離散時間形リカレントニューラルネットワークによる 人の顔の動的な 6 基本表情認識*

小 林 宏^{*1}, 原 文 雄^{*2}

Dynamic Recognition of 6 Basic Facial Expressions by Discrete-Time Recurrent Neural Network

Hiroshi KOBAYASHI and Fumio HARA

In order to develop an "Active Human Interface" that realizes emotive interactive communication between machine and human being, we have been investigating the recognition of human emotion from facial expressions. This paper deals with the machine recognition method of 6 basic dynamic facial expressions. First, we carry out psychological recognition tests of 6 basic dynamic facial expressions for 3 subjects to clarify the characteristics of human dynamic recognition of facial expressions. Then, we construct a recurrent neural network (RNN) equipped with a feedback layer and a differential input layer for developing the machine recognition method. Using the dynamic data of each facial expression and the human recognition result for 2 subjects, we train the RNN and recognition tests are performed by inputting the dynamic facial data of one subject not used in training the RNN. By comparing the results obtained with RNN to those of human recognition, the RNN is found to perform very similar dynamic recognition to that exhibited in human recognition of facial expressions.

Key Words: Active Human Interface, Robot-Human Communication, Recurrent Neural Network, Dynamic Facial Expressions, Machine Recognition

1. は じ め に

人間によりよく適応する優しい機械は、機械が人間の命令に従うだけでなく、人間と能動的にコミュニケーションできるインタフェースを必要とすると著者らは考え、Active Human Interface (AHI)という概念を提案した^{(1)~(6)}。

すなわち、AHI には、論理的な情報だけでなく心理的な情報を扱うことが必要と考え、そのために次の三つの機能を提案した⁽⁷⁾。

(1) 機械が人間の感情、強いては「心」を理解する(センサ部)。

(2) (1)での理解をもとに、機械が人間に対してどんな応答をすればよいかを考える(コントローラ部)。

(3) 機械が(2)の結果を人間にわかりやすいように表示伝達する(アクチュエータ部)。

さて、人間対人間のコミュニケーションにおいて、顔表情の重要性は広く認識されていることから⁽⁸⁾、著者らは AHI という新しいパラダイムを実現するため

のコミュニケーションメディアとして「顔表情」を取上げ、(1)~(3)のそれぞれについて検討した。(1)に関しては、静止顔画像を対象として、「驚き」、「恐怖」、「嫌悪」、「怒り」、「幸福」、「悲しみ」の 6 基本表情⁽⁹⁾のニューラルネットワーク(NN)による認識⁽¹⁾、NN による顔表情の正認識や誤認識について、学習済み NN の内部構造の特性を明らかにする研究として 6 基本表情の NN 認識における特性分析⁽²⁾、また 6 基本表情の強さの認識⁽³⁾、それらの混合表情の認識⁽⁴⁾について報告してきた。ここでとりあげた 6 基本表情は、Ekman らの 30 年にわたる研究の結果、万国共通に人間が表出、認識できるものとされている⁽⁹⁾。(2)に関しては、調和理論⁽¹⁰⁾を用いた人工感情モデルについて報告した⁽¹¹⁾。(3)に関しては、コンピュータグラフィックスにより人工感情に応じた顔表情を CRT 上に表出できるシステム⁽⁵⁾、および人間と同様な顔をもち表情を表出する「顔ロボット」について報告してきた⁽⁶⁾。

人間対人間のコミュニケーションでは、人間は相手の顔表情を連続して見ており、刻々と変わる相手の顔表情の微妙な動きにより敏感に相手の感情を読みとるものと思われる。つまり、顔表情は時間的な流れ、時間的文脈の中で敏感に、かつ正確に判断される⁽¹²⁾。そこ

* 原稿受付 1995 年 6 月 9 日。

^{*1} 東京理科大学工学部 (〒162 東京都新宿区神楽坂 1-3)。

^{*2} 正員、東京理科大学工学部。

で本研究では、人間と同様に刻々と変わる人間の顔表情から連続的に得られる情報を用い、過去の顔表情の影響や顔表情の動きの速さを考慮して、機械による人の顔の動的な6基本表情認識(動的顔表情認識)のアルゴリズムについて検討する。

心理学の分野において、動画を用いた動的顔表情認識については必要性が認められているものの、残念ながら工学に応用できるような研究報告がなされていない⁽¹²⁾。そこで本研究では最初に、6基本表情について人間の動的顔表情認識の特性を調べる。すなわち、「中立」(無表情)から「ある基本表情」(6基本表情の一つ)に連続的に変化する顔表情画像を用意し(各表情11画像、合計66画像)、それらが表情の変化に即して連続して人に表示し、(1)表情が表出されていく場合と、(2)消えていく、または弱まっていく場合とについて表情認識実験を行う。

次に、人の顔の動的顔表情認識特性を学習できる離散時間形リカレントニューラルネットワーク(Recurrent Neural Network: RNN)を構築する。RNNは中間層や出力層の信号が入力側へフィードバックされるネットワークで、学習アルゴリズムや構成方法などが数多く提案されている^{(13)~(18)}。RNNの学習方法は、離散時間学習⁽¹³⁾⁽¹⁴⁾と連続時間学習^{(15)~(18)}の2種類に分けられる。前者では基本的に、1時刻前、あるいはそれ以前のRNNの出力値をRNNの状態層にフィードバックし、それが次の時刻のRNNの出力値に影響を及ぼすようにRNNを構成する。このRNNの学習は、入力パターン、出力パターン、および1時刻前、あるいはそれ以前のRNNの出力値パターンをまとめて一つの学習パターンとすることで、フィードフォワードネットワークと同様にバックプロパゲーション法⁽¹⁰⁾で行われる。後者は、脳神経系のダイナミカルな情報処理の機構をニューラルネットモデルに埋め込み、それらのダイナミカルな情報処理の原理・機構を解明しようというもので、計算量が膨大になり、かつローカルミニマムに陥りやすいといわれている⁽¹⁷⁾。本研究では、過去の入力情報やその変化の速さを考慮することができ、ネットワークの構築、およびその学習が容易なことから前者を採用することにする。そして、RNNに人間の動的顔表情認識結果を学習させ、RNNによる動的顔表情認識テストと人間による認識結果とを比較し、RNNによる動的顔表情認識の特性を明確にする。

以降、2章で「中立」から「ある基本表情」に連続的に変化する人の顔表情画像の収集方法について述べ、3章で人間による動的顔表情認識の実験方法および結

果を示す。4章ではRNNにより顔表情を認識するための表情情報について説明し、5章でRNNによる動的顔表情認識の実験方法の説明と顔表情認識結果を示し、人間のそれと比較することによりその特性を浮き彫りにする。

2. 顔表情画像の収集

「中立」から「ある基本表情」まで連続的に変化する顔表情画像の収集は文献(1)、(3)、(4)と同様にEkmanらの手順にならった⁽⁹⁾。顔表情を表出してもらう顔画像提供者に、Ekmanらによる文献(9)の基本表情を表す特徴的写真を提示し、6基本表情を確認させる。その際、目・眉・口の特徴的な動きについて著者らが説明を加え、鏡を用いて自分のそれらの動きを確認させる。その後、顔画像提供者に顔表情を表出してもらう際、著者らは再度その顔表情表出のための顔の特徴的な動きを説明する。そして、顔画像提供者の表出する顔表情をCCDカメラで連続撮影してビデオテープに録画し、そのビデオテープより「中立」から「ある基本表情」に至る顔表情画像を1/30秒ごと収集した。

ところで、「中立」から「ある基本表情」が表出されるまでの時間は表情の種類により異なる⁽⁹⁾。例えば、「驚き」は一瞬で表れて消えるが、「悲しみ」は2、3日持続することもある。本研究で収集した1/30秒ごとの顔表情画像も、「ある基本表情」に至るまでの画像数が表情の種類ごとに異なる。そこで本研究においては、「中立」から「ある基本表情」に至る顔表情画像の中から、互いに顔画像間に違いが見られるような11画像(「中立」を含む)を、各表情ごとに著者らが選択した(合計66画像)。これは、「中立」から「ある基本表情」に至るまでの時間を1に規格化したことに相当することに注意する。以降、顔表情画像番号として、基本表情ごとに「中立」を零、「ある基本表情」を10とし、途中の顔表情画像にもこれにならって1、2、...、9と番号を付ける。このようにして3人の顔画像提供者(理工系大学4年の男子学生)の顔表情画像を収集した。説明の都

Table 1 Interval of each facial image -subject A-

	Facial image number											total
	0	1	2	3	4	5	6	7	8	9	10	
Surprise	1	1	1	1	1	1	2	2	2	3		15
Fear	1	1	1	1	1	1	1	1	4	4		16
Disgust	1	1	1	1	2	3	3	5	2	3		22
Anger	1	1	1	1	1	1	1	1	2	3		13
Happiness	1	1	1	1	2	2	2	2	3	3		18
Sadness	1	3	6	6	7	6	3	4	5	4		45

合上、これらの3人の顔画像提供者をA, B, Cとする。一例として、表1に顔画像提供者Aの「中立」から「ある基本表情」に至る0~10の顔表情画像間の時間間隔を1/30秒を単位として示す。この時間間隔は顔画像提供者によって若干異なることを注意しておく。

3. 動的顔表情認識実験

2章で述べたように、「中立」から「ある基本表情」に連続して変化する顔画像が11枚あり、これが6基本表情について収集されているので合計66枚の画像を、各表情ごとに表情の変化に即して連続的に表示した場合の動的顔表情認識の実験について説明する。この実験は、日常観察される(1)表情が表出されていく場合と、(2)表情が消えていく、または弱まっていく場合、とくに分けて次のように行われる。すなわち、(1)「中立」から「ある基本表情」に至るまで、あるいはその途中までを連続して表示して表情認識のテストを行う(upと表記)こと、(2)「中立」から「ある表情」まで連続的に表示し、さらにその表情が弱まって再び「中立」に至るまで、あるいはその途中までを連続的に表示して表情認識を行う(downと表記)ことにする。

3.1 実験方法

被験者：理工系大学4年男子学生17人

実験手続き：

① upのテスト 基本表情ごとに零から1まで、零から2まで、…、零から10まで(0→1, 0→1→2, …, 0→1→, , →10)と顔表情画像がCRTに連続して表示される動画(ビデオテープ)を10組用意する。各組での各表情画像はそれぞれ0.5秒ずつ表示される。人間の視覚認知実験結果によると、刺激が表示されてからそれをはっきりと認識するまでに、人間は平均480msを必要とするという研究報告があるため⁽¹⁹⁾、表情表出過程の途中までを被験者に見せて表情判断を行わせる本研究においては、実際の表情表出速度では速すぎて被験者がその表情を認識できないと考えられるため、本研究では1顔画像の提示時間を0.5秒とした。これにより、被験者は表情の変化もはっきりと認識できると考えられる。以上のような10組を6基本表情について用意し(合計60組)、それらの組をランダムに表示する。そして、動画が消えたときの表情が、7表情カテゴリー(6基本表情+無表情)のどれを表しているか被験者に判断させ、判断結果を回答用紙に記入させる。

② downのテスト 基本表情ごとに0~10まで約0.25秒ずつ連続して顔画像を表示した後、続いて10→9, 10→9→8, …, 10→9→, , →0と0.5秒ず

つ連続して顔表情画像が表示される動画(ビデオテープ)を10組用意する。upと同様に、この10組を6基本表情について用意し(合計60組)、それらの組をランダムに表示する。被験者には、その動画が6表情カテゴリーのどれを表しているか判断させ、その結果を回答用紙に記入させると同時に、顔表情画像が消えた時点で、判断した感情が消えたか否か(残っているか)を記入させる。

ところで、このdownのテストにおいて「中立」から「基本表情」までを表示したのは、上述のようにまず被験者に表情の種類を判断してもらう必要があったからである。また、その間の顔画像1枚の表示時間を約0.25秒としたのは、1秒間30フレームのビデオレートにおいて、著者らの所有する機器では1フレームごとに意図する画像を録画できるわけではなく、連続7フレームごとに画像を録画するものであることと、「中立」から「基本表情」まで約2.5秒で表示されれば、被験者は十分表情の種類を判断できると考えたからである。

3.2 実験結果 upの場合には、顔表情画像番号n(=1~10)ごとに、その表情カテゴリーに判断(認識)した被験者の割合、downの場合には、顔表情画像番号ごとに、判断(認識)した表情カテゴリーが残っていると判断した被験者の割合を求めた。

図1に、顔画像提供者Aの顔画像を用いた認識実験の結果を示す。紙面の都合上、BとCについては割愛する。図1中、upを実線で、downを破線で示す。各図において、横字が本来その顔表情画像が表していると著者らが判断する表情の種類で、縦に書かれた文字

Fig. 1 Dynamic recognition results by human being

が被験者に判断された表情の種類を表す。ただし、「無表情」は含まれていない。横軸の目盛りは実験に用いた顔表情画像の時間間隔(表1参照)で、横軸の右端の値は各表情での最大表出までの時間(1/30秒単位)を表す。縦軸はその表情に判断した人の割合を表す。

顔画像提供者が異なると、顔画像を収集する際の時間間隔が異なり各画像ごとの顔部位(目や口など)の変位量などが異なるにもかかわらず、3人の提供者において「驚き」、「嫌悪」、「怒り」、「幸福」については画像番号5ではup, downともに100%認識されていた。これに対して「恐怖」と「悲しみ」は、100%その表情に認識されるまでに、少なくとも画像番号8以上提示する必要があるという共通した傾向があることが明らかとなった。また、upとdownの違いは、「恐怖」以外の5表情については、同じ顔画像を表示しているにもかかわらず、3人の提供者の顔画像ともにupよりdownのほうが認識率が低かった(実線が破線の上方にずれている)。「恐怖」については、図1に示した顔画像提供者Aの場合だけ上記と同様にupがdownより認識率が高く、BとCについては若干ではあるが逆の傾向があった。この違いの詳しい分析については本研究の範囲外であると考え、特に行わないが、上述のように動的顔表情認識の全体的な傾向およびupとdownの違いについては、顔画像提供者に依存せず、ごく一部を除いて共通の傾向があることが明らかとなった。

4. RNNによる動的な顔表情認識のための表情情報

RNNで人の動的な顔表情を認識するために用いる情報(表情情報)は、文献(1),(3),(4)と同様に、図

Fig.2 Facial characteristic points (FCPs)

2に示す顔の特徴点の「中立」からの移動量を用いる。図2に示すように、特徴点はa1~a30の30点で、各特徴点は x 方向(水平方向)と y 方向(鉛直方向)の座標値をもっている。1顔画像につき表情情報の数は合計60である。なお、本研究では特徴点の座標値は既に得られているものとする。

5. RNNによる動的顔表情認識

5.1 離散時間形リカレントニューラルネットワーク 人間同士のコミュニケーションにおいて、人間は顔表情の微妙な変化を敏感に感じとっており、さらにはそれまでの、あるいは直前の表情の認識結果を無意識のうちに参考にして表情を認識していると思われる⁽¹²⁾。すなわち、

- (1) 表情の微妙な動き(移動量)を察知する。
- (2) 表情の動きの速さを察知する。
- (3) 過去の認識結果が現在の認識結果に影響を与える

の少なくとも一つの機能が人間には備わっていると仮定する。上記の3機能を備え、顔の表情情報を時系列で扱うことができる時間的離散形リカレントニューラルネットワーク(Recurrent Neural Network: RNN)の構造を図3に示す。このRNNでは、4章で述べたように顔の特徴点の「中立」からの移動量を表情情報として用いるので、(1)の機能は備わっている。また、(2)の機能を実現させるため、「直前の顔表情」からの特徴点の移動量も入力情報となるようにRNNを構成する(後述)。(3)過去の認識結果(出力層

Fig.3 Discrete time-type recurrent neural network

の出力)をネットワークにフィードバックすることで実現している。

RNN の構造を示す図 3 において、 t は現在の時刻を表し、 $Ii(t)$ と $Ok(t)$ はそれぞれ時刻 t の RNN 入力層(Input Layer)に入力する入力情報[(1)に相当]と RNN の出力値とを表す。入力情報は表情情報であり、 i は入力情報の番号である ($i=1\sim 60$)。また、 k は出力ユニットの数で、文献(1)、(3)、(4)と同様に 6 基本表情に対応しているため 6 ユニットとする ($k=1\sim 6$)。差分層(Differential Layer)は上記の(2)に相当し、「直前の表情」を表す表情情報 $Ii(t-1)$ からの特徴点の移動量 [$Ii(t)-Ii(t-1)$] が入力される。中間層は 50 ユニット ($j=1\sim 50$) で、入力層および差分層の出力値が入力され、中間層の出力値 $Hj(t)$ は出力層に入力される。中間層のユニット数(50 ユニット)は、20 ユニットから 10 ユニットずつ増やしてゆき、RNN の出力値が安定収束したときの値である。出力層の出力値は状態層(State Layer)にフィードバックされ、状態層の出力値 $Ck(t)$ は中間層とともに出力層に入力される[(3)に相当]。状態層のユニット数は出力層のそれと同数の 6 ユニットとし、1 対 1 に対応させる。状態層は出力層の出力値(状態)を保存する層であり、1 時刻前の出力ユニットの出力値 $Ok(t-1)$ を、現時刻の状態層の出力値 $Ck(t)$ とする。

この RNN では、入力層および差分層と中間層の間、中間層と出力層の間、状態量と出力層の間の結合係数を学習により獲得するものとする。

5・2 RNN の学習 RNN の学習には、1 章で述べたようにバックプロパゲーション法を用いる⁽¹⁰⁾。RNN は時系列情報を扱うため、時間的に連続した顔画像より求めた表情情報の変化の情報を、学習情報の 1 パターンとして RNN に学習させる必要がある。すなわち、 $Ii(t-1)$ から $Ii(t)$ への変化を RNN に学習させることにする。1 章で述べたように、本研究では、人間と同様の動的表情認識を RNN で実現することを目的としているので、人間による動的顔表情認識結果を RNN に学習させる。そこで、RNN の学習情報としては顔画像番号 0~10 に相当する表情情報を次々に RNN に入力し、図 1 に示した人間による認識結果を出力層の各ユニットへの教師信号とする。また、3・2 節で述べたように、表情が表出されていく場合(up)と、表情が消えていく場合(down)で人間による認識結果に違いが見られたため、本研究では、学習情報として、

- ① up
- ② down

③ up+down

の 3 種類を用いて RNN の学習を行う。

具体的には、①、②、③の各場合について、A、B、C の中の 2 人の動的表情認識結果を学習情報に用いる。①の場合には、一人の顔画像提供者について、各表情ごとに以下に述べる顔画像番号に相当する表情情報とそれに対応する教師信号の組(6 基本表情についての認識率の値)を学習情報とする。すなわち、時間的に連続した $Ii(t-1)$ から $Ii(t)$ への変化に相当する $0 \rightarrow 1$, $1 \rightarrow 2$, $2 \rightarrow 3$, ..., $9 \rightarrow 10$ の 10 パターンを学習情報とするので、一人の顔画像提供者については合計 60 パターン(10×6 基本表情)となる。同様に、② down の場合には、各表情ごとに $10 \rightarrow 9$, $9 \rightarrow 8$, $8 \rightarrow 7$, ..., $1 \rightarrow 0$ とし、一人の顔画像提供者についての学習情報の数は合計 60 パターンとなる。③は一人の顔画像提供者について、前述の両者を合わせた 120 パターンが学習情報となる。ここで、出力層の各ユニットに与える教師信号として、 $Ii(t)$ に対応した顔画像についての人間による動的表情認識率の値を与える。各顔画像に対応する人間による動的表情認識率の値%は、図 1 では 0~100 の範囲であるが、RNN の出力値が $-1\sim 1$ となっているため、この範囲に入るように比例配分したものをを用いる。また、状態層は 1 時刻前 ($t-1$) の認識結果を出力値とするため、 $Ii(t-1)$ に対応した顔画像についての人間による動的表情認識結果の値を与える。

3 人の顔画像提供者 A、B、C のうちの 2 人の認識結果を学習情報とし、残りの 1 人分の結果を認識テストに用いるので、3 人から 2 人を選び、残りの 1 人をテストに用いる組合せは 3 通りある。そこで、学習情報①~③のそれぞれについて、この 3 とおりの学習・認識テストを行う。2 人の顔画像提供者を RNN の学習に用いるので、学習情報のパターン数は、①の場合に 120、②の場合に 120、③の場合に 240 となる。学習は、RNN の出力値と教師信号との誤差二乗和が 0.01 以下になったときに終了することにした。

5・3 認識テスト 学習の終了した RNN に、学習に用いなかった顔画像提供者の表情情報を入力することにより RNN の性能評価のための認識テストを行う。RNN の出力値は $-1\sim 1$ の範囲であるが、これを 0~100 に比例配分し、それを RNN による認識結果とする。その結果を顔画像提供者(A、B または C)に対する人間の認識結果と比較し、RNN が人間とどの程度類似しているかを調べる。

RNN による動的顔表情認識テストの詳細は、以下のように手順で行う。顔画像提供者 2 人についての人

(1) Training information: ① up

(2) Training information: ② down

(3) Training information: ③ up+down

Fig. 4 Dynamic recognition results by RNN for subject A

の認識結果の学習を終了した RNN に、学習に用いなかった顔画像提供者の表情情報を、各表情ごとに up → down に相当するように、0 → 1 → 2 → , , → 10 → , , 1 → 0 の順序で入力する。顔画像提供者 A の顔画像を認識テストに用いた場合の認識結果を図 4 に示す。ここで (1) ~ (3) は、学習情報① ~ ③に対応している。図 4 中、実線が 0 ~ 10 (up) を RNN に入力した場合の結果、破線が実線の結果に続いて 10 ~ 0 (down) を RNN に入力した場合の結果である。図 4 の (1), (2) は (3) に比べて、その顔画像が本来表している表情以外の表情に認識する (誤認識する) 割合が高いことが分かる。また、学習していない側の表情情報、つまり、up を学習した場合では down の、down を学習した場合は up の、RNN の出力値の変化が滑らかでないことがわかる。これに対し図 4 の (3) の RNN の出力値の変化は、どちらの場合も RNN の出力値は滑らかに変化し、さらに図 1 (顔画像提供者 A の認識結果) に示した人間による認識結果とよく一致していることが見てわかる。他の顔画像提供者 B, C に対する RNN の認識結果についても同様のことがいえた。

両者の一致性を定量的に評価するために、顔画像が本来表している表情の種類、つまり図中において横に表示された表情と縦軸に表示された表情の種類が一致する場合について、各画像に対する人間と RNN の認識結果の誤差絶対値を求め、6 基本表情について平均し、単位として求めた。結果を表 2 に示す。学習情報の組① ~ ③についての RNN の性能を評価するために、up, down の場合の誤差の平均値をそれぞれ計算し、さらに両者の平均値を ave. として表 2 中に示した。まず、学習情報が① up と② down の場合、すべての顔画像提供者 (A, B, C) について、学習した表情過程 (up または down) の結果のほうが、学習していない表情過程 (down または up) の結果より誤差が小さいことがわかる。しかし、いずれの場合も①, ②よりも③の学習情報を用いたほうが誤差が小さく、③の場合、平均 (ave) 誤差が 8% 以下になっていることがわかる。

③の学習情報を用いた場合、RNN と人間の結果が

Table 2 Absolute error average between RNN and human

	Training information					
	① up		② down		③ up+down	
	up	down	up	down	up	down
A	11.0	23.5	26.5	9.6	6.8	6.6
B	18.1	18.5	34.8	22.2	6.9	7.9
C	5.2	25.0	19.0	16.0	5.7	8.8

Fig. 5 Comparison between RNN (solid line) and human (dotted line) for subject A

どの程度よく一致しているか確認するため、図5に、各表情に関して、RNNの認識結果(実線)と人間による認識結果(破線)を示す。これは、顔画像提供者Aに対応した結果である。またupは表情が表出されている場合であり、downは表情が消失される場合である。これより、学習情報として③を用いた場合、RNNによる動的顔表情認識結果は、人間による結果とかなりよく一致することがわかる。

5.4 静的なNNとの比較 時系列情報を考慮したRNNによる動的な表情認識結果と、それを考慮しないNNによる静的な表情認識結果とを比較する。静的な表情認識では、著者らがすでに報告した中間層2層の階層形NNを用いる⁽¹⁾⁽²⁾⁽⁴⁾、ところでRNNの場合は、時系列情報を扱うことができたので、一つの表情情報に対して、upとdownに対応する人間による認識結果を同時に学習することができた。ところが静的な認識の場合、一つの表情情報に対して教師信号は一つなので、両者を同時に学習することはできない。そこで、upとdownとに分け、それぞれの人間による認識結果をNNに学習させた(5.2節の学習情報①、②に相当)。RNNとの比較のため、このNNの学習情報はRNNと同様に2人の顔画像提供者の時系列表情情報とし、残りの1人の表情情報を認識テスト用とした。NNの学習では各表情ごと、顔画像0, 1, 2, , に相当する表情情報をNNにそれぞれ入力し、対応する人間による認識結果(upとdownに分ける)を教師信号とし、NNの出力値と教師信号との誤差二乗和が0.01以下になった場合、学習を終了した。

認識テストは、学習に用いなかった顔画像提供者の表情情報を、各表情ごとに、0, 1, 2, , 10と順番に、学習の終了したNNに入力して行った。upの場合に

Fig. 6 Static recognition results by NN for subject A

ついて、顔画像提供者AについてのNNの認識結果の例を図6に示す。RNNを用いた図4の(3)や図5と比較すると、NNの結果は「驚き」、「怒り」、「悲しみ」の場合のように、突発的に他の表情に認識することがあること、「幸福」や「怒り」のように100%認識されるまでに時間がかかる表情があること、がわかる。このような傾向は、他の学習情報を用いたNNの認識テストにおいても同様であった。時系列情報を考慮したRNNの場合にはこのようなことが発生しないことから、原因は時系列情報を考慮していないためと考えられ、RNNの有効性が示されたと考えられる。

6. ま と め

人間によりよく適応する優しい機械を実現するためのActive Human Interfaceの基礎研究として、RNNによる動的な顔表情認識を試みた。機械が人間による動的顔表情認識と同様な特性を示すことが大切と考え、まず、3人の顔画像提供者の顔画像を収集編集した動画を用いて、人間による動的な顔表情認識実験を実施した。特に、(1)表情が表出される場合(up)と、(2)表情が消えていく、あるいは弱まっていく場合(down)について実験を行った。その結果、全体的な認識の傾向およびupとdownの違いについては、顔画像提供者に依存せず、ほぼ共通の傾向があることが明らかとなった。次に、人間の表情認識の特徴と考えられる、(1)表情の微妙な動き(移動量)を察知する、(2)表情の動きの速さを察知する、(3)過去の設

結果が現在の認識結果に影響を与える、という 3 機能を備えたりカレントニューラルネットワーク(RNN)を構築した。人間による動的表情認識において、顔表情の表出される過程: up, down の違いによる認識の違いが見られたため、① up, ② down, ③ up+down, に相当する 3 種類の学習情報を用意した。そして、3 人の顔画像提供者のうち 2 人の表情情報とそれらに対する人間の認識結果を RNN に学習させ、その学習済みの RNN により残りの 1 人の表情情報を用いて RNN の認識テストを行い、その結果と人間による認識結果とを比較した。その結果、③ up+down の学習情報を用いた場合、各顔画像に対する人間と RNN の認識結果の違いは、平均 8% 以下となり、両者はかなりよく一致することがわかった。さらに、静的な NN との比較から、時系列情報を考慮した RNN の有効性を明確にした。

AHI の確立の観点から、機械がすべて自動的に、かつ実時間で人間の顔表情を認識するシステムの開発が今後の課題である。

文 献

- (1) 小林宏・原文雄, ニューラルネットワークによる人の基本表情認識, 計測自動制御学会論文集, **29** 1(1993), 112-118.
- (2) 小林宏・原文雄, 人の顔の 6 基本表情のニューラルネットワーク認識における特性分析, 機論, **61** 582, C(1995), 340-347.
- (3) 小林宏・原文雄, ニューラルネットワークによる人の顔の 6 基本表情の強さの計測に関する研究, 機論, **59** 567, C(1993), 177-183.
- (4) 小林宏・原文雄, ニューラルネットワークによる人の顔の混合表情の認識, 機論, **59** 567, C(1993), 183-189.
- (5) Hara, F. and Kobayashi, H., Computer Graphics for Expressing Robot-Artificial Emotions, *IEEE Int. Workshop on Robot and Human Com.*, (1992), 155-160.
- (6) 小林宏・原文雄・内田豪・大野宗久, アクティブ・ヒューマン・インタフェース(AHI)のための顔ロボットの研究, 日本ロボット学会誌学術論文, **12** 1(1994), 155-163.
- (7) Arbib, M. A., (著・金子隆芳訳), ニューラルネットと脳理論, (1992), 25-27, サイエンス社.
- (8) Raffler-Engel, W. V. (本名信行・ほか 2 名訳), ノンバーバル・コミュニケーション, (1981), 237, 大修館書店.
- (9) Ekman, P. and Friesen, W. V. (工藤力訳), 表情分析入門, (1988), 1-277, 誠信書房.
- (10) Rumelhart, D. E. and McClelland, J. L., *PARALLEL DISTRIBUTED PROCESSING*, (1986), 546-611, The MIT Press.
- (11) Mogi, S. and Hara, F., Artificial Emotion Model for Human-Machine Communication by Using Harmony Theory, *IEEE Int. Workshop on Robot and Human Com.*, (1992), 149-154.
- (12) 福井康之, 感情の心理学, (1990), 81, 川島書店.
- (13) Jordan M. I., Serial Order, A Parallel Distributed Processing Approach, *Institute for Cognitive Science Report*, 8604(1986), 1-37, UCSD.
- (14) 中川聖一・早川勲, シーケンシャルニューラルネットワークを用いた音声認識, 信学論文誌, **J74** D-II 9(1991), 1174-1183.
- (15) Williams, R. J. and Zipser, D., A Learning Algorithm for Continually Running Fully Recurrent Neural Networks, *Neural Computation*, **1**(1989), 270-280.
- (16) Pineda, F. J., Generalization of Back Propagation to Recurrent Neural Networks, *Physical Review Letters*, **59** 19, (1987), 2229-2232.
- (17) 銅谷賢治・吉沢修治, 時間パターンの記憶の神経回路網モデル, 信学論文誌, **J73**-D-II 8(1990), 1150-1157.
- (18) 内山匡・下原勝憲, リカレントネットワークのための実時間学習アルゴリズム, 信学論文誌, **J73** D-II 8(1990), 1158-1163.
- (19) Card, S. K., Moran, T. P. and Newell, A., The Model Human Processor: An Engineering model of human performance, *Handbook of perception and human performance*, **11**(1986), (45-1)-(45-35).