適に行なえるような実用的なマン・マシンインタフェースへの応用も可能であると考えられる。尚、本論文の感情認識システムは、その論理部において最大値を与える NN 或いは HMM が対応する感情を認識結果として出力する仕組みとなっているため、認識精度は NN 或いは HMM の学習結果に依存している。従って、認識率を向上させるためには NN 或いは HMM の学習精度の改善が重要であるといえる。また、それに加えて論理部も学習を行なうネットワーク構造を用いて階層型の学習・認識システムを構築する事で認識精度を改善できる可能性も考えられるが、これらは今後の課題である。

5. 結 言

本研究では、音声に含まれる人間の感情を機械によって認識する手法として、ニューラルネットを用いた認識システムの特性について検討した。意識的な感情で発話される音韻バランスの取れた日本語により構成された大型感情音声データベースを用いて、システムの学習と認識実験を行った。更に、感情認識システムの性能評価のため、隠れマルコフモデルを用いた認識システムによる実験とヒトによる主観評価実験を行った。実験の結果、8 感情の男女平均で、ニューラルネットにより 51%、隠れマルコフモデルにより 32%、ヒトにより 55% の認識率が得られた。提案するニューラルネットによる感情認識システムは、話者独立、コンテキスト独立の条件下でヒトによる感情認識とほぼ同程度であり、感情認識に有効である事を示した。

謝 辞

本研究は、A. Solomide 氏、J. Nicholson 氏が ATR 知能映像通信研究所において行なった研究を基礎にしたものであり、ここに感謝の意を表する。

文 献

- [1] 黒岩隆夫, “ノンバーバルインタフェース,” オーム社, (1994).
- [2] R. Picard, “Affective Computing,” The MIT Press, (1998).
- [3] J. Sato and S. Morishima, “Emotion Modeling in Speech Production using Emotion Space,” Proc. of RO-MAN'96, (1996), 472-477.
- [4] G. W. Cottrell and J. Metcalfe, “EMPATH: Face, Emotion, and Gender Recognition Using Holons,” Advanced in Neural Information Processing Systems, Vol.3, (1991), 564-571.
- [5] M. Rosenblum, Y. Yacoob, and L. Davis, “Human Expression Recognition from Motion Using Radial Basis Function Network Architecture,” IEEE Transactions on Neural Networks, Vol. 7, No. 5, (1996), 1121-1120.
- [6] C. R. Wren, B. P. Clarkson, and A. Pentland, “Understanding Purposeful Human Motion,” Proc. of FG 2000, (2000), 378-383.
- [7] D. M. Gavria, “The Visual Analysis of Human Movement: A Survey,” Computer Vision and Image Understanding, Vol. 73, No. 1, (1999), 82-98.
- [8] 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,” 電子情報通信学会総合大会講演論文集 A-399, (1996), 400.
- [9] 新井貴美子, 米倉達広, “音声(挨拶言葉)に含まれる情緒の分析に関する一検討,” 電子情報通信学会総合大会講演論文集 A-144, (1997), 314.
- [10] J. Nicholson, K. Takahashi, and R. Nakatsu, “Emotion Recognition in Speech Using Neural Networks,” Neural Computing and Applications Journal, Vol. 9, No. 4, (2000), 290-296.
- [11] L. R. Rabiner and B. H. Juang, “An Introduction to Hidden Markov Models,” IEEE ASSAP Magazine, (1986 - 1), 4-16.
- [12] 坂口竜己, 大谷淳, 岸野文郎, “隠れマルコフモデルによる顔動画像からの表情認識,” テレビジョン学会誌, Vol. 49, No. 8, (1995), 1060-1067.
- [13] T. Stamer and A. Pentland, “Real-Time American Sign Language Recognition From Video Using Hidden Markov Models,” Technical Report 375, MIT Media Laboratory,

- (1996).
- [14] I. R. Murray and J. L. Arnott, "Toward the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion," *Journal of Acoustical Society of America*, Vol. 93, No.2, (1993), 1097-1108.
- [15] G. Klasmeyer and W. F. Sendmeier, "Objective Voice Parameters to Characterize the Emotional Content in Speech," *Proc. of ICPhS95*, Vol. 1, (1995), 182.
- [16] S. Mozziconacci, "Prosodic Signs of Emotion in Speech: Preliminary Results from a New Technique for Automatic Statistical Analysis," *Proc. of ICPhS95*, Vol. 1, (1995), 250.
- [17] S. McGiloway, R. Cowie and E. D. Cowie, "Pitch Variations and Emotions in Speech," *Proc. of ICPhS 95*, Vol. 1, (1995), 178.
- [18] K. R. Scherer, "How Emotion is Expressed in Speech and Singing," *Proc. of ICPhS95*, Vol. 3, (1995), 90.
- [19] J. D. Markel and A. H. Gray, "Linear Prediction of Speech," *Springer-Verlag*, (1976).
- [20] 坂和正敏, 田中雅博, "ニューロコンピューティング入門," 森北出版, (1997).
- [21] G. Hatano, N. Okuda, and H. Tanabe (Eds.), "AFFECTIVE MINDS," *Elsevier*, (2000), 183-202.
-