

ニューラルネットによる人の基本表情認識†

小 林 宏*・原 文 雄**

The Recognition of Basic Facial Expressions by Neural Network

Hiroshi KOBAYASHI* and Fumio HARA**

We propose the concept of Active Human Interface(AHI) that makes the machine(computer and/or robot) respond to human being more actively and for establishing the new paradigm to realize the AHI, as the first step of this study, we investigate the method of machine recognition of human emotions.

This paper deals with the neural network method of human emotion recognition from facial expressions. Facial expressions were categorized into 6 groups (Surprise, Fear, Disgust, Anger, Happiness and Sadness), and obtained CCD camera-acquired data with respect to facial characteristic points relating to 3 components of face (Eyebrows, Eyes and Mouth). Then we generated the position information and shape information about the 6 basic facial expressions for 30 clients. These information were input into the Input units of the 4-layered neural network and network learning was carried out by back propagation algorithm. The neural network recognition system of facial expressions showed a high recognition rate up to 80% to 6 basic facial expressions for both the position and shape information and particularly the system showed a smaller rate of mis-recognition between some of 6 basic expressions.

Key Words: active human interface, robot-human communication, facial expressions, neural network, pattern recognition

1. はじめに

コンピュータやロボット(以降、総称して機械とする)がいろいろな作業を遂行するほかに、人間と同じように感性や感情をもち、喜んだり、悲しんだり、人を思いやったりすることが出来たら、機械と人とのコミュニケーションにおいて人間は機械に親しみをを感じるだろう。また、機械が人間の機械自身の考えを表示し、それによる人間の反応から人間の性格などを学習してゆくことが出来たら、人間は機械をアイデンティティのある一つの存在として認め、両者のコミュニケーションを「心」と「心」の通ったものにしてすることができるだろう。このように人間に対して応答し働きかける優しい機械を作るために

は、人間に積極的に働きかけるという意味で、特に Active Human Interface(アクティブ・ヒューマン・インタフェース: AHI)という概念が必要で、そのための新しいパラダイムを確立することが大切と考えられる。機械と人間とのコミュニケーションにおいてこのような能動的なインタフェースは、

1. 機械が人間の感情を認識する
2. 1.で得られた認識をもとに機械が人間に対してどんな反応をすれば良いか考える
3. 機械が人間にわかりやすいように2.の結果を表示する

の3段階の処理過程が少なくとも必要と考えられる。

さて上述のような Active Human Interface(AHI)のための第1段階として、機械が人間の感情を認識する方法を検討するとき、まず人間と人間とのコミュニケーションの場合を参考にする。Mehrabian(1968)によると、コミュニケーションにおいて、メッセージのわずか7%が言語で伝達され、55%が顔の表情によって伝達されるといわれる¹⁾。また、Ekman らによると、ある感情を表示するための特定の身体動作というものはないが、お

† 信学ヒューマンコミュニケーション研究会(1991・1), 信学秋期大会シンポジウム講演会(1991・9), International Joint Conference of Neural Network (1991・11)で発表

* 東京理科大学工学部 新宿区神楽坂 1-3

** Faculty of Engineering, Science University of Tokyo, Shinjyuku-ku, Tokyo

(Received February 10, 1992)

(Revised July 22, 1992)

のおのの感情に固有の顔の形というものがあり、感情は主に顔に表わされるとされている²⁾。このように人間対人間のコミュニケーションにおいて顔の表情の認識は相手の感情を知るために非常に重要な要素であるといえる。そこで本研究では、人間の感情を機械が認識するとき、人間の顔表情から認識することがまず大切であると考えられるため、人の顔表情を認識する方法を詳しく検討する。人の顔表情から感情を認識するこのような技術は、心理学の分野においても人間行動の研究のための有力なツールとなるだろう。たとえば、人間がどのように感情を体験しているのか調べるとき、機械が一定の基準に基づいて顔表情より感情を判断することにより、その定量的評価を可能にすることができる。

機械により顔表情から感情を認識する現在までの研究は、大別すると2通りの方法で行われている。一つは、表情筋の動きを推定して表情の分析と認識を行うもので、もう一つは顔の特徴点や特徴量によるものである。前者の表情筋の動きの推定方法には、オプティカルフローにより表情筋の動きを検出する方法³⁾と、顔面筋の44の基本動作を表わす Action Unit (AU)⁴⁾の動きを推定する方法^{5)~7)}がある。間瀬らはオプティカルフローを用いて、幸福・怒り・驚き・嫌悪の4表情の認識を行い、1人の被験者について85%の識別率を得ている⁸⁾。彼らは表情の識別にK-最近傍法という方法を用い、その人の各感情の典型的な顔動画像から得られた特徴ベクトルと、ある顔の動画像から得られたベクトルとの距離により識別している。AUを用いる方法は、現在のところ顔画像の再合成のための表情の分析を行うにとどまり^{9)~11)}、AUの動きから表情の認識は行われていない。

顔の特徴点や特徴量から感情を認識する研究として、Suwa ら⁹⁾は、1人の被験者について喜・怒・怖・悲・妬の5種類の感情の認識を行い、平均79%の識別率を得ている。そこでの特徴量として、眉間の近傍の三角形の面積、口の開き具合を示す長さ、鼻の中心と上唇の左右両端の上方の点を結ぶ直線が水平方向に対してもつ角度の四つを取り上げた。標準顔のこれら特徴量とある表情におけるものとのユークリッド距離を用いて表情の識別を行っている。寺村ら¹⁰⁾は、正面顔について30点、横顔について22点の特徴量から、表情の変化を考慮した個人の識別を行ったが、表情としては“笑い”だけにとどまっている。石井¹¹⁾らは、1人の被験者について幸福・怒り・悲しみ・驚きの四つの感情の認識を行っている。そこでは、眉の角度、目の開き具合、口の曲率、口の縦横比の四つのパラメータを用い、各感情に対応する基準値とある表情から得られたものとの修正ユークリッド距離により識別を行っている。これらの研究では、識別のための

ユークリッド距離の基準が明確でない。

このように顔表情から感情を認識するための研究はあるものの、認識する感情の種類は5種類までに留まり、その手法も「距離」による識別法に基づいている。したがって、この分野の研究はいまだ十分でないといえる。また、ニューラルネットによる識別の研究はなく、さらに、顔表情の認識について、アクティブ・ヒューマン・インタフェース(AHI)の立場からの研究は皆無に等しいといえる。

AHIとして、人の感情を顔表情から認識するとき、喜怒哀楽のような明らかな表情ばかりでなく、その強さ、混合表情などの認識も必要である。しかし本論文では、その第一段階としてEkman らの研究²⁾による驚き・恐怖・嫌悪・怒り・幸福・悲しみの6種類の基本的な感情の認識を対象とする。なお、この6感情の人による認識は万国共通とされている²⁾。本論文では、表情から感情を認識する方法としてニューラルネットワーク(以降NN)を用いることにする。NN¹²⁾は人間の脳の情報処理機能を模倣した並列計算処理であり、人間らしい表情認識を行うために適しているものの一つと考えられ、顔の表情を特徴づけるデータと表情識別の結果との非線形な写像を可能にしてくれるものと予想される。また、NNに入力する情報は、顔の特徴点の2次元座標とその座標から計算される特徴量を用い、ある表情から得られたそれらの情報と無表情(中立)から得られたそれらとの差とする。本論文では識別対象を1人に限定することなく、NNの学習とそれによる認識テストのために30人分の顔画像を用意し、NNに複数の人の表情情報を学習させ、その結果を用いて複数の人の表情の認識を行う。

以下、2章で六つの基本感情を表わす顔画像のサンプリング方法について述べ、3章で顔の特徴点とその特徴点からNNの学習・認識に用いる入力情報への変換方法を示す。4章で本研究で使用するNNとその学習について説明し、5章で6基本感情の認識結果と考察を示す。本論文でのNNによる表情認識の方法は、6基本感情の認識に関して最大で91.2%の正解率を得ており、表情認識をするための訓練を受けた人間と同等な識別能力¹⁵⁾をもつものであることが判明したことも示す。

2. 顔画像のサンプリングについて

本研究では、NNの学習・認識のための6基本感情を表わす顔表情の収集は、Ekman ら²⁾の手順を倣った。まず被験者にEkman らによる文献2)の顔表情の特徴的写真を提示し、これらの感情を表わす表情を確認させる。その際、眉・目・口の特徴的な動きについて著者らが説明を加える。被験者に各感情を表わす表情を表出してもらう際、著者らは再度その感情表出のための顔の特

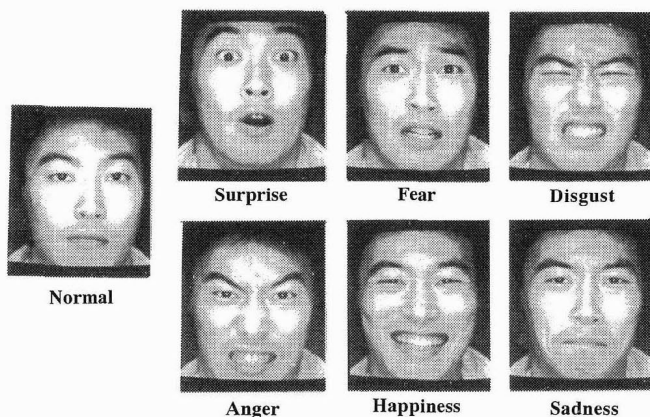


Photo. 1 Sample of 6 facial expressions

微的な動きを説明した。被験者の表出するあらゆる顔表情を CCD カメラで連続撮影し、ビデオテープに録画した。その後、各感情が最も良く表われていると思われる 1 コマを著者らと被験者がビデオテープより選び、それをその被験者の該当する感情を表わす顔表情とした。

本研究では、このようにして 20 歳前半の日本人 20 人分の 6 基本感情についての顔画像(男:16 人, 女:4 人)を得た。これを Kobayasi's data とする。さらに Ekman らによる文献²⁾よりコーカシア人 10 人分の顔画像(男:5 人, 女:5 人)を得(これを Ekman's data とする), 合計 30 人分の顔画像を用意した。ただし, Ekman's data の 10 人のうち 5 人は, 6 感情を表わす顔画像が全て揃っていない。一例として, ある日本人学生の 6 基本表情と無表情(中立)の写真を Photo. 1 に示す。

3. 顔の特徴点と表情情報

顔面筋の基本動作を表わす 44 の AU⁴⁾のうち, 特に 6 基本感情を表わすのに必要な AU は, 頬・顎・しわの動きを表現する五つの AU 以外は, 全て眉・目・口の動きに関連するものである。したがって眉・目・口の 3 部位の変化を表現する情報を用いることにより, 顔の表情が良く表現されると考えられる。そこで, これらの 3 部位に注目し, 特徴点の指定時に, 比較的誤差が入りにくく安定してかつ容易に同定できる点, すなわち, 3 部位と皮膚との境で明らかに色が変わる所から特徴点を選定した。Fig. 1 に本論文で用いる顔の 30 特徴点を示す。Xb-Yb 座標系は絶対座標系を表わし, X-Y 座標系は本研究における特徴点の位置を求めるための座標系である。

本論文では, これらの 30 個の特徴点から 2 通りの情報を得る。一つは, 特徴点を X-Y 座標系の 2 次元座標で表わす情報であり, もう一つはその情報を組合せ加工し,

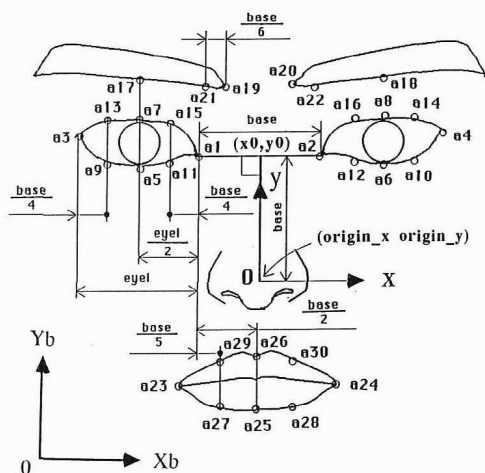


Fig. 1 Model of facial characteristic point

3 部位の形状を表わす情報で, 人間が表情認識の際に用いているだろうと思われる情報を定量的に表わしたものである。ある表情の顔画像から得られたこれらの情報と無表情(中立)の顔より得られた情報との差を NN への入力情報とした。このようにして求めた入力情報のうち, 前者は特徴点の位置の移動量を表わすものなので, 位置情報と呼び, 後者は 3 部位の形状の変化を表わす情報であるので形状情報と呼ぶことにする。本研究でこれらの 2 種類の情報を用いたのは, 入力情報の違いによる認識結果を比較し, 表情認識における NN の特性を検討するためである。

以下, 上述した情報を得るための手順を示す。なお, 特徴点の自動的抽出方法は本論文の範囲外であると考え特に検討はしない。それゆえ本論文では, 特徴点の絶対座標系における座標値は前もって抽出されているものとする。

3.1 位置情報の決定

Fig. 1 に示す特徴点 ai ($i=1\sim 30$) の絶対座標系における座標を (x_{bi}, y_{bi}) と表わす。これを X - Y 座標系に直すために、原点 ($origin_x, origin_y$) を絶対座標で表わす。初めに、どんな表情においても変化しないと考えられる左右の目頭間の距離 $base$ を求める。

$$base = \sqrt{(x_{b2} - x_{b1})^2 + (y_{b2} - y_{b1})^2} \quad (1)$$

水平軸と両目頭を結ぶ線の傾き、つまり顔全体の傾きを反時計回りを正として θ とすると、 θ は次式で与えられる。

$$\theta = \tan^{-1} \frac{y_{b2} - y_{b1}}{x_{b2} - x_{b1}} \quad (2)$$

目頭間の中心の座標 (x_0, y_0) は、

$$x_0 = (x_{b1} + x_{b2})/2, \quad y_0 = (y_{b1} + y_{b2})/2 \quad (3)$$

と表わせるので、これより ($origin_x, origin_y$) は

$$origin_x = x_0 + base * \sin \theta \quad (4)$$

$$origin_y = y_0 - base * \cos \theta$$

となる。求めたい特徴点 ai の X - Y 座標系における座標を (x_i, y_i) とすると、

$$x_i = x_{bi} - origin_x, \quad y_i = y_{bi} - origin_y \quad (5)$$

と表わせる。被験者の顔と CCD カメラとの距離によって撮影された顔の大きさを補償するため、どんな表情でも変わることがない値 $base$ でこれらを割る。

$$x_i = x_i / base, \quad y_i = y_i / base \quad (6)$$

このようにして得られた位置情報から、無表情(中立)顔についての位置情報 (n_x_i, n_y_i) を減じることにより新しい位置情報を得る。

$$x_i = x_i - n_x_i, \quad y_i = y_i - n_y_i \quad (7)$$

さらに各情報を絶対値の最も大きい情報値で割り、NN に入力する位置情報とした。

一例として、**Photo. 2** に示す顔画像より得られた位置情報(Position Information)を **Table 1** に示す。

3.2 形状情報の決定

(5)式で得た特徴点の座標 (x_i, y_i) により、3 部位の形状を表わす 21 個の形状情報を求める。(8)~(28)式にこれらの情報を求める計算式を示す。なお便宜上、各情報を ($data\ k: k=1\sim 21$) と表記する。これらの形状情報



Photo. 2 Sample of facial expression

Table 1 Sample of position information

Data No.	Data	Data No.	Data	Data No.	Data
1	0.0	21	-0.036	41	-0.048
2	0.0	22	-0.026	42	0.320
3	0.0	23	0.033	43	0.127
4	0.0	24	-0.114	44	0.204
5	0.038	25	0.010	45	0.106
6	0.097	26	0.053	46	-0.455
7	-0.103	27	-0.168	47	-0.035
8	0.115	28	-0.017	48	-0.458
9	0.069	29	0.008	49	0.022
10	-0.056	30	0.008	50	-1
11	-0.108	31	-0.062	51	-0.056
12	-0.082	32	-0.044	52	-0.017
13	0.010	33	0.093	53	-0.011
14	0.014	34	0.224	54	-0.884
15	-0.131	35	-0.152	55	0.024
16	0.023	36	0.163	56	-0.905
17	0.105	37	-0.013	57	0.014
18	-0.052	38	0.244	58	0.063
19	-0.072	39	0.127	59	0.084
20	-0.052	40	0.174	60	0.076

Table 2 Meaning of data

Data No.	Meaning
1	Distance between the most outer points of eyes
2, 3, 4	Openess of eyes
5, 6, 7	Movement of upper eyelids
8, 9, 10	Movement of lower eyelids
11	Distance between inner ends of eyebrows
12, 13, 14	Rise of eyebrows
15	Inclination of eyebrows
16	Width of mouth
17, 18	Openess of mouth
19, 20	Movement of the upper lip
21	Movement of the lower lip

は、各部位の縦と横の幅、距離、弛みや堅さ(たとえば下瞼や唇の張り具合)などを表現できるよう考慮して選んだ。**Table 2** にこれらの情報がどのような意味をもっているか示す。

$$data\ 1: |x_4 - x_3| \quad (8)$$

$$data\ 2: |y_7 - y_5| + |y_8 - y_6| \quad (9)$$

$$data\ 3: |y_{13} - y_9| + |y_{14} - y_{10}| \quad (10)$$

$$data\ 4: |y_{15} - y_{11}| + |y_{16} - y_{12}| \quad (11)$$

$$data\ 5: |y_7 - y_1| + |y_8 - y_2| \quad (12)$$

$$data\ 6: |y_{13} - y_1| + |y_{14} - y_2| \quad (13)$$

$$data\ 7: |y_{15} - y_1| + |y_{16} - y_2| \quad (14)$$

$$data\ 8: |y_5 - y_1| + |y_6 - y_2| \quad (15)$$

$$\text{data } 9: |y9 - y1| + |y10 - y2| \quad (16)$$

$$\text{data } 10: |y11 - y1| + |y12 - y2| \quad (17)$$

$$\text{data } 11: |x20 - x19| \quad (18)$$

$$\text{data } 12: |y17 - y7| + |y18 - y8| \quad (19)$$

$$\text{data } 13: |y19 - y1| + |y20 - y2| \quad (20)$$

$$\text{data } 14: |y21 - y1| + |y22 - y2| \quad (21)$$

$$\text{data } 15: |y21 - y17| + |y22 - y18| \quad (22)$$

$$\text{data } 16: |x24 - x23| \quad (23)$$

$$\text{data } 17: |y26 - y25| \quad (24)$$

$$\text{data } 18: |y29 - y27| + |y30 - y28| \quad (25)$$

$$\text{data } 19: |y26 - y24| + |y26 - y23| \quad (26)$$

$$\text{data } 20: |y29 - y23| + |y30 - y24| \quad (27)$$

$$\text{data } 21: |y27 - y23| + |y28 - y24| \quad (28)$$

これらを 3.1 節と同様に (base) で割り、さらに、顔の傾きを補償するために $(\cos \theta)$ で割る。

$$\text{data } k = \text{data } k / (\text{base} * \cos \theta) \quad (29)$$

このようにして得られた (data k) を、無表情 (中立) の顔について得られた形状情報で減じることにより新しい形状情報を得る。さらに、3.1 節と同様に、絶対値の最も大きい情報値で割ったものを NN に入力する形状情報とした。

一例として、Photo. 2 より求めたこの形状情報 (Shape Information) を Table 3 に示す。

4. ニューラルネットワークの学習

4.1 ニューラルネットワーク

顔表情の認識についての学習と認識テストのために用いるニューラルネットワークを Fig. 2 に示す。これは 4 層構造のバックプロパゲーションモデル¹³⁾である。このような 4 層の階層型ネットワークモデルは、必要なだけ中間層のユニット数を用いれば、ユニット間の結合を適当に設定することにより、任意の非線形の連続関数をいくらかでも精密に近似できることが知られている¹⁴⁾。そのため、入力と出力の間の非線形な写像変換が高精度に行われ、顔表情の入力情報と表情識別の 2 値的な結果との非線形な変換、すなわち表情の識別にも高性能が期待できる。このような理由により、本研究では 4 層の階層型ネットワークを用いた。

4.2 学習条件

Fig. 2 に示すネットワークにおいて、入力層のユニット数は入力の情報数に対応させる。したがって、位置情報であれば 60 ユニットの、形状情報であれば 21 ユニットの。また出力層のユニット数は 6 基本感情に対応させ 6 とし、見かけ上、上から驚き、恐怖、嫌悪、怒り、幸福、悲しみを表すユニットとする。中間層 1 層、2 層とも 100 ユニットの。この数は、中間層のユニット数を 30 から 10 ずつ増やしてゆき、次節で説明する表

Table 3 Sample of shape information

data	1	0.028	data	12	0.499
	2	0.476		13	0.958
	3	0.605		14	-0.281
	4	0.348		15	0.942
	5	1.000		16	-0.340
	6	0.453		17	0.478
	7	0.231		18	0.500
	8	0.769		19	0.718
	9	0.192		20	0.465
	10	0.261		21	0.790
	11	0.051			

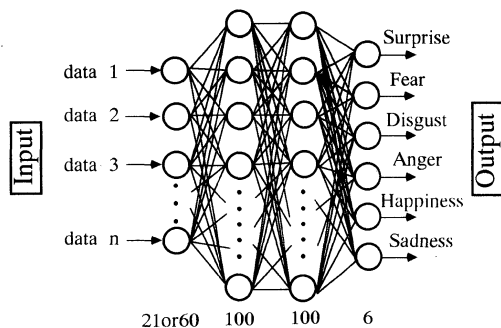


Fig. 2 Neural network

情の正しい認識率 (正解率) が安定収束したときのユニット数である。

NN の学習の際、教師信号の与え方は、学習させる入力情報が表わす感情と同じ感情を表わす出力ユニットの教師信号を 1 とし、ほかの出力ユニットの教師信号を 0 とする。たとえば驚きの表情を学習させる場合の教師信号は、(1, 0, 0, 0, 0, 0) となる。出力関数にはシグモイド関数を用いた。学習させる表情に対応する全入力情報 (全パターン) についての教師信号と、出力層ユニットの出力値との自乗和誤差が 0.0001 以下になった場合に学習を終了した。なお、本研究において、学習回数は最大でも 20,000 回を超えていない。

4.3 学習用情報の構成

Table 4 に、NN の学習に使用した情報について、被験者の人数と学習用情報名を示す。学習用情報名の数字は人数を表わしている。位置情報および形状情報について、それぞれ学習用入力情報をランダムにその数だけ選び、NN の学習を行う。2 章で述べたように本論文で扱う人数は 30 人に限られており、表情の認識テストには学習に用いなかった被験者の情報を用いる関係上、NN の学習に用いる顔の被験者数は最大でも 30 人の半数の 15 人とした。

Table 4 Learning data constitution

Number from Ekman's data	Number from Kobayasi's data	Name of learning data
5	0	5F
0	5	5J
5	5	10M
5	10	15M

5. 認識結果

5.1 認識正解率の定義

4.3節で述べた学習用情報を用いて NN を学習させ結合係数の定まった NN に、学習に用いなかった被験者の顔表情の位置情報および形状情報を入力する。その結果得られた NN の出力値により、被験者の表情が正しく認識されているか、すなわち正解かどうか判断する。この際正解とは、出力層 6 ユニットのうち最も出力値の大きいユニットに対応する感情を、この NN の認識した感情とし、その感情が被験者と著者らが前もって確認した顔表情の表わす感情と一致するなら正解とする。

Photo.2 より求められた形状情報 (Table 3) を、5F の形状情報で学習された NN に入力したときの出力結果を Table 5 に示す。これより、この NN による認識結果は「驚き」であることがわかる。Photo.2 は読者にも判断されるように、「驚き」の表情であると前もって確認されているので、NN の認識結果は正解となる。

正解率は、認識テストで正解した顔の総数を、認識させた顔の総数で割って定義する。ただし、被験者 30 人分の情報しかないので、NN を学習させた被験者数により認識テストに用いた顔の数は異なってくる。

5.2 認識結果

位置情報、形状情報に関して、Table 4 に示す学習用情報の種類別の正解率を Table 6 に示す。この表より、位置情報でも形状情報でも、ほぼ似たような正解率が得られており、NN への入力情報の形態は認識結果に大きく影響しないといえよう。NN の学習に被験者 15 人の位置情報を用いた場合では 83.8%、形状情報を用いた場合では 91.2% の正解率が得られている。一方、学習用情報 5J を用いた場合の正解率は、位置、形状情報ともほかの学習用情報 (5F, 10M, 15M) を用いた場合の正解率より低く、70% 程度である。これは、一般的にいわれるように日本人の顔表情が豊かでないことに起因すると思われるが、詳細な検討は今後の課題とする。Bassill (1979)¹⁵⁾によると、顔表情から感情を識別するための訓練を受けた心理学専攻の 20 人の学生に、6 人の俳優の 6 基本感情を表わす顔画像をモニタに表示し識別させたところ、正

Table 5 Sample of output

Output unit	Output
Surprise	0.999858
Fear	0.000010
Disgust	0.000010
Anger	0.000010
Happiness	0.000010
Sadness	0.000010

Table 6 Correct recognition ratio unit : %

Learning data name	Position Information	Shape Information
5F	76.4	85.8
5J	70.7	66.4
10M	80.9	79.0
15M	83.8	91.2

解は約 87% であった。これより、本論文での NN による基本感情の認識の方法は、表情認識の訓練を受けた人と同等な識別能力をもつことがわかる。

5.3 認識の特性

上述の 2 種類の入力情報の形態について、5F, 5J, 10M, 15M の学習用情報で NN を学習し、その NN を用いて認識テストを行った結果をまとめ、認識率の 6 基本感情に関する分布、すなわちどの感情にどれだけ誤認識しているかの分布を Table 7, Table 8 に示す。たとえば、Table 7 では、NN に「驚き」の顔表情の位置情報を入力したとき、NN が「驚き」と認識した正解が 92.9%、「恐怖」に 2.4% 誤認識をしていることを示す。これらの表より、位置情報を用いた場合では「恐怖」を「驚き」に、形状情報では「驚き」を「恐怖」に、さらにどちらの情報を用いても「嫌悪」を「怒り」や「悲しみ」に、「悲しみ」を「嫌悪」に誤認識しやすく、「嫌悪」と「悲しみ」については正解率も 60% 前後で低いことが読み取れる。参考となる文献は入手していないが、これらの誤認識されやすい感情に対応する顔表情では眉・目・口の形状が似ており、やはり人間でも誤りやすいと思わ

Table 7 Recognition condition—Position information— unit : %

		Input emotion					
		Surprise	Fear	Disgust	Anger	Happiness	Sadness
Recognized emotion	Surprise	92.9	17.8	0	1.3	3.9	6.2
	Fear	2.4	69.9	0	0	0	2.5
	Disgust	0	2.7	57.1	9.1	0	14.8
	Anger	0	2.7	19.5	76.6	1.3	7.4
	Happiness	4.7	1.4	2.6	5.2	94.8	1.2
	Sadness	0	5.5	20.8	7.8	0	67.9

Table 8 Recognition condition—Shape information—
unit: %

		Input emotion					
		Surprise	Fear	Disgust	Anger	Happiness	Sadness
Recognized emotion	Surprise	85.9	4.1	0	1.3	0	0
	Fear	14.1	86.3	1.3	0	5.2	4.9
	Disgust	0	1.4	62.3	7.8	3.9	23.5
	Anger	0	8.2	13.0	85.7	0	11.1
	Happiness	0	0	0	0	90.9	0
	Sadness	0	0	23.4	5.2	0	69.5

れ, NN による誤認識の結果は人間に近いものであると判断される. Table 7 および 8 では, 5 F, 5 J, 10 M, 15 M の学習用情報による認識テストの結果を総和しているため, これらの表の数値は正解率の低い 5 F や 5 J の影響を反映していることに注意する. しかし, 入力情報として位置情報および形状情報のいずれを用いても, ほとんど同様の分布結果が得られていることがわかる.

6. ま と め

著者らは, 人間と機械とのより良いコミュニケーションを実現するために, アクティブ・ヒューマン・インタフェース (Active Human Interface, AHI) という概念を提案した. 本論文ではその実現の第 1 段階として, 機械が人間の感情を認識するための方法を検討し, 人間の顔表情から, 眉, 目, 口の 3 部位についての特徴点および特徴量を定義し, それらの中立的顔からの変化量により, 4 層のニューラルネットワーク (NN) の学習を行い, 感情を認識する方法を示した. 認識テストの結果より, この方法の有効性と機械による人間の感情認識の可能性を示した.

本論文で提案した顔表情から感情を認識する方法は, NN を用いることにより, 認識対象をある特定の 1 人に限定する必要がなく, 複数の人間の顔表情の認識を可能にした. さらに, 6 基本感情の認識については, 学習情報として 15 人分の形状情報を用いることにより 91.2% の正解率を得ており, 表情認識の訓練をした人と同等の認識能力があることが明らかになった.

当面の課題としては, 5.2 節, 5.3 節で述べたような正解率が低い場合や誤認識が発生する理由など, 認識の特性を分析することが挙げられる. 顔表情の認識に関しては, 6 基本感情の強さおよび, 複数の感情が混ざりあった混合感情の認識についての検討, 動画画像からの表情認識の実施が今後の課題である.

参 考 文 献

- 1) 本名, ほか訳: ノンバーバル・コミュニケーション, p 237, 大修館書店 (1981)

- 2) P. Ekman and W. V. Friesen 著, 工藤力訳: 表情分析入門, 1/277, 誠信書房 (1987)
- 3) 間瀬健二: オプティカルフロー抽出による表情筋の動作検出, 電子情報通信学会技術研究報告, PRU89-128, 17/24 (1990)
- 4) P. Ekman and W. V. Friesen: Facial Action Coding System, Consulting Psychologist Press (1977)
- 5) 崔, 原島, 武部: 知的画像符号化における合成規則による表情の分析, 電子情報通信学会技術研究報告, PRU89-55, 148/155 (1989)
- 6) 崔, 原島, 武部: 知的画像符号化における頭部の動きと顔面の動き情報の高精度推定, 電子情報通信学会技術研究報告, PRU90-68, 1/8 (1990)
- 7) 崔, 原島, 武部: 顔の 3 次元モデルを用いた顔面表情の分析, 電子情報通信学会論文誌 D-II, J74-D-II, 6, 766/777 (1991)
- 8) 間瀬, 前田, 末永: 表情動画画像からの感情の認識の 1 手法, 電子情報通信学会技術研究報告, PRU91-24, 95/101 (1991)
- 9) M. Suwa et al.: A Preliminary Note on Pattern Recognition of Facial Emotional Expression, The 4th International Joint Conference on Pattern Recognition 408/410 (1978)
- 10) 寺村, 茅: 顔の表情とその認識に関する研究, 電子情報通信学会技術研究報告, PRL81-72, 41/47 (1981)
- 11) 石井, 岩田: コンピュータ画像処理を利用した顔の表情の自動認識, 日本機械学会論文集, 52-483, 2989/2992 (1986)
- 12) 甘利俊一監訳: PDP モデル, 産業図書 (1989)
- 13) 中野, ほか: ニューロコンピュータ, 28/84, 技術評論社 (1989)
- 14) 船橋, ほか: ニューラルネットによる連続写像の近似表現について, 電子情報通信学会技術研究報告, MBE-88-9 (1988)
- 15) J. N. Bassill: Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of Face, Journal of Personality and Social Psychology, 37-11 2049/2058 (1979)

[著 者 紹 介]

小 林 宏 (学生会員)



1990 年, 東京理科大学工学部機械工学科卒業. 92 年同大学大学院博士課程前期課程修了. 現在, 同大後期課程在学中. 日本学術振興会特別研究員. ヒューマン・インタフェース, ロボットのダイナミクスと制御の研究に従事. 電子情報通信学会等の会員.

原 文 雄 (正会員)



1970 年, 東京大学大学院産業機械専門課程 (博士) 修了, 同年東京大学生産技術研究所講師, 71 年東京理科大学工学部助教授, 83 年同教授. 流体関連振動とその制御, ヒューマン・インタフェース, ファジィ制御系の安定性解析, ロボットのダイナミクスと制御の研究に従事 (工学博士). 75 年~76 年西ドイツ・フンボルト財団奨励研究員としてカールスルーエ大学に留学. 日本機械学会, ASME, 原子力学会, 電子情報通信学会, 日本ファジィ学会などの会員.