

自発的表情における表情表出テンポとリズムの抽出

須藤 尚志[†] 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 株式会社 〒011-0945 秋田県秋田市土崎港西 3-9-15 チャレンジオフィスあきた 304
E-mail: †{b09a047,ksato,madokoro}@akita-pu.ac.jp, ††sakura@smart-d.jp

あらまし 本論文では、表情表出テンポとリズムという枠組みを提案する。我々は特に自発的な表情表出の繰返しにより生成される覚醒度の時系列変化に着目し、心理的ストレスが表情の表出プロセスに与える影響を解析する。本研究では無表情から表情表出を経て無表情に戻る覚醒度の時系列変化を1テンポ、テンポを複数回繰り返したものを1リズムと定義する。我々はLeft-to-Right型の隠れマルコフモデルHMMs (Hidden Markov Models) を用いることで、表情表出テンポとリズムを取得する。表情表出テンポとリズムは、心理的ストレスの中でも外部刺激に伴う一過性ストレスに強く影響を受けると考えられる。評価実験では、被験者6名に対して平常状態、快刺激を与えた状態、不快刺激を与えた状態の表情画像データセットを取得し、各被験者の表情表出テンポとリズムを抽出し解析することにより、一過性ストレスが表情の表出プロセスに与える影響を明らかにする。

キーワード 隠れマルコフモデル, 覚醒度, 自己組織化マップ, ファジー適応共鳴理論, 表情空間チャート.

Extraction of Expressive Tempos and Rhythms with Spontaneous Facial Expressions

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

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

^{††} SmartDesign Co.,Ltd 304 Challenge office, 3-9-15 west tsuchizakiminato, Akita City, 011-0945 Japan
E-mail: †{b09a047,ksato,madokoro}@akita-pu.ac.jp, ††sakura@smart-d.jp

Abstract This paper presents a new framework to extract tempos and rhythms from facial expressions. We specifically examine time-series changes of arousal levels obtained from iteration of spontaneous facial expressions. Moreover, we analyze expressive processes affected by psychological stress. We define time-series changes of arousal levels from the maximum expressions to a neutral expressions as one tempo. Moreover, we define some tempos as one rhythm. We obtain facial expressions tempos and rhythms using left-to-right Hidden Markov Models (HMMs). We consider that facial expressions tempos and rhythms are strongly affected by temporal stress with external stimulus in psychological stress. For the evaluation experiment, we created an original facial expressions dataset in three states (normal, pleasant, unpleasant) of six subjects. Subsequently, we demonstrate expressions processes affected by temporal stress for extracting and analyzing facial expressions tempos and rhythms in each subject.

Key words HMMs, Arousal levels, SOMs, Fuzzy ART, Facial expressions spatial charts.

1. はじめに

顔は様々な情報を発信しており、人間は表情が与える多彩な情報から相手の意思や感情を機敏に察することができる。特に、親しい友人や家族などの間柄なら、体の調子やその時の機嫌なども表情から感じ取ることができる。このように、人間のコミュニケーションにおいて表情の果たす役割は大きい。次世代のMMI (Man-Machine Interface) として発展が期待され

ている [1]。赤松 [2] は人間の顔には静的多様性と動的多様性が存在すると述べている。静的多様性とは、顔の構成要素である目、鼻、口、耳などの形状、大きさ、配置といったように、個人が有する多様性を意味する。人物を同定したり、性別などの属性や印象を判定できるのは、このような顔が持つ静的な特徴の多様性によるからである。一方、動的多様性とは、時々刻々と変化する「喜び」や「怒り」といった内的な情動の発露として無意識のうちに、あるいは相手に対するメッセージとして何

らかの感情を伝達するために、人間が顔面の表情筋を動かすときの多様性を意味する。これによって、顔を構成するパーツの形状や全体の配置が変化し、真顔からの変位として表情が表出される。顔画像処理の中でも、表情を扱う場合には、静的多様性と動的多様性の両面から扱う必要がある。

我々は、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、各表情の豊かさを覚醒度で表し表情空間を組織化・可視化することを目的として、表情の動的多様性を表現するための表情空間チャートという枠組みを提案した [3]。表情空間チャートの生成手順は自己写像特性を有する SOMs (Self Organizing Maps) を用いて表情表出に伴う表情空間のカテゴリを抽出し、安定性と可塑性を併せ持ち適応的にカテゴリを分類できる Fuzzy ART (Adaptive Resonance Theory) ネットワークを用いて一定の粒度で再分類することによりカテゴリを統合する。更に、Fuzzy ART から得られるカテゴリから、無表情画像が含まれるカテゴリを基準として、Russell の円環モデル [4] における覚醒度とそれぞれのカテゴリを対応付けることにより、表情表出時の覚醒度の動的多様性を表情空間チャートとして可視化した。

人間は動いているもの、音を発しているものなど身の回りのものすべてにリズムを感じることができる。また、人間は会話のリズム、生活のリズムというように日常生活を営む上で何らかのリズムを感じている [5]。このような生体リズムはパーソナルテンポ (精神テンポ) に基づいている。パーソナルテンポとは個人に固有の生体リズムのことであり、話したり歩くといった日常の生活行動において、特に制約のない自由な行動場面で自然に表出される個人固有の速さのことを言う。表情表出は日常的生活行動の一つであるため、個人に固有のリズムが存在すると考えられる。そこで、本研究では表情からストレス状態を推定するために、人間が自ら進んで発する自発的表現の繰り返しの覚醒度の時系列変化に着目し、表情表出テンポとリズムという枠組みを提案する。表情表出テンポとリズムは、無表情から表情表出を経て無表情に戻る覚醒度の時系列変化を 1 テンポ、テンポを複数回繰り返したものを 1 リズムと定義する。本研究では、表情表出テンポの抽出に Left-to-Right 型の隠れマルコフモデル HMMs (Hidden Markov Models) を用いる。HMMs は時系列データからパターンの生起部を抽出しカテゴリとして分類する手法として、音声認識や各種信号処理分野で幅広く用いられており、覚醒度の生起パターンである表情表出テンポの抽出に対して最適と考えられる。ストレス反応は身体 (心拍数変化、血圧変化)、心理 (落ち込み、イライラ)、行動 (酒量増加、じっとしてられない) に表れると言われている [6]。ここで、表情はストレス反応の中でも行動に分類される。そのためストレス状態の異なる、表情表出テンポとリズムを解析することで、個人固有の表情変化からストレス状態の推定が可能になると考えられる。本研究では、人間のストレス状態に起因する表情の複雑性や曖昧性を客観的に表現するための表情表出テンポとリズムを抽出する手法を提案する。

2. 周辺研究

表情の動的側面に着目した研究として、平山 [7] らは、顔パーツの運動を、区間を単位として表し、表情におけるタイミング構造を記述する表現形式として表情譜を提案している。平山らは、Ekman [8] が提唱した FACS (Facial Action Coding System) における 46 種類の基本単位 (Action Unit: AU) では描写できる表情が静的なものに留まり、時間的な描写ができないとして、特徴ベクトル系列からボトムアップに線形システム (モード) を求めている。ここでモードとは顔パーツの様々な静止状態や運動状態 (例えば、口パーツでは開く、開いたまま、閉じる、閉じたままのモード集合の要素がある) を表している。平山らの手法では、まず始めに表情時系列画像に対して AAM (Active Appearance Model) を用いて各眉、各目、鼻、唇、顔の下半分の輪郭の計 58 点の特徴点を追跡し、それぞれの顔パーツの特徴ベクトル系列を得る。次に、得られた特徴ベクトルからモードへの自動文節化を行い、自発的表現と非自発的表現の表情譜を獲得している。実験結果では得られた 2 つの表情譜におけるタイミング構造に着目して、2 つの表情がどの程度分離されるかを解析している。解析結果では 2 つの表情で口を動かす筋肉と頬を持ち上げる筋肉の動くタイミングに差異がみられている。このように表情譜では表情表出のタイミング構造を記述するために、時系列画像及びモードの時間分解能は高く設定されている。しかしながら、表情表出を表現するモードは空間分解能が低く、表情表出強度の差異の解析においては表情譜では十分な結果が得られないと考えられる。

大塚 [9] らは、異なった表情筋の動き (弛緩、収縮、静止、伸長) に対応した状態を遷移する HMMs により表情の動きをモデル化し、Ekman らの基本 6 表情 [8] (「喜び」、「怒り」、「悲しみ」、「嫌悪」、「恐れ」、「驚き」) を個別に抽出する手法を提案している。大塚らは、FACS の AU は目および口の周囲に分布していることに注目し、それらの周囲の動きベクトルを分析することによって表情の認識を試みている。大塚らの手法では、まず始めに表情時系列画像 (20 例において 2 種類の表情表出) から勾配法 [10] により目および口の周囲の動きベクトルを取得する。次に、縦横成分に 2 次元フーリエ変換を施し 15 次元の特徴ベクトルの時系列を取得する。この特徴ベクトルの時系列を入力として、Left-to-Right 型 HMMs を適用し個別表情を抽出している。このときの表情の正誤判定は、実験者が確認し、実際の表情表出区間において対応する表情が抽出された場合を正解に、いずれの表情も抽出されなかった場合と異なった表情が一回でも抽出された場合は誤解答としている。実験結果は 20 例の中で抽出された 40 表情のうち、正確に抽出された表情の数は 36 であり、90% の抽出率が得られている。しかしながら、正誤判定は、表情表出区間内に対応する表情が表出されていれば正解としているため、表情の表出及び終了の正確な区間が抽出されているとは限らない。

平山らが表情表出の強度としてモードを用いたのに対して、我々のアプローチでは覚醒度を用いる。覚醒度は無表情から表情最大表出までを段階的に表すため、表情表出の分解能はモー

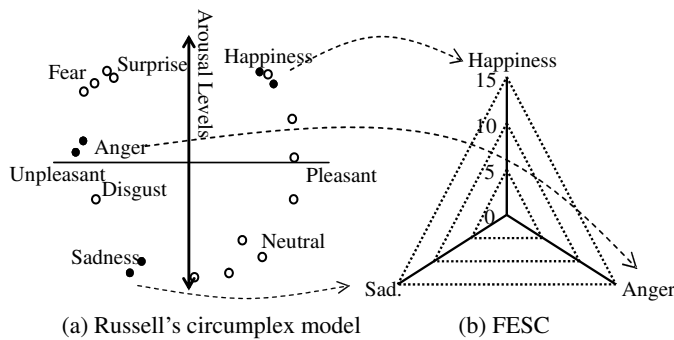


図 1 Russel の円環モデルと表情空間チャートの対応関係。

Fig. 1 Correspondence relationship between Russell's circumplex model and FESC.

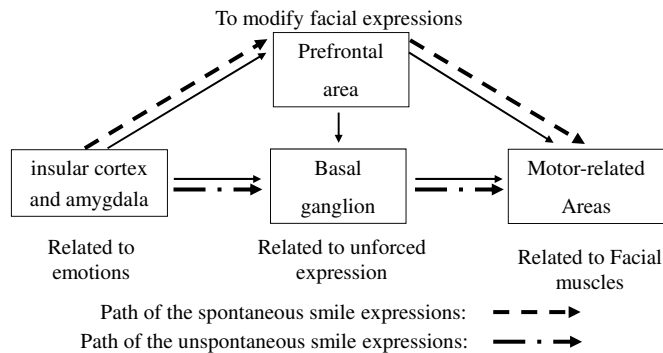


図 2 自発的な表情と非自発的な表情の発現パス。

Fig. 2 Paths of spontaneous facial expressions and unspontaneous facial expressions.

ドに比べ高くなる。そのため、表情表出強度の差異に着目した解析が可能となる。また、平山らが顔のパーツの特徴点のベクトル系列、大塚らが目および口の周りの動きベクトルによって実験を行っているのに対して、我々のアプローチは、表情の見え方ベースで実験を行う。見え方ベースでの実験の利点は、特徴点の抽出及び追跡が不要であるため、特徴点ではなく全体の見え方で判断する、人間の認知に近い条件で実験が可能なことである。また、大塚らが正誤判定を実験者の主観で行ったのに対して、我々は表情表出終了フレームの GT (Ground Truth) を設定し、HMMs で抽出されたフレームと GT のフレームとが ± 5 フレームの範囲内で一致しているか否かで正誤の判定を行っている。この判定方法により HMMs で表情表出の終了フレームを抽出できているかを客観的に判断することができる。

3. 表情表出テンポとリズム

3.1 覚醒度と表情空間チャート

我々は前報で、表情空間を定量化するための指標として、覚醒度という概念を導入した [3]。覚醒度とは、図 1(a) に示す Russell の円環モデルの縦軸である覚醒次元に注目して、その度合いを定量化した数量である。Russell の円環モデルでは、全ての感情は快の次元 (快・不快) と覚醒の次元 (覚醒・眠気) の 2 次元で表される平面上に円環状に布置される。Russell の円環モデルの覚醒次元に注目し、個々の表情空間の基準となる無表情から表情が表出されるまでの顔パターンの幾何学的変化

を位相変化として定量化した値を覚醒度と定義する。ここで、表情画像の位相変化から覚醒次元が抽出できるのは、本研究では自発的な表情を扱っているからと考える。表情画像中には快の次元の要素も含まれているが、自発的に表出した表情であるため、快・不快に関係なく (すなわち、心理的状态を抑えて) 表情を形成するために表情筋を動かしていると考えられる。つまり、自発的な表情表出を扱った実験では、表出される表情のパターンは覚醒の次元に対して強く対応付くと考えられる。一方、自発的な表情を扱う本実験では、快の次元の影響を直接的に扱うことが難しいと考えられるため、心理学分野で用いられている専用のストレスシートを使って評価する。

次に、表情空間とは、表情から認知される様々な感情の意味的な対極性に注目して、各表情を空間的に配置した概念である [2]。表情空間には、表情を表出する際に生じる顔パーツの変化の物理的パラメータと、認知される感情の心理的パラメータとの対応関係が表現されている。物理的パラメータは、人物ごとに大きさや形状が異なる顔パターン上で、表情表出によって生じる顔面変形の種類や大きさを、一定の基準に基づいて記述する必要がある。そこで我々は、表情空間を記述し各表情を構成する覚醒度のパターンを表現する手法として、表情空間チャートという枠組みを提案し、表情の物理的パラメータの記述・表現を試みた [3]。Russell の円環モデルの中から、第一象限に位置する「喜び」、第二象限に位置する「怒り」、第三象限に位置する「悲しみ」の 3 表情を対象とした。表情空間チャートの例を図 1(b) に示す。各軸は各表情の覚醒度の最大値となる。覚醒度の最大値間を結ぶことで、表情空間チャートが生成される。

3.2 表情表出テンポとリズムの定義

Blair [11] 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) 表情を非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 [12] 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働き、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 2 に示すように、非自発的な表情と自発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、覚醒度の時系列変化を利用して、人間が創り出す表情表出リズムの可視化を目指す。ここで、表情表出テンポとリズムという言葉の定義については、無表情から表情表出を経て無表情に戻る覚醒度の時系列変化を 1 テンポ、テンポを複数回繰り返したものを 1 リズムとする。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

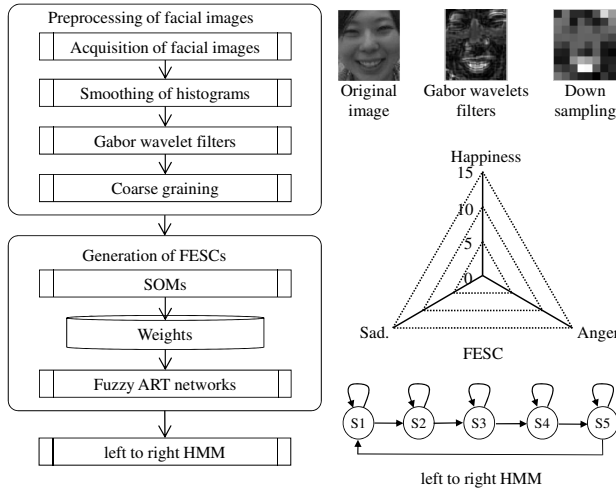


図3 提案手法の全体構成.

Fig. 3 Procedures and whole architecture of our method.

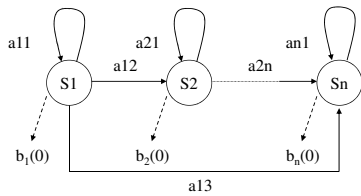


図4 Left-to-Right型HMMsの構成.

Fig. 4 Structure of Left-to-Right HMMs.

ム」としてモデル化することにより、ストレス状態に起因する表情の複雑性や曖昧性を客観的に表現する。

4. 提案手法

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。このため、赤松は、個々人の表情表出の特性に応じて修正していく適応的な学習メカニズムが必要と述べている [2]。したがって、本研究では、被験者が自発的に表出した表情を対象として、時間軸方向への圧縮による正規化と表情表出における位相変化を抽出するために SOMs を用いて表情パターンの分類を行う。更に、SOMs により分類した表情画像を、安定性と可塑性を併せ持った適応的学習アルゴリズムである Fuzzy ART を用いて再分類する。SOMs は、予め決められた写像空間の中で相対的にカテゴリ分類を行うが、Fuzzy ART は、ビジランスパラメータで制御された一定の粒度のもとでカテゴリ分類を行うため、長期間に及ぶ時系列データに対しても、同じ基準で分類することができる。提案手法の処理手順を図 3 に示す。以下では、前処理、SOMs によるカテゴリ分類、Fuzzy ART によるカテゴリの統合及び覚醒度の抽出 [3]、HMM による表情表出テンポの抽出について個別に説明する。

4.1 覚醒度の時系列変化の取得

まず、表情時系列画像に対して、輝度値の正規化によって、照明条件などによる濃淡値の影響を軽減する。更に、ヒストグラムの平滑化によって、画像の明瞭化、コントラストの調整を

行う。次に、特徴表現法として Gabor Wavelets 特徴の方位選択性により、目、眉、口、鼻といった表情表出を特徴づける部位を強調する。Gabor Wavelets 変換した表情時系列画像に粗視化処理を行うことで、ピクセル情報量の圧縮と、顔位置の微小なずれの影響を緩和する。取得した粗視化画像の輝度値情報を SOMs により学習し、顔画像を類似した特徴ごとに、15 ユニットに振り分ける。続いて、類似したユニットを Fuzzy ART で同一のカテゴリに統合する。無表情を表情表出の 1 段階目、Fuzzy ART によって統合された、各カテゴリに属する顔画像の平均画像と無表情との相関を取り、最も近いものを表情表出の 2 段階目、2 段階目の画像を基準に以下同様に相関を取り、並び替えることで表情表出レベルを定量化した覚醒度を得る。最後に、原画像の時系列変化と覚醒度を対応付け、覚醒度の時系列変化を得る。

4.2 HMMs による表情表出テンポの抽出

HMMs は、音響信号からの音韻の推定を行い単語を認識する手法として、まず音声認識分野において使用された。また、武田 [13] らは、HMMs によって MIDI 音楽の楽譜追跡と自動伴奏を行った。このように HMMs は、時系列データからパターンの生起部を抽出し、カテゴリとして分類する手法として確立されている。我々はデータセットとして、表情の時系列画像を対象としており、覚醒度の生起パターンとして、表情表出テンポという枠組みを提案している。そのため、表情表出テンポを抽出するためには、HMMs の使用が最も適していると判断した。

HMMs 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。HMMs は抽出対象カテゴリ毎に用意し、各 HMMs を各カテゴリの学習用シンボル系列で学習させる。学習とは、学習用シンボル系列を発生する確率が高い状態遷移確率とシンボル出力確率の 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 [14] 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。HMMs 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 $S1; S2; \dots; Sn$ とする。ここで、 $S1$ を表情表出時の初期状態、 $S2; \dots; Sn-1$ を中間状態、 Sn を最終状態と呼ぶ。状態 Si ($i=1; \dots; n$) 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。

状態遷移確率 A_{ij} : 状態 Si から次の時刻に状態 Sj に遷移する確率で、Left-to-Right 型 HMMs 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。

$$A_{ij} = 0, \quad \text{if } j < i \quad (1)$$

$$0 \leq A_{ij} \leq 1, \quad \text{if } j \geq i \quad (2)$$

$$\sum A_{ij} = 1 \quad (3)$$

シンボル出力確率 $B_i(O)$: 状態 Si においてシンボル系列 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、離散的なシンボルに確率を割り付ける離散分布を用いる。

大塚 [9] らは、表情表出における表情の状態は、無表情→表情表出→無表情という遷移によって成り立っており、このときの表情筋の動作は、弛緩→収縮→静止→伸長→弛緩という動作であるとした。大塚らの手法では、HMMs の状態と表情筋の状態が初期値で対応付けされている条件下で、HMMs で遷移可能な状態数を 5 と設定し、表情筋の状態を HMMs によってモデル化している [9]。それに対して本研究では、HMMs の状態数を 3~10 の範囲で変化させて、式 (4) に示す抽出率の最も高くなる最適値を求めてから、その状態数を用いて実験を行う。状態数を変化させ最適化することで、ストレス刺激を与える条件下での表情筋の動きを最適に表現する状態数を得ることができる。

また、我々は HMMs による抽出の正誤判定として、実験者が原画像において無表情に戻っていると判断したフレームを GT として、HMMs で抽出されたフレームが GT のフレームと ±5 の範囲で一致していれば抽出成功、そうでなければ抽出失敗とした。この正誤判定における抽出率を式 (4) に示す。

$$A = \frac{x_1, x_2, x_3}{C} \times 100[\%]$$

$$x_1, x_2, x_3 = \begin{cases} 1, & E \subseteq R \pm 5 \\ 0, & \neg(E \subseteq R \pm 5) \end{cases} \quad (4)$$

なお、式 (4) における A は抽出率、C は表情表出回数、E は HMMs で抽出した表情表出終了フレーム、R は実際の表情表出終了フレームである。

5. データセット

5.1 表情画像データ

本研究では、長期間に渡る表情変化を扱うために、独自のデータセットを構築した。また、表情の撮影と同時に、専用のストレスシートを用いて心理的ストレスレベルを測定した。

5.2 表情画像データ

表情認識用の顔画像データセットは、大学や研究機関において様々な撮像条件で作成したものが、インターネット等を通じて一般に公開されており、特に性能比較のために活用されている。静止画像に関しては、表情認知の心理実験で視覚刺激とする目的で収集された Ekman&Friesen のデータセット [8] が有名である。動画画像では、Cohn Kanade [15] や Ekman Hager [16] などの表情データセットが有名であり、評価及び比較実験用として幅広く利用されている。また、近年では、Pantic らが公開している MMI Facial Expression Database [17] が著名である。これらのデータセットは、被験者数としては十分な人数が確保された横断的なデータセットであるものの、各被験者に対して 1 回のみ撮影に留まっており、同一被験者に対して長期的に表情を撮影したデータベースは存在しない。そこで我々は、同一人物の特定表情を長期間に収集した縦断的なデータセットを独自に構築した。

前報で提案した表情空間チャート [3] では、「喜び」、「怒り」、「悲しみ」の 3 表情を対象にして表情空間の可視化を行ったた

め、表情画像データセットはその 3 表情で構築されていた。本研究では、その 3 表情の中の「喜び」の表情とストレスの関係性に注目し、快刺激を与えストレスが低い状態の「喜び」、不快刺激を与えストレスが高い状態の「喜び」を対象として、6 名の被験者に対して表情画像を取得した。刺激を与える方法として、快・不快の情動を与える情動喚起ビデオを事前に選定し、被験者はそれを視聴した直後に表情表出を行った。被験者の内訳は、男子大学生 6 名 (B = 21 歳, A, D, F = 22 歳, C, D=23 歳) である。撮影期間は全ての被験者において 3 週間 (1 週間間隔) とした。撮影環境は、室内の一角にカーテンで仕切られた表情撮影用スペースにおいて、被験者の頭部がフレーム中に含まれる状態で正面顔画像を撮影した。あらかじめ被験者には、頭部をあまり動かさないで表情を表出するように指示して撮影したため、一定の範囲内に顔領域が収まっているが、微小な変動に対しては、Haar-like 特徴と Boosting を用いた顔領域のトラッキング手法を用いて補正した。撮影の容易性と今後の応用展開を考慮して、カメラは市販の USB カメラ (Logicool 社製 QcamOrbit) を用いた。被験者には 10 秒間を目安に自分のタイミングで表情表出を 3 回繰り返し、3 回表出し終えたら無表情を継続するよう指示した。カメラのサンプリングレートは毎秒 15 フレームに設定し、1 セット 200 フレームの画像列から構成されるよう設定した。

5.3 ストレス測定方法

表情に影響を与えると考えられる心理的ストレスとして、慢性的ストレスと一過性ストレスがある。慢性的ストレスは人間が日常から抱えているストレスであり、一過性ストレスは一時的な刺激を与えられたときに生じるストレスである。本研究では被験者毎に固有の慢性的ストレスがある状態に対して、一過性ストレス刺激を与えるため、慢性的ストレス状態を測定するために POMS (Profile of Mood States) を、一過性ストレス状態を測定するために唾液アミラーゼ試験によるストレス測定を行った。

慢性的なストレス反応測定尺度として、気分、感情、情緒といった主観的側面からの評価が行える POMS がある。POMS は、日常生活で経験する心理的ストレス反応を、簡易かつ多面的に測定することができる質問紙から構成されている。ストレス反応としては、日常的に体験する各種ストレスサによって引き起こされる、憂うつ・不安や怒り (情動的反応)、無気力や集中困難 (認知的反応)、仕事の能率の低下 (行動的反応) などがある。測定内容としては、緊張・不安、抑うつ・落ち込み、怒り・敵意、活気、疲労、混乱の 6 つの気分尺度を同時に評価することが可能である。

POMS は 65 項目の質問に対して、過去 1 週間の気分を表すのに最も当てはまるものとして、「まったくなかった」から「非常に多くあった」の 5 件法で記入し、それぞれに 0~4 ポイントの得点が与えられる。得点は 6 つの気分尺度の項目に分類し、尺度ごとに合計点 (素得点) を算出する。素得点に対して標準化得点を算出し、各々の尺度で標準化得点が高いほど気分の状態が甚だしいことを表現する。POMS に対して POMS 短縮版がある。POMS 短縮版は解答方法、採点方法は同じであり質

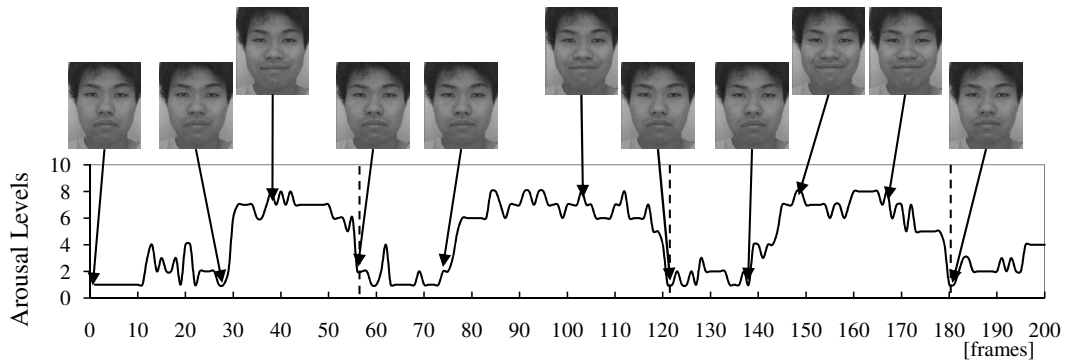


図 5 表情表出に伴う覚醒度の時間変化 (被験者 A, 1 週目, 快刺激後の喜び) .

Fig. 5 Time-series changes of arousal levels (Subject A, first week, happiness after stimulated comfortably).

問項目を 65 項目から 30 項目に短縮しながら, POMS と同様の結果が得られる. また, 項目数を減らすことで被験者の負担感を軽減することが可能となる. 質問用紙の記入による負担自体が, ストレス要因になることも考えられるため, 負担を軽減するという理由で, 本研究では POMS 短縮版を用いることとした. 本実験では, 被験者 6 名を対象として, 表情撮影 (3 週間) を行うと同時に POMS で心理的ストレスの測定を行った. POMS への記入は表情撮影前に行い, 表情の表出に影響を与えないようにするために, 得点は被験者に提示していない.

一過性のストレス反応測定方法として唾液アミラーゼ試験がある. 唾液アミラーゼ活性は生体反応として快状態ならば低く, 不快状態ならば高く検出される. 山口 [18] らは, 一過性の外部刺激を受けた時のストレス反応として, 唾液アミラーゼ活性がストレス評価のための有効な物質であることを確認している. 一過性の刺激に対するストレス反応の評価として適切であるため, 本研究では情動喚起ビデオ視聴直後に対する一過性ストレス反応の測定方法として唾液アミラーゼ試験を用いた.

6. 評価実験

本章では始めに, 表情表出区間と覚醒度の対応関係を確認する. 次に, HMMs の状態数の最適化を行う. 続いて, 最適化された状態数を用いて, 覚醒度の時系列変化に対して HMMs を適用することで得られる, 表情表出テンポの抽出結果を示す. 最後に, 刺激付与状態時の HMMs により抽出された表情表出終了フレームにおける覚醒度から, 無表情領域の広がりの変異を比較する.

6.1 表情表出区間と覚醒度の対応関係

覚醒度の抽出結果の一例として, 被験者 A (22 才, 男性) の 1 週目における, 喜びの覚醒度の時系列変化を図 5 に示す. 横軸は画像のフレーム数であり, 各画像セットとも 200 フレームから構成されている. 縦軸は覚醒度を示す. 事前に覚醒度の時系列変化から, 表情表出が終了しているフレーム位置 (55, 124, 181) を確認し, 図中に縦の破線で示している. 被験者は撮影 1 セット (10 秒間) で 3 回の表情表出を繰り返しており, 表情表出の終了のタイミングが覚醒度の変化として表現されている. また, 11 箇所のフレームにおける原画像と覚醒度を比較

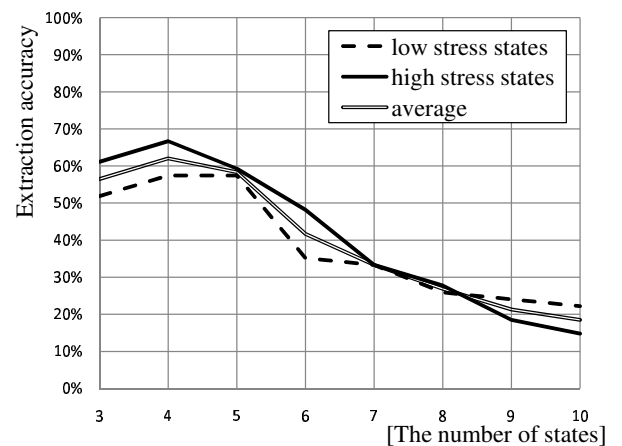


図 6 状態数の推移における抽出率.

Fig. 6 Extraction accuracy on transition of the number of states.

すると, 表情表出に対応づくように覚醒度が変化していることが分かる. 無表情フレームが覚醒度 1~4 の間で推移している部分は, 覚醒度における無表情領域が, 覚醒度 1 だけではなくいくつかの段階で広がり, 構成されていることを表現している. 覚醒度における無表情領域の広がり, 被験者が無表情状態でも無意識下で微細な表情筋の動きがあることにより生じていると考える.

6.2 HMMs の状態数の最適化

大塚らは, 表情筋の弛緩→収縮→静止→伸長→弛緩の 5 つの動作に対応づけて, HMMs の状態数を 5 に設定することで表情筋の動作をモデル化した [9]. 本研究では, ストレス刺激を与えた時の表情筋の動きを, 最適に表現する状態数を取得してから実験を行うために, 全対象者の表情画像データに対して, 快ビデオ視聴後をストレスが低い状態, 不快ビデオ視聴後をストレスが高い状態として, 各々の状態において HMMs の状態数を 3~10 の範囲で変化させて抽出率の比較を行った. その結果を図 6 に示す. 実験結果では, 状態数を 4 に設定した時の平均抽出率が最も高く, 状態数 4 をピークに, それ以降の状態数では平均抽出率が低下している. また, 状態数 5 や 7 の時に比べて, 状態数 4 の時ではストレスが低い状態と高い状態とで, 抽出率の差が顕著に見られたため, ストレス状態が表情表出に何

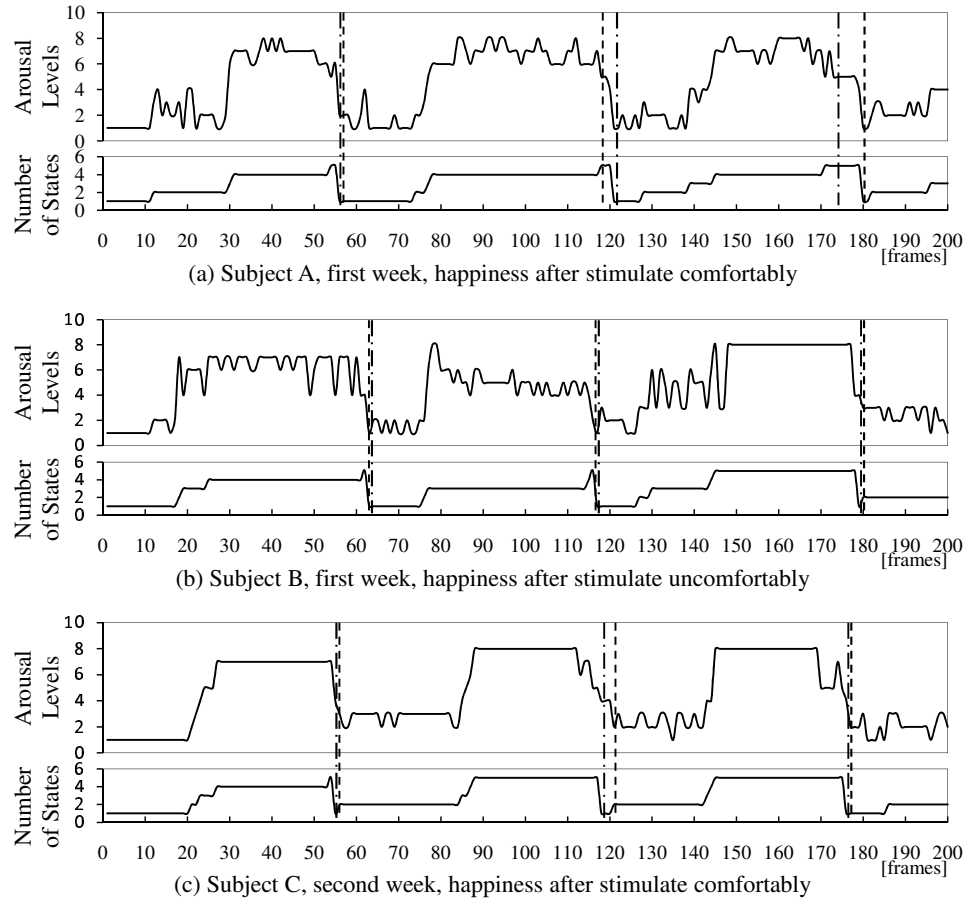


図 7 表情表出テンポの抽出結果 (被験者 A, B, C) .

Fig. 7 Extraction results of facial expressions tempos (Subject A, B, C).

表 1 表情表出テンポの抽出率.

Table 1 Extraction accuracy of facial expressions tempos.

| Subject | low stress states[%] | high stress states[%] |
|---------|----------------------|-----------------------|
| A | 88.9 | 77.7 |
| B | 55.6 | 100 |
| C | 77.8 | 66.6 |
| D | 66.7 | 33.3 |
| E | 33.3 | 66.6 |
| F | 22.2 | 55.6 |
| average | 57.4 | 66.7 |

らかの影響を及ぼしていることが、抽出率の差として表れていると考える。これらの2つの理由から本研究では状態数を4に設定した。

6.3 表情表出テンポの抽出結果

被験者 A, B, C の3例において、HMMs を適用して表情表出テンポを抽出した結果を図7に示す。図中上部の座標軸は覚醒度の時系列変化を表しており、図中下部の座標軸は状態数の推移を表している。また、実験者が原画像から表情表出が終了していると判断したフレーム位置を GT (Ground Truth) として破線で表す。HMMs により抽出された表情表出が終了していると判断されたフレーム位置は一点鎖線で表す。図7 (a) に示す被験者 A の抽出結果について、GT のフレーム数が (57, 120, 180) であるのに対して、HMMs により抽出された表情

表出終了時のフレーム数は (56, 118, 173) であるため、3テンポ目の抽出フレームが GT に対して7フレーム差であり、抽出に失敗している。次に、図7 (b) に示す被験者 B の抽出結果について、GT のフレーム数が (63, 116, 180) であるのに対して、HMMs により抽出された表情表出終了時のフレーム数は (62, 116, 180) であるため、実際の表情表出に対して、表情表出終了フレームは最大で±1フレームの範囲内で抽出されている。最後に、図7 (c) に示す被験者 C の抽出結果について、GT のフレーム数が (57, 122, 177) であるのに対して、HMMs により抽出された表情表出終了時のフレーム数は (55, 118, 176) であるため、実際の表情表出に対して、表情表出終了フレームは最大で±4フレームの範囲内で抽出されている。

次に、被験者 A~F の3週間の表情データについて、ストレスが低い状態、ストレスが高い状態時のそれぞれの表情表出テンポの抽出率を表1に示す。3週間の抽出率の平均を取ると、ストレスが低い状態での抽出率の平均が57.4%であるのに対し、ストレスが高い状態での抽出率の平均は66.7%である。よって、ストレスが高い状態での抽出率の方が9.3%高い。ストレスが高い状態での抽出率の方が高くなったのは、ストレスが高い状態になることで、ストレス反応が適度な緊張時の反応として表れ、表情表出と無表情を明瞭に表出し分けたことで、そのテンポを HMMs が抽出したためと考える。

表 2 HMMs 抽出フレームにおける覚醒度平均.

Table 2 Average of arousal levels on a frame extracted by HMMs.

| | low stress states | high stress states |
|---------------------------|-------------------|--------------------|
| average of arousal levels | 2.6 | 2.2 |

6.4 覚醒度における無表情領域の広がり

渡邊ら [19] は、無表情は視覚的情報空間では 1 点に表されるのではなく、広がりを持っていることを示唆している。我々は無表情領域の広がりがストレスの差異により影響を受けると考えた。そこで、被験者に対して快刺激を与えることで、ストレスが緩和されストレスが低い状態になり、不快刺激を与えることで、ストレスが増大しストレスが高い状態になるとして、被験者全員のデータについて刺激付与状態時において、HMMs により抽出された（表情表出から無表情に戻ったと判断された）表情表出終了フレームを調べ、そのフレーム時の覚醒度を取得し、快刺激付与状態時と不快刺激付与状態時とで、各々の覚醒度の平均を求めた。その結果を表 2 に示す。ストレスが低い状態での HMMs 抽出フレームにおける覚醒度平均は 2.6、ストレスが高い状態での HMMs 抽出フレームにおける覚醒度平均は 2.2 となった。このことから、ストレスが高い状態の方がストレスが低い状態に比べて、覚醒度における無表情領域が小さいと言え、ストレスが高い状態では無表情の多様性が乏しくなることが示唆される。ストレスが高い状態の方が無表情領域の広がりが小さくなったのは、ストレスが高い状態では表情筋の活動性が抑制され無表情での微細な表情筋の活動が少なくなったためと考える。

7. まとめ

本研究では表情表出テンポとリズムという枠組みを用いて、表情表出における覚醒度の時系列変化とストレスとの関係性について検証した。データセットとして被験者 6 名に対して、平常状態、快刺激を与えた状態、不快刺激を与えた状態の表情画像データセットを取得し、各被験者の表情表出テンポを抽出した。実験結果として、HMMs の状態数の最適値は 4 であり、快刺激付与状態時の平均抽出率は 57.4%、不快刺激付与状態時の平均抽出率は 66.7%であった。ストレス状態の差異による無表情の覚醒度の広がりの変化について、ストレスが低い状態よりもストレスが高い状態の方が無表情領域の広がりが小さくなった。

今後の課題として、ストレス測定のための POMS と唾液アミラーゼ試験について、被験者数が少なく、ストレスと測定値の傾向を見るためには不十分だったため、今後は被験者を増やし、ストレスと測定値に傾向が見られるかを実験する。また、本研究では、表情表出テンポの抽出と実験を重点的に行ったが、表情表出リズムという観点から、ストレスが、複数のテンポの各々のフレーム数と、それらの関連性にどのような影響を与えるのかを明らかにすることが課題である。更に、表情表出リズムの差異と、脳内の表情発現パスの関連性を明らかにすることも課題である。

Acknowledgment

実験データの取得に際し、被験者として長期に渡り顔画像の撮影に協力して頂きました本学の 6 名の学生諸氏に深く感謝申し上げます。

文 献

- [1] M. Pantic, L. J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," *IEEE Trans. PAMI*, Vol.22, No.12, pp.1424-1445, Dec. 2000.
- [2] 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その自動解析-", *信学誌*, Vol.85, No.9, pp.680-685, Sep 2002.
- [3] H. Madokoro, K. Sato, A. Kawasumi, and S. Kadowaki, "Facial Expression Spatial Charts for Representing of Dynamic Diversity of Facial Expressions," *IEEE Int'l Conf. System, Man, and Cybernetics, (SMC)*, Oct. 2009.
- [4] J.A. Russell and M. Bullock, "Multidimensional Scaling of Emotional Facial Expressions: Similarity From Preschoolers to Adults," *Journal of Personality and Social Psychology*, vol. 48, pp.1290-1298, 1985.
- [5] 延谷直哉, 仲谷善雄, "パーソナルテンポを基とした音響リズム支援による会話支援システム," *情報処理学会第 71 回全国大会*, pp.4-227~4-228, Mar. 2009.
- [6] 熊野宏昭, "ストレスの評価," http://hikumano.umin.ac.jp/Stress_Assess.pdf, pp.1-6.
- [7] 平山高嗣, 川嶋宏彰, 西山正統, 松山隆司 "表情譜: 顔パーツ間のタイミング構造に基づく表情の記述," *ヒューマンインタフェース学会論文誌*, pp.271-281, May 2007.
- [8] P. Ekman and W. V. Friesen, "Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues," *Malor Books*, 2003.
- [9] 大塚尚宏, 大谷敦, "連続した表情シーケンス画像からの HMM を用いた個別表情抽出に関する検討," *電子情報通信学会技術研究報告*, pp.17-24, Nov. 1997.
- [10] B. K. P. Horn and B. B. Schunck, "Determining optical flow," *Artificial Intelligence*, Vol.17, pp.185-203, 1981.
- [11] R.J.R. Blair, "Facial expressions, their communicatory functions and euro-cognitive substates," *Philos. Trans. R. Soc. Lond.*, B358, pp.561-572, 2003.
- [12] 独立行政法人理化学研究所監修, "理研探偵団: オンリーワンに挑む研究者の素顔に迫る," pp.266-277, 2006.
- [13] 武田晴登, 西本卓也, 嵯峨山茂樹, "HMM による MIDI 演奏の楽譜追跡と自動伴奏," *情報処理学会研究報告*, pp.109-116, Aug. 2006.
- [14] 電子情報通信学会編, "確率モデルによる音声認識," *コロナ社*, pp.29-66, 1988.
- [15] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *IEEE Face and Gesture*, pp.46-53, 2000.
- [16] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. "Measuring facial expressions by computer image analysis. *Psychophysiology*," 36:253-264, 1999.
- [17] M. Pantic, M.F. Valstar, R. Rademaker and L. Maat, "Web-based Database for Facial Expression Analysis," *Proc. IEEE Int'l Conf. Multimedia and Expo, Amsterdam, The Netherlands*, Jul. 2005.
- [18] 山口昌樹, 金森貴裕, 金丸正史, 水野康文, 吉田博, "唾液アミラーゼ活性はストレス推定の指標になり得るか," *医用電子と生体工学: 日本 ME 学会雑誌*, pp.46-51, Sep. 2001.
- [19] 渡邊伸行, 山田寛, 巖島行雄, "無表情と知覚される顔に関する基礎的研究: 線画表情を用いた表情の視覚的情報空間における無表情空間の同定," *電子情報通信学会技術研究報告*, pp.13-18, Jan. 2002.