表情譜: 顔パーツ間のタイミング構造の記述とその自動獲得

西山正紘[†]　川嶋宏彰[†]　松山隆司[†]

概要 ヒューマンインタフェースへの応用の観点から，表情認識・生成をはじめとする顔表情に対する研究が活発に行われている．従来の表情認識の研究では，主に表情を感情に基づく基本的なカテゴリー (喜び・驚き・恐怖・怒り・嫌悪・悲しみ) に分類することにより検討されている．しかし，実際に人がコミュニケーションを行う際には，より粒度の細かな分類を行っていると考えられる．本研究では，顔を独立した動きが観測される領域 (顔パーツ) に分け，表情変化を各顔パーツの動的な運動の組み合わせによって生じるものと考える．そのために，まず顔パーツの変化を静止状態もしくは収束性の運動を行う時間範囲 (区間) を単位として表す．そして，顔パーツ間の運動のタイミング，すなわちこれら区間の間の時間関係を用いることで，表情をより詳細に理解する枠組を提案する．本論文では，この顔パーツ間の運動のタイミングを記述する表現形式を「表情譜」と呼び，顔画像系列を入力として表情譜を自動獲得する手法について述べる．表情譜の表情理解における有効性を，意図的・自発的な笑顔を対象として評価した．

Facial Expression Recognition based on Timing Structures in Faces

MASAHIRO NISHIYAMA ,[†] HIROAKI KAWASHIMA [†]
and TAKASHI MATSUYAMA[†]

Abstract This paper presents a method for interpreting facial expressions based on temporal structures among partial movements in facial image sequences. To extract the structures, we propose a novel facial expression representation, which we call a facial score, that is similar to a musical score. The facial score enables us to describe facial expressions as spatio-temporal combinations of temporal intervals; each interval represents a simple motion pattern with the beginning and ending times of the motion. Therefore, we can classify fine-grained expressions from multivariate distributions of temporal differences between the intervals in the score. In this paper, we provide a method to obtain the score automatically from input images using bottom-up clustering of dynamics. Our experiment shows the effectiveness of the method by separating smiling expressions into intentional and spontaneous categories using the obtained scores.

1. はじめに

1.1 研究背景

人と人とのコミュニケーションにおいて，顔における表情は，非言語情報を効果的に伝達することのできるメディアとして重要な役割を果たしている．例えば，私たちは表情を通じて自分の心理状態を伝達することができ，その一方で，表情から相手の心理状態を読み取ることができる．

ヒューマンインタフェースへの応用を目的として，表情認識・生成システムの研究が行われているが，これらの研究では，表情の記述形式として，Ekman らが開発した FACS（Facial Action Coding System）における AU（Action Unit）を利用したものが主である[4]．AU は，解剖学的に独立し，視覚的に識別可能な表情動作の最小単位として設定されており，FACS とは，これら AU の組み合わせで表情を記述する手法である．しかし，FACS には，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いう限界が存在する[3]．さらに，AU は，多くの表情を人間が観察し，主観的に分類したものであるので，それによって表現しきれないような表情動作も実際には存在するのではないかと考察される．

表情生成の入力，もしくは表情認識の出力として，どのような表情の分類を設定するかという問題も重要である．従来の研究では，主に表情を静止画をもとにして基本的なカテゴリー（喜び・驚き・恐怖・怒り・嫌悪・悲しみ・軽蔑）に分類することにより検討されている[3]．しかし，実際の表情は，意図的に作られたものもあれば自発的に表出されたものもあり，同じカテゴリー内でもさまざまな種類に分類できるほど多様かつ微妙なものである．人は，コミュニケーションを行う際に，より粒度の細かな分類を行っていると考え

† 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

られるが，その分類は表情の静的な要因のみからは困難であり，刻々と変化する相手の表情の微妙な動きを観察することにより行っているものと思われる．しかし，表情変化の時間的な要因の検討は，その重要性が指摘されながらも主として技術的な困難から多くはなされていない．

1.2 提案手法

以上より，従来の表情認識・生成システムの研究には以下の問題点がある．

- 表情の動的側面を十分に用いていない
- 表情の分類が感情に基づく基本カテゴリーに留まっている

そこで，本論文では，表情変化を顔の構成要素（顔パーツ）それぞれの時間的な運動によって生じるものと考える．そして，タイミング構造から得られる情報を利用して表情をより詳細に理解する枠組を提案する．ここで，タイミング構造とは，ある 2 つの区間がどのような時間関係で発生し終了するのかといった構造を表すものと定義する．また，区間とは，静止状態や収束性の運動のような単純な変化を行う事象の時間範囲を表すものとし，開始時刻（始点），終了時刻（終点），及び運動パターン（モード）のラベルを属性として持つものとする．本論文では，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を，音符と音符のタイミングの芸術である音楽を記述する楽譜に準えて「表情譜」と呼ぶ．

このような区間を単位とした記述では，区間のモードをどのように定義するかが重要である．AU では，それによって表現しきれないような表情動作が存在するのではないかと考察されるため，AU を表情動作の基本単位として受け入れることには問題がある．そこで，顔表情の特徴を表す特徴ベクトル系列からボトムアップにモードを求めていく手法を取る．これによって，AU では表現しきれない表情動作も表現可能となる．

以上をまとめると，本論文で提案する表情譜は以下の特徴を持つ．

- 区間を単位とした表情のタイミング構造の記述が可能
- 区間のモードとして学習データからボトムアップに抽出された運動パターンの利用が可能

表情から表情譜を獲得し，表情譜で記述されるタイミング構造から表情を理解する流れを以下に示す（図 1 参照）．

(1) 表情から顔の特徴を表す特徴ベクトル系列を抽出する
(2) 特徴ベクトル系列を用いて顔の運動をモードに分節化し，表情譜を獲得する
(3) 表情譜で記述されるタイミング構造から有用な情報を抽出し，表情を理解する

以上の処理を自動化することができれば，表情認識システム等に応用でき，コンピュータがより詳細に人間の表情を読み取ることができるようになると考える．

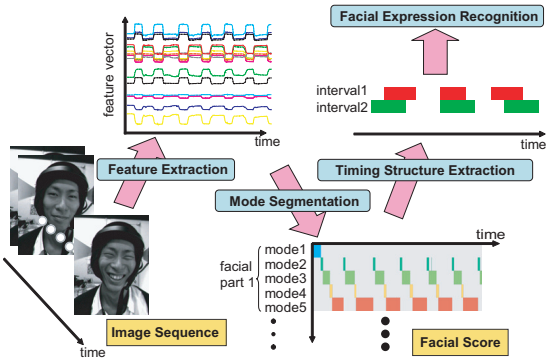

図 1 表情譜を用いた表情理解の流れ．表情から顔の特徴を表す特徴ベクトル系列を抽出し，それらを用いて顔の運動をモードに分節化し，表情譜を獲得する．そして，表情譜で記述されるタイミング構造から有用な情報を抽出し，表情を理解する．

ゆえに，本研究の目的は，

(1) 表情譜の自動獲得手法を提案し，
(2) 表情譜が表情理解においてどの程度有効であるかを評価する

ことである．評価実験で対象とする表情は，人間同士のコミュニケーションにとっては重要な分類であるが従来は同じ分類とされることが多かった，意図的な笑いと自発的な笑いの表情とし，これら両者のタイミング構造の比較を行う．

次章で関連研究について述べ，3 章では，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，表情譜を導入する．4 章では，入力として顔画像系列を与えた時に，出力として表情譜を自動獲得する手法について述べる．5 章では，実際に撮影した顔画像系列から表情譜を自動獲得し，笑顔を対象として表情譜の有効性の評価を行う．最後に，6 章では本論文の結論を述べる．

2. 関連研究

心理実験として，表情映像を被験者に見せて評価することにより，以下のような時間的要因に関する知見が得られている．Bassili は，顔に黒化粧を塗り，その上に白い点を特徴点として塗った表情映像を撮影することにより，顔の特徴点の運動のみによってある程度の表情の分類が可能であることを示した[2]．しかし，運動のどのような成分が分類に影響するのかは明らかにしていない．運動の成分をより直接的に扱った研究として，小山らは，コンピュータを用いて目と口の動きの時間関係を制御した表情映像を作成し，その時間的差異に基いて笑いを快の笑い・不快の笑い・社交の笑いに分類できることを示した[5]．一方で，時間的な要因としては動作の開始時刻の差のみを扱っており，動作の終了時刻の差や継続時間，動作の滑らかさ，変化

の線形・非線形性などの要因も今後検討すべきであると考察している．

このように，心理学的にも顔の時間的な運動が表情理解に重要な役割を果たしていることが示されており，特に顔の器官別にその運動を考察することが有効であると示唆される．それを踏まえて次章では，本論文ではどの顔部位を扱うか，どのような運動を扱うか，どのような時間関係を扱うかについて述べる．

3. 表情譜の設計

3.1 表情譜の定義

表情譜とは，顔の各構成要素がどのようなパターンで，どのような時間関係で運動するかを記述する表現形式である．ここで，以下の用語を定義する．

顔パーツと顔パーツ集合: 顔パーツとは，空間的に分離可能な顔の構成要素のことを表す．表情譜で記述する顔パーツの個数を N_p とした時，顔パーツ集合を $\mathcal{P} = \left\{P_1, ..., P_{N_p}\right\}$ で定義する．例えば，顔パーツ集合の要素としては，口，右目，左目，右眉，左眉等が考えられる．

モードとモード集合: モードとは，静止状態や収束性の運動のような単純な変化を行う事象のことを表す．顔パーツ $P_a\,(a \in \{1, ..., N_p\})$ におけるモードの個数を N_{m_a} とした時，顔パーツ P_a におけるモード集合を $\mathcal{M}^{(a)} = \left\{M_1^{(a)}, ..., M_{N_{m_a}}^{(a)}\right\}$ で定義する．例えば，口パーツにおけるモード集合の要素としては，開く，開いたまま，閉じる，閉じたまま等が考えられる．

区間と区間集合: 区間とは，静止状態や収束性の運動のような単純な変化を行う事象の時間範囲を表す．顔パーツ P_a における時系列データが T 個あり，その時系列データが N_{k_a} 個の区間で表されるとした時，顔パーツ P_a における区間集合を $\mathcal{I}^{(a)} = \left\{I_1^{(a)}, ..., I_{N_{k_a}}^{(a)}\right\}$ で定義する．また，区間 $I_k^{(a)}\,(k \in \{1, ..., N_{k_a}\})$ は始点 $b_k^{(a)} \in \{1, ..., T\}$，終点 $e_k^{(a)} \in \{1, ..., T\}$，及びその区間を表現するモードのラベル $m_k^{(a)} \in \mathcal{M}^{(a)}$ を属性として持つ．

表情譜: 表情譜とは，全ての顔パーツにおける区間集合の集合である．つまり，表情譜を $\{\mathcal{I}^{(1)}, ..., \mathcal{I}^{(N_p)}\}$ で定義する．表情譜の概念図を，図 2 に示す．図の縦軸は顔パーツとそのモードを表す軸，横軸は時間軸である．そして，各顔パーツ毎にその運動状態の遷移を，モードを単位として時間軸に沿って記述する．図では，各パーツ毎に異なるモードを縦軸に沿って表示している．よって，表情譜を用いることにより顔パーツ間のタイミング構造の記述が可能となる．

3.2 表情譜における顔パーツ

タイミング構造から得られる情報を利用して表情を理解，表現するという観点から考えると，動きのタイミングの差異が表れる領域同士を別の顔パーツとして扱うべきである．従って，動きに常に共起性のある顔の領域同士は，1 つの顔パーツとして扱ってよいと言える．つまり，appearance-base で独立した動きが観測される領域を顔パーツとして考える．

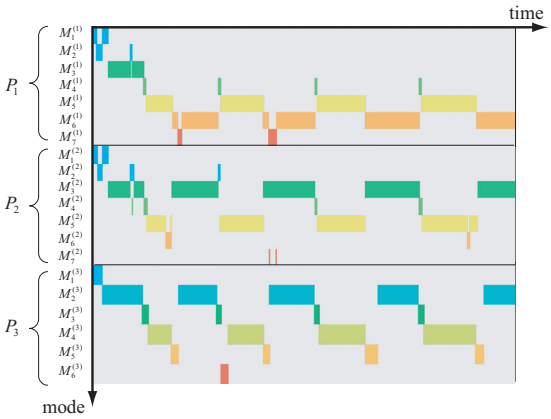

図 2 表情譜．縦軸は顔パーツとそのモードを表す軸，横軸は時間軸である．各顔パーツ毎にその運動状態の遷移を，モードを単位として時間軸に沿って記述する．

Ekman らは，基本的感情（驚き・恐怖・嫌悪・怒り・幸福・悲しみ）が表情に現れる様子の違いを，appearance-base で独立した動きが観測される顔の 3 領域（眉の周辺部分・目の周辺部分・口の周辺部分）の組み合わせにより解明した[6]．本論文では，この 3 領域に着目するとともに，眉と目の周辺部分に関しては左右の各顔パーツをそれぞれ別のパーツとして扱う．これは，実際の表情において眉や目の運動が非対称に起こる表情が観察されるので，別々に扱うことでより微妙な表情の変化を記述することが可能であると考察されるためである．

着目する 5 領域に関して，どのような特徴量（特徴ベクトル）を設定すれば微妙な表情を表現できるかという問題も重要である．本論文では，動きの情報を直接に扱える特徴点の座標値を特徴ベクトルとして使用する（図 5(a) 参照）．皺は微妙な表情を理解するために有効な情報を提供すると考えられるが，特徴点を利用すればこれらを間接的に表現することが可能である．例えば，笑う時にできる頬皺は鼻の特徴点の動きに着目すれば間接的に表現することができる．

ゆえに，顔パーツ集合 $\mathcal{P}$ の要素は，右眉，左眉，右目，左目，鼻，口とする．また，顔パーツ P_a の特徴ベクトル $z^{(a)}$ は，顔パーツ P_a における特徴点の数を n_{p_a}，p 番目 $(p \in \{1, ..., n_{p_a}\})$ の特徴点における座標値を $\left(x_p^{(a)},\, y_p^{(a)}\right)$ とすると，

$$z^{(a)} = \left(x_1^{(a)},\, y_1^{(a)}, ..., x_{n_{pa}}^{(a)},\, y_{n_{pa}}^{(a)}\right)^{\mathsf{T}} \qquad (1)$$

という $2n_{p_a}$ 次元列ベクトルとして表せる．

3.3 表情譜におけるモード

顔パーツの運動は，ある静止状態から運動状態に，そして，運動状態からは別の静止状態，もしくは別の運動状態に遷移するという形の状態遷移の繰返しで記述できると考える．ここで，運動状態とは，周期運動や特徴ベクトルの値が急激に増加していくような運動ではなく，速度の符号が変わらずかつ一定の値に収束していくような運動のみが行われる状態を指すことにする．例えば，口の開閉の運動は，閉じている（静止状態），開く（運動状態），開いている（静止状態），閉じる（運動状態），閉じている（静止状態）というように記述できる．

このような顔の運動の単位を表現したものに，FACS における AU がある．しかし，AU は，多くの表情を人間が観察し，主観的に分類したものであるので，それによって表現しきれないような表情動作が存在するのではないかと考察される．例えば実際の人間の表情では，目を閉じる（AU43）という動作は，動作の速度や強度に関して幅広く変化する．そこで，本論文では AU をモードとして設定せず，顔表情の特徴を表す特徴ベクトル系列からボトムアップにモードを求めていく手法を取る．これによって，AU では表現しきれない表情動作も表現可能となる．

本論文では，映像の変化をいくつかの線形動的システムで表現する区間に分けることができるという仮定の下，実際に撮影された映像から，線形動的システムで表現可能なモードを抽出する．顔パーツ P_a におけるモード $M_i^{(a)}$ $(i \in \{1, ..., N_{m_a}\})$ の状態方程式は，次式で表される．

$$z_t^{(a)} = F^{(a,\,i)} z_{t-1}^{(a)} + f^{(a,\,i)} + \omega_t^{(a,\,i)} \tag{2}$$

ここで，$z_t^{(a)}$ は時刻 t における特徴ベクトルである．$F^{(a,\,i)}$ は遷移行列であり，モード毎に異なる．$f^{(a,\,i)}$ はバイアス項である．$\omega^{(a,\,i)}$ はプロセスノイズであり，平均ベクトル $\mathbf{0}$，共分散行列 $Q^{(a,\,i)}$ の正規分布に従うとする．

線形システムは周期的，振動的な運動も表現可能である．しかし，先に述べたように，我々は静止状態もしくは収束性の運動状態のみをモードとして抽出したい．そこで，4.2 節では，式 (2) における遷移行列 F の固有値に制約を加え，静止状態もしくは収束性の運動状態のみをモードとして抽出する手法について述べる．

3.4 表情譜におけるタイミング構造

前節までに定義した表情譜を用いることで，顔パーツ間の運動の関係，すなわちタイミング構造を表現することが可能となる．ここでは，いったん表情譜が求まった場合に，そこから抽出可能なタイミング構造の表現方法について考察を行う．

3.4.1 分布によるタイミング構造の表現

一般に，2 つの区間 I_i, I_j の時間関係は，区間の始点 b_i, b_j ，終点 e_i, e_j の前後関係 {前，後，同時} に注目すれば，図 3 に示すように 13 通りに分類可能である[1)7)]．しかし，実際に表情を理解する上では，単なる前後関係だけでは不十分であり，区間がどの程度の時間差で開始，終了するのかといったずれの程度が重要となる．したがって本論文では，図 3 の 13 通りの関係を拡張し，区間の始点・終点の時間差の分布を用いたタイミング構造の表現方法を提案する．

まずはじめに，2 つの区間 I_i, I_j のタイミング構造を 1 次元空間の分布で表現すると，$H(b_j - b_i), H(e_j - e_i), H(b_j - e_i), H(e_j - b_i)$ の 4 個の分布で表現できる．ここで，$H(r)$ は r を変数とする 1 次元空間の分布とする．同様に，2 次元空間の分布で表現すると，$H(b_j - b_i, e_j - e_i), H(b_j - b_i, b_j - e_i), H(b_j - b_i, e_j - b_i) H(e_j - e_i, b_j - e_i), H(e_j - e_i, e_j - b_i), H(b_j - e_i, e_j - b_i)$ の 6 個の分布で表現できる．ここで，$H(r_1, r_2)$ は r_1, r_2 を変数とする 2 次元空間の分布とする．その例として，横軸を始点の差，縦軸を終点の差とする分布 $H(b_j - b_i,\ e_j - e_i)$ を図 4 に示す．同様に 3 次元以上の空間の分布も表現できる．

さらに，3 つ以上の区間のタイミング構造も表現できる．例えば，3 つの区間 I_i, I_j, I_k のタイミング構造を 1 次元空間の分布で表現すると，$H(b_j - b_i)$, $H(b_k - b_j)$ 等のような 12 個の分布で表現できる．同様に 2 次元以上の空間の分布も表現できる．

3.4.2 表情譜から抽出するタイミング構造

実際にこれらの分布を考える際には，どの区間の組み合わせを扱うかが重要となる．本論文の評価実験では，その組み合わせを以下のように定める．まず，顔パーツ P_a における区間 $I_k^{(a)}$ と，P_a 以外の全ての顔パーツ P_b $(b \neq a, b \in \{1, ..., N_p\})$ における $I_k^{(a)}$ と時間的に最も近い区間の組み合わせに注目する．これらの区間は $I_{l^*}^{(b)} \left(l^* = \arg\min_l \mathsf{IntervalDist}\left(I_k^{(a)},\ I_l^{(b)} \right) \right)$ で求められる．ここで，区間同士の距離 IntervalDist は次式で表されるものとする．

$$\mathsf{IntervalDist}\left(I_k^{(a)},\ I_l^{(b)} \right) = \left| b_k^{(a)} - b_l^{(b)} \right| + \left| e_k^{(a)} - e_l^{(b)} \right| \tag{3}$$

次に，求められた区間の組み合わせにおいて，そのタイミング構造を 2 次元空間の分布で表現する．この分布がクラスターを形成していれば，基本的な感情に基づくカテゴリーより詳細な表情のカテゴリーを表現することができたといえる．

4. 表情譜の自動獲得

この章では，入力として顔画像系列を与えた時に，出力として表情譜を自動獲得するための方法に関して述べる．

4.1 顔パーツの特徴点の座標値の抽出

与えられた顔画像系列から顔パーツの特徴点を検出するために Active Appearance Model（AAM）[9)] というモデルを利用する．AAM とは，shape（特徴点の座標

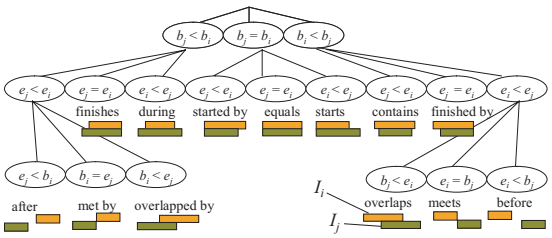

図 3 2 つの区間の始点・終点の時間関係による分類．2 つの区間 I_i，I_j の時間関係は，始点 b_i，b_j と終点 e_i，e_j の前後関係に注目すれば 13 通りに分類できる．

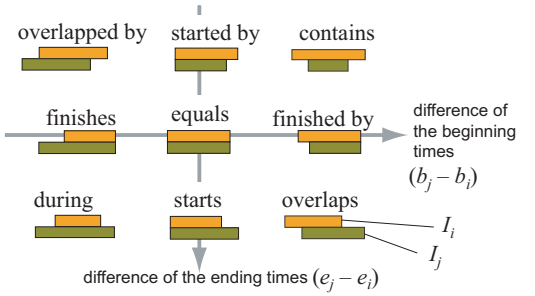

図 4 2 つの区間の始点の差・終点の差で表される 2 次元空間の分布．横軸は 2 つの区間 I_i，I_j の始点の差 $b_j - b_i$，縦軸は終点の差を $e_j - e_i$ を表す．

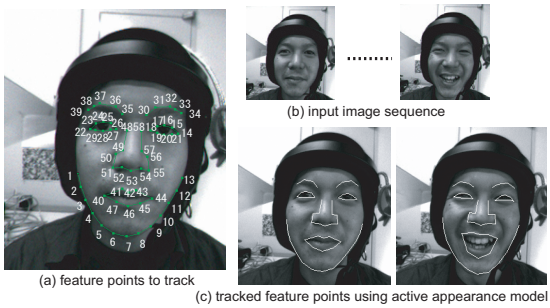

図 5 (a) AAM のモデル構築で用いる学習データ．(b) 評価実験で撮影した画像系列．(c) AAM を用いて検出した特徴点を表示した画像系列．

値）と grey-level（輝度値）の相関をパラメタとして持つ統計学的なモデルであり，このモデルを利用して高速かつ安定したマッチングを行うことが可能である．

4.1.1 AAM のモデル構築と学習

AAM のモデルを構築するための学習データとして，特徴点を配置した画像を用意する．学習データの例を，図 5 (a) に示す．画像上にある点が特徴点であり，数字は特徴点のラベルを表す．以下では，画像の特徴点の座標値をベクトルとして表したものを shape ベクトル s，平均 shape で覆われた領域内の輝度値をベクトルとして表したものを grey-level ベクトル g と呼ぶことにする．

まず，shape ベクトル s と grey-level ベクトル g の系列をそれぞれ独立に主成分分析すると，各学習データの s，g は次式で近似される．

$$s = \bar{s} + U_s c_s \ , \ g = \bar{g} + U_g c_g \tag{4}$$

ここで，$\bar{s}$，$\bar{g}$ はそれぞれ shape と grey-level の平均ベクトル，U_s，U_g は主成分を列ベクトルとして並べた列直交行列，c_s，c_g は主成分の係数であり，それぞれ shape パラメタベクトル，grey-level パラメタベクトルと呼ぶ．shape の変化と grey-level の 変化には相関があるので，これらパラメタベクトル c_s，c_g を連結させたもの c を定義し，この c で表されるベクトル系列に対して主成分分析をさらに行うことで次式のモデルが得られる．

$$\begin{bmatrix} W_s c_s \\ c_g \end{bmatrix} = c = \begin{bmatrix} V_s \\ V_g \end{bmatrix} d = V d \tag{5}$$

ここで，W_s は shape パラメタの重みを表す対角行列であり，これにより shape と grey-level 間の単位の違いを吸収している．また，V は固有ベクトル系列，d は shape と grey-level の両方を制御するパラメタベクトルである．モデルの線形性に注目すれば，shape ベクトル s と grey-level ベクトル g は，d の関数として次式で表せる．

$$s = \bar{s} + U_s W_s^{-1} V_s d \ , \ g = \bar{g} + U_g V_g d \tag{6}$$

ここで，$V = (V_s^{\mathsf{T}}, V_g^{\mathsf{T}})^{\mathsf{T}}$ である．つまり，d を与えると式 (6) より画像の grey-level ベクトル g と shape ベクトル s が求まる．そして，この g と s を用いて画像を合成することができる．AAM の学習では，モデルと学習データ画像の間の grey-level の残差ベクトルとモデルパラメタの修正ベクトルの関係を学習する．

4.1.2 AAM を用いた探索

探索対象画像と上述のモデルが与えられた時，マッチングは，探索対象画像とモデルから合成された画像の間の grey-level の残差を最小化する最適化問題として考えることができる．つまり，マッチングによって得られるモデルパラメタ d^* は，

$$d^* = \arg\min_d |g_u - g_v|^2 \tag{7}$$

により表せれる．g_u は探索対象画像の grey-level ベクトルであり，g_v はモデルパラメタから合成された画像の grey-level ベクトルである．得られた最適解 d^* と式 (6) より shape ベクトル s^* が求められる．そして，s^* の成分の中から各顔パーツの座標値に相当するものを選び，それを特徴ベクトルとして得る．

4.2 モードへの分節化

4.1 節で抽出された顔パーツの特徴ベクトル系列をモードへと分節化するために，特徴ベクトル系列が，線形システムで表現可能な区間からなると仮定する．本論文では，階層型クラスタリングを用いて，系列を構成する線形システム集合の各パラメタを，各システムで表される区間への分節化と同時に推定する手法を提案する．この手法は，各顔パーツ毎の系列に独立に適用される．

各モード内は，それぞれ状態方程式が 式 (2) で表さ

れるような線形動的システムによって表現される．ここで，特徴量の値がほとんど変化しないような静止状態や,「はじめは早い変化であり，次に静止していく」ような収束性の運動を単位として分節化を行うために，式 (2) における遷移行列 F に制約を加え，全ての固有値の絶対値を 1 より小さくしたような線形システムを考える．以下では，まず最大固有値の絶対値に制約を加える方法について述べ，次に線形動的システムのクラスタリング手法について述べる．このとき，説明を簡便にするために顔パーツ P_a のものであることを示す添字 a を省略して述べる．

4.2.1 制約付き線形システム同定

特徴ベクトル系列 $z_1^{(i)},..,z_T^{(i)}$ から遷移行列 $F^{(i)}$ を計算するためにまず，$Z_0^{(i)} = \left[z_1^{(i)},...,z_{T-1}^{(i)}\right]$，$Z_1^{(i)} = \left[z_2^{(i)},...,z_T^{(i)}\right]$ と置く．このとき，$F^{(i)}$ の同定は，各時刻における自乗予測誤差を最小にする問題と考えることができる．

$$F^{(i)*} = \arg\min_{F^{(i)}} \left\| F^{(i)} Z_0^{(i)} - Z_1^{(i)} \right\|^2 \quad (8)$$

これを行列方程式として微分法を用いて解くことで $F^{(i)}$ は，

$$F^{(i)*} = \lim_{\delta^2 \to 0} Z_1^{(i)} Z_0^{(i)\mathsf{T}} \left(Z_0^{(i)} Z_0^{(i)\mathsf{T}} + \delta^2 I \right)^{-1} \quad (9)$$

と求められる．ここで，I は $2n_{p_a}$ 次元の単位行列であり，δ は正の実数値である．δ を 0 に収束させずに適当な正の実数値にすることで $F^{(i)}$ の固有値を 1 より小さくする．

4.2.2 モードの自己組織化

3.3 節で述べたように，顔パーツの運動は，静止状態と収束性の運動状態に分けることが出来る．そこで，まずは特徴ベクトルの一次微分のゼロ交差点においておおまかに分節化を行う．

次に，モード間に距離を定義することで，階層型クラスタリングに基づいて動的システムのクラスタを併合していき，モードの数を減らしていく．はじめに，それぞれの区間が別のモードに従うとしてモードのモデルパラメタの同定を行う．そして，全ての区間同士の距離を計算し，その中で最も近い 2 つのモードを 1 つのモードとして併合し，新たにそのモードのモデルパラメタ及び他の区間との距離を計算する．この併合の反復処理は，モードの数が 1 つになるまで繰り返される．付録に階層型クラスタリングのアルゴリズムを示す．

4.2.3 モード間の距離

モード間の距離尺度としては，次の式で表される予測誤差 E を距離尺度として用いる．

$$E\left(M_i||M_j\right) = \frac{1}{C} \sum_{I_k \in \mathcal{I}_i} \sum_{t=b_k}^{e_k} \left({E_t^{(i|j)}}^2 - {E_t^{(i|i)}}^2 \right) \quad (10)$$

ここで，C は区間集合 $\mathcal{I}_i$ に含まれる区間 I_k の区間長の総和であり，これによって時間的な正規化を行う．また，$E_t^{(i|j)}$ は，

$$E_t^{(i|j)} = F^{(j)} z_{t-1}^{(i)} + f^{(j)} - z_t^{(i)} \quad (11)$$

で表されるものとする．式 (10) は，モード M_i と M_j に関して非対称であるため，これを相互に評価することで以下のような対称な距離を定義する．

$$\mathrm{Dist}\left(M_i,\ M_j\right) = \left\{ E\left(M_i||M_j\right) + E\left(M_j||M_i\right) \right\} / 2 \quad (12)$$

5. 評 価 実 験

この章では，4 章で述べた方法を用いて，実際に撮影した顔画像系列から表情譜を自動獲得し，表情譜の有効性について評価を行う．

5.1 映像キャプチャ

入力顔画像系列は，2 人の人物の意図的に作った笑いと自発的に表出された笑いを，解像度 240 × 320, フレームレート 60fps で撮影したものを使用した．撮影は，頭部の動きが生じた場合でも正面顔の撮影を行うために，ヘルメット前方にカメラを固定したカメラシステムを用いた．被験者には無表情から始めて，当該の表情を表出した後は無表情に戻すように指示した．意図的な笑いは，被験者に作り笑いをするように指示をして撮影した．自発的な笑いは，被験者と向かい合った位置に立った協力者が被験者を笑わせて撮影した．そして，1 つの映像を撮影する時はどちらかの笑いのみを表出するように指示し，両方の笑いが混合しないようにした．撮影した顔画像系列の一部を，図 5 (b) に示す．

5.2 表情譜の自動獲得

撮影された顔画像系列に対して，4.1 節で述べた方法を用いて各顔パーツの特徴点のトラッキングを行った☆．この時 AAM のモデルで用いる特徴点の数は，各眉に 5 点，各目に 8 点，鼻に 11 点，唇に 8 点，顔の下半分の輪郭に 13 点の合計 58 点とした（図 5 (a) 参照）．その結果，各眉 10 次元，各目 16 次元，鼻 22 次元，口 16 次元の特徴ベクトル系列を得た．

AAM によって検出された特徴点を表示した顔画像系列の一部を，図 5 (c) に示す．図 5 (c) の画像は，図 5 (b) の画像と同時刻のものである．これらの図を比較すると，表情の変化に伴う特徴点の変化が精度良く検出されていることが読み取れる．

次に，得られた特徴ベクトル系列に対して，各パーツ毎に 4.2 節で述べた方法を用いてモードへの分節化を行った．その結果，意図的な笑いと自発的な笑いの表情譜を獲得した．

得られた表情譜の一例として，自発的な笑いにおける表情譜のうち，口のパートを図 6 に示す．図の上段は特徴点の x 座標の値，中段は特徴点の y 座標の値，下段は異なるモードを縦軸として表している．横軸は

☆ 特徴点のトラッキングには，Stegmann（Technical University of Denmark）の AAM-API を用いた[8]．

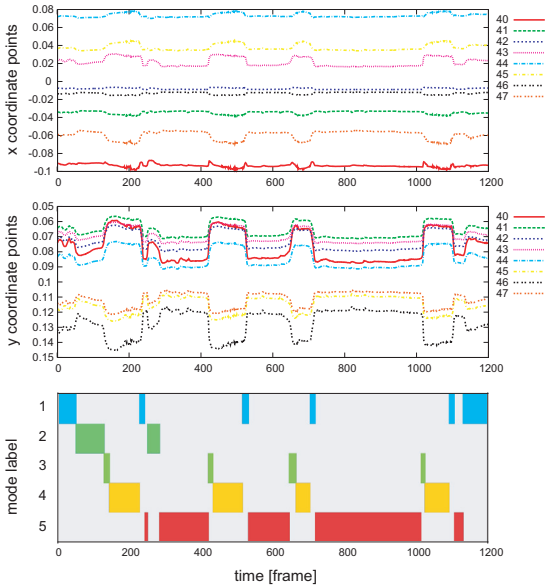

図 6 自発的な笑いにおける分節化の結果（口）．縦軸は，上段は特徴点の x 座標値，中段は特徴点の y 座標値，下段は異なるモードを表す．横軸は時間軸である．上段，中段における凡例の数字は，図 5 (a) の特徴点のラベルを表す数字に対応している．

時間軸である．この図より，無表情の状態，笑っている状態を始め，笑いの開始時の動作や笑いの終了時の動作がそれぞれ異なるモードとして分節化されていることが読み取ることができる．

5.3 意図的な笑いと自発的な笑いにおけるタイミング構造の比較

得られた表情譜を用いて意図的な笑いと自発的な笑いにおけるタイミング構造の比較を行った．今回はその一例として，笑いの開始時（begin smiling）のモードに着目して，口・鼻・左目の 3 つの顔パーツのモードがどのような時間関係で発生し終了するかに関して考察を行った（図 7 参照）．モードの標本数はそれぞれの表情に対して 20 個とした．

各顔パーツ間の時間関係を計算した結果，口と鼻の始点の差を横軸，鼻と左目の始点の差を縦軸とする 2 次元空間の分布 $H(b_{nose} - b_{mouth},\ b_{leye} - b_{nose})$ を考えた時に 2 表情を最も良く分離することができた．ここで，b_{mouth}，b_{nose}，b_{leye} はそれぞれ口，鼻，左目の始点を表す．その分布を図 8 に示す．これらの図より意図的な笑いと自発的な笑いにおける分布が 2 被験者の間で類似しており，それぞれクラスタを形成していることが読み取れる．よって，表情譜から抽出されるタイミング構造は，意図的な笑いと自発的な笑いという 2 表情を分類するのに有効であるといえる．

自発的な笑いでは口を動かす筋肉が動かされ，その動きに付随して頬を持ち上げる筋肉が動くのに対し，意図的な笑いではそれらの筋肉を意図的に制御して笑顔を作ろうとするために，口を動かす筋肉と頬を持ち

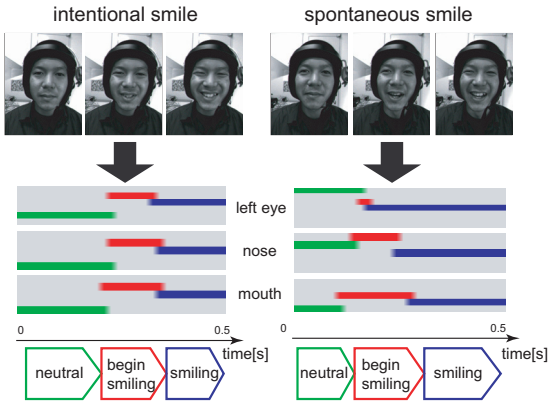

図 7 意図的な笑いと自発的な笑いの表情譜の比較．

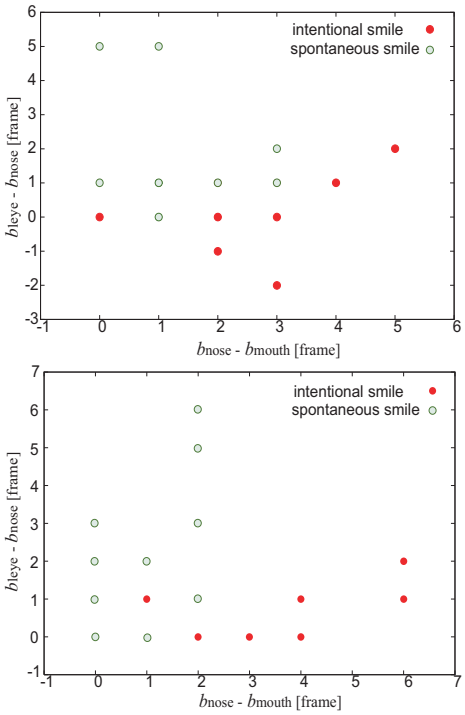

図 8 笑い開始時のモードにおける口と鼻，鼻と左目の始点の差の分布．横軸は口と鼻の始点の差 $b_{nose} - b_{mouth}$，縦軸は鼻と左目の始点の差 $b_{leye} - b_{nose}$ を表す．上段と下段の図はそれぞれ別の被験者のものである．

上げる筋肉が同時に動かされると考えられる．それゆえに，2 表情間で観測された差異が生じるのではないかと考察される．

6. 結　　論

本論文では，表情変化を顔パーツそれぞれの時間的な運動によって生じるものと考え，そのタイミング構造を記述する表現形式として「表情譜」を定義し，表情をより詳細に理解する枠組を提案した．

そして，実際に撮影した顔画像系列から表情譜を自動獲得し，顔パーツの運動が，線形システムで表される区間を単位として記述されることを示した．さらに，この表情譜を用いることで，意図的な笑い及び自発的な笑いを分離できることを確認し，表情理解においてタイミング構造を用いることの有効性を示した．

しかし，以下の項目については本論文では十分に扱うことができなかったため，今後の課題とする．

タイミング構造と表情理解の因果性: 表情譜からこのようなタイミング構造が得られた時は表情をこのように理解できるという，タイミング構造と表情理解の因果性を確認するため，様々な表情から得られる表情譜を分析する必要がある．

タイミング構造の個性: 本論文で示した 2 表情間におけるタイミング構造の差異は，今回の 2 人の被験者が持つ差異が偶然に類似していた可能性がある．つまり，他の被験者においては別の種類の差異が観測される場合や，一方で全く差異が観測されない場合も考えられる．これに関しては，さらに被験者数を増やして今後の考察を行う必要がある．

文脈という時間的要因: 本論文では，表情理解における時間的要因としてタイミング構造に着目した．しかし，実際に私たちは，会話等の文脈から得られる情報も踏まえて表情を読み取っていると考察できる．つまり，モードの前後関係という時間的要因に関しても今後の考察を行う必要がある．

謝辞: 本研究の一部は，科学研究費補助金 13224051 及び 16700175 の補助を受けて行った．

参 考 文 献

1) Allen, J. F.: Maintaining Knowledge about Temporal Intervals, *Communications of the ACM*, Vol. 26, No. 11, pp. 832–843 (1983).
2) Bassili, J. N.: Facial Motion in the Perception of Faces and of Emotional Expression, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 4, No. 3, pp. 373–379 (1978).
3) Essa, I. A. and Pentland, A. P.: Facial Expression Recognition using a Dynamic Model and Motion Energy, *Proc. 5th IEEE International Conference on Computer Vision '95*, pp. 360–367 (1995).
4) li Tian, Y., Kanade, T. and Cohn, J. F.: Recognizing Action Units for Facial Expression Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 97–115 (2001).
5) Nishio, S., Koyama, K. and Nakamura, T.: Temporal Differences in Eye and Mouth Movements Classifying Facial Expressions of Smiles, *Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 206–211 (1998).
6) P.Ekman and W.V.Friesen: *Unmasking the Face*, Prentice Hall (1975).
7) Pinhanez, C. and Bobick, A.: Human Action Detection using PNF Propagation of Temporal Constraints, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 898–904 (1998).
8) Stegmann, M. B., Ersboll, B. K. and Larsen, R.: FAME - A Flexible Appearance Modelling Environment, *Informatics and Mathematical Modelling, Technical University of Denmark* (2003).
9) T.F.Cootes, G.J.Edwards and C.J.Taylor: Active Appearance Model, *Proc. European Conference on Computer Vision*, Vol. 2, pp. 484–498 (1998).

付 録

A.1 階層型クラスタリング

Algorithm 1 階層型クラスタリング

for $i \leftarrow 1$ to N **do**
  $M_i^{(a)} \leftarrow \mathsf{Identify}\left(I_i^{(a)}\right)$
end for
for all pair$\left(M_i^{(a)},\ M_j^{(a)}\right)$ where $M_i^{(a)},\ M_j^{(a)} \in \mathcal{M}^{(a)}$ **do**
  $\mathsf{Dist}\,(i,\ j) \leftarrow \mathsf{CalcDistance}\left(M_i^{(a)},\ M_j^{(a)}\right)$
end for
while $N \geq 2$ **do**
  $(i^*,\ j^*) \leftarrow \arg\min_{(i,\ j)} \mathsf{Dist}\,(i,\ j)$
  $\mathcal{I}_{i^*}^{(a)} \leftarrow \mathsf{MergeIntervals}\left(\mathcal{I}_{i^*}^{(a)},\ \mathcal{I}_{j^*}^{(a)}\right)$
  $M_{i^*}^{(a)} \leftarrow \mathsf{Identify}\left(\mathcal{I}_{i^*}^{(a)}\right)$
  erase $M_j^{(a)*}$ from $\mathcal{M}^{(a)}$
  $N \leftarrow N - 1$
  for all pair$\left(M_i^{(a)*},\ M_j^{(a)}\right)$ where $M_j^{(a)} \in \mathcal{M}^{(a)}$ **do**
    $\mathsf{Dist}(i^*,\ j) \leftarrow \mathsf{CalcDistance}\left(M_{i^*}^{(a)},\ M_j^{(a)}\right)$
  end for
end while

$M^{(a)}$ や $I^{(a)}$ 等に見られる添字 a は顔パーツ P_a のものであることを示す添字である．Identify は 4.2.1 項で述べたシステム同定法を表し，区間内にある特徴ベクトルデータを用いて，モードのモデルパラメタ $\theta_i^{(a)} = \left\{F^{(a,\,i)},\ f^{(a,\,i)},\ Q^{(a,\,i)},\ z_{init}^{(a,\,i)}\right\}$ を同定する．階層型クラスタリングでは，時間的に離れた位置にある（互いに重なりを持たない）区間であっても，同じモードで表現されることがある．そこで，モード $M_i^{(a)}$ によって表現される区間の集合を $\mathcal{I}_i^{(a)}$ としている．CalcDistance は，モード間の距離を求める処理であり 4.2.3 項で定義する．MergeIntervals によって 2 つの区間集合は併合され，得られた区間集合からモードのモデルパラメタを再同定する．