

## 一過性のストレス刺激が意図的な表情表出に与える影響

佐藤 和人<sup>†</sup> 間所 洋和<sup>†</sup> 門脇 さくら<sup>‡</sup>

Kazuhiro Sato Hirokazu Madokoro Sakura Kadowaki

## 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。

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 [3] [4]。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を表す。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、表情と心理状態との関連性を明らかにするために、ストレス刺激後の意図的な表情を繰り返し表出する動作に着目した表出テンポとリズムという枠組みを提案する。表出テンポとリズムは、表情の表出程度 (真顔からの変位量を数値化したもの) をラベル化した表出強度 [5] の時系列変化において、無表情から表情表出を経て無表情に戻る区間を 1 テンポ、テンポを複数回繰り返したものを 1 リズムと位置付ける。本研究では、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、表出強度の生起パターンである表出テンポの抽出が可能であると考えられる。ス

トレス反応は身体面 (心拍数変化, 血圧変化), 心理面 (落ち込み, イライラ), 行動面 (酒量増加, じっとしてられない) に表れる [6]。ここで、表情はストレス反応の中でも行動面に分類される。そのためストレス状態の異なる、表出テンポとリズムを解析することで、将来的に、個人固有の表情変化から快・不快といった心理状態の推定が可能になると考える。本研究では、人間の心理的ストレス状態に起因する表情表出時の複雑性や曖昧性を客観的に表現するための基盤として、表出テンポとリズムという枠組みを提案する。

## 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、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、始めに時系列表情画像に対して 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、各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・非自発的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認められる。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いため、中間表情を表す表出強度の差異の解析においては十分な結果が得られないと考える。

大塚 [8] 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 [9] (「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 (Facial Action Coding System) の AU (Action Unit) 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。この手法では、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、

<sup>†</sup> 秋田県立大学, Akita Prefectural University<sup>‡</sup> SmartDesign 株式会社, SmartDesign Corp.

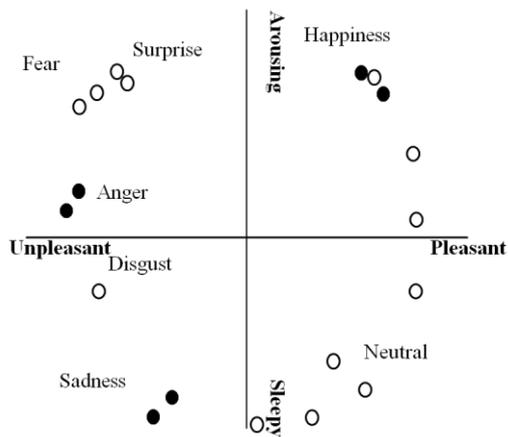


図 1 Russel の円環モデル

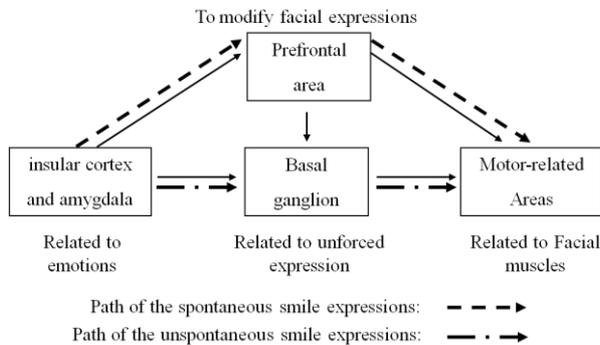


図 2 意図的な表情と自然な表情の発現パス

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抽出としている。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 表情であり、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、表情の表出と終了の正確な区間が抽出されているとは限らない。

### 3. 表出テンポとリズム

#### 3.1 表情の表出強度

我々[5]は、個人固有な表情空間を定量化するための指標として、表出強度という概念を提案した。表出強度とは、図 1 に示す Russell の円環モデル[11]において、縦軸の覚醒次元と横軸の快次元の両方の特徴を併せ持つ情報である。具体的には、目や口、眉などの顔パーツの動きに伴う変化を位相変化として抽出する。ここで、位相とは、集合の各要素のつながり方を規定している構造である[4]。本研究では、抽出された位相変化を特徴としてカテゴリ化し、表出度合いに応じて並び替えることによって表出強度が得られる。すなわち、本研究における表出強度は、覚醒次元と快次元の両者の特徴を含む情報となる。Russell の円環モデルでは、全ての感情は快次元(快・不快)と覚醒次元(覚醒・眠気)表される 2 次

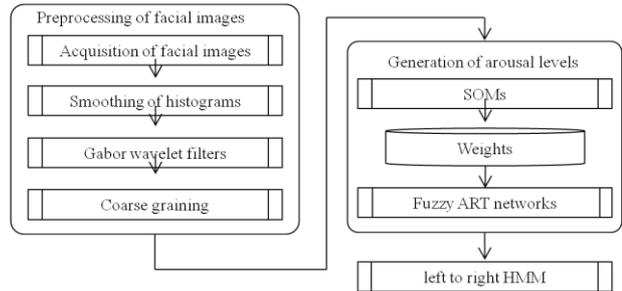


図 3 提案手法の全体構成

元平面上に円環状に布置される。したがって、本研究で対象とする意図的な表情の表出では、快の次元の影響を表情画像に対して直接的に扱うことが難しいと考えられるため、一過性の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て評価する。

#### 3.2 表出テンポとリズムの定義

Blair [12]は、表情表出には(1)感情を生み出す部位(扁桃体、島質など)、(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に関連があることを指摘している。山口[13]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、自然な表情と意図的な表情を修飾する場合では、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、表出強度の時系列変化を利用して、人間が創り出す表情表出テンポとリズムの可視化を目指す。ここで、表出テンポとリズムとは、表出強度の時系列変化において、無表情から表情表出を経て無表情に戻る区間を 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。人間が意図的に表出する表情は、顔の静的多様性と動的多様性に基づき人それぞれ固有な空間を形成しており、表情のダイナミクスは「表情筋が創り出す顔パターンの時系列な位相変化」として捉えることができる。すなわち、顔パターンの位相変化を表出強度で組織化・可視化し、表情表出テンポとリズムという枠組みを用いて、人間の心理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ことを試みる。

### 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、被験者が意図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

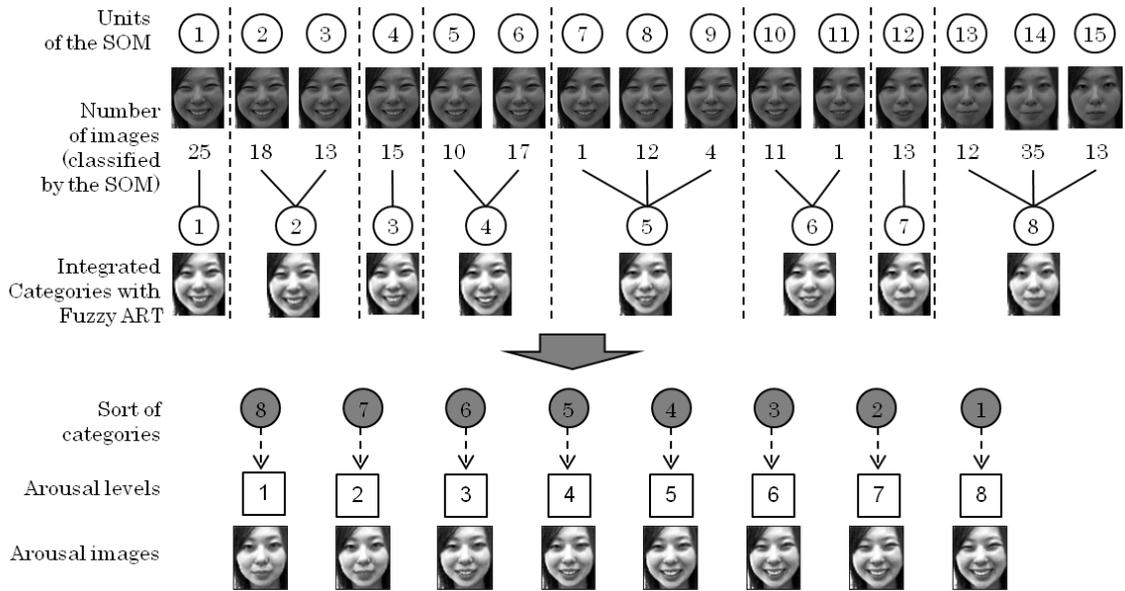


図 4 表情の表出強度の取得手順

(Self Organizing Maps) を用いて表情パターンの分類を行う。さらに、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、表出強度の時系列変化の抽出、HMMsによる表出テンポの抽出について個別に説明する。

#### 4.1 表出強度の時系列変化の取得

前処理として、時系列表情画像に対して輝度値の正規化を行い、照明条件などによる濃淡値の影響を軽減する。また、ヒストグラムの平滑化によって、画像の明瞭化とコントラストの調整を行う。さらに、特徴表現法として Gabor Wavelets 特徴の方位選択性により、目、眉、口、鼻といった表情のダイナミクスを特徴づける顔部分を強調する。Gabor Wavelets 変換した時系列表情画像に粗視化処理を行うことで、情報量の圧縮と顔画像を撮影する際に発生する微少な位置ずれの影響を緩和する。

次に、表出強度の時系列変化を取得する手順の詳細を図 4 に示す。始めに、粗視化処理を施した時系列表情画像の輝度値情報を SOMs により学習し、表情の位相変化が類似する顔画像ごとに 15 個の写像ユニットに分類する。次に、これら 15 個の写像ユニットの中でも類似したユニットを Fuzzy ART で同一のカテゴリに統合する。さらに、Fuzzy ART によって統合されたカテゴリを無表情(真顔)から最大表出まで並び替えることで、表情の表出程度を定量的にラベル化した表出強度を取得する。最後に、時系列表情画像の各フレームと表出強度の対応付けを行ない、表出強度の時系列変化を生成する。

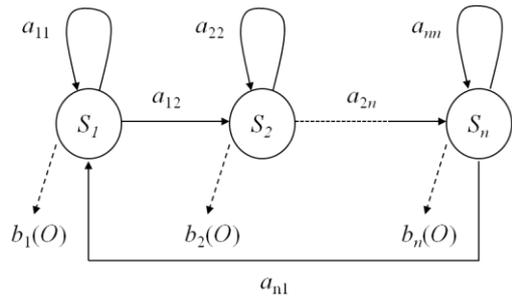


図 5 Left to Right 型 HMMs の構成

#### 4.2 HMMs による表出テンポの抽出

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[14]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。本研究で用いるデータセットは、時系列表情画像を対象としており、表出テンポは表出強度の生起パターンから成る。そのため、我々は表出テンポを抽出するために HMMs を用いる。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 は抽出対象毎に用意し、各 HMMs を各抽出対象の学習用シンボル系列で学習させる。HMMs における学習とは、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 [15]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。本研究で用いる HMMs の構成は、図 5 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。

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

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

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

シンボル出力確率  $Bi(O)$ : 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。

## 5. データセット

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。実験プロトコルの詳細を図 6 に示す。1 回の実験は、平常時の step1、快ビデオ視聴時の step2 および不快ビデオ視聴時の step3 で構成される。図 6 に示すように、快・不快状態を引起す情動喚起ビデオを視聴するタスクを被験者に与え、一過性の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。また、情動喚起ビデオは、視聴時間が約 3 分の快ビデオ (3 種類の漫才映像) と不快ビデオ (2 種類の残酷映像とインプラント手術映像) を用意し、視聴した際の主観評価 (5 段階評価) も実施した。なお、全ての被験者に対して、研究倫理規定に基づき事前に実験内容を十分説明し、被験者の自由意志により書面により実験参加の同意を得た。さらに、特定の被験者からは、実験参加の同意と併せて顔画像掲載の許諾に関する同意も得ている。

### 5.1 表情画像データ

表情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 [9] が有名である。動画画像では、Cohn Kanade [16] や Ekman Hager [17] 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 [18] 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。

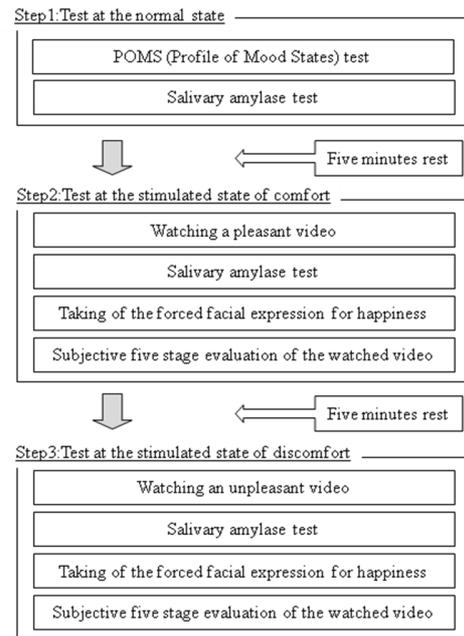


図 6 実験プロトコル

Ekman [9] らが提案した基本 6 表情は、「喜び」、「怒り」、「悲しみ」、「嫌悪」、「恐れ」、「驚き」である。本研究では、基本 6 表情の中で、自発的に最も表出しやすいと考えられる「喜び」の表情に注目し、快状態時の「喜び」、不快刺激状態時の「喜び」を対象として、20 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 10 名 (J=20 歳, B,G,H,I=21 歳, A,E,F=22 歳, C,D=23 歳)、女子大学生 10 名 (K,M,O,P=20 歳, L,Q,R,S,T=21 歳, N=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、微少な変動に対しては、テンプレートを画像上で移動させながら表情画像と比較し、次いで画像間の差分情報を利用することによる、テンプレートマッチングによって補正した。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 20 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 10 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。

### 5.2 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

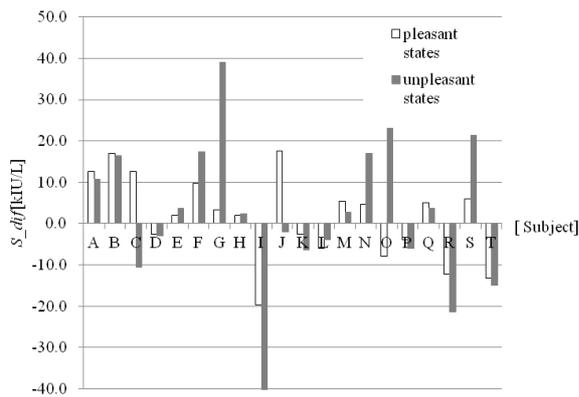


図 7 唾液アミラーゼ活性

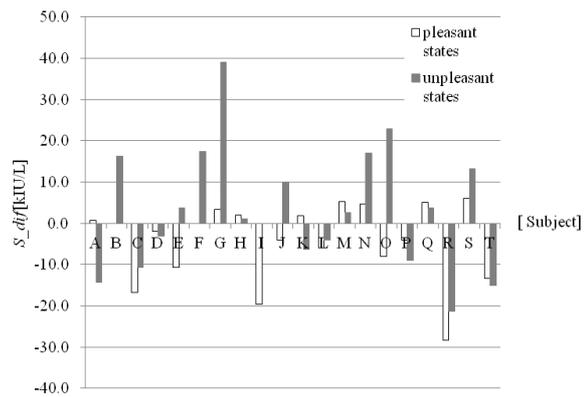


図 8 選択された唾液アミラーゼ活性

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が与えられたときに生じるストレスである。本研究では被験者に対して一過性ストレス刺激を与えるため、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。一過性のストレス反応測定方法の 1 つ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[19]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。本研究では、一過性の外部刺激として情動喚起ビデオを取り上げ、視聴直後における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、快・不快刺激として与える情動喚起ビデオの有効性を検証する。次に、表情の表出テンポ抽出時における HMMs の状態数の最適化を行う。続いて、最適化された状態数のモデルを用いて、表出強度の時系列変化から得られる表出テンポの抽出精度を検証する。最後に、快・不快のストレス刺激が表情の表出リズムに与える影響を分析する。

### 6.1 快・不快刺激の有効性

唾液アミラーゼ試験により、快・不快刺激として用いたビデオ視聴時における情動喚起要因の有効性を検証した。唾液アミラーゼ活性は快適な状態ならば下降、不快な状態ならば上昇することが示されている[19]。よって、平常時の唾液アミラーゼ活性値を  $S_{normal}$ 、ビデオ視聴後の唾液アミラーゼ活性値を  $S_{stimu}$  とすれば、平常時とビデオ視聴後の唾液アミラーゼ活性値の差  $S_{dif}$  は、次式で定義される。

$$S_{dif} = S_{stimu} - S_{normal} \quad (4)$$

快ビデオ視聴後は

$$S_{dif} < 0 \quad (5)$$

不快ビデオ視聴後は

$$S_{dif} > 0 \quad (6)$$

となれば、ビデオは有効であると評価できる。

被験者 A~T の 20 名において求めた  $S_{dif}$  の結果を図 7 に示す。ここで、被験者 B および C の唾液アミラーゼ活性の結果が先行研究[19]の結果と異なる原因は、被験者ごとに快・不快ビデオに対する感じ方が違うためと考える。そのため、被験者の主観評価が高いデータのみを対象として、唾液アミラーゼ活性を算出することとした。

主観評価の得点方法は、快ビデオ視聴時は「1:全く快適でない~5:非常に快適である」、不快ビデオ視聴時は「1:全く不快でない~5:非常に不快である」の 5 段階評価とした。主観評価から、ビデオが快・不快刺激として有効に働いていると考えられる、評価 4 および 5 のみに着目した場合の唾液アミラーゼ活性を図 8 に示す。この結果から、快状態時の平均  $S_{dif}$  は  $-2[\text{kIU/l}]$ 、不快状態時の平均  $S_{dif}$  は  $5[\text{kIU/l}]$  となり、快・不快刺激として情動喚起ビデオが有効に働いたことが確認された。

### 6.2 HMMs のパラメータ検出

大塚[8]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[8]。それに対して本研究では、HMMs の状態数と初期状態遷移確率を変化させ、式(7)に示す抽出率が最も高くなる最適値を求めてから実験を行う。したがって、一過性の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パラメータ (HMMs の状態数と初期状態遷移確率) 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、評価者 3 名が表情を撮影した動画から無表情に戻っていると判断したフレームの平均値を GT (Ground Truth) として設定し、評価者間のばらつきを考慮して GT フレームに対して  $\pm 5$  フレーム内の範囲を抽出成功範囲とした。この正誤判定における抽出率を式(7)に示す。

$$x_1, x_2, x_3 = \begin{cases} 1, E \in R \pm 5 \\ 0, -(E \in R \pm 5) \end{cases}$$

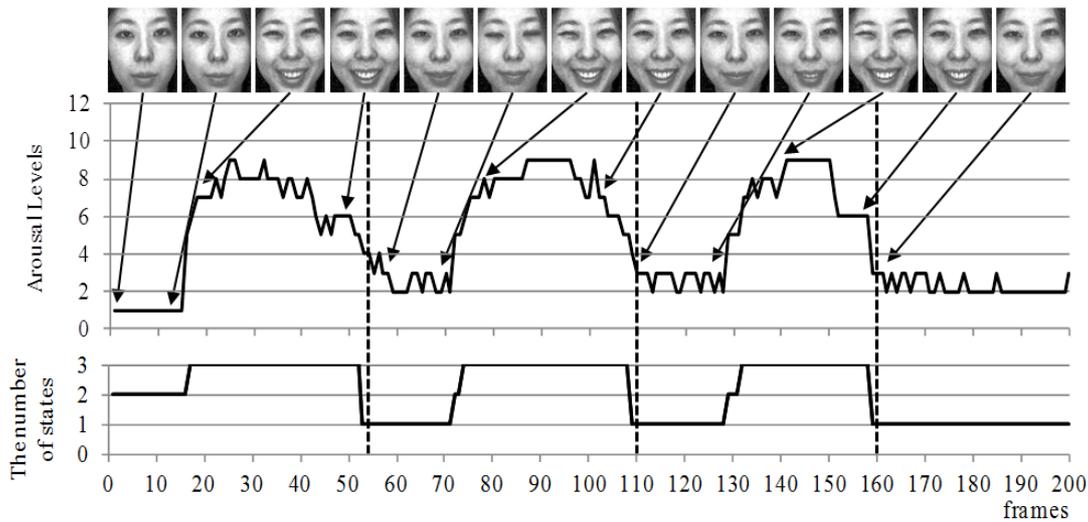


図 10 表情表出に伴う表情の表出強度の時間変化 (被験者 J, 1 週目, 快刺激の喜び)

$$A = \frac{x_1, x_2, x_3}{C} \times 100[\%] \quad (7)$$

なお、式(7)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 GT として求めた表情表出の終了を示すフレームである。

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必要がある。快・不快状態において、HMMs の状態数と自己遷移確率  $a$ 、次の状態に遷移する確率  $b$  を変化させて、抽出率の比較を行った。その結果を図 9 に示す。

実験結果では、状態数を 3 に設定した時の平均抽出率が最大であり、状態数 3 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、自己遷移確率  $a$  は 0.70、次の状態に遷移する確率  $b$  は 0.30 の時、平均抽出率が最大となった。以上の検討結果から、本研究では HMMs の状態数を 3、自己遷移確率  $a$  を 0.70、次の状態に遷移する確率  $b$  を 0.30 に設定した。

### 6.3 表情表出区間と表出強度の対応関係

表出強度の抽出結果の一例として、被験者 J (20 才, 女性) の 1 週目における、「喜び」表出強度の時系列変化を図 10 に示す。縦軸は表出強度、横軸は時系列表情画像のフレーム数であり、各画像セットとも 200 フレームで構成されている。事前に表出強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 (53, 110, 160) 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 (20 秒間) で 3 回の表情表出を繰り返しており、表情表出の終了を示すタイミングが表出強度の変化として表現されている。また、無表情を基準に表情表出を経て無表情に戻る時系列変化の様子を、13 箇所のフレーム NO の原画像と表出強度を比較すると、表情表出に対応付くように表出強度が変化していることが分かる。

### 6.4 表出テンポの抽出結果

HMMs を適用して表出テンポを抽出した結果として、被験者 A, C, J, K, Q, S の 6 例の表出テンポを図 11 に

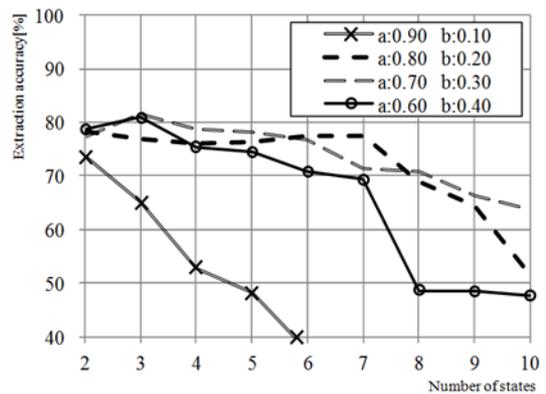


図 9 状態数の推移における抽出率

示す。各々の図中上部の座標軸は表出強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、評価者 3 名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の平均値を GT として破線で示す。また、評価者間のばらつきを考慮して、GT フレームに対して  $\pm 5$  フレーム内の範囲を抽出成功範囲として灰色の網掛けで示す。

被験者 A, C, K, Q, S については、HMMs による抽出フレームが抽出成功範囲内に全て含まれているため、抽出率 100%となっている。被験者 J については、GT のフレーム NO が (40, 71, 98) 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 NO は (60, 76, 88) であるため、実際の表情表出に対して、2 テンポ目のみが抽出に成功している。図中上部の表出強度の時系列変化を見ると、表情表出と対応していないことが分かる。その主な原因として、原画像を評価者が目視で確認しても、表情表出の程度が非常に小さいため、表情表出区間の分割が困難であることが挙げられる。本研究は、見え方ベースで表情データを扱うため、人間の目視による表情表出区間の同定が難しい場合、表出テンポの自動抽出は非常に困難になることが考えられる。そのため、

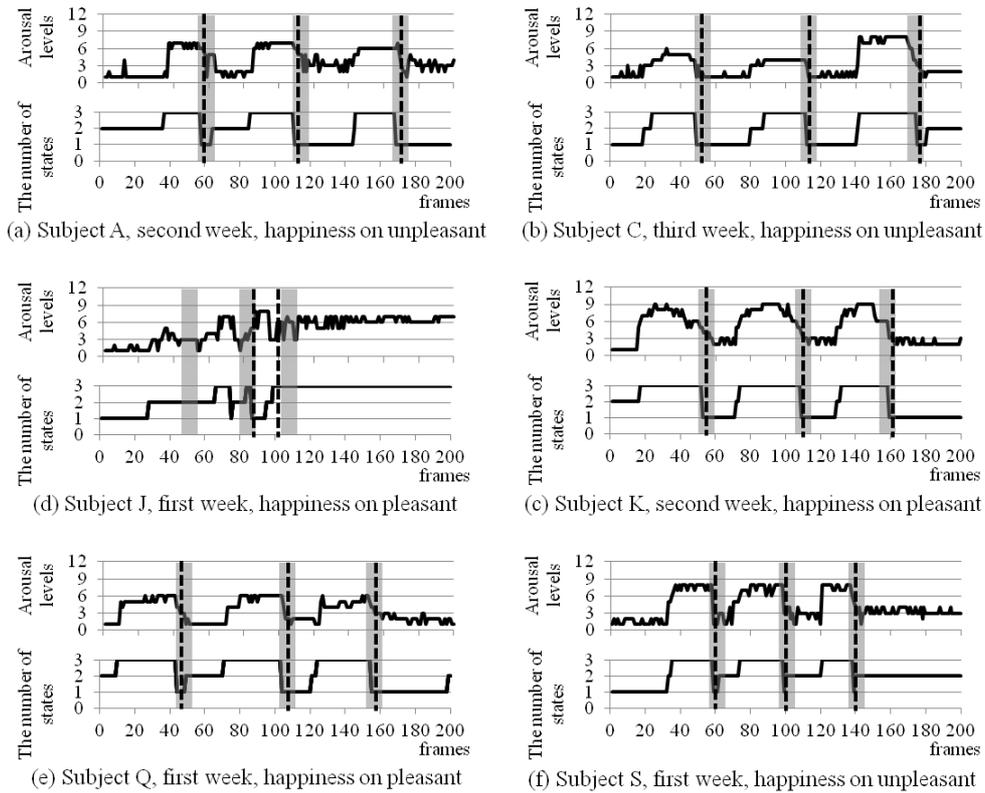


図 11 表出テンポの抽出結果

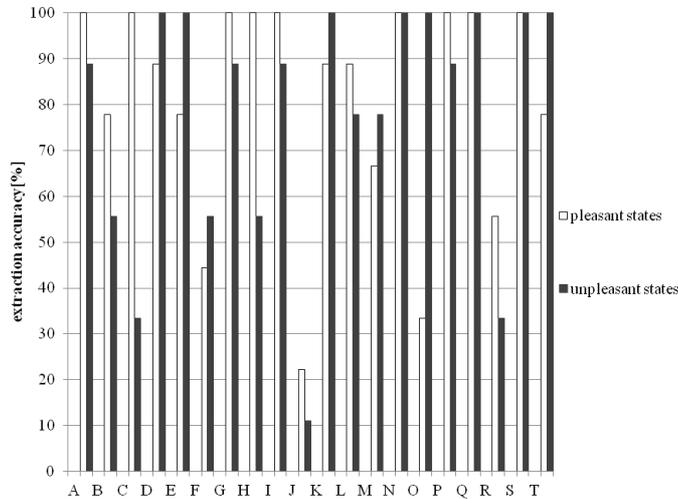


図 12 表出テンポの抽出率

表情データを取得する際に、被験者が表出し得る最大強度の表情を表出するよう指示を徹底する必要がある。

次に、被験者 A~T (20 例) の 3 週間の表情データについて、各被験者における表出テンポの抽出率を図 12 に示す。3 週間の抽出率の平均を取ると、快状態時における抽出率の平均が 81.1%、不快状態時で 77.8%である。図 12 (d) のように、目視による表情表出区間の同定が困難な事例を含めても、全体で 79.5%の抽出率が得られた。

### 6.5 快・不快状態が表出リズムに及ぼす影響

被験者 G の快ビデオ視聴後における「喜び」表出強度の時系列変化と表出テンポの抽出結果を図 13 に示す。抽出された 1 つ目のテンポは 60 フレーム、2 つ目のテンポは 57 フレーム、3 つ目のテンポは 36 フレームで構成されている。このように、リズムを構成する 3 テンポにばらつきが存在する。そこで、快状態時と不快状態時において、被験者 A~T までのテンポを構成する平均フレーム

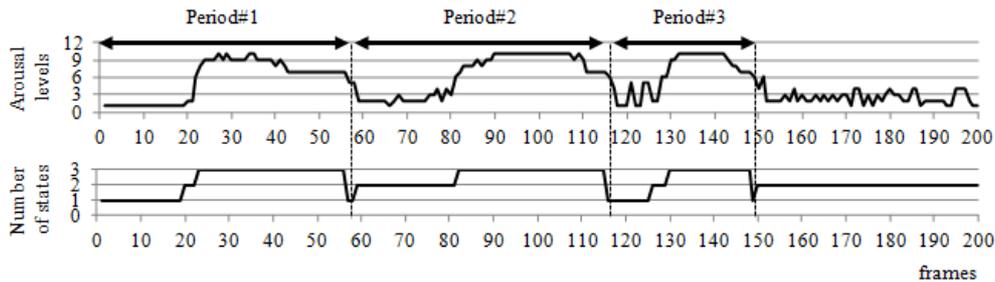


図 13 表情の表出強度の時系列変化と表出テンポの抽出 (Subject G)

表 1 1 テンポを構成する平均フレーム数及び標準偏差

	pleasant states	unpleasant states
average of frames	49.1	49.2
standard deviation	6.1	8.4

数と、標準偏差を求めることで、快・不快状態と表出リズムの関連性を考察する。

被験者全体の表出テンポを構成する平均フレーム数と 3 週間分のテンポの標準偏差を表 1 に示す。1 テンポを構成する平均フレーム数は、快状態時が 49.1 フレーム、不快状態時が 49.2 フレームであり、快・不快状態は 1 テンポを構成する平均フレーム数に影響を与えないと考える。また、1 テンポを構成するフレーム数の標準偏差は、快状態時が 6.1 フレーム、不快状態時が 8.4 フレームであり、不快状態時の方が、1 テンポを構成するフレーム数がばらつく傾向が見られた。すなわち、被験者全体の傾向として、不快ビデオ視聴による一過性のストレス刺激によって、リズムの構成要素である表出テンポにゆらぎが発生することを定量的に示した。以上の結果から、人間の心理状態を推定する指標の一つとして、意図的に表出する「喜び」の表情の繰返し動作に着目した表出テンポとリズムの解析が有効であると考えられる。

## 7. むすび

本研究では、表情の表出テンポとリズムという枠組みを用いて、意図的な表情表出における表出強度の時系列変化と快・不快といった人間の心理状態との関係性について検証した。被験者 20 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、各被験者の表出テンポを抽出した。実験結果として、快刺激付与状態時の平均抽出率は 81.1%、不快刺激付与状態時の平均抽出率は 77.8%であった。快・不快状態が表出リズムに与える影響について、快状態時よりも不快状態時の方が、1 テンポを構成するフレーム数のばらつきが大きくなる傾向が得られた。以上から、人間の心理状態を推定する指標の一つとして、表出テンポとリズムを用いた解析が有効である可能性が示された。

今後は、心理的ストレス反応測定尺度の 1 つである POMS を用いて、被験者の慢性的ストレス状態を含めた解析が必要である。また、表出テンポとリズムのゆらぎを定量化することにより、自然な表情と意図的な表情の発現パスの違いを明らかにする予定である。

## 参考文献

- [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] 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.
- [4] 大石周平, 尾田政臣, "話者間の精神テンポの差がコミュニケーションの円滑化に及ぼす影響," 電子情報通信学会技術研究報告, pp.31-36, 2005.
- [5] 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.
- [6] 熊野 昭, "ストレスの評価," [http : //hikumano.umin.ac.jp/StressAssess.pdf](http://hikumano.umin.ac.jp/StressAssess.pdf), pp.1-6.
- [7] 平山高嗣, 川嶋宏彰, 西山正紘, 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] 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.
- [9] P. Ekman and W. V. Friesen, "Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues," Malor Books, 2003.
- [10] B. K. P. Horn and B. B. Schunck, "Determining optical flow," Artificial Intelligence, Vol.17, pp.185-203, 1981.
- [11] 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.
- [12] R.J.R. Blair, "Facial expressions, their communicatory functions and euro-cognitive substates," Philos. Trans. R.Soc. Lond., B358, pp.561-572, 2003.
- [13] 山口陽子, "リズムが刻む脳のコンテキスト情報," 電学論 C, vol.128, no.8, pp.1068-1071, Aug.2000.
- [14] 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.
- [15] 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.
- [16] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," IEEE Face and Gesture, pp.46-53, 2000.
- [17] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Measuring facial expressions by computer image analysis. Psychophysiology," 36:253-264, 1999.
- [18] 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 Nether-lands, Jul. 2005.
- [19] 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.