2004 年情報論的学習理論ワークショップ
2004 Workshop on Information-Based Induction Sciences (IBIS2004)
Tokyo, Japan, November 8 - 10, 2004.

力学系集合の自己組織化に基づく唇映像の構造化

Finding Structure in Lip Image Sequences based on Self-Organization of Dynamical Systems

川嶋宏彰* 堤公孝* 松山隆司*
Hiroaki KAWASHIMA Kimitaka TSUTSUMI Takashi MATSUYAMA

Abstract: This paper addresses the problem of multivariate time series segmentation by assuming an interval-based temporal structure in the observed time series data. Hybrid systems that consist of continuous and discrete state transition models have been proposed to model complex and continuously changing dynamical events such human motion and utterance. We introduce a new self-organization algorithm of linear dynamical systems to find structural representation of multivariate time series data. Experimental results on lip image sequences show that our algorithm can find an interval-based structure of time series data and can self-organize hidden dynamical systems behind the image sequences.

Keywords: 時系列データ，分節化，唇映像，力学系，自己組織化

1 時系列データに潜む力学系の構造

音声認識に比べ，映像によるジェスチャーなどの視覚的な動きの認識は，まだ十分な性能が得られていないのが現状である．この理由の一つには，音声では自然言語という構造がすでに与えられており，音素や単語集合，さらには単語の関係を決める文法などの言語モデルを，学習時にトップダウン的に導入することができたことがある．一方，映像・画像からの変化情報の認識は，認識対象が手話やジェスチャー，人の歩行動作，表情や唇の動きなど非常に多岐に渡るため，イベント（ここでは時間的な幅を持った変化の情報）のモデル化は人手で行われることが多い．例えば，唇の動きであれば口形素，表情であれば Action Unit などの繰り返し出現する単純な変化の組み合わせで記述されることになるが，この音素に相当するような構成要素（以下，要素イベント）を，認識対象に応じてそれぞれ人手で用意することは，手間がかかり，実際には個人差や文化によって大きな差があるために，認識率を下げる一因になりうる．

したがって，映像などの時系列データにおいては，多くの事例から自動的に構造を見つける手法の方が適していると考えられる．つまり，統計的な学習を経て，扱いたい時系列の要素イベントを，何らかの制約に基づいて自動的に組織化・分節化し，これら要素イベント間の構造を学習することで，複雑で多様な対象を表現できることが期待される．そこで本論文では，映像の変化をいくつかの線形な力学系で表現できる区間に分けられるという仮定を設ける．このとき，個々の力学系は，単純な変化である要素イベントを系のパラメタとして表現することが可能である．すなわち，ここで扱う映像は，背後に複数の力学系をそれぞれモードとしてもち，これらのモードを遷移していくことで，複雑な映像を表現するような生成モデルを仮定する（図 1）．本論文の目的は，大量の映像を入力することで，これら力学系のパラメタと，その個数を同時に効率よく推定するアルゴリズムを提案することである．いったん時系列データを表現する力学系の集合を見つけることで，これらを単位として，正則文法や文脈自由文法で受理されるような時系列データの構造記述が可能となる．

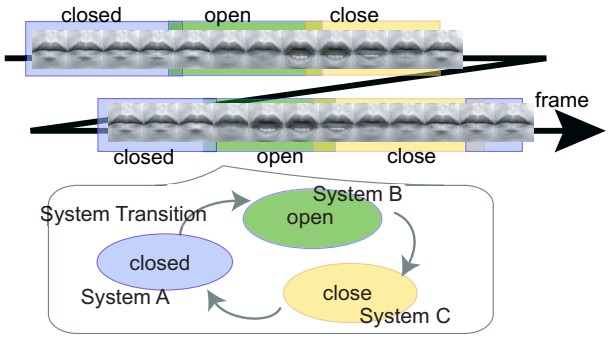

図 1: 複数の力学系による映像の生成モデルの例

本論文では，まず，人の動きの分節化に適した線形シ

*京都大学 情報学研究科, 606-8501 京都市左京区吉田本町
Grad. Sch. of Informatics, Kyoto Univ., Yoshida-Honmachi, Sakyo, Kyoto, 606-8501, tel. 075-753-4768,
e-mail:kawashima@i.kyoto-u.ac.jp, tsutsumi@vision.kuee.kyoto-u.ac.jp, tm@i.kyoto-u.ac.jp

ステムの制約の設け方を，2 節にて提案する．続いて 3 節では，大量の時系列データが与えられたときに，分節化と力学系のパラメタ推定を同時に行うための探索アルゴリズムについて述べる．提案アルゴリズムの評価は，発話時の唇映像を用いて行い，唇の変化がいくつかの力学系の集合によって分節化できることを示す．さらに，より上位の力学系間の遷移モデルを用いて唇映像の生成を行う（4 節参照）．

関連研究: あらかじめ離散化された記号列の分節化手法としては，予測接尾木を構築してオートマトンに変換する手法 [1] や，マルチグラムモデル [2] などが提案されている．一方，本手法では多変量の時系列データを直接扱うことが可能である．

時系列データを複数の線形な力学系で区分的に表現する手法については古くから提案されているが，最近よく用いられるものに，Dynamic Bayesian Network の枠組みで提案された SwitchingLinear Dynamical System(SLDS) などの手法が挙げられる [3, 4, 5]．しかし，SLDS を含むこれらの手法では，実際のデータを扱う上で，モデルとしての問題点とパラメタ推定時の問題点がある．

モデルとしての問題点： SLDS は力学系内の状態遷移だけでなく力学系間の遷移までもが全て物理的時間でモデル化されているため，同じ力学系に留まる時間が短ければ短いほど尤度が高いという不自然なモデルになっている．これに対し我々は，ひとつの要素イベントにとどまる持続時間を明示的にモデル化可能な Hybrid Dynamical System[6] を提案しており，力学系間の遷移モデルの一例として，評価実験において利用する．

パラメタ推定時の問題点： 時系列を表現するための力学系の個数をどのように決定するかが問題となる．さらに，EM アルゴリズムのような局所最適化手法を用いる場合，初期値依存性の問題がより顕著になる．このため，多くの場合は最適解に近い初期値を人手で与える必要があり，実際の応用が少ない理由はこのパラメタ推定の難しさによるところが大きい．これに対し提案手法は，イベントの分節化と必要な力学系の数をボトムアップ的に同時に推定することが可能であり，初期値依存性の問題を解決できる．

2 力学系とその制約付き同定法

力学系とは，あるシステムの状態が一定の法則のもとで時間とともに変化するとき，その変化の法則を表す数学モデルのことをさす．ただし，ここでの力学系とは入力を持たない自律系をさし，ある初期値を与えれば，時間とともに時系列を生成できるようなモデルを考える．はじめに，以下の用語と記法を定義しておく．

観測データ: マイクやカメラでキャプチャーしたデータを特徴抽出することで時系列特徴ベクトル（観測ベクトル）が得られる．観測ベクトルの得られるタイミングは，カメラのサンプリングレートなどに従うものとし，離散時刻 t で得られる観測ベクトルを y_t のようにあらす．観測ベクトルの定義される空間を観測空間と呼ぶ．

力学系とその状態空間: 各要素イベントを表現する力学系はひとつの共通な n 次元実数ベクトル空間 $\mathbf{R}^n$ を状態空間として持つものとする．

力学系集合: 与えられた時系列データは，複数の力学系を切り替えることで表現する．このとき力学系の個数を N とし，力学系集合を $\mathcal{D} = \{D_1, ..., D_N\}$ で定義する．

2.1 線形システム

各区間内はそれぞれ線形な力学系によって表現される．力学系 D_i の状態方程式および観測方程式は次式で表される．

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

$$y_t = H x_t + \upsilon_t \quad (2)$$

ここで x_t は時刻 t における状態ベクトルである．$F^{(i)}$ は遷移行列であり，力学系ごとに異なる．また，H は観測空間と状態空間を結びつける観測行列である．本来は，異なる力学系には異なる状態空間を設計することが可能であり，その場合は観測行列を $H^{(i)}$ のようにそれぞれの力学系について用意する必要がある．しかし，今回は力学系モデルのパラメタを減らすために，全力学系で状態空間を共通と仮定する．このとき，観測行列も全力学系で共通となる．

$\omega^{(i)}$，υ はプロセスノイズおよび観測ノイズである．これらは平均ベクトル $\mathbf{0}$，共分散行列 $Q^{(i)}$ および R の正規分布にそれぞれ従うとする．以上をまとめると，次式の確率密度関数を考えることになる．

$$P(x_t|x_{t-1}, d_t = D_i) = \mathcal{N}(F^{(i)} x_{t-1}, Q^{(i)})$$

$$P(y_t|x_t, d_t = D_i) = \mathcal{N}(H x_t, R)$$

ここで，$d_t = D_i$ は時刻 t でシステム D_i にしたがって状態を遷移させていることを意味する．$\mathcal{N}(a, B)$ は平均 a，共分散 B の多次元ガウス関数を表す．すると，通常のカルマンフィルタと同様にガウス・マルコフ過程を仮定することになるため，$t-1$ まで観測が得られた条件の下で，一期先の状態および観測を推定することができ

る．以下，系列 $y_a, ..., y_b$ を y_a^b のように表す．

$$P(x_t|y_1^{t-1}, d_t = D_i) = \mathcal{N}(x_{t|t-1}^{(i)}, V_{t|t-1}^{(i)})$$
$$P(y_t|y_1^{t-1}, d_t = D_i) = \mathcal{N}(Hx_{t|t-1}^{(i)}, HV_{t|t-1}^{(i)}H^{\mathsf{T}} + R)$$

ここで $x_{t|t-1}^{(i)}$ と $V_{t|t-1}^{(i)}$ はカルマンフィルタの更新式に基づいて更新される．

2.2 動きのクラスと制約の関係

時系列モデルの自由度は，扱いたい時系列に応じて決める必要がある．我々は，ある区切られた時間範囲では，時間的に隣り合う状態間に線形な関係があるという大きな制約を設けた．しかし，たとえば画像中の appearance-base の特徴量を用いる場合，線形なモデルでもなお，非常に幅広い画像の変化（2 次元画像上でのアフィン変換や非線形変換を含む）が可能である [7, 8]．

そこで，より安定に力学系を組織化していくために，線形システムのパラメタにさらに制約を設けることを考える．ここで，設ける制約は求めたい系の特性に合わせて決める必要があり，個々の力学系（連続系）でどのような時間的変化を扱いたいか（どのような変化で分節化したいか）だけでなく，より上位の力学系間の遷移構造（離散系）を含めた設計方針にも依存する．

以下では，特に式 (1) 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．これは,「はじめは早い変化であり，次第に静止していく」ような単調な変化を，分節化・組織化していきたいからである．例えば唇の開閉運動であっても，ひとつの周期性を持った力学系で表現するのではなく,「開く」と「閉じる」は別の力学系で表現し，その間の遷移関係はより上位の離散系で表現することにする．人の生成する動きの多くは，このような単調性を持った滑らかな動きの変化区間を単位として構成されると考えられる．

2.3 制約付き線形システム同定

まずはじめに，制約がない場合の一般的な F の同定方法について述べる．次に，制約を加えた一般化逆行列を導入することによって，F の固有値を 1 より小さな値に設定する手法について述べる．

なお，本来ならば観測行列 H も含めて同定しなければならないが，以下ではいったん部分空間同定法 [9] などで H および状態空間中の状態ベクトル系列が得られていることを仮定する．

制約なしの線形システムの同定：状態系列 $x_1, .., x_T$ から遷移行列 F を計算することを考える．まず，$X_0 = [x_1, ..., x_{T-1}]$，$X_1 = [x_2, ..., x_T]$ と置く．このとき，F の同定は，各時刻における二乗予測誤差を最小にする問題と考えることができる．

$$F^* = \arg\min_F ||FX_0 - X_1||^2 \tag{3}$$

これを行列方程式として微分法を用いて解くことで，

$$F^* = X_1X_0^{\mathsf{T}}(X_0X_0^{\mathsf{T}})^{-1} = X_1X_0^+ \tag{4}$$

となる．X_0^+ は X_0 の一般化逆行列である．ただし，$x_1, ..., x_{T-1}$ は少なくとも次元数以上の独立なベクトルが存在すると仮定する．系列の長さが短いなどの場合は，式 (3) の解は一意に決まらないが，$F^* = X_1X_0^+$ により最小ノルムを与える特殊解が求まる．

制約を加える場合の線形システムの同定：制約を加える場合は，適当な正の実数値 δ を設定し，式 (4) を以下のように変更する．

$$F_\delta^* = X_1X_0^{\mathsf{T}}(X_0X_0^{\mathsf{T}} + \delta^2 I)^{-1} \tag{5}$$

ここで I は n 次元の単位行列である．これは，極限に基づく一般化逆行列の定義 [10] において，$\delta \to 0$ の極限を途中で止めた場合に相当する．

$(X_0X_0^{\mathsf{T}} + \delta^2 I)^{-1}$ は，状態空間において状態系列の分布を主成分分析したときに，各主成分軸を分布が白色化する方向にスケーリングする役割を持つ．δ を大きくするほど，各軸で小さな値にスケーリングされるため，遷移行列 F_δ^* の固有値は全体として小さくなっていく．問題は，F^* の構造がどの程度保たれるかであるが，定性的には，δ に状態系列分布の標準偏差と同程度の値を与えることで，状態系列の分布中で，あまり変化しない共通の軸や，over fitting を起こすような細かすぎる情報を持った軸を抑制しながら，一期先の状態を回帰予測する上で有効な F^* の構造を保つことができる．

2.4 予備実験〜制約付き同定法の有効性

/mamamama/と連続で発話した人の唇を 30fps で撮影し，色相と水平エッジを用いて唇の切り出しと低解像度化を行った．その後，全フレームを用いて KL 変換を行うことで，各フレーム 16 次元の特徴ベクトル系列を得た．なお，観測行列は $H = I$ とした．次に唇が閉じ

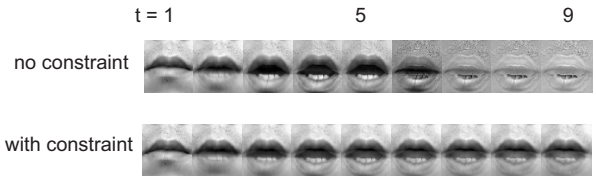

図 2: 制約付き線形システム同定の有効性（唇を閉じた状態から生成された系列．可視化の際の画素値のスケーリングには，全フレームで固定の値を用いている．）

た状態から開く区間を人手で 4 箇所（各 4 から 5 フレーム）切り出し，式 (4) および式 (5) を用いてそれぞれ遷移行列 F を求めた．

制約のない場合およびある場合で，第 1 固有値の絶対値はそれぞれ 2.98, 1.03 であった．次に，唇が閉じた状態を初期状態 x_1 とし，$x_t = Fx_{t-1}$ を $t = 9$ まで繰り返すことによって得られた系列を図 2 に示す．制約のない場合はいくつかの固有値が 1 を越えており，特に学習時の区間長を超えたところから状態のノルムが急激に発散し，元の画像の構造が失われていく．一方，第 1 固有値が 1 に近くなるように制約を加えた場合では，状態の変化速度は比較的緩やかであり，生成された画像では，唇を開いた状態を保ち続けること分かる．これより，2.3 節で述べた制約付き同定が，学習した系列の付近という短い区間だけでなく，系全体を安定にして汎化を促す上でも有効であることが分かる．

ただし，線形システムでは十分長い時間が立てば発散もしくは収束してくため，同じ力学系が持続しうる時間分布をモデル化し推定しておくことで，時間が立てば別の力学系に遷移するような仕組みも別途必要となる [6].

3　力学系集合の自己組織化

時系列データを複数の力学系でモデル化する際には，大量の学習用観測データが与えられたときに，次の 2 つの問題を同時に解く必要がある．

1. 力学系集合の推定
2. 観測データ系列の分節化

いったんどの区間がどの力学系に従うかが分かれば，同じ力学系に従う区間を集めてきてシステムの同定（パラメタ推定）を行うことで，システムのパラメータを求めることができる．しかし，この区間の分節化を行うには，あらかじめ力学系の集合が分かっている必要があるという卵と鶏の問題を抱えている．そこで，本手法では力学系の間に距離を定義することで，ボトムアップ的に力学系を併合し組織化していく手法をとる．

3.1　問題の定式化と制約条件

以下，学習に用いる観測データは 1 つの長い系列であるとして説明を行う．複数の系列から学習する場合は，系列の始点と終点で何らかの分節化が終了していると考えれば，これらを接続した系列を考えても問題なく，1 つの系列における議論が同様に成り立つ．

学習用の観測データ $y_1, ..., y_T$ が与えられたときに，

- 力学系集合: $\mathcal{D} = \{D_1, ..., D_N\}$
- 分節化された区間集合: $\mathcal{I} = \{I_1, ..., I_K\}$

を同時に推定したい．なお，力学系の個数 N や区間の個数 K は未知である．ここで，力学系 D_i はパラメタ $\theta_i = \{F^{(i)}, Q^{(i)}\}$ を属性として持ち，力学系の集合の推定とは必要な力学系の個数と，各力学系のパラメタを推定することを指す．区間 I_k は始点 $s_k \in \{1, ..., T\}$，終点 $e_k \in \{1, ..., T\}$ およびその区間を表現する力学系のラベル $d_k \in \mathcal{D}$ を属性として持つ．(記法の便宜上，力学系の実体とラベルに同じ D_i を使う.)

このとき，全区間についての尤度 $\mathcal{L}$ をできるだけ大きくすると共に，力学系の個数 N ができるだけ少なくなるように $\mathcal{D}$ と $\mathcal{I}$ を推定する．

$$\mathcal{L} = P(\mathcal{I}|\mathcal{D}) = \prod_{k=1}^{K} P(I_k|d_k, \mathcal{D}) = \prod_{k=1}^{K} P(y_{s_k}^{e_k}|d_k, \mathcal{D})$$

区間 I_k で観測された系列 $y_{s_k}^{e_k}$ が同じ力学系 D_i に従うと仮定すると，区間 I_k における尤度 $L_k^{(i)}$ は，2.1 節の式を用いて以下のように計算できる．

$$L_k^{(i)} = P(I_k|d_k = D_i) = \prod_{t=s_k}^{e_k} P(y_t|y_{s_k}^{t-1}, d_t = D_i) \quad (6)$$

なお，各力学系の初期分布は正規分布を仮定する．

分節化とは集合 $\mathcal{I}$ を決定する操作とする．ただし $\mathcal{I}$ の要素の区間は，それぞれが 1 つの力学系で（高い尤度を持つという意味で）表現可能とする．ラベリングとは，各区間がどの力学系で表現されているかというラベルを与える操作とする．最も一般的な場合（区間同士が包含関係にある場合も含む）には，図 3 上段のラティスを構成する全てのノードの組み合わせ（$2^{T(T+1)/2}$ 通り，ラベリングも含めると $N^{T(T+1)/2}$ 通り）という，非常に多くの可能性について探索する必要がある．しかし，唇のような人体の一部の動きに着目した場合，複数の力学系が時間的に重なるのは一時的であると考えられるため，次のような制約を設ける．

- 区間の最小の長さは $l_{\min}$ である．
- 区間の最大の長さは $l_{\max}$ である．
- ある区間が，異なる力学系で表される別の区間に包含されることはない．
- 区間同士の重なりは $l_{\min} - 1$ 時刻許す．

すると，探索範囲は大幅に狭まり，例えば図 3 下段のトレリスにおいてパスを 1 つ決めれば，分節化が一つ決まることになる．

3.2　時系列データの分節化と力学系集合の同時推定アルゴリズム

区間の最小値 $l_{\min}$ を与えることで，固定長 $l_{\min}$ の区間が $T - l_{\min} + 1$ だけ得られたとする．このとき，これ

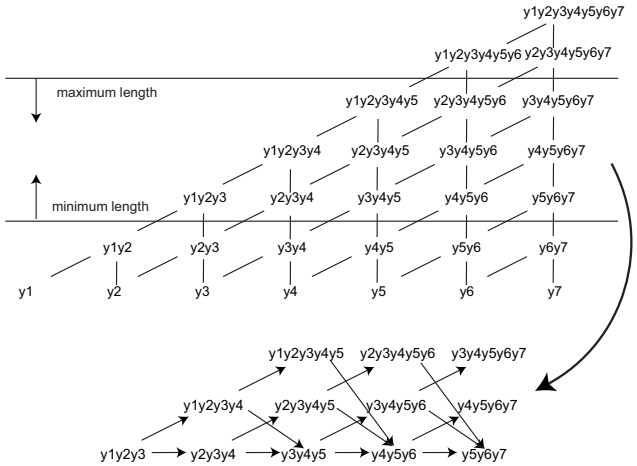

図 3: 分節化の候補となる全ての区間集合のラティス（上段）と，前向き併合のための可能な探索パス（下段）（系列長 $T = 7$, 最短区間長 $l_{\min} = 3$, 最長区間長 $l_{\max} = 5$, 重なり 2 の場合. 1 つのパスを決めると分節化が 1 つ決まる.）

ら区間を併合していく際に，次の 2 つの併合アルゴリズムを定義する．

1. 前向き併合 (forward merge)
2. 最近傍併合 (nearest merge)

前向き併合は時間的な連続性を用いるアルゴリズムであり，図 3 下段のトレリスにて，パスを一つ決めることにより，分節化を行うアルゴリズムである．固定長 $l_{\min}$ で同定した力学系を用いて次の時刻の固定長区間を予測し，尤度が閾値 $\mathrm{Th}_{\mathrm{merge}}$ を下回らないならば区間を延ばしていく．オンライン学習への拡張が容易，計算量が少ないなどの長所がある反面，あらかじめ閾値をうまく設定しておかなくてはならず，望ましい区間を得るためには，閾値を少しずつ変えながら何度も試行を繰り返す必要がある．

一方，最近傍併合は力学系としての近接性を用いるアルゴリズムであり，あらかじめ力学系の間に距離を定義しておき，最も近い 2 つの力学系を 1 つの力学系に順に併合していく．これと同時に，それぞれの力学系によって表現されていた区間集合も併合される．これはモデルベースの階層型クラスタリング [11] を，力学系という時系列モデルへ拡張したものと考えることもできる．オフライン学習を前提としたものであり，あらかじめ全ての区間について力学系のパラメタを推定しておかなくてはならないため，力学系の個数に応じて計算量が二乗のオーダで大きくなる．一方で，ボトムアップ的に併合を行うため，力学系間や力学系内の尤度やエントロピーなどが，力学系の数に対してどのように変化するかを利用することで，力学系の個数を決定することが可能である．

本論文で提案する併合アルゴリズムは，この前向き併合と最近傍併合を段階的に用いて，両者の短所を互いに補うものである（図 4）．つまり，はじめに，前向き併合を用いて多くの区間が得られるようなおおまかな閾値で分節化を行っておき，次に，得られた力学系の集合に対して最近傍併合を適用してその個数を減らしていく．これにより，前段の前向き併合では閾値を厳密に調整必要がなく，かつ後段の最近傍併合では計算量を大幅に減らすことができる．さらに，ボトムアップ的な併合時にその個数を推定することも可能となる（3.4 節参照）．

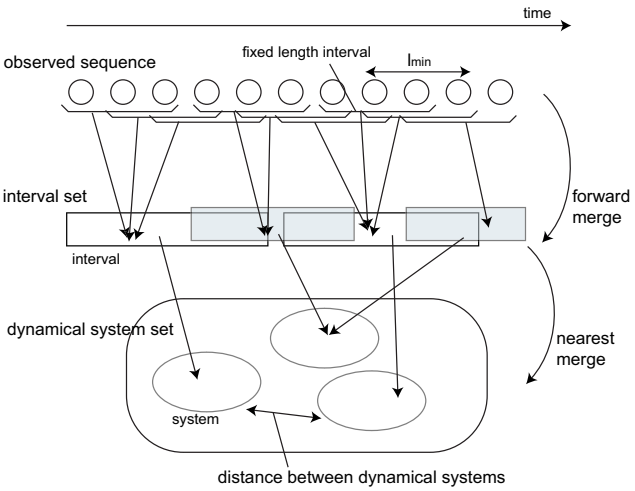

図 4: 観測データ系列から区間へ（前向き併合），区間から動的システムへ（最近傍併合）の 2 段階併合アルゴリズム

以下，前向きおよび最近傍併合アルゴリズムを示す．

Algorithm 1 前向き併合アルゴリズム

$N \leftarrow 1$;　$s_1 \leftarrow 1$;　$e_1 \leftarrow l_{\min}$
$D_1 \leftarrow \mathrm{Identify}(I_1)$
for $t \leftarrow 2$ to $T - l_{\min} + 1$ **do**
　$s_{\mathrm{next}} \leftarrow t$;　$e_{\mathrm{next}} \leftarrow t + l_{\min} - 1$
　$L \leftarrow \mathrm{CalcLikelihood}(D_N, I_{\mathrm{next}})$
　if $L < \mathrm{Th}_{\mathrm{merge}}$ or $e_N - s_N + 1 \geq l_{\max}$ **then**
　　$N \leftarrow N + 1$;　$s_N \leftarrow s_{\mathrm{next}}$;　$e_N \leftarrow e_{\mathrm{next}}$
　　$D_N \leftarrow \mathrm{Identify}(I_N)$
　else
　　$I_N \leftarrow \mathrm{MergeIntervals}(I_N, I_{\mathrm{next}})$
　　if $L > \mathrm{Th}_{\mathrm{update}}$ **then**
　　　$D_N \leftarrow \mathrm{OnlineUpdate}(D_N, I_{\mathrm{next}})$
　　end if
　end if
end for

Identify は 2.3 節で述べたシステム同定法を表し，区間内にある観測データを用いて，システムパラメタ $\theta_N = \{F^N, Q^N\}$ を同定する．CalcLikelihood は 式 (6) の尤度計算を表し，現在の力学系 D_N で 1 期先の固定長区間 I_{next} を推定する．なお，観測データ内のモデル誤差

の分布を考慮しない場合は，単に予測誤差などを利用することもできる．MergeIntervals は 2 つの区間を併合する処理である．前向き併合の場合は $e_N \leftarrow e_{\rm next}$ となり，元の区間が後方に延びていく．尤度 L が閾値 $\mathrm{Th}_{\rm update}$ を超えた場合は，区間 $I_{\rm next}$ の観測データが有効であるとして OnlineUpdate を行う．これは力学系のパラメタの逐次更新であり，Greville の定理 [12] 等によって実現される．前向き併合アルゴリズムの終了時では，全ての区間が別の力学系で表現されているため，区間数 K は N に一致する．

Algorithm 2 最近傍併合アルゴリズム

```
for i ← 1 to N do
  D_i ← Identify(I_i)
end for
for all pair(D_i, D_j) where D_i, D_j ∈ 𝒟 do
  Dist(i, j) ← CalcDistance(D_i, D_j)
end for
while N ≥ 2 do
  (i*, j*) ← arg min_(i,j) Dist(i, j)
  ℐ_i* ← MergeIntervals(ℐ_i*, ℐ_j*)
  D_i* ← Identify(ℐ_i*)
  erase D*_j from 𝒟
  N ← N − 1
  for all pair(D*_i, D_j) where D_j ∈ 𝒟 do
    Dist(i*, j) ← CalcDistance(D_i*, D_j)
  end for
end while
```

最近傍併合では，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，同じ力学系で表現されることがある．そこで，力学系 D_i によって表現される区間の集合を $\mathcal{I}_i$ としている．CalcDistance は，力学系間の距離を求める処理であり次節で定義する．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力学系のパラメタを再同定する．

3.3 力学系間の距離

力学系の間の距離尺度としては，(a) パラメタの直接比較，(b) いったん併合した際の尤度の減少率 [13]，および (c) 分布間距離に基づく定義 [14] などが挙げられる．

線形システムでは，2 節の制約を加えてもなおパラメタの自由度が大きく，特に本手法のようなボトムアップに併合を行う場合では，その初期段階で汎化が十分行われない（学習時系列データが系全体ではなく局所的に表現されている）ことがある．したがって，(a) のようにパラメタを直接比較する評価は望ましくない．(b) は理想的な条件ではうまく機能するが，すべての併合の可能性について尤度計算を行う必要があり，非常に計算量が多くなる．経験的には (c) に基づく距離定義の方が計算量が少なく，かつ安定に力学系を形成できる．したがって，ここでは分布間距離のひとつである，Kullback-Leibler (KL) divergence を距離尺度として用いる．

$$KL(D_i||D_j) = \sum_{I_k} P(I_k|D_i) \log \left(\frac{P(I_k|D_i)}{P(I_k|D_j)} \right)^{\frac{1}{|I_k|}} \quad (7)$$

$$\sim \frac{1}{|\mathcal{I}_i|} \sum_{I_k \in \mathcal{I}_i} \{\log P(I_k|D_i) - \log P(I_k|D_j)\} \quad (8)$$

ここで，$|I_k|$ は I_k の区間長 $e_k - s_k + 1$ であり，これによって時間的な正規化を行う．$|\mathcal{I}_i|$ は区間集合 $\mathcal{I}_i$ に含まれる区間の区間長の総和 $\sum_{I_k \in \mathcal{I}_i} |I_k|$ である．式 (7) の総和の直後における I_k の条件付き生起確率は $P(I_k|D_i) \sim |I_k|/|\mathcal{I}_i|$ と近似した．式 (8) は力学系 D_i と D_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．

$$Dist(D_i, D_j) = \{KL(D_i||D_j) + KL(D_j||D_i)\}/2 \quad (9)$$

3.4 力学系の個数の決定基準

一般に力学系の個数が増えれば増えるほど，モデル化の精度は上がる．しかし，それに伴って計算のコストが上がるだけでなく，over-fitting が起こる可能性があるため，力学系の個数をある程度小さくする必要がある．よく用いられる基準として MDL などがあるが，一般的な基準のみで完全に自動で決定するのは困難であり，目的によってはあらかじめ一定の範囲内にしぼり込まれている場合も多い．そこで，人手で力学系の個数を決める範囲を与えておき，その範囲内で，比較的近い力学系は併合されているが，次に N を減らすと，離れた力学系も併合されるような N を，力学系間の距離の変化が極小となる N として取り出すことにする．

前節で述べた手法では，最近傍併合を繰り返すことで，最終的にはひとつの力学系が得られるが，その併合の過程で力学系間距離の重み付け和 $\mathcal{H}(N)$ を計算しておく．

$$\mathcal{H}(N) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} P(D_i)P(D_j)Dist(D_i, D_j) \quad (10)$$

ここで，力学系の事前確率として評価実験では $P(D_i) = |\mathcal{I}_i| / \sum_i |\mathcal{I}_i|$ を用いた．$\mathcal{H}(N)$ は N に関して単調増加するため，その差分が，与えられた範囲 $\mathcal{N}$ 内で極小となる N を候補とし，目的に応じて半手動で N を決定する．もしくは単純に最小値を取る N とする．

$$N^* = \arg\min_{N \in \mathcal{N}} \{\mathcal{H}(N+1) - \mathcal{H}(N)\} \quad (11)$$

4　評価実験

4.1　唇映像の分節化と力学系の組織化

繰り返し要素イベントが出現する単純な例として/matsui/を 3 回連続で発話した人の唇を 60fps で撮影し，2.4 節と同様の手法で各フレーム 8 次元，長さ 341 の特徴系列を得た．なお，予備実験と同様に観測行列 $H = I$ とした．得られた特徴系列を図 5 の上段に示す．

まず，前向き併合アルゴリズムによって 41 の区間および力学系が得られた．次に，最近傍併合アルゴリズムを用いて，力学系同士の距離が近い区間を併合した．この様子を図 5 の中段に示す．横軸は時間軸であり，異なる色は異なる力学系によって表現される区間である．このときの力学系数決定のための評価値の変化を図 6 に示す．グラフより，$\mathcal{N} = \{2, ..., 10\}$ では $N = 3, 5$ のとき極小となる．参考までに，力学系間の距離関係を多次元尺度法を用いて可視化したものを図 7 に示す．

$N = 5$ のときの各区間内における変化の様子を，図 5 下段に示す．これより，代表的なものとしては「閉じたまま静止」,「閉じた状態から開く」,「開いた状態から口をすぼめる」,「すぼめた状態から横に広げる」という 4 つの要素イベントが確認できる．このように，線形な力学系で表現できる範囲をひとつの区間として分節化することで，唇の動きを比較的人間の直感に近い力学系として組織化することが可能である．

4.2　力学系間の遷移モデルの導入

力学系の数およびパラメタが決まれば，力学系間の遷移モデルを導入して，より複雑なモデルを設計していくことができる．その一例として，ここでは力学系をひとつの離散状態と考え，これら状態間を遷移していく Hybrid Dynamical System[6] を用いることにする．なお，区間同士の重なり部分においては，音声における調音結合のように何らかの形で力学系に変調が加わる可能性があるが，今回はそのような過渡的なモデル化は行わないものとする．

まず，前節で組織化されたパラメタを初期値とし，EM アルゴリズムを行うことで，離散状態間の遷移確率と持続時間分布を推定した．この結果得られた離散状態の遷移図を図 8 に示す．今回の実験では繰り返し同じ単語を発話したため，ほぼ周期的な遷移になっていることが分かる．次に，学習されたシステムにおいて自律的に状態を遷移させることで，どのような系列を生成することができるかを調べた．初期値として唇を閉じた状態を与え，その後は一切外部からの入力を与えずに，システムの状態を遷移させた．このとき，離散状態の遷移によって力学系間のマクロな変化が決まり，これと並行して，各区間内では個々の力学系の状態遷移によって観測ベクトル系列が生成された．これを元の画像空間に写像したものを図 9 に示す．図から，学習に用いた唇の動きとほぼ同様の映像が生成されていることが分かる．学習系列に比べ，全体的に画像がぼけたように劣化しているのは，観測ベクトル間に単純マルコフ性を仮定した（観測行列 $H = I$ とした）ため，状態に動的な情報が十分保持されていないことが考えられる．より動的な情報を扱うには，連続する複数時刻の観測ベクトルから状態および観測行列を計算することが必要となる [9].

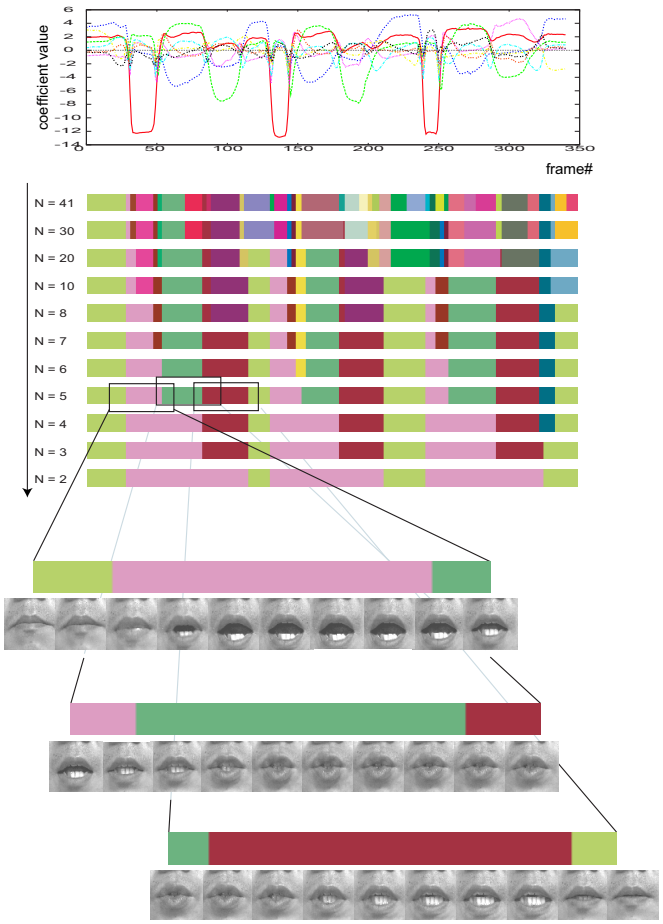

図 5: 力学系が形成される様子（上段は画像の各フレームを KL 変換して得られた特徴系列）

5　結論

複数の力学系の遷移で表現されるような時系列モデルを対象として，大量の学習用時系列データが得られた際に，力学系の集合および時系列の分節化を同時に推定する手法を提案した．提案手法を実際の唇映像を用いて評価を行った結果，学習用の唇映像を与えるだけで，唇の動きの変化を複数の力学系の遷移として組織化できることを確認した．

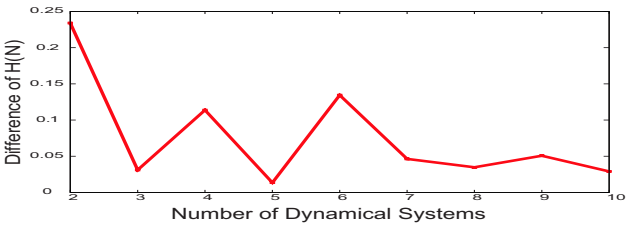

図 6: 力学系数決定のための評価値 $\mathcal{H}(N+1) - \mathcal{H}(N)$

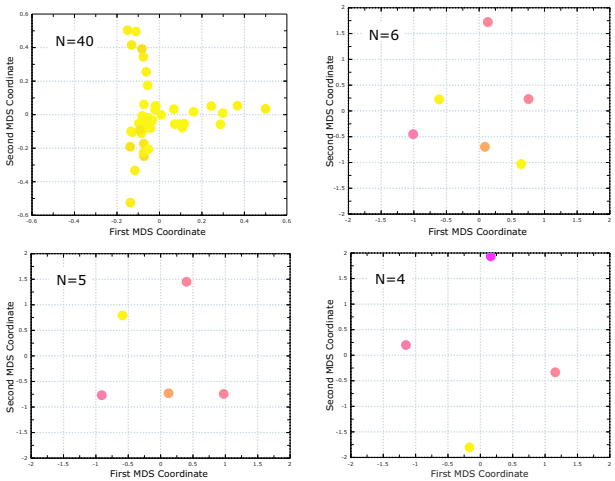

図 7: 力学系の距離関係を多次元尺度法を用いて 2 次元平面上にプロットしたもの．$N = 40, 6, 5, 4$ のとき．(学習時系列中で占める割合が高いほど明度を暗くしている)．$N = 6$ では一部の力学系が近距離にあるが $N = 5$ では互いに離れている．

本論文では，主に唇映像などの人の動きを対象としているため，力学系としては単純な線形システムを用いた．センサの入力データから適切な特徴量を抽出することで，区分的に線形システムで表現可能な時系列データが与えられれば，画像以外の時系列データであっても分節化や力学系の組織化が可能であると考えられる．しかし，一般の時系列データ（例えば音声の子音）には非線形な力学系によって表現する方が望ましいものも多い．また，力学系のみで複雑な身体動作の構造を学習しようというアプローチもある [15]．どのような自由度を持つ力学系を用い，さらに上位の離散的な状態遷移にどのような構造を入れるかは，モデル化したい対象の物理特性や設計方針に合わせて選択していく必要があり，今後より詳細な検討が必要である．

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] D. Ron, Y. Singer, and N. Tishby. The power of amnesia: Learning probabilistic automata with variable memory length. *Machine Learning*, Vol. 25, , 1996.

[2] S. Deligne and F. Bimbot. Inference of variable-length linguistic and acoustic units by multigrams. *Speech Communication*, Vol. 23, pp. 223–241, 1997.

[3] Z. Ghahramani and G. E. Hinton. Switching state-space models. *Technical Report CRG-TR-96-3, Dept. of Computer Science, University of Toronto*, 1996.

[4] V. Pavlovic, J. M. Rehg, and J. MacCormick. Learning switching linear models of human motion. *Proc. of Neural Information Processing Systems*, 2000.

[5] B. N. A. Blake, M. Isard, and J. Rittscher. Learning and classification of complex dynamics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 9, pp. 1016–1034, 2000.

[6] 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.

[7] R. Rao. Dynamic appearance-based recognition. In *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 540–546, 1997.

[8] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto. Dynamic textures. *International Journal of Computer Vision*, Vol. 51, No. 2, pp. 91–109, 2003.

[9] P. V. Overschee and B. D. Moor. A unifying theorem for three subspace system identification algorithms. *Automata*, 1994.

[10] A. Albert. *Regression and The Moore-Penrose Pseudoinverse*. Academic Press, 1972.

[11] S. Zhong and J. Ghosh. A unified framework for model-based clustering. *Journal of Machine Learning Research*, Vol. 4, No. 11, pp. 1001–1037, 2003.

[12] A. Ben-Israel and T. N. E. Greville. *Generalized Inverses : Theory and Applications*. Springer, 2nd edition, 2003.

[13] T. Brants. Estimating hmm topologies. *Logic and Computation*, 1995.

[14] B. H. Juang and L. R. Rabiner. A probabilistic distance measure for hidden markov models. *AT & T Technical Journal*, Vol. 64, No. 2, pp. 391–408, 1985.

[15] M. Okada, K. Tatani, and Y. Nakamura. Polynomial design of the nonlinear dynamics for the brain-like information processing of whole body motion. *Proc. of International Conference on Robotics and Automation*, 2002.

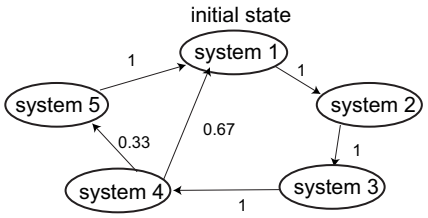

図 8: 学習された力学系間の遷移構造

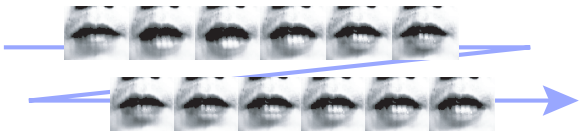

図 9: 学習されたシステムにより生成された唇映像