

表情の動的多様性に基づく表情表出リズムの抽出

佐藤 和人[†] 間所 洋和[†] 門脇さくら^{††}

[†] 秋田県立大学システム科学技術学部 〒015-0055 秋田県由利本荘市土谷字海老ノ口 84-4

^{††} SmartDesign 有限責任事業組合 〒011-0945 秋田県秋田市土崎港西 3 丁目 9-15

E-mail: [†]{ksato,madokoro}@akita-pu.ac.jp, ^{††}sakura@smart-d.jp

あらまし 本論文では、心理的ストレスレベルに起因する表情の覚醒度合いとその時間的遷移が作り出す時系列パターンを表情表出リズムと定義し、顔の静的多様性と動的多様性に着目した個人固有な表情空間チャートを利用して、人間が創り出す表情表出リズムの可視化する手法を提案する。本手法は、SOM (Self-Organizing Maps) と Fuzzy ART (Adaptive Resonance Theory) の 2 種類の教師なしニューラルネットワークをハイブリッド化して得られる表情の覚醒度から表情空間チャートを生成し、更に Fuzzy ARTMAP で学習することにより表情空間をモデル化するためのコードブックを作成し、HMM (Hidden Markov Model) を用いて表情の時系列パターンの変化から表情表出リズムを抽出する。18 歳から 22 歳までの男子学生 5 名と女子学生 5 名の計 10 名を対象とした評価実験では、HMM が推定した状態遷移系列は表情表出に伴う表情変化の範囲を適切に捉えており、表情表出プロセスにおける中間表情の覚醒パターンを抽出する際に HMM の状態数を用いて制御可能であることを確認した。

キーワード 自己組織化マップ、ファジー ARTMAP、隠れマルコフモデル、表情空間チャート、表情表出リズム

Extraction of Facial Expression Rhythms Based on Dynamic Diversity of Facial Expressions

Kazuhito SATO[†], Hirokazu MADOKORO[†], and Sakura KADOWAKI^{††}

[†] Faculty of Systems Science and Technology, Akita Prefectural University 84-4 Aza Ebinokuchi Tsuchiya, Yurionjo City, 015-0055 Japan

^{††} SmartDesign LLP 3-15-9 Tsuchizakikou Nishi, Akita City, 011-0945 Japan

E-mail: [†]{ksato,madokoro}@akita-pu.ac.jp, ^{††}sakura@smart-d.jp

Abstract This paper presents an extraction method of facial expression rhythms, which are defined for time-series patterns created by changes of facial expressions and arousal levels caused by psychological stress levels, using individual facial expression charts focused on static diversity and dynamic diversity of faces. The proposal method consists of three steps: the first step is to create facial expression charts produced by arousal levels that is obtained by the results of hybrid learning of Self-Organizing Maps (SOM) and Fuzzy Adaptive Resonance Theory (ART) networks; the second step is to make codebooks using Fuzzy ARTMAP networks for modeling of facial expression spaces; and the third step is to extract facial expression rhythms using Hidden Markov Model (HMM) from time-series patterns of facial expressions. We developed for an original facial expression dataset that consists of five male students and five female students from 18 through 22 years old at our university. We applied our method to the dataset and obtained the result of the state transitions to be estimated by HMM is extracted the rage of the facial expression changes. Moreover, we conformed the method can control the patterns of arousal levels in mixed facial expressions at the expressed process using the number of transition states on HMM.

Key words SOM, Fuzzy ART, HMM, Facial expression charts, Facial expression rhythms

1. ま え が き

赤松 [1] は、顔が持つ多様性を静的多様性と動的多様性に分

類しており、個人認識では、顔をパターンとして扱う静的多様性で十分な結果が得られるが、表情認識では、静的多様性に加えて顔パターンの時間変化としての動的多様性の両方を同時に

扱う必要性を指摘している。表情変化の時間的要因の検討に関する研究として、Bassili [2] は、表情の運動成分のみによって表情分類の可能性を提示し、Leonard ら [3] は、喜びの感情を表す笑いの表情の識別に顔全体の時間的変化が影響することを示した。しかし、いずれにおいても顔面運動のどのような成分が分類に影響するのは明らかにされていない。Ekman ら [4] は、笑いを3種類 (Felt Smile, False Smile, and Miserable Smile) に分類し、各表情にどの表情筋が用いられているかを考察している。しかし、時間的な要因の影響については、表情変化の速度の影響を想定するに留まっている。Pilowsky ら [5] は、笑いを表出した時の目や口の動きの時間経過に着目し、それらの動く量を計測しているものの、時間的な要因についてはほとんど言及していない。

一方、西尾ら [6] は、目と口の動きの開始時点の時間差から、笑いの種類 (快の笑い, 不快の笑い, 社交の笑い) を分類する基準を導出している。しかし、なぜ時間差が笑いの分類に影響を及ぼすのかは明らかになっていない。表情の動的多様性を扱った最近の研究としては、太田らの顔構成要素モデルに基づく手法 [7]、川嶋らの表情譜 [8] が挙げられる。太田らは、顔全体の大局的な動きを検出する手法では、目のまばたきや発話による口の変形などのような個々の顔構成要素の細かな局所の変動を検出することはできないと述べている [7]。川嶋らは、Ekman らの FACS (Facial Action Coding System) で記述できる表情は静的なものに留まり、時間的な表情変化の描写はできないと指摘している [8]。いずれも特徴点ベースの手法であるが、AU (Action Units) は用いず、独自に設定した特徴点を用いている。人間は、顔の大局的な構造の中から局所的な構成要素やその動きをとらえて表情を認識することができる。これは、人間は顔の個々の要素を特徴点の動きとしてとらえているのではなく、全体的な表情変化の中から局所的な表情変化を位相変化としてとらえているからと考える。

2. 表情の個人差と動的多様性

2.1 表情データベース

実際の人間による表情表出は、複数の感情に対応する表情がブレンドされた中間的な表情であったり、口元は微笑みながら目は泣いているといった複数の感情を並列的に表出した表情であったりすることが多い。顔という対象の形状が人物それぞれで異なるように、表情表出の仕方、例えば、ある感情をどの程度の大きさの顔面変形として表情に表出するかについては個人差がある。すなわち、表情の個人差には、表情変化の土台となる顔そのものの造りの個人差と、FACS でも規定されている個々の表情筋の動きの強弱やそのバランスの違いといった表情表出のダイナミックな特性における個人差とが含まれる。

多様な表情の顔画像を集めたデータベースとして、静止画像では Ekman & Friesen の顔画像データセット [9]、動画画像では Ekman & Hager による FACS データセット [10] と Cohn & Kanade による顔表情データセット [11] が挙げられ、認識性能の比較評価を行うための実験用データとして利用されている。性能評価では、各カテゴリの感情を明確に表出しているとみな

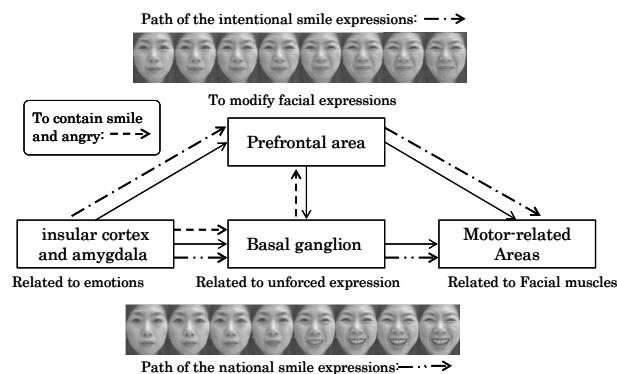


図1 自然な表情と意図的な表情の表情発現パス
Fig. 1 Expression paths of natural smile and intentional smile.

せる表情の顔画像データが学習及びテストサンプルとして大量に必要なにもかかわらず、これらのデータベースでは、同一人物・同一表情のサンプル数が十分とは言えない。したがって、多くの研究では、研究者や研究機関が独自に収集した表情の画像データを用いて認識実験を行っているのが現状である。

2.2 表情空間と表出リズムの定義

表情表出時の顔面の移動量、形状の変化を高速度カメラで調べた研究 [12] によれば、意図的にコントロールされた表情と、情動喚起による自発的に生じた表情との間では、目、眉、唇の各部の動きには時間的に微妙なずれが存在すること、顔面上の各点の変位量は必ずしも時間の一次関数によって記述できるとは限らないことが確認されている。さらに、顔面変位量を制御する時間関数の与え方によって、表情から人間が認知する感情の強度や表情の自然らしさやわざとらしさの印象は変化し、最適な表情が得られるとされた場合の条件も表情の種類に応じて変化していることが確認されている [13]。

Blair [14] は、表情表出には (1) 感情を生み出す部位 (扁桃体、島皮質など)、(2) 表情を自発的に形作る部位 (大脳基底核)、(3) 周囲の状況に合わせて表情を修飾する部位 (前頭前野)、(4) 実際に表情筋を動かす運動関連領域が必要であると指摘している。山口 [15] らは、「脳はリズムで経験を記憶する」、つまり、特定の脳波に合わせて神経細胞が協調して動き、経験が記憶されると述べている。また、知覚認識では、速いリズムの脳波であるガンマ波に合わせて神経細胞が同時に活動することが解明されつつある。これらの研究から、表情の表出プロセスには、脳の中の神経細胞が刻むリズムが関与していることが推察される。すなわち、図1に示すように、自然な表情と意図的に表情を修飾する場合には、表情発現パスの違いにより覚醒する表情の時系列パターンに変化が発生すると考える。

本研究では、顔の静的多様性と動的多様性に着目した個人固有な表情空間チャート [16] を利用して人間が創り出す表情表出リズムの可視化を目指す。情動の発露として現れる自然な表情は、(1) 人それぞれ固有な空間を形成しており、(2) 表情の動的変化を「表情筋が創り出す顔パターンの位相変化」として捉え、(3) 各表情の覚醒度で空間を組織化・可視化し、(4) 時系列な顔パターンの位相変化を「表情が奏でるリズム」としてモデル化することにより、心理的ストレスレベルに起因する表情の複雑

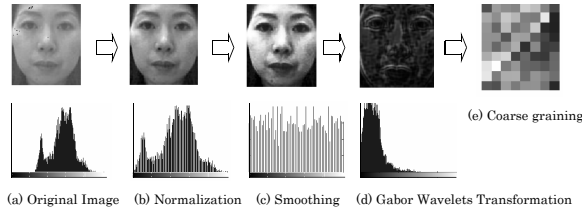


図 2 入力画像の前処理
Fig. 2 Preprocessing steps for input images.

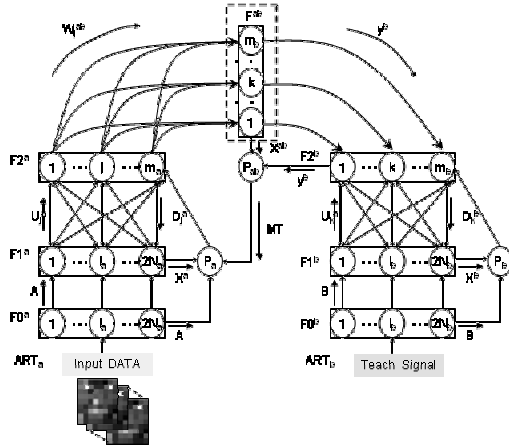


図 3 Fuzzy ARTMAP ネットワーク構成
Fig. 3 Network architecture of Fuzzy ARTMAP.

性や曖昧性を客観的に表現する。

3. 表情表出リズムの抽出法

自己組織化マップ (Self-Organizing Maps, SOM) [23] と Fuzzy ART (Adaptive Resonance Theory) [20] の 2 種類の教師なしニューラルネットワークをハイブリッド化して得られる表情の覚醒度を Fuzzy ARTMAP [21] で学習することにより、表情空間をモデル化するためのコードブックを生成し、隠れマルコフモデル (HMM: Hidden Markov Model) [22] を用いて表情の時系列変化から表情表出リズムを抽出する。表情表出リズムは、心理的ストレスレベルに起因する表情の覚醒度合いとその時間的遷移が作り出すパターンであり、処理対象を顔全体、顔上部、顔下部に分類することにより、3 種類の表情表出リズムを抽出することができる。

3.1 顔画像の前処理

人間は体の大きさが違うように顔の大きさも異なるため、同じ撮影距離で取得した顔画像であっても顔の大きさは異なる。また、撮影時にできるだけ水平 (正面顔) になるように被験者に指示しても、上下左右の動きなどある程度の傾きは避けられない。すなわち、顔画像から表情を自動認識するためには、シーン中から顔を自動検出し、顔の位置や大きさを正規化した顔画像を抽出することが不可欠である。しかしながら、一意に決定できる正規化の基準が存在するわけではなく、研究目的に合わせた基準での正規化が必要となる。顔認識における特徴表現法として、画像の各点の濃淡値が有する大局的分布を主成分分析によって次元圧縮した Eigenface 法 [17] と、人間の視覚野の特性を近似した Gabor フィルタの出力に対する空間的サン

プリングによって得られる Gabor Wavelets 特徴 [18] は、先駆的な位置付けにありその有効性が示されている。一般的に、顔の表情を真顔 (無表情) に限定した場合、Eigenface 法よりも Gabor Wavelets 特徴の方が顔認識のための特徴表現法として優れている。しかしながら、認証する顔の表情変化を許容した場合、Gabor Wavelets 特徴を用いた認証が Eigenface 法の精度を下回ることが報告されている [19]。この実験結果の重要な側面は、Gabor Wavelets 特徴が、人物の違いによる顔の差異よりも表情の違いによる顔パターンの変化を顕著に検出できる特徴表現法であることを示した点にある。

本研究では、図 2 に示すように照明条件がコントロールされた室内での正面顔を対象とし、実験に使用する全ての顔画像に対して濃淡値の正規化を行い、照明条件などによる濃淡値の影響を軽減することに主眼を置く。また、特徴表現法として、顔の動的多様性に基づく顔パターンの幾何学的な位相変化を Gabor Wavelets 特徴の方位選択性により検出し、粗視化処理により撮影時の微妙な位置ズレの影響を抑えることができる。

3.2 Fuzzy ARTMAP アルゴリズム

ARTMAP は 2 つの ART をマップフィールドによって結合した教師あり学習システム [21] で、本研究ではアナログ入力に対応可能な Fuzzy ART を基本構成要素とする Fuzzy ARTMAP を用いる。Fuzzy ARTMAP のネットワーク構成を図 3 に示す。Fuzzy ARTMAP において、Fuzzy ART_a にサンプルデータ a を、Fuzzy ART_b に教師データ b を入力した場合、各 Fuzzy ART の入力に対応するカテゴリ情報がマップフィールド F^{ab} に供給される。 F^{ab} ではサンプルデータと教師データの対応付けが評価され、その成否に基づいてネットワーク構造を変更することで教師あり学習が実現される。

Fuzzy ART_a と Fuzzy ART_b において、サンプルデータ a と教師データ b の属するカテゴリがニューロンの J と K であると判断された場合、Fuzzy ART_a からは荷重 w_j^{ab} に等しい信号が、Fuzzy ART_b からは $F2_b$ 層の活性状態 Y^b が F^{ab} に伝達される。 F^{ab} において下式が成立する場合、サンプルデータ a と教師データ b の対応付けが正しいと判断される。

$$X^{ab} \equiv Y^b \wedge w_j^{ab} > \rho_{ab} Y^b \quad (1)$$

不成立の場合、マッチトラッキング (MT) と呼ばれる処理が実行される。MT とは ART_a のニューロンを強制的にリセットした上で、警戒パラメータ ρ_a を僅かに増加させることであり、上式を満足するニューロンが ART_a で活性化するまで警戒パラメータ ρ_a を増加する。また、サンプルデータ a と教師データ b が正しく対応付けられたとき、荷重は次式によって更新され、警戒パラメータ ρ_a は初期値に戻る。

$$w_j^{a(new)} = \beta^a (A \wedge w_j^{a(old)}) + (1 - \beta^a) w_j^{a(old)} \quad (2)$$

$$w_k^{b(new)} = \beta^b (B \wedge w_k^{b(old)}) + (1 - \beta^b) w_k^{b(old)} \quad (3)$$

$$w_j^{ab(new)} = \beta^{ab} X^{ab} + (1 - \beta^{ab}) w_j^{ab(old)} \quad (4)$$

ただし、 β^a β^b β^{ab} は学習率である。

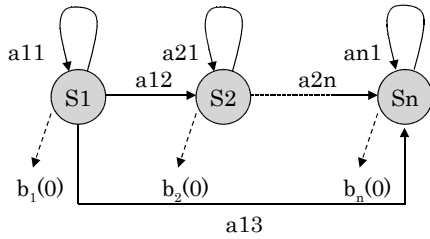


図 4 Left-to-Right 型 HMM の構成
Fig. 4 Structure of left-to-right HMM.

3.3 HMM アルゴリズム

HMM は複数の状態を持ち、それら相互の状態間の遷移確率により定義される単純マルコフモデルに、各状態からの複数シンボルの出力確率を加えたモデルである。HMM は抽出対象カテゴリ毎に用意し、各 HMM を各カテゴリの学習用シンボルの系列で学習させる。学習とは、学習用シンボルの系列を発生する確率が高い状態遷移確率とシンボルの出力確率の 2 つのパラメータを推定することである。また、学習は Baum-Welch アルゴリズム [22] を用いて、出力尤度の変化が十分に小さくなり、パラメータが収束したと考えられるまで繰り返す。

HMM の構成は、図 4 に示すような Left-to-Right 型とし、状態を状態数 n に対して左から順に S_1, S_2, \dots, S_n とする。ここで、 S_1 を表情表出時の初期状態、 S_2, \dots, S_{n-1} を中間状態、 S_n を最終状態と呼ぶ。状態 $S_i (i = 1, \dots, n)$ の状態確率の更新値を求めるために以下の確率を定義する。

(1) 状態遷移確率 A_{ij} : 状態 S_i から次の時刻に状態 S_j に遷移する確率で、Left-to-Right 型 HMM では自己および右側の状態への遷移のみが可能であることから、以下の拘束条件が成り立つ。

$$A_{ij} = 0, \quad \text{if } j < i \quad (5)$$

$$0 \leq A_{ij} \leq 1, \quad \text{if } j \geq i \quad (6)$$

$$\sum A_{ij} = 1 \quad (7)$$

(2) シンボルの出力確率 $B_i(O)$: 状態 S_i においてシンボルの系列 O を出力する確率密度分布で、ここでは音声認識の分野で良く用いられている離散的なシンボルの確率を割り付ける離散分布を用いる。

4. 評価実験

本実験の被験者は、18 歳から 22 歳までの男子学生 5 名と女子学生 5 名の計 10 名で、撮影環境は一般的な室内（通常の蛍光灯による照明下）とし、極端な陰影が付かないように考慮した。表情の種類は、心理的ストレスレベルとの関連性が顕著に表れやすい「笑い」、「怒り」、「悲しみ」の 3 表情を対象とし、表情画像は、被験者が意図的に各表情を表出し、かつ、無表情から表情の表出を交互に繰り返した動画（毎秒 10 フレームで約 20 秒間撮影）を 200 フレームの静止画像列（8 ビットのグレースケール）に変換した。

4.1 表出リズムの抽出結果

意図的に「笑い」の表情を表出した場合において、200 フ

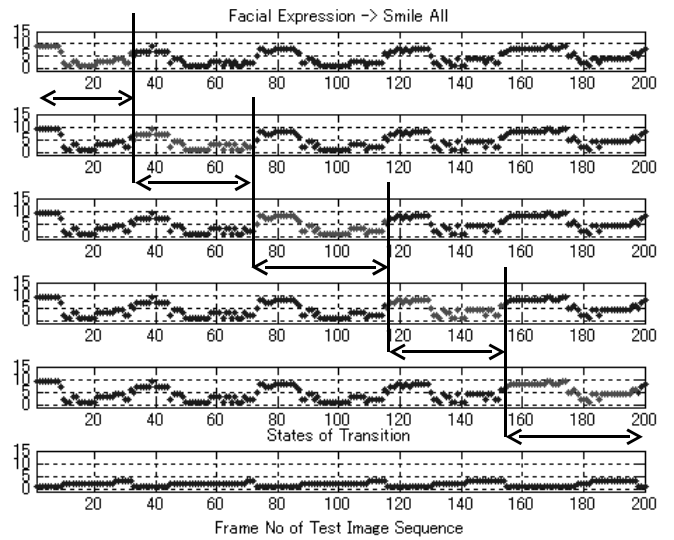


図 5 状態数 3 の抽出結果
Fig. 5 Extraction results of three states.

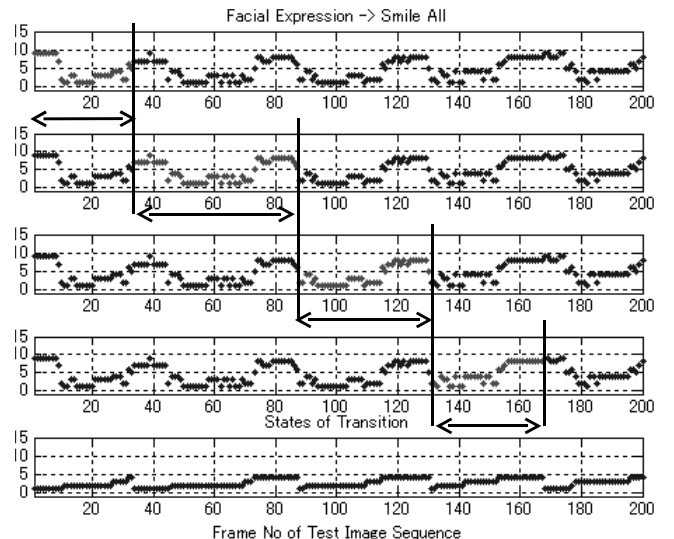


図 6 状態数 4 の抽出結果
Fig. 6 Extraction results of four states.

レームの静止画像列に対して前処理を行った後、表情の覚醒度を Fuzzy ARTMAP で学習することにより「笑い」の表情空間をモデル化するためのコードブックを生成する。次に、HMM の状態数を 3 から 5 まで変化させ、このコードブックを用いて表情の時系列変化から 3 種類（顔全体、顔上部、顔下部）の表情表出リズムを抽出した。顔全体の抽出結果を図 5, 6, 7 に示す。図 5 は状態数が 3、図 6 は状態数が 4、図 7 は状態数が 5 の HMM が推定した状態遷移系列に対応する 200 フレームのコードブック（表情の覚醒度）をラベル化し、上段から抽出順に整理したものである。また、図中の各段における横軸はフレーム数を示しており、最下段は HMM が推定した状態遷移系列を表している。なお、HMM の状態数に応じて抽出（ラベル化）されるフレーム範囲が変化するため、表情の覚醒度（コードブック）が刻む時系列パターンの基本単位である表情表出リズムが異なる。

図 8 は、200 フレームの静止画像列の中から、状態数が 5 の

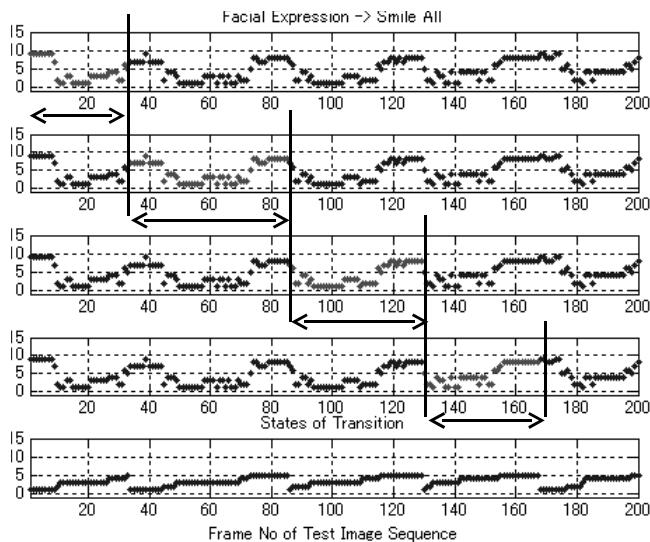


図 7 状態数 5 の抽出結果
Fig. 7 Extraction results of five states.

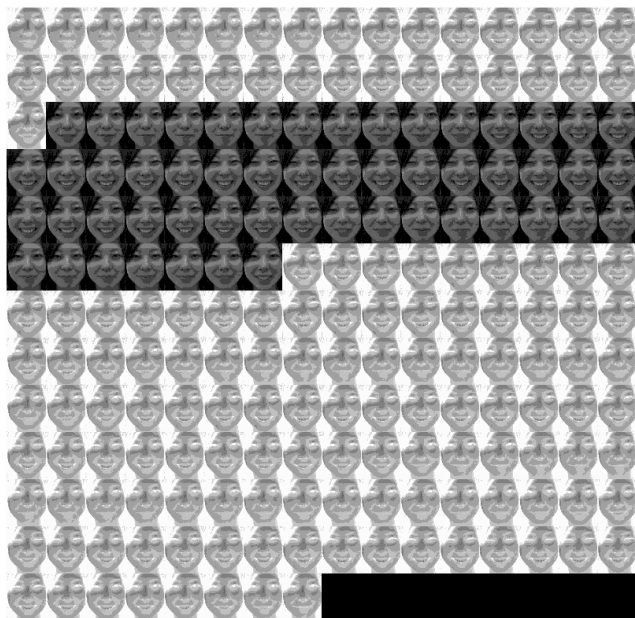


図 8 表情画像のサムネイル表示
Fig. 8 Thumbnail images of facial expressions.

HMM が推定した状態遷移系列を基に、2 番目にラベル付けされた「笑い」の表情表出リズムに対応する顔表情画像の時系列変化を選択し、サムネイル表示したものである。図中、左上がフレーム No.1、右下がフレーム No.200 に対応する。HMM が推定した状態遷移系列は、表情表出に伴う表情変化の範囲を適切に捉えている。

4.2 実験結果の考察

表情表出リズムと HMM の状態数の関連性を考察する。図 5, 6, 7 から理解できるように、HMM の状態数は、表情のコードブックが刻む時系列パターンの抽出範囲に変化を与えている。これらの結果は、表情表出プロセスにおける中間表情の覚醒パターンを抽出する際、HMM の状態数を用いて制御可能であることを示している。すなわち、HMM の状態数を可変にすることにより、単調な表情表出リズムから意図的で複雑な表情表出

リズムまで抽出可能であると考えられる。また、目と眉で特徴付けられる顔上部や口元の変形で特徴付けられる顔下部のリズムデータを併用することにより、心理的ストレスレベルの推定が期待できる。

5. まとめ

本研究では、表情の覚醒度を Fuzzy ARTMAP で学習することにより、表情空間をモデル化するためのコードブックを生成し、HMM を用いて表情の時系列変化から表情表出リズムを抽出する手法を提案した。意図的な「笑い」の表情に着目し、表情表出プロセスにおける中間表情の覚醒パターンを抽出する際、HMM の状態数を用いて制御可能であることを確認した。

文 献

- [1] 赤松茂, "人間とコンピュータによる顔表情の認識 [I]-コミュニケーションにおける表情とコンピュータによる自動解析-", 信学誌, Vol.J85, No.9, pp.680-685, 2002.
- [2] J.N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *Journal of Personality and Social Psychology*, vol.37, no.11, pp.2049-2058, 1979.
- [3] G.M. Leonard, K.K.S. Voeller, and J.M. Kuldau, "When 's a smile a smile? Or how to detect a message by digitizing the signal," *Psychological Science*, vol.2, no.3, May 1991.
- [4] P. Ekman and W.V. Friesen, "Felt, false and miserable smiles," *Journal of Nonverbal Behavior*, vol.6, no.4, pp.238-253, 1982.
- [5] I. Pilowsky, M. Thornton, and B.B. Stokes, "Towards the quantification of facial expressions with the use of a mathematic model of the face," in *Aspects of Face Processing*, eds. H.D. Ellis, M.A. Jeeves, F. Newcombe and A. Young, pp.340-348, Martinus Nijhoff, Dordrecht, 1986.
- [6] 西尾修一, 小山謙二, "目と口の動きの時間的差異に基づく笑いの分類基準," 信学論, J80-A, no.8, pp.1316-1318, 1997.
- [7] 太田寛志, 佐治斉, 中谷広正, "顔面筋に基づいた顔構成要素モデルによる表情変化の認識," 信学論 (D-II), Vol.J82-DII, No.7, pp.1129-1139, Jul 1999.
- [8] 川嶋宏彰, 西山正紘, 松山隆司, "表情譜: タイミング構造に基づく表情の記述・生成・認識", FIT2005 (第 4 回情報科学技術フォーラム) 情報科学技術レターズ, pp.153-156, 2005.
- [9] P. Ekman and W. V. Friesen, *Pictures of facial affect*, Human Interaction Laboratory, Univ. of California Medical Center, San Francisco, 1976.
- [10] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying Facial Actions," *IEEE Trans. PAMI*, vol.21, no.10, Oct. 1999.
- [11] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis," *Proc. Int'l Conf. Face and Gesture Recognition (FG2000)*, pp.46-53, Mar. 2000.
- [12] 内田英子, 四倉達夫, 森島繁生, 山田寛, 大谷淳, 赤松茂, "高速度カメラを用いた顔面表情の動的変化に関する分析," 信学技法, HIP99-76, pp.1-6, March 2000.
- [13] 尾田正臣, 磯野勝宣, "表情認知の時間関数特性についての一検討," 信学技法, HCS01-43, pp.67-72, Jan. 2002.
- [14] R.J.R. Blair, "Facial expressions, their communicatory functions and neuro-cognitive substates," *Philos. Trans. R. Soc. Lond.*, B358, pp.561-572, 2003.
- [15] 独立行政法人 理化学研究所 監修, "理研探偵団: オンラインに挑む研究者の素顔に迫る", pp.266-277, 2006.
- [16] 川澄晃, 佐藤和人, 間所洋和, "表情の動的多様性に基づく表情パターンの分類法," 信学技法 (PRMU2008-12), 2008.
- [17] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. of Cognitive Neuroscience*, vol.3, no.1, pp.71-86, 1991.

- [18] L. Wiskott, J.M. Fellous, N. Kruger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," IEEE Trans. on PAMI, vol.19, no.7, pp.775-779, 1997.
- [19] 加藤博之, 田中健一, 赤松茂, "被認証顔の属性の多様性に対する個人認証システムのロバスト性の評価 - GaborJet 法と Eigenface 法の比較 -," 信学技法, HI2005-16, pp.87-92, 2005.
- [20] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System," Neural Networks, vol.4, pp.759-771, 1991.
- [21] Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Transactions on Neural Networks, 3, 698-713.
- [22] 電子情報通信学会編, "確率モデルによる音声認識," pp.29-66, コロナ社, 1988.
- [23] T. Kohonen, Self-organizing maps, Springer Series in Information Sciences, 1995.