

表情譜: タイミング構造に基づく表情の記述・生成・認識

Facial Expression Description, Generation, and Recognition based on Timing Structure

川嶋 宏彰[†]
Hiroaki Kawashima

西山 正紘[†]
Masahiro Nishiyama

松山 隆司[†]
Takashi Matsuyama

1. 表情のタイミング構造

表情は動的な変化である。脳内からの指令が表情筋を制御することで、目や眉毛、頬や口などの運動が、時間的・空間的に複雑に組み合わさって表情を生成する。人と人とのコミュニケーションにおいて、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。すなわち、生得的・発達の獲得した表情の変化パターンをプロトコルとして利用することで、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。

マンマシ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、従来の研究では、主に表情を基本的な感情カテゴリ（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 [5]。しかし、実際の表情は、意図的に制御されて生成されるものや、感情などによって自発的に生じられるものがあり、これらが同時に混ざり合って表情が生まれることも多い。つまり、人は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で、基本的な感情カテゴリに比べてより粒度の細かな分類を行い、相手の内部状態の推定していると考えられる。

従来の研究で用いられる表情の記述形式としては、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 [7]。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。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ため、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。

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。

- 表情の分類が感情に基づく基本カテゴリに留まる
- 表情の記述単位が主観的に選択されている
- 表情の動的側面を十分に用いていない

表情譜—区間に基づく表情の記述— 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する。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・生成する枠組を提案する。ここで、タイミング構造とは、ある複数個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

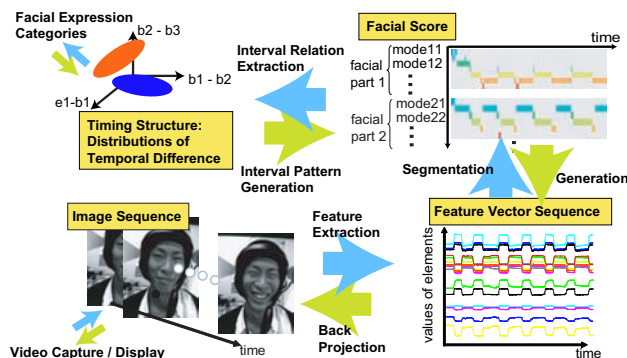


図 1: 表情譜を用いた表情認識・生成の流れ。

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 では表現できない運動が多く、またその獲得方法自体がまだ研究対象であるため、詳細なタイミング構造を記述する上では、AU を記述単位とするのは適切ではない。そこで本論文では、ある顔パーツに注目したときに、そのパーツに現れる要素的な運動を、それぞれ異なる線形システムでモデル化し、そのパーツの全体の運動を、複数の線形システムの活性化パターンによって表現できるものと仮定する。そして、顔表情の特徴を表す特徴ベクトル系列から、ボトムアップに線形システム（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運動の記述が可能となると共に、得られた区間から逆に特徴系列を生成できる。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動的構造の記述が可能
- 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
- 表情の認識と生成を同一の枠組みで行うことが可能

表情譜に基づく表情の認識と生成 本研究で提案する枠組みは、図 1 に示すような、映像と表情カテゴリの双方向のプロセスからなる。映像から表情を認識する際は、まず、表情から顔の特徴を表す特徴ベクトル系列を抽出する。次に、特徴ベクトル系列を用いて顔の運動をモードに節分化し、表情譜を獲得する。最後に、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、表情を認識する。実際に表情認識を行う際には、この認識プロセスに先立って、あらかじめ獲得した表情コーパスから、モードおよびタイミング構造（区間同士の時間関係の分布）を得る必要がある。表情の生成は、この逆のプロセスによって可能である。すなわち、いったん学習されたタイミング構造に基づいて表情譜を生成し、さらにこの表情譜に沿って表情映像を生成する。

[†] 京都大学情報学研究所, Grad. Sch. of Informatics, Kyoto Univ.

本論文では以上のプロセスのうち、表情譜の (1) 自動獲得手法の提案, (2) 映像生成における有効性の評価, および (3) 表情解析における有効性の評価を行うことを目的とする. 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, 意図的な笑いとお spontaneous 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.

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, 表情理解の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]. 運動の成分をより直接的に扱った研究として, 小山らは目と口の運動の時間的差異に着目して, 快の笑い・不快の笑い・社交の笑いの分類について検討を行っている [9]. 人間の表情を直接計測した研究として, 内田らや高橋は基本 6 表情において意図的な表情とお spontaneous な表情における動的変化 (特徴点の変位) の違いを示している [12, 8]. 本論文ではこれらの知見を踏まえて, いったん映像を要素に還元し, それら要素間の時間的關係から表情という動的な事象を記述・生成・認識する枠組みの実現を目指す.

2. 表情譜とその自動獲得

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, 表情譜を以下で定義する.

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 $\mathcal{P} = \{P_1, \dots, P_{N_p}\}$ 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 $P_a (a \in \{1, \dots, N_p\})$ 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 $\mathcal{M}^{(a)} = \{M_1^{(a)}, \dots, M_{N_{m_a}}^{(a)}\}$ 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 $\mathcal{I}^{(a)} = \{I_1^{(a)}, \dots, I_{N_{k_a}}^{(a)}\}$ で定義する. また, 区間 $I_k^{(a)} (k \in \{1, \dots, N_{k_a}\})$ は始点 $b_k^{(a)} \in \{1, \dots, T\}$, 終点 $e_k^{(a)} \in \{1, \dots, T\}$, 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 $\{\mathcal{I}^{(1)}, \dots, \mathcal{I}^{(N_p)}\}$ 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.

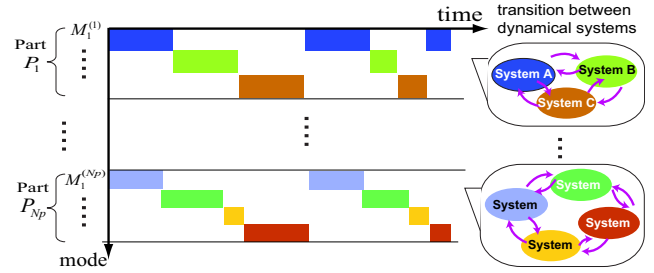


図 2: 表情譜 (縦軸: モード, 横軸: 時間). 各顔パーツ毎に運動状態の時間的遷移をモードを単位として記述する. これは各パーツでの線形システムの活性化パターンに一致する.

2.1 表情譜における顔パーツ

顔パーツをどう定義するかは重要な問題であるが, タイミング構造を利用して表情を記述するには, 少なくとも視覚的に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べきである. Ekman らは, 基本的感情が表情に現れる様子の違いを, 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 [4]. 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, 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. 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 ($p \in \{1, \dots, n_{p_a}\}$) の特徴点における座標値を $(x_p^{(a)}, y_p^{(a)})$ とすると, 以下の $2n_{p_a}$ 次元列ベクトルとして表せる.

$$z^{(a)} = (x_1^{(a)}, y_1^{(a)}, \dots, x_{n_{p_a}}^{(a)}, y_{n_{p_a}}^{(a)})^T \quad (1)$$

パーツの自動追跡 与えられた顔画像系列から顔パーツの特徴点を追跡するために Active Appearance Model (AAM) [3] を利用する. AAM とは, 特徴点の座標値と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. 直前のフレームで探索された特徴点を初期値として各フレームで AAM による探索を行い, 顔の特徴点を追跡する.

2.2 表情譜におけるモード

顔パーツの運動は, 様々な静止状態や運動状態 (モード) の間の遷移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, 始めは急速に変化し, 次第に一定の値に収束していくような運動 (以下では収束性の運動と呼ぶ) のみが行われる状態を指すことにする.

本論文では, 各モードをそれぞれ異なる線形システムでモデル化する. そして, 特徴ベクトルの変化は, 複数の線形動的システムを順に活性化しながらモデル化できると仮定する. すなわち, 各パーツごとに, 文献 [10] のような Hybrid Dynamical System を生成モデルとして用いる. 各パーツを構成するモードは, 実際に撮影された映像から線形動的システムの集合として抽出する. 顔パーツ P_a におけるモード $M_i^{(a)} (i \in \{1, \dots, N_{m_a}\})$ の状態方程式は, 次式で表される.

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

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, $f^{(a, i)}$ 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。一般に線形システムは収束性、発散性のいずれの変化も表現可能であるが、式 (2) における遷移行列 F の固有値の絶対値が 1 以下という制約を加えることで、静止状態もしくは収束性の運動のみをモードとして抽出する。

モードの自動抽出 モードの抽出は各パーツで独立に行う。まず、モード M_i, M_j 間の距離を、モード間の予測誤差に基づいて定義する。 M_j の線形システムを用いて、 M_i で表現されている系列 $z_{t-1}^{(i)}, z_t^{(i)}$ の予測誤差を $\epsilon_t^{(ij)} = F^{(j)} z_{t-1}^{(i)} + f^{(j)} - z_t^{(i)}$ のように計算する。さらに、 M_i 自身で予測した場合の誤差を考慮し、 $E(M_i || M_j) = \frac{1}{C} \sum_{I_k \in \mathcal{I}_i} \sum_{t=b_k}^{e_k} (\|\epsilon_t^{(ij)}\|^2 - \|\epsilon_t^{(ii)}\|^2)$ を求める。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、これにより時間的な正規化を行う。 $E(M_i || M_j)$ は 2 つの線形システム間の KL divergence の近似になっており、 i, j に関して非対称であるため、 M_i, M_j 間の距離は $\{E(M_i || M_j) + E(M_j || M_i)\} / 2$ により定義する。

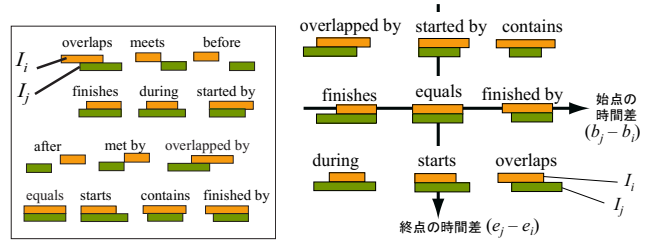
モード抽出処理は、はじめに、特徴点の速度の零交差点を用いて大まかに分節化を行い、各区間で線形システムを同定する。次に、最も近い 2 つの線形システムと、それらによって表現される区間集合を併合し、併合された区間集合からシステムを再度同定する。階層的クラスタリングに基づいてこの併合を繰り返すことで、線形システム (モード) の数を減らしていく。同定の際には、先に述べた固有値の制約を設ける。得られたモデルで生成した系列と、元の系列との誤差を各併合段階で計算し、誤差の急激な変化を検出することで、適切なモード数を決定する (詳細は文献 [10, 11] を参照のこと)。

2.3 表情譜から抽出可能なタイミング構造

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(a) に示すように 13 通りに分類可能である [1]。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なく、区間がどの程度の時間差で開始、終了するのかといった程度の重要性となる。したがって本論文では、図 3(a) 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、 $H(b_j - b_i, e_j - e_i), H(b_j - b_i, b_j - e_i)$ をはじめとする 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 $H(b_j - b_i, e_j - e_i)$ の空間を図 3(b) に示す。同様に 3 次元以上の空間の分布や、3 つ以上の区間の間のタイミング構造も表現できる。

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、顔パーツ P_a における区間 $I_k^{(a)}$ に注目したとき、 P_a 以外の全ての顔パーツ $P_b (b \neq a, b \in \{1, \dots, N_p\})$ における $I_k^{(a)}$ と時間的に



(a) 13 通りの時区間関係 (b) 始・終点による 2 次元空間の例

図 3: (a) 区間の始点の差・終点の差による 13 通りの分類。(b) 横軸に 2 つの区間 I_i, I_j の始点の差 $b_j - b_i$ を、縦軸に終点の差 $e_j - e_i$ をとった 2 次元空間。

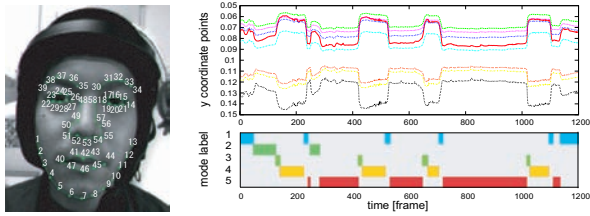
最も近い区間 $I_{l^*}^{(b)} (l^* = \arg \min_l \text{IntervalDist}(I_k^{(a)}, I_l^{(b)}))$ の組み合わせを用いる。ここで、区間同士の距離は $\text{IntervalDist}(I_k^{(a)}, I_l^{(b)}) = |b_k^{(a)} - b_l^{(b)}| + |e_k^{(a)} - e_l^{(b)}|$ と定義する。次に、求められた区間の組み合わせにおいて、そのタイミング構造を 2 次元分布で表現する。この分布が複数のクラスタを形成していれば、タイミング構造が表情の分類に利用可能であるといえる。

3. 評価実験

実際に撮影した顔映像を用いて表情譜の有効性を評価した。 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、その後は再び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。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。

表情譜の自動獲得 撮影された顔画像系列に対して、AAM を用いて各顔パーツの特徴点の追跡を行った [6]。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 4(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。次に、得られた特徴ベクトル系列に対して、各パーツ毎に 2.2 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 4(b) に示す。

表情譜による映像生成 表情譜は、各顔パーツの運動を線形システムの活性化パターンとして表現しているため、逆に特徴ベクトル系列を生成することが可能である。生成された特徴点を、AAM によって画像空間に逆投影することで、映像生成が可能となる。図 5 に、自発的に 2 回笑顔が生じたときの表情譜、および生成した映像中で特徴的な 6 フレームを取り出して示す。これより、例えば口の表情譜からは、無表情の状態、笑っている状態、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れる。



(a) 追跡する特徴点 (b) 口の特徴ベクトル系列 (y座標のみ表示) と表情譜

図 4: (a) AAM のモデル構築と追跡で用いる特徴点 (追跡精度を高めるために顎の点も用いる). (b) 自発的な笑いにおける口の特徴系列と獲得した表情譜. 上段縦軸: 特徴点の y 座標値 (x 座標は省略). 下段縦軸: モード. 上下段横軸: 時間軸.

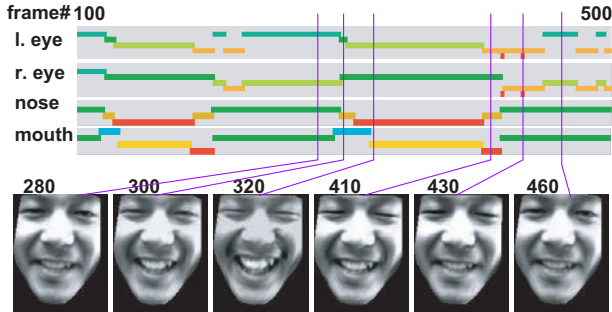


図 5: 表情譜から生成された映像の一部 (左右眉毛の譜は省略)

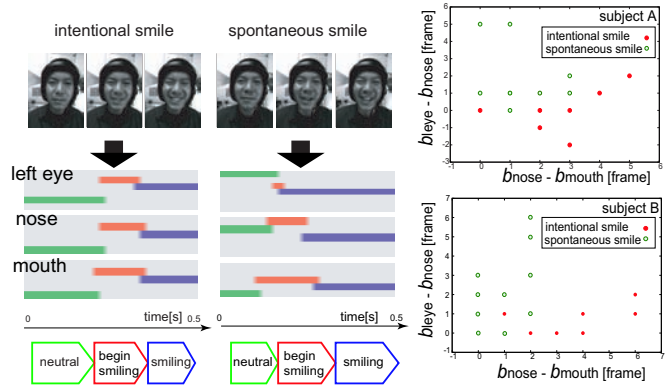
意図的・自発的な笑いにおけるタイミング構造の解析 得られた表情譜を用いて意図的な笑いと自発的な笑いにおけるタイミング構造の比較を行い、2つの表情がどの程度分離されるかを解析した。今回はその一例として、笑いの開始時のモードに着目して、口・鼻・左目の3つの顔パーツのモードがどのような時間関係で発生し終了するかに関して考察を行った (図 6(a) 参照)。モードの標本数はそれぞれの表情に対して 20 個とした。

各顔パーツ間の時間関係を 2.3 節で述べた 2 次元分布で表現した結果、図 6(b) に示すような、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 $H(b_{nose} - b_{mouth}, b_{leye} - b_{nose})$ を用いることで、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。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、それぞれの表情がクラスを形成していることが読み取れる。これは、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上げる筋肉が同時に動かされると考えられる。その結果、2 表情間で観測された差異が生じると考察される。

以上より、表情譜から抽出されるタイミング構造は、意図的・自発的な笑顔を始めとする詳細な表情を分類するのに有効であることが示唆される。

4. 結論

本論文では、顔パーツ間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、映像データからボトムアップに抽出される運動を用いて、表情生成と認識を同一の枠組みで実現する手法を提案した。表情譜のようなタイミング構造に基づく枠組みは、表情のみならず一般の動作理解、さらには映像の符号化・圧縮に適用す



(a) 意図的な笑いと自発的な笑いの表情譜 (b) 2 表情の分布の比較

図 6: 意図的・自発的な笑いの開始時におけるタイミング構造の比較。(b) は笑い開始時のモードにおける始点の時間差分布。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。上段と下段は異なる被験者。

ることが可能である。今回はタイミング構造のみに注目して表情の分類を試みたが、実際には特徴点に変化するスケールや加速度、持続時間などにも重要な情報が存在するはずである。また、タイミング構造を考える上では、構成要素 (パーツ) をどのように定義するかが問題となる。表情理解という観点からは、表情の得られた文脈 (会話内容や相手の表情) を同時に記述することが不可欠である。今後はこれらの問題に取り組むとともによりの多くデータから提案手法の有効性を評価する必要がある。

謝辞: 本研究の一部は、科学研究費補助金 13224051 および 16700175 の補助を受けて行った。

参考文献

- [1] J. F. Allen. Maintaining knowledge about temporal interval. *Commun. of the ACM*, Vol. 26, No. 11, pp. 832–843, 1983.
- [2] J. N. Bassili. 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] 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.
- [4] P. Ekman, W. V. Friesen (工藤力訳). 表情分析入門. 誠信書房, 1987.
- [5] I. A. Essa and A. P. Pentland. Facial expression recognition using a dynamic model and motion energy. *Proc. of Intl. Conference on Computer Vision*, pp. 360–367, 1995.
- [6] M. B. Stegmann, B. K. Ersboll, and R. Larsen. FAME - a flexible appearance modelling environment. *Informatics and Mathematical Modelling, Technical University of Denmark*, 2003.
- [7] Y. Tian, T. Kanade, and J. F. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 97–115, 2001.
- [8] 高橋直樹. FACS を用いた表情の時系列分析とその展望. 対人社会心理学研究, No. 2, pp. 75–82, 2002.
- [9] 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, Vol. J80-A, No. 8, pp. 1316–1318, 8 1997.
- [10] 川嶋宏彰, 堤公孝, 松山隆司. 動的イベントの分節化・学習・認識のための hybrid dynamical system. 第 3 回情報科学技術フォーラム (FIT), pp. 175–178, 2004.
- [11] 川嶋宏彰, 堤公孝, 松山隆司. 力学系集合の自己組織化に基づく唇映像の構造化. 情報論的学習理論ワークショップ (IBIS), pp. 86–93, 2004.
- [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. 信学技報, HIP99-76, pp. 1–6, 3 2000.