

時区間に基づく動的イベントの記述と学習

川嶋宏彰 平山高嗣 松山隆司

<http://vision.kuee.kyoto-u.ac.jp/~hiroaki/research/>

マルチメディアイベントの認識

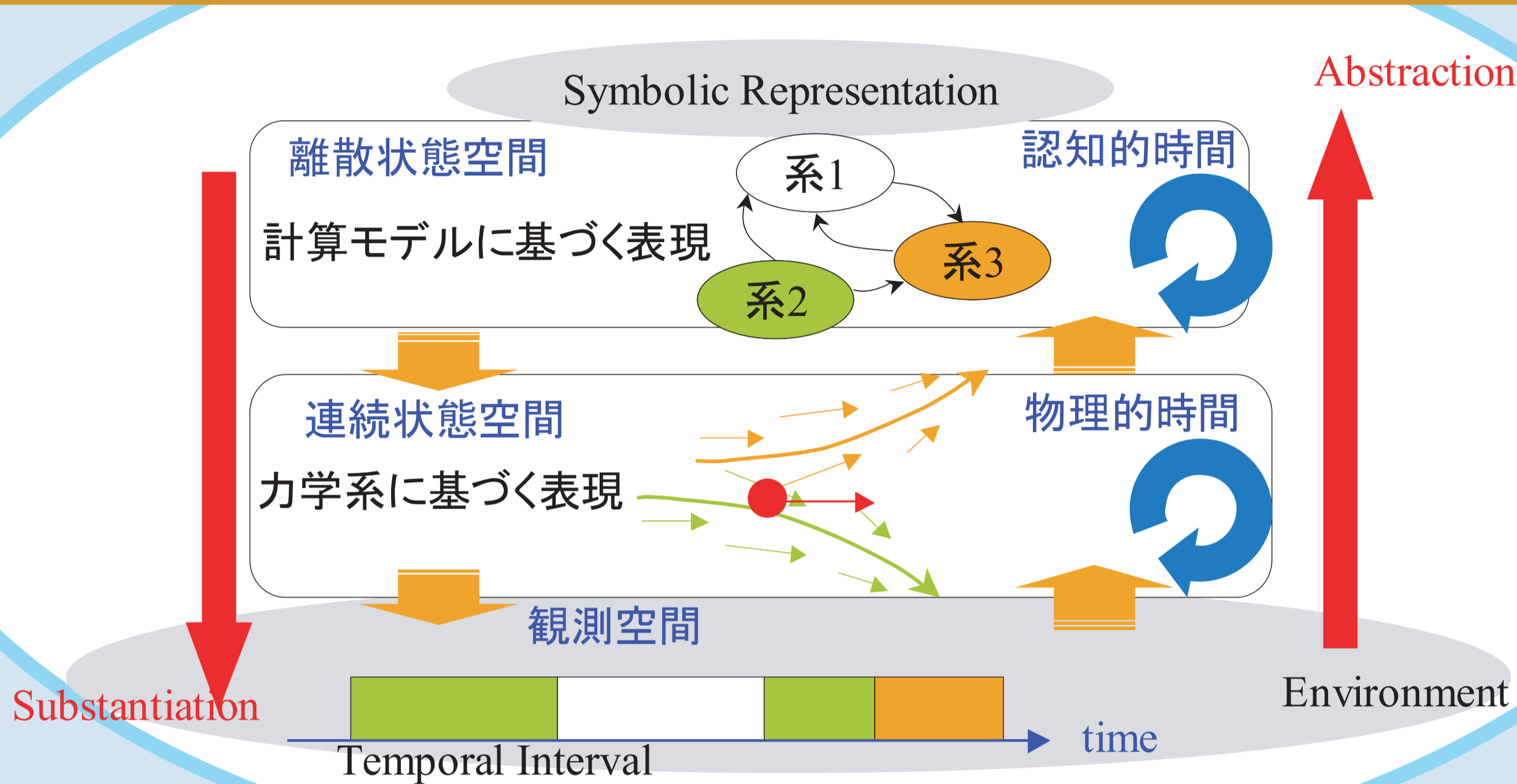
1. 人のしぐさや表情・発話, ジェスチャーなどの動的なイベントを理解する
2. 人間にとって自然な情報提示・応答のタイミングを抽出する

解決すべき課題

モダリティ間の時間関係 (発話と唇の動きがどの程度ずれてもよいかなど) や息の合った, 間合いの取れた状態をいかに表現するか

人間が自然に感じるような「タイミング・間」を記述学習する枠組みが必要

自律ダイナミクスに基づくイベントの区間記述



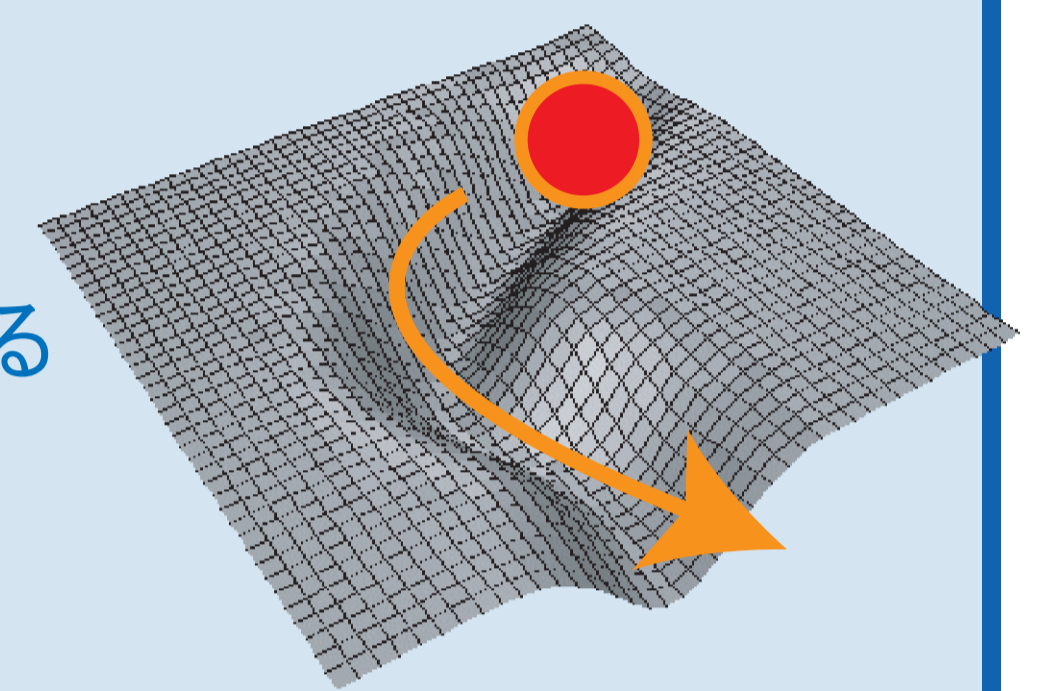
1. 「同じダイナミクスに従う時間範囲」は同じ区間として定義する

2. 動的なイベントを区間の時間的な前後関係や包含関係として記述する

ひとつのダイナミクスは単純な変化のパターンを状態遷移によって生成・予測できる

$$\text{モデル: 線形システム } x_{t+1} = Fx_t + w_t$$

$$z_t = Hx_t + v_t$$



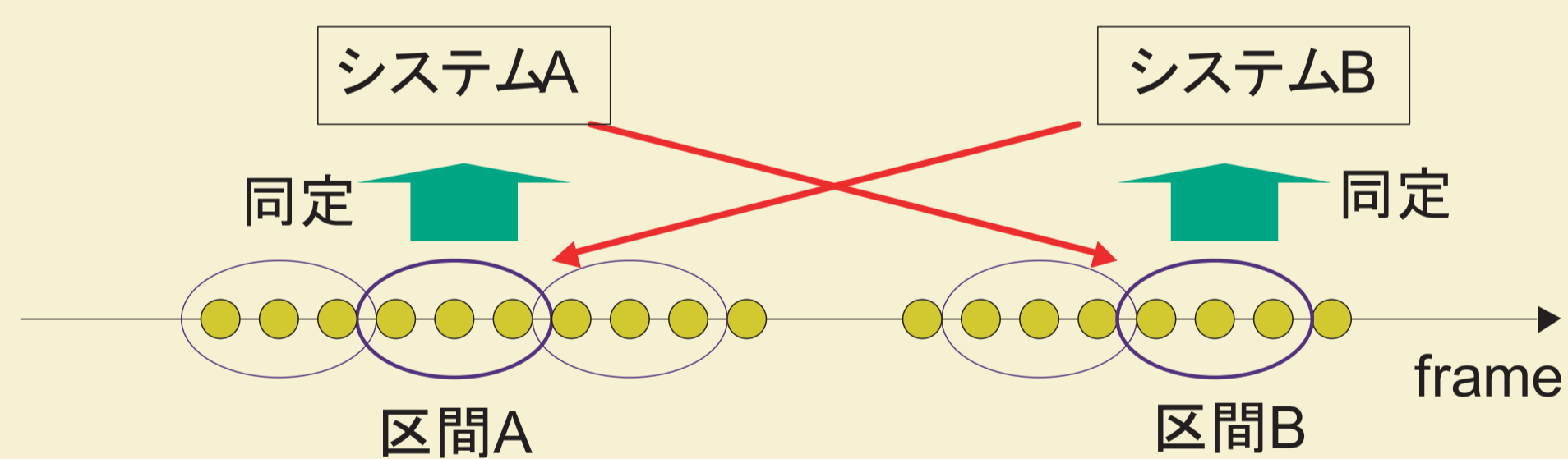
・複雑な変化を単純な変化の組み合わせで表現可能

・区間同士の関係によって時間幅やタイミングの学習が可能

1. ダイナミクスの自己組織化 観測される時系列データに内在するダイナミクスを見つけ出す

(i) 観測データを細かな時間で分割して多くのダイナミクスを求める

(ii) より近いダイナミクスを持つ区間を順に併合していく

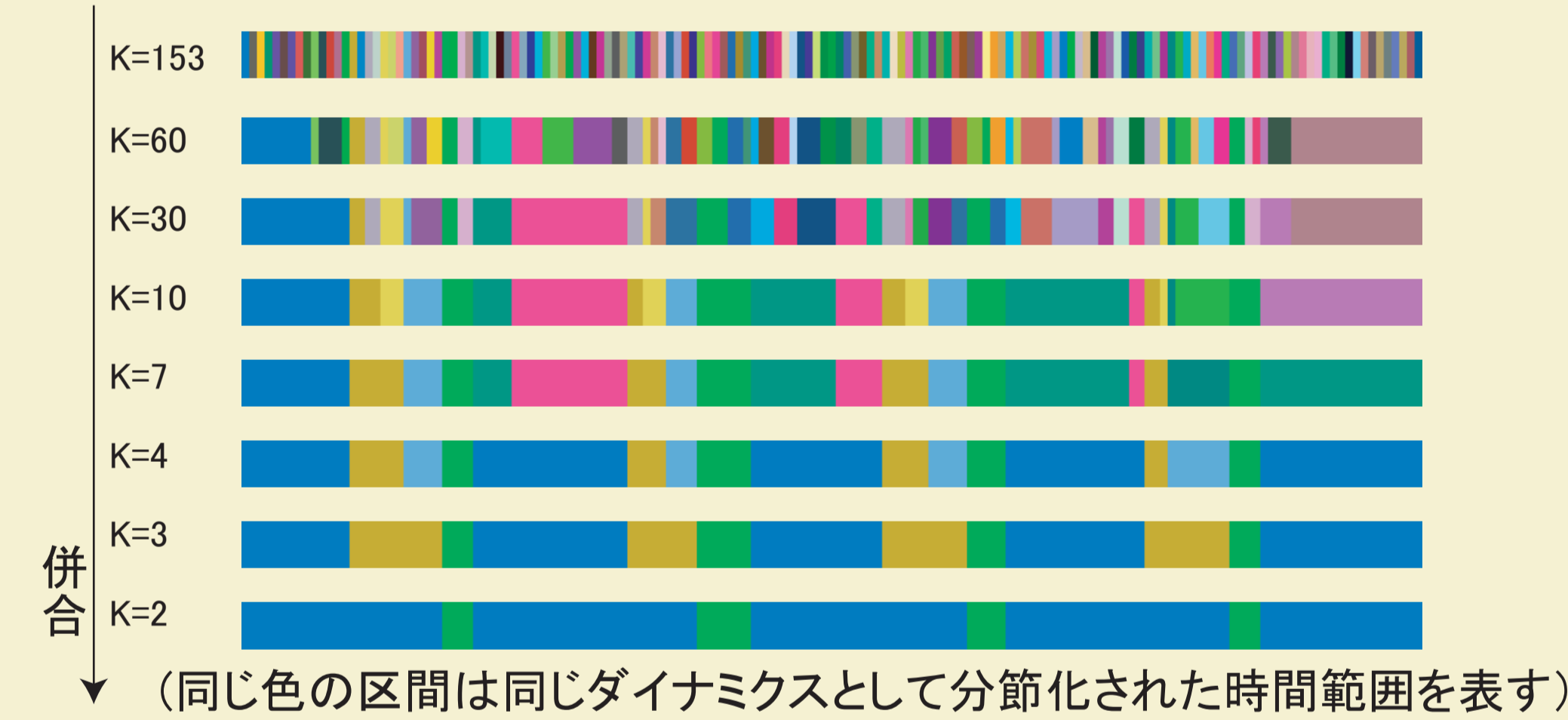


(iii) 併合した区間でダイナミクスを再度同定

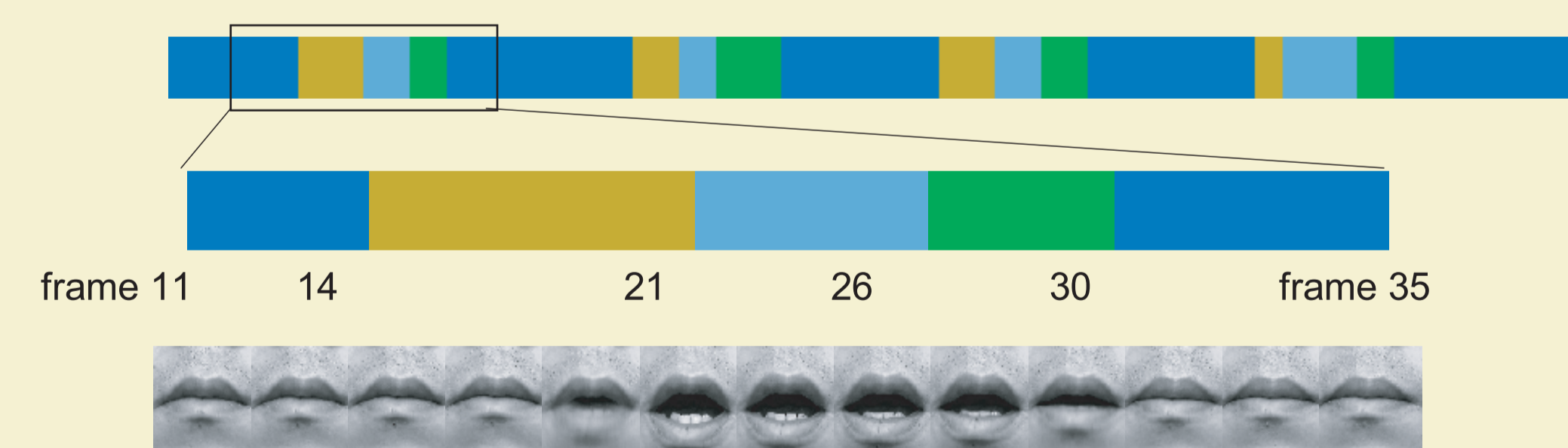
(iv) 区間内の予測誤差が大きく変化する時点で併合を停止

唇の動きのダイナミクスが組織化される様子

入力: 発話/mamamama/ の唇映像

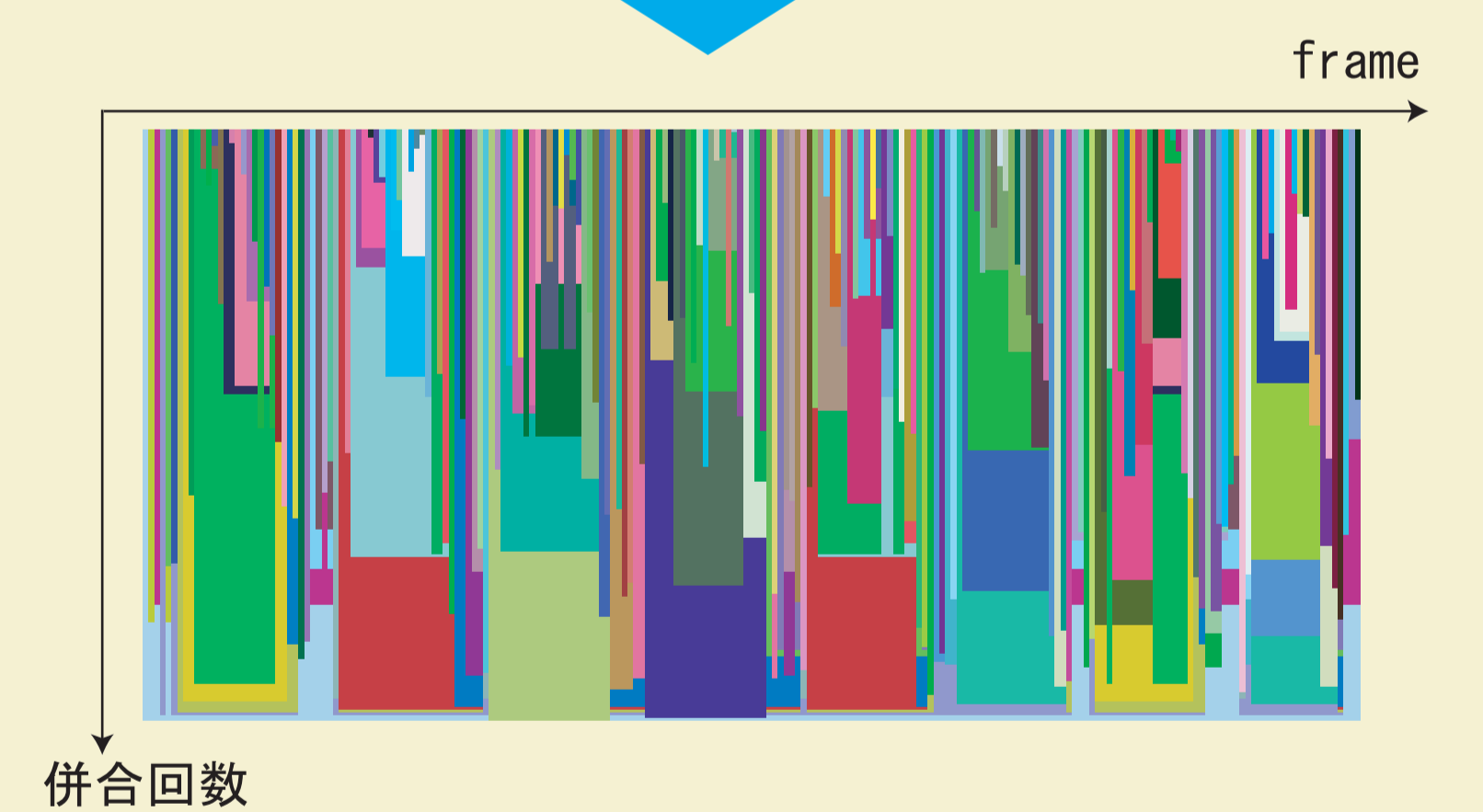
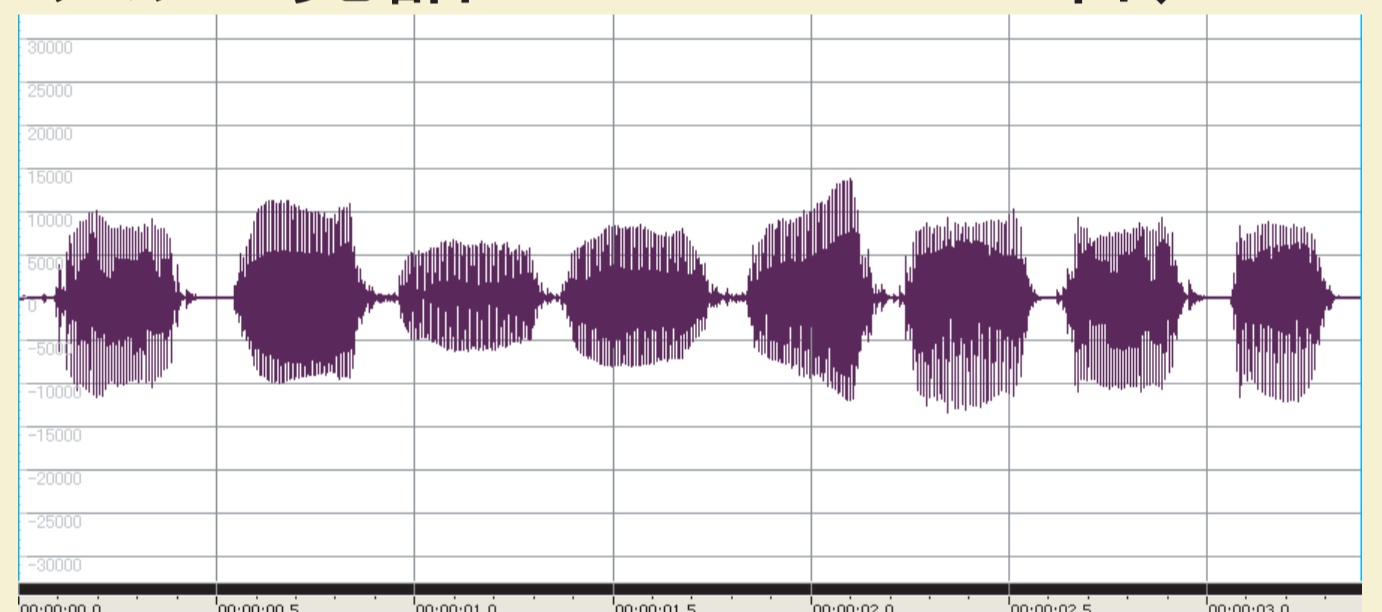


組織化された4つのダイナミクス



音声発話の組織化

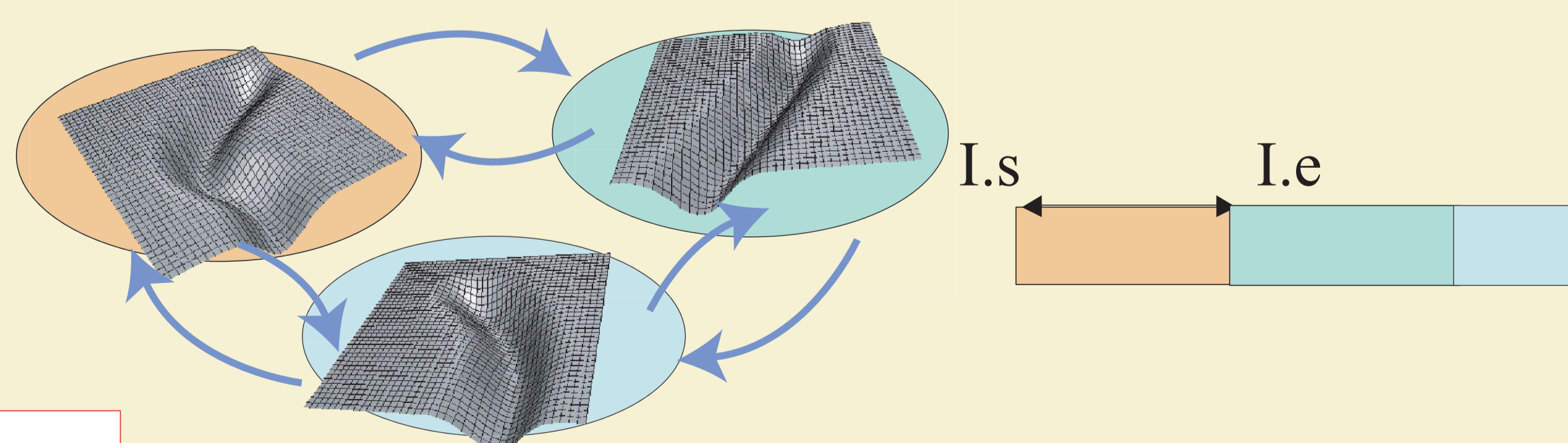
入力: 発話/aeiueoao/ の音声



2. ダイナミクス間遷移の区間記述

分節化された区間から区間への遷移則を学習

学習するパラメータ (a) 同じ区間に留まる持続時間分布 (b) 区間から区間への遷移確率



E step

現在のパラメータに基づき, 学習時系列を区間に分割しラベルを振る

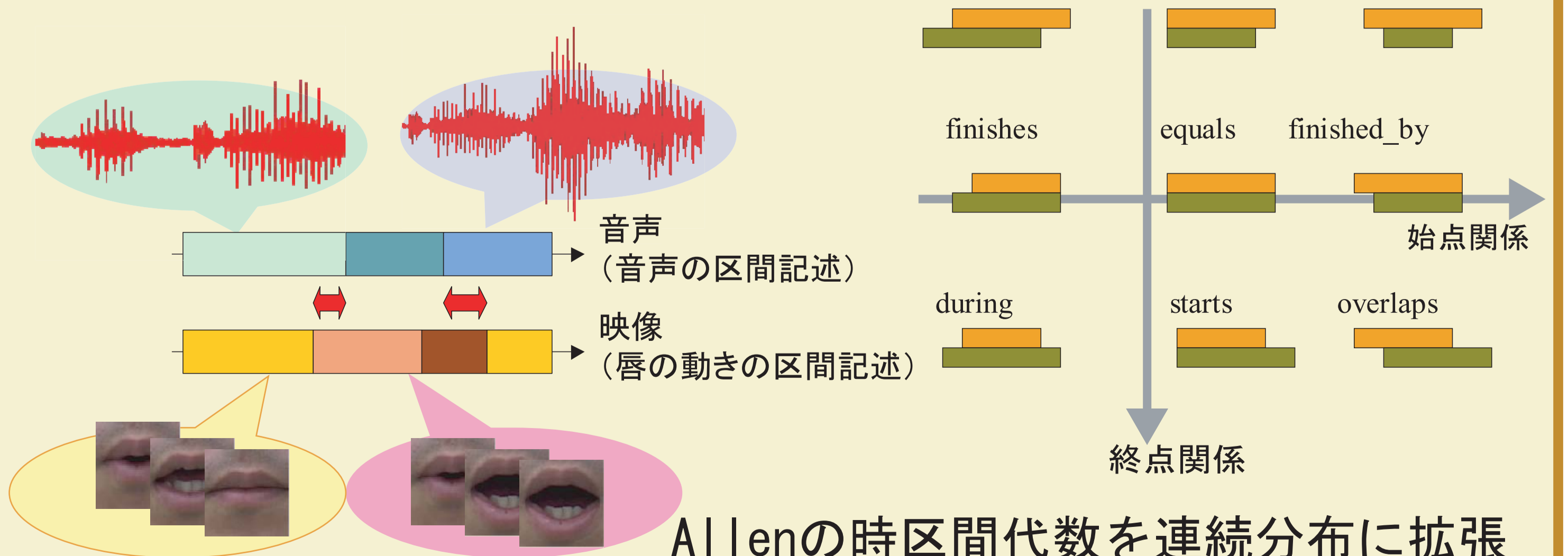
M step

各区間および区間遷移に関するパラメータを更新

3. 区間に基づくマルチメディアイベント記述

モダリティ間の時間関係や会話の「間」を時区間で記述

区間の始点と終点の関係を学習する



Allenの時区間代数を連続分布に拡張