

『『ダイナミクスに基づく情報処理の諸相』特集号』

解説

ハイブリッドダイナミカルシステムによる動的事象のモデル化と認識

川嶋 宏彰, 松山 隆司*

1. はじめに

実世界における人の動きや、表情変化、発話などの動的な事象を、カメラやマイクによって得られる計測データから認識するうえで、何らかの数理モデルを導入することは、特に時系列パターンの、クラスとしての広がりや表現するためには有効であり、おもに 80 年代後半から、動的な事象の様々なモデル化手法が提案されている [1]。

記号処理の観点からは、有限状態オートマトンやペトリネットのような離散事象系 (discrete-event system) を、比較的抽象度の高い (たとえば文脈を踏まえた) 行動解析などに応用することが提案されている。しかしながら、記号をどのように定義するか、実際の信号と記号とをどのように対応させるかといった問題があり、さらに扱える記号は、人が認識でき手動で定義できる粒度・規模にとどまるといった限界があった。一方で、信号処理の観点からは、線形システムやリカレントニューラルネットワークなどの力学系 (dynamical system) モデルがしばしば用いられる。動作のダイナミクスを微分方程式や差分方程式として、信号レベルで詳細に記述できるため、単純な動作の追跡やテキスト変化の認識には適しているものの、判断・操作などの (人為的) 事象が関与する場合にはモデル化が困難であった。

そこで我々は、離散事象系と力学系の混在系である、ハイブリッドダイナミカルシステム (hybrid dynamical system, HDS) を用いて、これらの問題の解決を図ってきた (具体的なモデルは次章にて述べる)。HDS として、制御系の分野では hybrid automata [2] などがよく知られているが、一方で、音声やグラフィクス、コンピュータビジョンなどのメディア分野では、HMM (hidden Markov model) と線形システムの混合モデルをはじめとする様々なモデルが、音声や人の動きの認識・生成を目的として提案されている [3][4][5]。これら HDS の多くは、信号中の要素的な変化パターン (以下ではモードと呼ぶ) を表現できるような力学系を複数用意し、これら力学系間の順序構造を、上位の離散事象系によってモデル化する。このとき、モードを表現する具体的な力学系モデルの存在により、複雑なメディア信号は、力学的性質が切り替

わる時点 (以下、力学的分節点) にて自動的に分節化でき、逆に HDS から複雑な信号を生成できるなど、信号レベルに接地しながら記号処理を扱うことが可能となる。

しかし、どのように力学系集合を定め、システム同定を行うかという学習の困難さが、これら HDS を実際に用いる際の障害となっている。これに対して著者らは、計測された信号から、複数の線形システムを自動的に抽出・クラスタリングする手法を提案しており (3. 章参照のこと)、本手法により、実際にカメラやマイクから得られた音声や映像、そこから抽出されたテキストや形状変化といった各種メディア信号から HDS を同定し、信号をその力学的分節点に基づいて分節化することができる。

このとき、各メディア信号をそれぞれ異なる HDS でモデル化することで、異なるメディア信号間に現れる変化パターンの間の、共起性、同期や遅延といった時間的構造を、力学的分節点の間の時間関係として記述することができ、実世界での動的な事象を詳細にモデル化し、認識することが可能となる。4.、5. 章ではその具体的な応用事例として、タイミングに基づく表情の記述・認識、および音声と映像の間のタイミング構造モデルに基づくリップシンク (音に合った映像・動作生成) を紹介する。

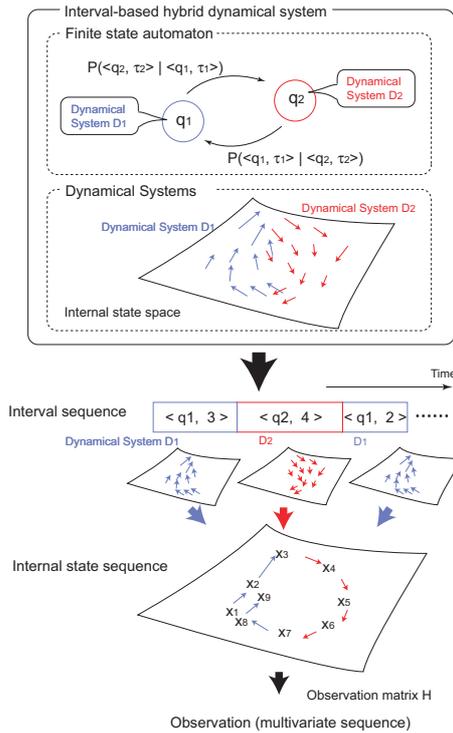
2. 時区間ハイブリッドシステム

2.1 システムアーキテクチャ

我々の用いる HDS は、図 1 に示すような 2 層からなり、第 1 層 (図 1 における 1 段目の点線矩形) は複数の離散状態間の確率的遷移をモデル化する有限状態確率オートマトン、第 2 層 (図 1 における 2 段目の点線矩形) は内部状態空間を共有する複数の線形システム $D = \{D_1, \dots, D_N\}$ である。これら 2 層を統合するために「時区間」(図 1 の中央) を導入する。離散状態の遷移では、各離散状態の順序のみが決まるため、各離散状態が活性化される物理的時間長が必要となる。そこで、1 つの時区間 (以後、単に区間と呼ぶ) には、属性としてオートマトンの離散状態 q_i と、 q_i が持続する物理的な時間長 τ を持たせる。このとき、離散状態 q_i を線形システム D_i と対応付けることで、オートマトンが、内部状態空間におけるダイナミクスの切り替わりを制御できるようになる。以下では、このような HDS を時区間ハイブリッド (ダイナミカル) システム (Interval-based HDS, IHDS) と呼ぶ。

* 京都大学大学院 情報学研究科

Key Words: hybrid dynamical system, timing, segmentation, facial expression recognition, motion generation



第 1 図 時区間ハイブリッドシステム

(1) IHDS による信号の生成

IHDS は確率的生成モデルであり、いったん 3. 章で述べるアルゴリズムで学習されると、ベクトル系列を生成することが可能となる。まず、第 1 層のオートマトンが各離散状態を、状態遷移確率に基づいて順に活性化する (図 1 中段)。次に、活性化された離散状態に対応する線形システムが、そのダイナミクスによって内部状態を遷移させ、これによって観測空間における信号を生成できる (図 1 下段, 2.2 節)。このとき、各離散状態の持続長が対で生成されるようオートマトンが拡張されているため、物理時間軸上では区間系列が得られる (2.3 節)。

(2) IHDS による信号の分節化

観測系列 (信号そのもの、もしくは特徴ベクトル系列) が与えられると、IHDS は、信号のどの期間でどの線形システムを活性化させると、元の系列を最もよく表現できるかを、尤度に基づいて計算する。これによって、観測系列は、線形システムの切り替わりによって分節化され、区間系列に変換することができる。

(3) 用語と記号の定義

内部状態と観測ベクトル: すべての線形システムは n 次元ベクトル空間を内部状態空間として共有する。時刻 t の内部状態を、内部状態空間中の一点 $x_t \in \mathbf{R}^n$ 、観測を、観測空間中のベクトル $y_t \in \mathbf{R}^{n_y}$ で表す。

離散状態および線形システム: オートマトンは離散状態集合 $Q = \{q_1, \dots, q_N\}$ を持ち、各離散状態 $q_i \in Q$ が、それぞれ一つの線形システム D_i に対応する。各線形システムは、内部状態の変化をそれぞれ異なるダイナミクスとして表現する。

持続長と区間: 各線形システムが活性化される持続長を $\tau \in \mathcal{T} \triangleq [l_{\min}, l_{\max}]$ とする。離散状態 q_i と持続長 τ の対 (ラベル付き持続長) $\langle q_i, \tau \rangle \in Q \times \mathcal{T}$ を、以下では区間と呼ぶ。

2.2 線形システム

線形システムはガウス・マルコフ過程に従うとし、線形システム D_i の内部状態遷移、および内部状態の観測への写像を、それぞれ

$$x_t = A^{(i)}x_{t-1} + b^{(i)} + \omega_t^{(i)} \quad (1)$$

$$y_t = Cx_t + v_t, \quad (2)$$

のようにモデル化する。ここで、 $A^{(i)}$ は遷移行列、 $b^{(i)}$ はバイアス、 C は観測行列である。 $\omega_t^{(i)}$ と v_t はそれぞれプロセスノイズおよび観測ノイズであり、いずれも平均 0 の多変量ガウス分布でモデル化される。内部状態はすべての線形システムによって共有されているため、線形システム D_i は $A^{(i)}$ 、 $b^{(i)}$ 、およびノイズ $\omega_t^{(i)}$ の共分散行列をパラメタとして持つことになる。

2.3 区間に基づく離散状態遷移

2.1 節で述べたように、確率オートマトンでは、離散状態遷移を、以下のように区間に基づいて拡張する。まず、生成される区間系列 I_1, \dots, I_K に単純マルコフ性を仮定する (添字 k は区間の生成順序を表し物理時間 t とは独立である)。さらに単純化のために、区間同士は互いにギャップやオーバーラップがないものとする。このとき、区間 $\langle q_j, \tau \rangle$ が区間 $\langle q_i, \tau_p \rangle$ の後に起こる確率

$$P(I_k = \langle q_j, \tau \rangle | I_{k-1} = \langle q_i, \tau_p \rangle), \quad (3)$$

を区間遷移確率と呼ぶ。ただし実際には学習時のパラメタ削減のため、離散状態遷移確率 $P(q_j | q_i)$ と、持続長分布 $P(\tau | \tau_p, q_i, q_j)$ の関数に分離してモデル化する。

この区間遷移確率に基づいて区間を順に決定していくことで、区間系列を生成できる。一方、観測信号が得られた際に、2.2 節の線形システムと本節の区間遷移確率に基づいて尤度計算をし、観測信号 (ベクトル系列) を分節化できるが、具体的方法については紙面の都合から文献 [6] に委ねる。

3. 時区間ハイブリッドシステムの学習法

3.1 学習の困難さ

観測ベクトル系列 (もしくは系列の集合) だけが学習データとして与えられたときに、ここから前章で述べた IHDS をどのように同定することができるだろうか。まず、線形システムの個数が既知である場合を考える。このとき、IHDS における各線形システムのパラメタを推定するには、与えられたベクトル系列を、異なる線形システムで表現されるべき区間に分節化しておく必要がある。一方で、この分節化を正しく行うには、パラメタが与えられた線形システムの集合が必要となり、「卵と鶏」

の問題となる。

このような問題を解くには, expectation-maximization (EM) アルゴリズムがうまく働くことが知られているが, EM アルゴリズムは初期値依存性が強く, 特にモデルが複雑になった場合には, 最適解に近い初期値を与える必要がある。さらに, 一般には線形システムの個数(モード数)をあらかじめ決めることが困難な場合が多い。そのため多くの既存手法では, 線形システムの個数や, およそのパラメタ(初期値)を手動で定めることで学習を単純化しており, 実問題への適用は限られていた。以下では, これらの学習時の問題を解決するために, EM アルゴリズムの前段に線形システムのクラスタリングを行う, 二段階学習法 [6] について述べる。

3.2 二段階学習法

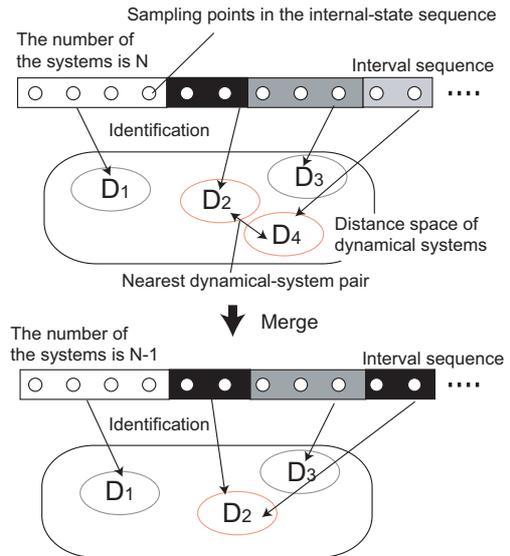
学習データとして与えられた観測ベクトル系列は, あらかじめ内部状態系列に変換されていることを仮定し, これを以下では学習系列と呼ぶ。内部状態系列と式 (2) の観測行列 C などのパラメタを同時に推定する方法としては, 部分空間同定法などを利用できるが, 後述の応用事例では, 多変量自己回帰モデルを用いて, $x_t = [y_{t-p+1}^T, \dots, y_t^T]^T, C = [O, \dots, O|I]$ のように観測系列と内部状態とを直接対応させている。

第一段階: 線形システムの階層的クラスタリング

与えられた学習系列を表現するのに必要な, 線形システムの数と, それぞれの大まかなパラメタを, 線形システムのクラスタリング(図2)によって推定する。

- (1) 学習系列を, 適当な基準で比較的短い区間に分節化し, 分節化された各区間で, それぞれ線形システム(ここでは式 (1))を同定する(初期分節化)。
- (2) すべての線形システム対での距離を, Kullback-Leibler (KL) divergence に基づいて計算する。線形システム D_i と D_j の距離 $Dist(D_i, D_j)$ を考える場合, KL divergence は D_i と D_j に関して非対称となるため, これらの平均をとるものとする(なお, 後で紹介する応用事例では, KL divergence は近似的に計算している)。
- (3) 最も近い線形システム対を併合する。すなわち, それぞれの線形システムによって表現されていた区間集合から, 1つの線形システムを再同定する。
- (4) (2) と (3) における併合の反復を, 線形システムが1つになる前に適当な基準によって止める。

なお, (4) の併合停止(線形システムの個数決定)には, 情報量基準などの利用が考えられるが, モデルが複雑な場合はその自由度をしばしば適切に評価できない。そこで本手法では, 推定された線形システムによる生成系列と, 実際に与えられた学習系列との誤差を評価し, 線形システムの併合によって誤差が急激に変化する直前の線形システム集合を, アルゴリズムの出力としている(複数の候補がある場合は人手によりいずれかを選ぶ)。



第2図 線形システムの階層的クラスタリング

第二段階: EM アルゴリズムによるパラメタ調整

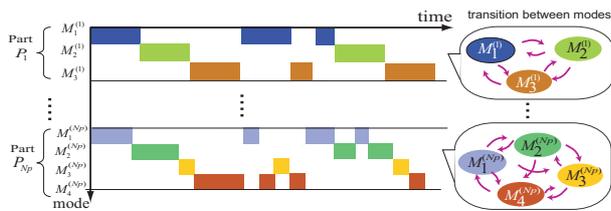
一段階目で得られた線形システムの個数を, この段階では固定し, EM アルゴリズムを適用する。一段階目の線形システムのクラスタリングによって, 各線形システムのパラメタは大まかに推定されている。これにより, EM アルゴリズムの初期値依存性が解決されることになる。EM アルゴリズムの流れは以下のとおりである。ただし, ここでは計算量削減の観点から, 厳密な EM アルゴリズムではなく, 最も尤度の高い分節化だけを考慮するという近似を行っている。

- (1) 一段階目で得られている線形システムのパラメタを用いて, 与えられているすべての学習系列を分節化する。
- (2) 分節化結果を用いて, 線形システムのパラメタを更新するとともに, オートマトンにおける区間遷移確率(離散状態遷移確率および持続長分布のパラメタ)を求める(M-step)。
- (3) 再び, すべての学習系列を分節化する(E-step)。その時点での IHDS の学習系列に対する尤度を計算する。
- (4) (2) と (3) を, (3) における尤度の更新が閾値以下となるまで繰り返す。

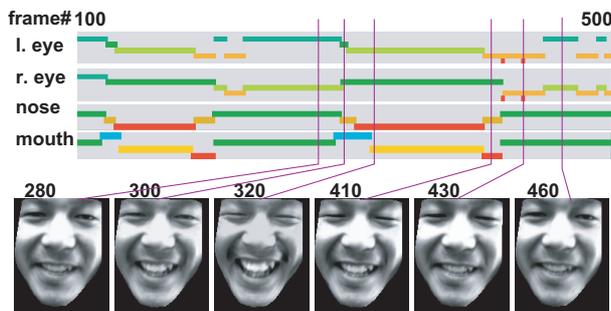
以上により, 線形システムとオートマトンのパラメタが推定される。この学習法は, 実際のメディア信号へ適用することが容易であり, 以降の章では, 各種メディア信号間の時間的構造を解析する応用事例を紹介する。

4. 事例1: タイミングに基づく表情認識

これまでの表情認識・生成研究では, 主として感情に基づく基本的カテゴリ(喜び・驚き・恐怖・怒り・嫌悪・悲しみ・軽蔑)に表情を分類することが検討されてきた。しかし, 実際の表情は, 意図的に制御されて生成される



第3図 表情譜（縦軸：モード，横軸：時間）．各パーツでの線形システムの活性化パターンを表す．



第4図 表情譜から生成された映像例（眉毛動作の譜は省略）

ものや、感情・気分などによって自発的に表出されるものがあり、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よりもずっと粒度の細かな分類を行い、相手の内部状態およびその変化を推定していると考えられる。

従来の表情の記述形式としては、Ekmanらが開発したFacial Action Coding System (FACS) がよく利用されている[7]。しかし、これはAction Unitと呼ばれる顔の各パーツの見えや動きの記述単位を単に組み合わせるものであるため、動きの間の時間的構造を詳細に扱うことができないという問題がある。そこで本研究では、表情は、顔パーツにおける運動のタイミング構造によって生じるものと考え、新たな表情の記述方法を提案した[8]。

4.1 表情譜：時区間に基づく表情記述

顔の各パーツの運動をそれぞれ異なるIHDSによってモデル化する。このとき、図3に示すように、線形システムの力学的分節点によって得られる時区間の集合を、音符と音符のタイミング構造を記述した楽譜になぞらえて「表情譜」と呼ぶことにする。すなわち、ここで音符に相当するものは、目が閉じるといった比較的単純な動き（モード）を表現するような力学系ということになる。

映像から表情譜を得る流れは次のようになる。まず、映像から各顔パーツの動きを、たとえばActive Appearance Model (AAM)[9]などを用いて追跡する。次に、ある顔パーツの特徴点ベクトル系列に対して3.章で述べたIHDSの学習法を適用することで顔の運動をモードに分節化する。これを全ての顔パーツに対して行うことで、たとえば実際に自発的に2回笑った際の顔映像からは、図4上段のような表情譜が得られる。

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、表情譜獲得時の処理を逆にたどることで、線形システムから特徴点ベクトル系列を生成し、さらに、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4下段に、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、たと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。

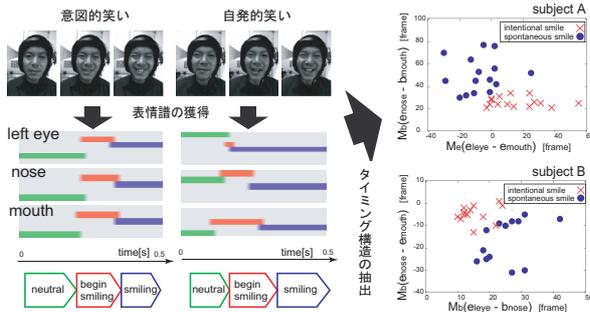
4.2 表情譜による表情認識

では表情譜を用いることで、はじめに述べたような粒度の細かな表情認識は可能になるだろうか。たとえば、笑い顔というものは普段のコミュニケーションの中でも頻繁にみられるが、同じ笑い顔であっても、「意図的な笑い」と「自発的な笑い」といった2クラスに分類できる。そこで、表情表出時に観測される、口・鼻・目といった顔パーツ間のタイミング構造、すなわち運動開始・終了タイミングやその持続時間を解析することで、これら2つの「笑い」を識別することを試みた。

まず認識プロセスに先立って、あらかじめ「意図的な笑い」と「自発的な笑い」の表情データベースを、被験者を撮影することでそれぞれ作成しておく。各表情のデータベースには、1被験者あたり、およそ30から50程度の表情表出の映像が含まれる。その後、各表情映像をそれぞれ表情譜へと変換し、特に笑い開始時と笑い終了時のそれぞれの動きに相当するモード M_b , M_e を、左目・鼻・口のそれぞれのパーツから取り出す(図5(左))。そして、これらモードの開始・終了時刻に注目し、これら分節点間の時間差を軸とするような空間中で、2つの表情の分布をそれぞれ得ることで、各表情のタイミング構造を学習することができる。

このとき、笑い方は人によって大きく異なり、この個人差は識別にとって無視できない。そこで、タイミング構造の軸（分節点間の時間差）のうち、これら2表情を分離しやすい軸をあらかじめ見つけておくことにする。具体的には、各被験者にて、2表情の分布間のマハラノビス汎距離を最大とするような2軸を選んでおく。すると、図5(右)に示すように、2つの表情の分布を分離するのに適した空間を被験者ごとに得ることができる。

表情認識は、新たに得られた表情を表情譜に変換し、このときのタイミング構造が、すでに得られている2つの表情のどちらの分布に近いかをみることで、意図的、自発的な笑いのいずれであったかを識別することができる。識別手法としてはSupport Vector Machineを用い、leave-one-out法によって評価を行った結果、6被験者のいずれでも、「意図的な笑い」では80-100%、「自発的な笑い」では79-96%という高い精度で識別できることが確認された。



第 5 図 タイミング構造の分析による表情の分類

5. 事例 2: 複数メディア間の同期構造

複数のメディア信号における同期構造モデルには、特徴量同士の相関を直接扱うものや、いったん HMM などの状態を考え、異なるメディアの状態間での構造を扱うもの [10] があるが、基本的にはフレームを単位として、同時刻もしくは隣接時刻での共起性を扱う。しかし実際には、異なるメディアの変化パターン間に、これらのモデルでは直接表現しきれない構造が現れる。例えば、音声の破裂音/pa/と母音/a/を比べると、破裂音と唇動作の開始時刻はほぼ同期するのに対し、母音に対しては、唇の動きが若干先行することが多く、その時間差にもばらつきがある。また、ピアノなどの楽器演奏では、実際の音に対して演奏者の手や体の準備的動作が入り、豊かな情報を観客に与えている。そこで、前章での単一（視覚）メディアにおけるタイミング表現を発展させ、映像や音声などのマルチメディア信号において、あるメディアの変化パターンと別のメディアの変化パターンとの間で許容される系統的時間差を確率的に表現できるようにした、新たなタイミング構造モデルを紹介する [11]。

5.1 マルチメディア・タイミング構造モデル

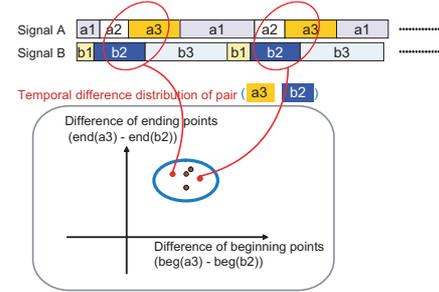
まず、マイクやカメラなどの複数センサを用い、演奏や発話といった動的事象を同時計測することで各種メディア信号を得る。これらメディア信号を、それぞれ異なる IHDS によってモデル化することで、各信号は線形システム（モード）の切り替わりとして表現できる。

次に、それぞれ異なるメディアに現れる 2 つのモードが、どの程度の時間差で開始し、終了するかを、それら 2 つの時間差を軸とする 2 次元空間（図 6）において、実際の観測モード対から分布として統計的に学習する。

各モード対の時間関係をこのような分布としてそれぞれ表現したものを「マルチメディア・タイミング構造モデル」と呼ぶ。このモデルを用いることで、発話状態の高精度な認識や、与えられた音響信号に合った自然な映像（唇の動き）の生成（リップシンク）を実現でき、本稿ではこのうち後者の映像（動作）生成について紹介する。

5.2 音響信号からの映像・動作生成

音響特徴系列（信号 S_a ）からそれと同期した映像特徴系列（信号 S_b ）を生成する流れは以下ようになる。



第 6 図 マルチメディア・タイミング構造モデルの学習

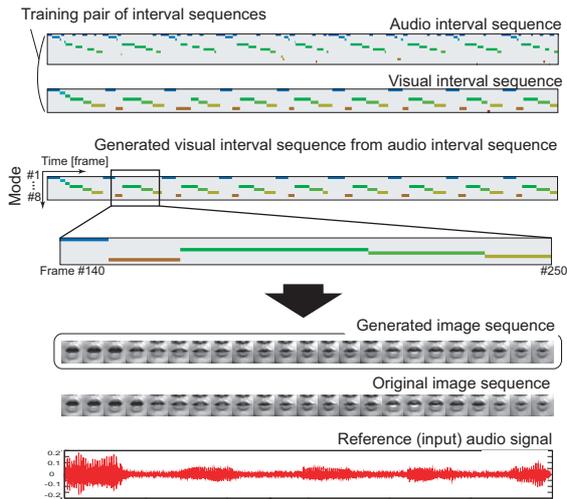
- (1) 入力メディア信号 S_a を区間系列 $\mathcal{I}^{(a)} = [I_1^{(a)}, \dots, I_{K_a}^{(a)}]$ へ分節化する。この分節化には、音響特徴系列であらかじめ学習された IHDS を利用できる。
- (2) メディア信号 S_a の区間系列 $\mathcal{I}^{(a)}$ から別のメディア信号 S_b の区間系列 $\mathcal{I}^{(b)} = [I_1^{(b)}, \dots, I_{K_b}^{(b)}]$ を生成する。この変換は、学習済みのタイミング構造モデルを用いて行うことができる。
- (3) 生成された区間系列 $\mathcal{I}^{(b)}$ からメディア信号 S_b を生成する。この信号生成には、映像特徴系列であらかじめ学習された IHDS を利用できる。

このうちステップ (1), (3) は 2. 章ですでに概略を述べたが、ステップ (2) については、あらかじめ前節で述べた方法によって学習したタイミング構造モデル Φ を用い、メディア信号 S_a の区間系列 $\mathcal{I}^{(a)}$ が与えられたときに、この区間系列とともに生じ得るもう一方のメディアの区間系列 $\mathcal{I}^{(b)}$ のうち、最も高い確率をとるものとして推定でき、以下のような最適化問題として定式化できる。

$$\hat{\mathcal{I}}^{(b)} = \operatorname{argmax}_{\mathcal{I}^{(b)}} P(\mathcal{I}^{(b)} | \mathcal{I}^{(a)}, \Phi) \quad (4)$$

これは動的計画法を用いて解くことができ、 S_a および S_b の長さを T とすれば、具体的には時間範囲 $[1, T]$ において分節化された区間数 K_b 、および各区間の終了時刻 e_k とそのモード $m_k (k=1, \dots, K)$ を決めることができる。

実際に、音声を入力した際に得られる映像が、元の映像とどの程度一致するかを検証するために、まず、/a//i//u//e//o/を 9 回発話した際のマルチメディア・タイミング構造を学習した（学習に用いた区間系列を図 7 の 1, 2 段目に示す）。次に、学習時に用いた音声の区間系列 $\mathcal{I}^{(a)}$ を入力として、映像の区間系列 $\mathcal{I}^{(b)}$ を生成した結果を図 7 の 3 段目に示す。その後、生成された映像の区間系列 $\mathcal{I}^{(b)}$ と、あらかじめ学習された映像の IHDS とを用いて、映像特徴系列（画像集合の主成分系列）を生成した。最後に、主成分分析における固有ベクトルとの線形和を計算し、各フレームの特徴ベクトルを画像化し、映像を生成した。このうち、フレーム 140 から 250 までを、5 フレーム間隔で図 7 の 5 段目に示す。学習に用いた画像系列を 6 段目に示すが、両者の唇の動きはほぼ同期していることが分かる。さらに、同じ時間範囲における音声信号（図 7 の一番下）と比較すると、音声の開始



第7図 音声信号を入力とする唇映像の生成

に先行して唇が動くなど、詳細かつ自然な時間的構造をタイミング構造モデルにより表現できることが分かる。

6. おわりに

人の発話や動きといった動的事象を、ハイブリッドシステムとしてモデル化することで、その力学的分節点を抽出し、これら分節点の間の時間関係(タイミング構造)を利用して、人の表情や発話状態を学習・認識・生成する手法を紹介した。著者らは、これらの研究と並行して、発話応答の間合いやしぐさのタイミングと、意図や興味といった心的状態との関係を解析しており[12]、そこで得られた知見をハイブリッドシステムに基づいてモデル化することで、人と時間感覚を共有するようなインタフェースシステムの実現を目指している。このとき、本解説で紹介した学習や認識だけでなく、認識された人の状況や意図に基づいて、情報システム側の行動(たとえば人への反応や動きかけ)をいかに適切に時間調整するかといったタイミング制御法が現在の課題である。

謝辞 本研究の一部は科学研究費補助金18049046の補助を受けて行った。

参考文献

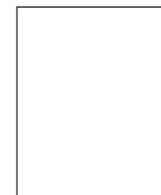
- [1] 川嶋宏彰, 西村拓一. コンピュータビジョンにおける時系列パターン認識. 情報処理学会研究報告(2006-CVIM-154), pp. 197-209, 2006.
- [2] R. Alur, C. Courcoubetis, N. Halbwachs, T.A. Henzinger, P.-H. Ho, X. Nicollin, A. Olivero, J. Sifakis, and S. Yovine. The algorithmic analysis of hybrid systems. *Theoretical Computer Science*, Vol. 138, No. 1, pp. 3-34, 1995.
- [3] M. Ostendorf, V. Digalakis, and O. A. Kimball. From HMMs to segment models: A unified view of stochastic modeling for speech recognition. *IEEE Trans. Speech and Audio Process*, Vol. 4, No. 5, pp. 360-378, 1996.
- [4] C. Bregler. Learning and recognizing human dynam-

ics in video sequences. *Proc. Int. Conference on Computer Vision and Pattern Recognition*, pp. 568-574, 1997.

- [5] Y. Li, T. Wang, and H.-Y. Shum. Motion texture: A two-level statistical model for character motion synthesis. *Proc. SIGGRAPH*, pp. 465-472, 2002.
- [6] H. Kawashima and T. Matsuyama. Multiphase learning for an interval-based hybrid dynamical system. *IEICE Trans. Fundamentals*, Vol. E88-A, No. 11, pp. 3022-3035, 2005.
- [7] P. Ekman and W. V. Friesen. *Unmasking the Face*. Prentice Hall, 1975.
- [8] 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*, Vol. 9, No. 2, pp. 201-211, 2007.
- [9] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance model. *Proc. European Conference on Computer Vision*, pp. 484-498, 1998.
- [10] M. Brand, N. Oliver, and A. Pentland. Coupled hidden Markov models for complex action recognition. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 994-999, 1997.
- [11] 川嶋宏彰, 松山隆司. 時区間ハイブリッドダイナミカルシステムを用いたマルチメディア・タイミング構造のモデル化. *情報処理学会論文誌*, Vol. 48, No. 12, pp. 3680-3691, 2007.
- [12] 松山隆司, 川嶋宏彰, 平山高嗣. 時間と時間感覚に対する感性の情報処理. *電子情報通信学会誌*, Vol. 92, No. 11, pp. 952-954, 2009.

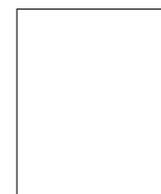
著者略歴

かわしま ひろあき
川嶋 宏彰 (非会員)



2001年京大大学院・情報学・修士課程了。2007年より同大学院講師。博士(情報学)。時系列パターン認識、ハイブリッドシステム、実世界インタラクションの研究に従事。2004年FIT論文賞, 2005年船井ベストペーパー賞, 2007年FITヤングリサーチアワード。電子情報通信学会, ヒューマンインタフェース学会, 情報処理学会, IEEE 会員。

まつやま たかし
松山 隆司 (非会員)



1976年京大大学院・工学・修士課程修了。同大助手, 東北大助教授, 岡山大学教授を経て, 平7京大教授。現在, 同大情報環境機構長, 副理事。画像理解, 3次元ビデオ, ヒューマンコミュニケーション, エネルギーの情報化の研究に従事。1980年情報処理学会創立20周年記念論文賞, 1990年人工知能学会論文賞, 1993年情報処理学会論文賞, 1994年電子情報通信学会論文賞, 1995年第5回国際コンピュータビジョン会議 Marr Prize, 2009年文部科学大臣表彰科学技術賞。国際パターン認識連合, 情報処理学会, 電子情報通信学会フェロー。