静的・動的特徴の明示的な関係によりHMMから導出されるトラジェクトリモデル

全　炳河　　徳田　恵一　　北村　正（名工大）

A Trajecotry Model Derived from HMM with Explicit Relationship between Static and Dynamic Features

Heiga Zen, Keiichi Tokuda, Tadashi Kitamura
(Nagoya Institute of Technology)

背景

音声認識システムの代表的枠組み

特徴量　　　：MFCC+ΔMFCC

音響モデル：連続分布HMM

言語モデル：単語*N*-gram

HMMの問題点

(1) 状態内での時間的変化をモデル化できない

(2) 出力ベクトル間の時間的独立性を仮定

⇒　様々な音響モデルが提案

背景

(1) 状態内での時間的変化をモデル化できない

多項式近似HMM, 隠れ動的モデル (HDM),

声道共振周波数モデル, HMM-trajectory法など

(2) 出力ベクトル間の時間的独立性の仮定

条件付きHMM, フレーム間相関HMM,

部分隠れマルコフモデル (PHMM), セグメントモデル,

ダイナミックベイジアンネットワーク (DBN) など

背景

動的特徴量の利用（古井;1986）

・特徴ベクトルに動的特徴量を追加

・認識性能が大幅に改善

・時間方向の依存性をモデル化するための一手法

動的特徴量の計算

前後数フレームの静的特徴量から回帰係数として計算

$\Rightarrow$ 静的特徴・動的特徴系列間には確定的な関係が存在！

HMM + (MFCC・ΔMFCC) の枠組みでは無視！

発表内容

HMMから導出されるトラジェクトリモデル

1. 静的特徴・動的特徴間の関係をHMMに導入

   ⇒ HMMをトラジェクトリモデルとして再定式化

2. Viterbi学習形式のパラメータ更新式の導出
3. 準最適状態系列選択アルゴリズムの提案
4. 連続音声認識実験による評価

発表内容

HMMから導出されるトラジェクトリモデル

1. 静的特徴・動的特徴間の関係をHMMに導入

   ⇒ HMMをトラジェクトリモデルとして再定式化

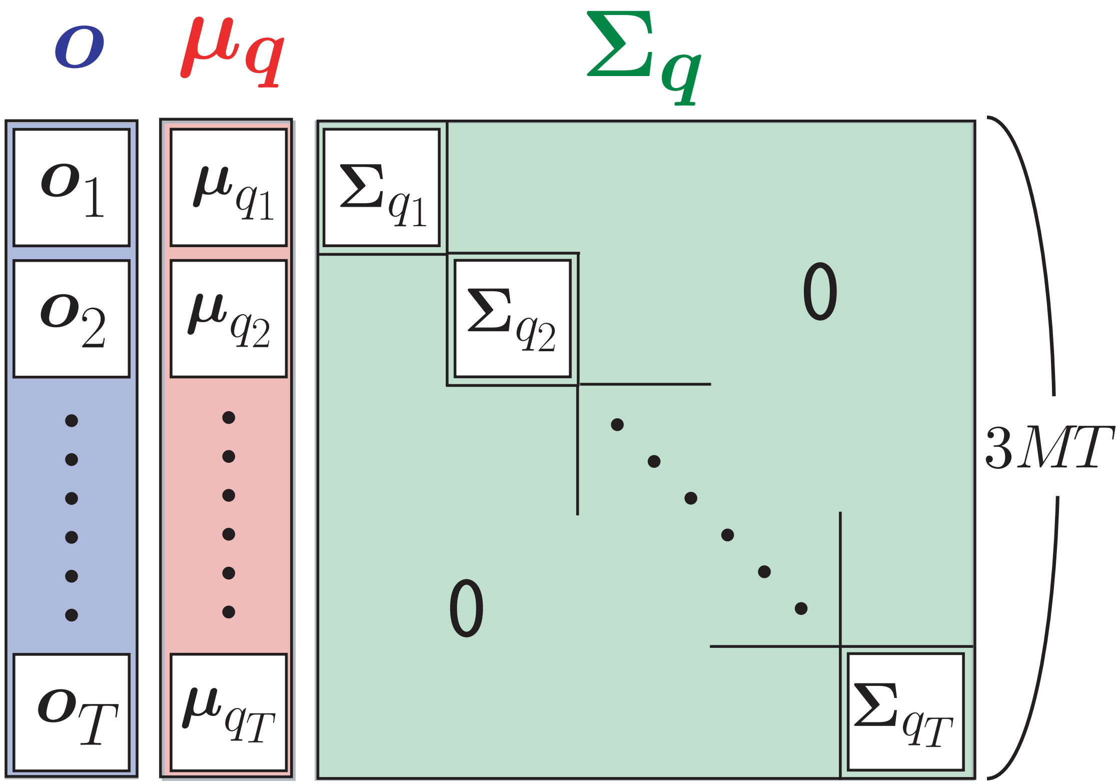
2. Viterbi学習形式のパラメータ更新式の導出
3. 準最適状態系列選択アルゴリズムの提案
4. 連続音声認識実験による評価

トラジェクトリHMMの導出 (1)

長さTの観測系列$\boldsymbol{o}$に対するHMMの尤度

$$P(\boldsymbol{o} \mid \lambda) = \sum_{\text{all } \boldsymbol{q}} P(\boldsymbol{o} \mid \boldsymbol{q}, \lambda) P(\boldsymbol{q} \mid \lambda)$$

$$P\left(\boldsymbol{o} | \boldsymbol{q}, \lambda\right) = \mathcal{N}\left(\boldsymbol{o} | \boldsymbol{\mu}_{\boldsymbol{q}}, \boldsymbol{\Sigma}_{\boldsymbol{q}}\right)$$ ($3MT$次元のガウス分布)



$\boldsymbol{o}_t$: 観測ベクトル ($3M$次元)

$\boldsymbol{q}$: 状態系列

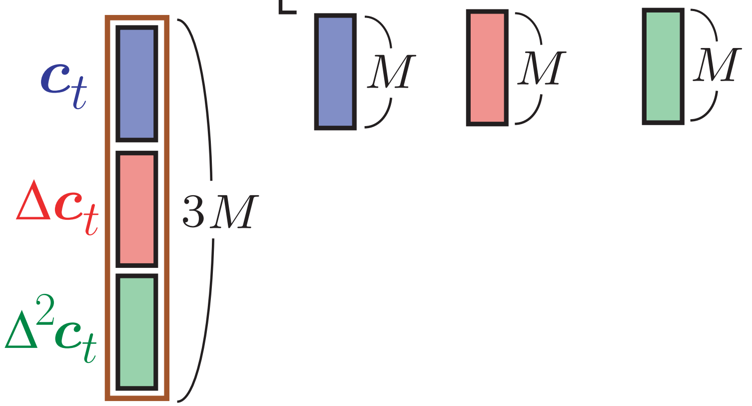
$\boldsymbol{\mu}_{q_t}$: 状態q_tの平均ベクトル

$\boldsymbol{\Sigma}_{q_t}$: 状態q_tの共分散行列

トラジェクトリHMMの導出 (2)

観測ベクトルが静的特徴と動的特徴から成る場合

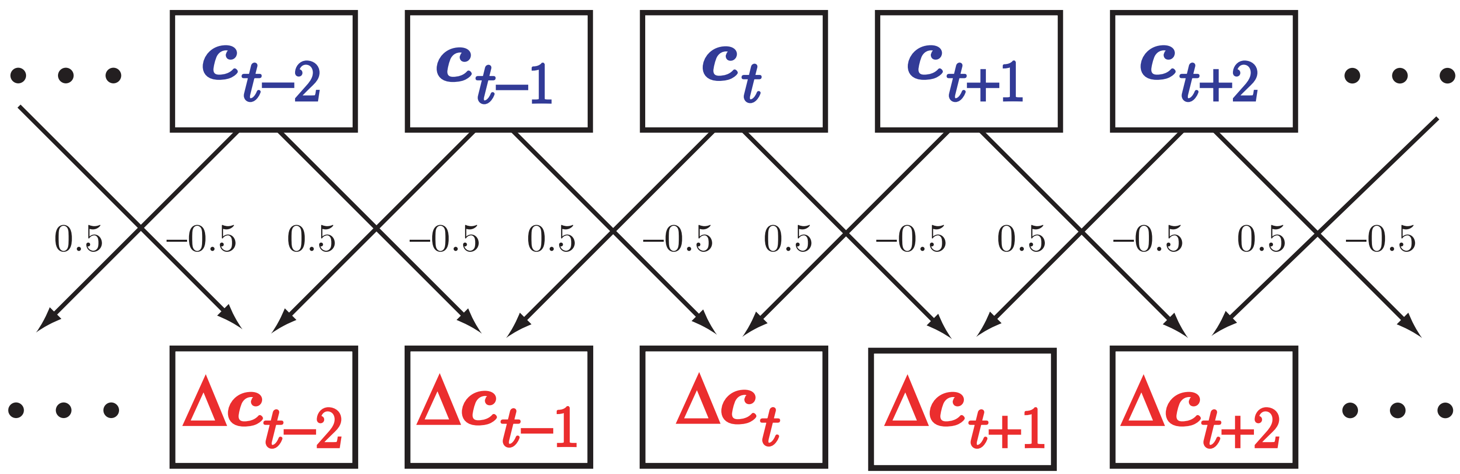
$$\boldsymbol{o}_t = \left[\boldsymbol{c}_t^\top \ \ \Delta\boldsymbol{c}_t^\top, \Delta^2\boldsymbol{c}_t^\top\right]^\top$$



- $\boldsymbol{c}_t$ ： 静的特徴ベクトル
- $\Delta\boldsymbol{c}_t$ ： 1次動的特徴ベクトル
- $\Delta^2\boldsymbol{c}_t$ ： 2次動的特徴ベクトル

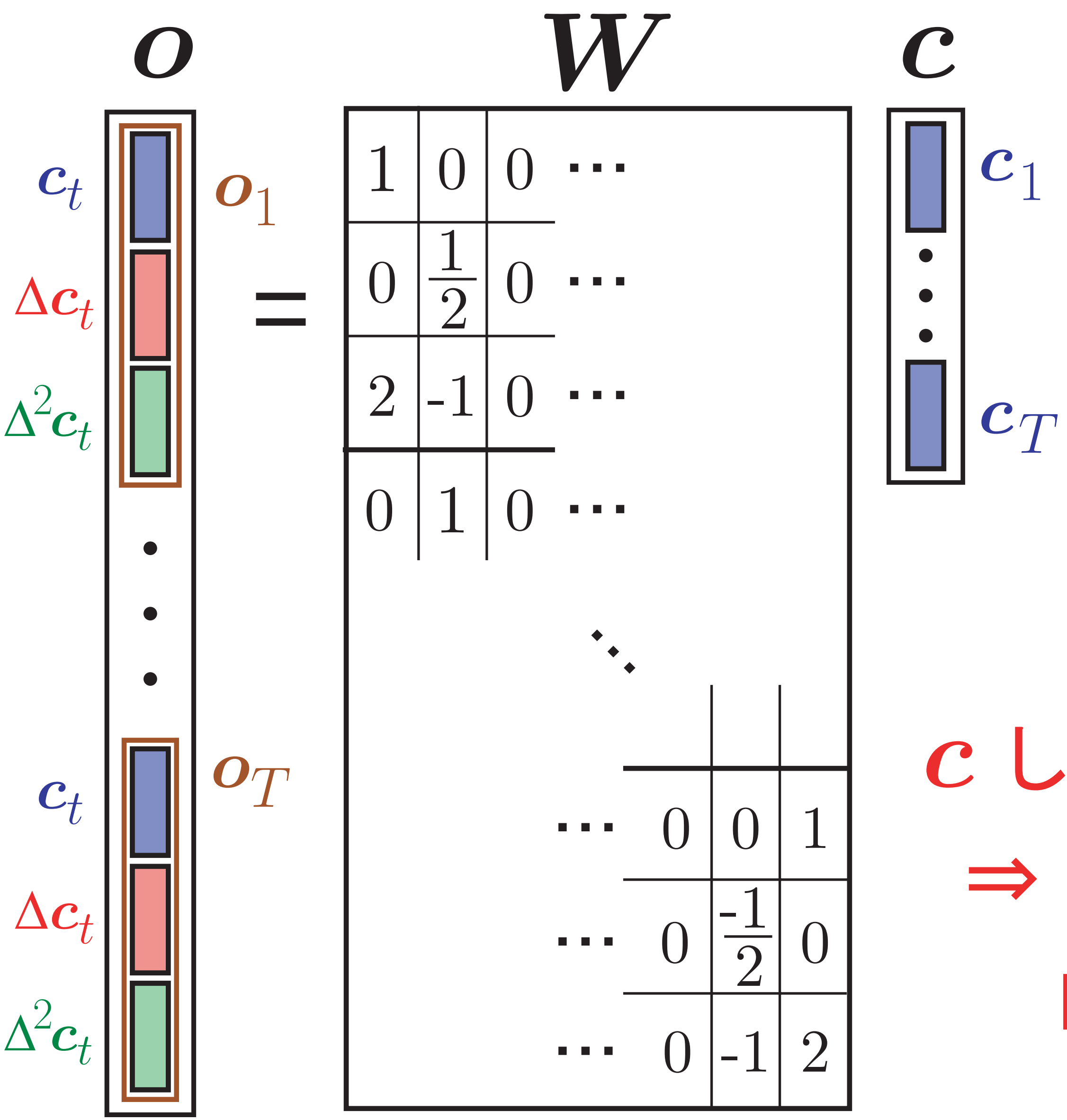
各時刻の動的特徴は前後数フレームの静的特徴から計算

（例） $\Delta\boldsymbol{c}_t = 0.5\boldsymbol{c}_{t+1} - 0.5\boldsymbol{c}_{t-1}$



トラジェクトリHMMの導出(3)

$\boldsymbol{o}$ と $\boldsymbol{c}$ の関係を行列形式で表せば,



$\boldsymbol{W}$: 動的特徴計算時の係数を要素とする窓行列 $(3MT \times MT)$

$\boldsymbol{c}$ しか自由度はない！

⇒ $\boldsymbol{c}$ のみ確率変数とみなしてHMMの尤度関数を変形

トラジェクトリHMMの導出(4)

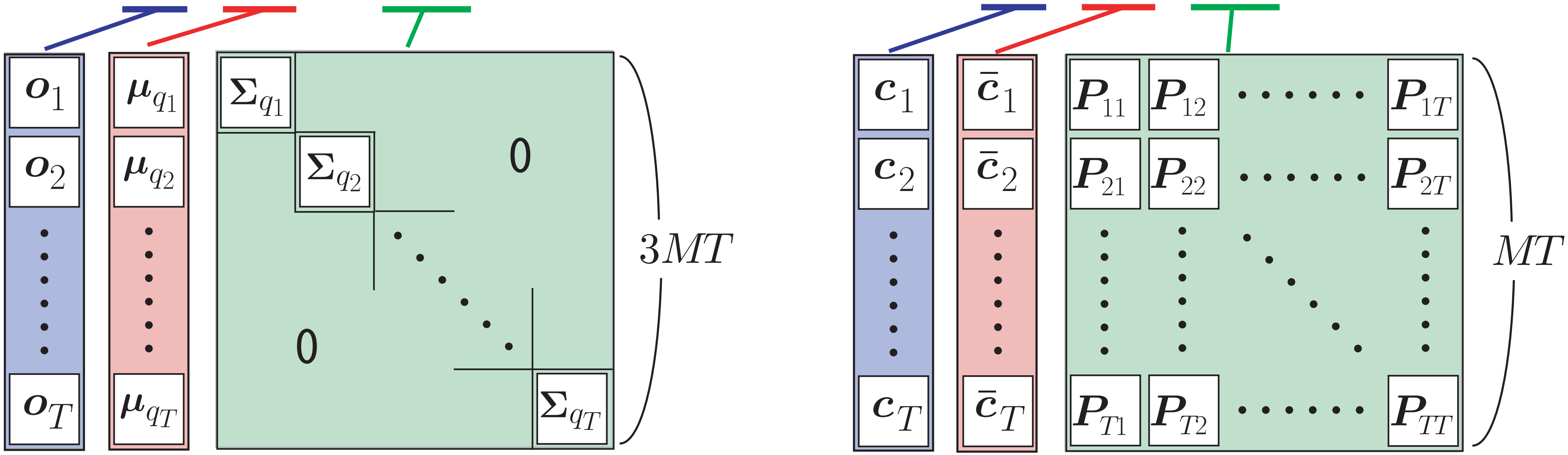
$$P\left(\boldsymbol{W}\boldsymbol{c} \mid \boldsymbol{q}, \lambda\right) = \mathcal{N}\left(\boldsymbol{W}\boldsymbol{c} \mid \boldsymbol{\mu}_{\boldsymbol{q}}, \boldsymbol{\Sigma}_{\boldsymbol{q}}\right)$$
$$= K_{\boldsymbol{q}} \cdot \mathcal{N}\left(\boldsymbol{c} \mid \bar{\boldsymbol{c}}_{\boldsymbol{q}}, \boldsymbol{P}_{\boldsymbol{q}}\right)$$

$\bar{\boldsymbol{c}}_{\boldsymbol{q}}$: 平均ベクトル(MT)

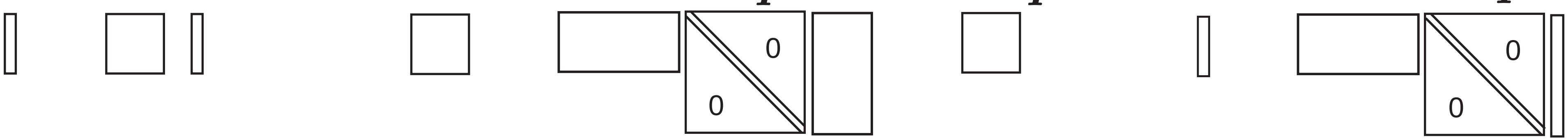
$\boldsymbol{P}_{\boldsymbol{q}}$: 共分散行列(MT× MT)

$K_{\boldsymbol{q}}$: $\boldsymbol{c}$ によらない定数

$$\mathcal{N}\left(\boldsymbol{o} \mid \boldsymbol{\mu}_{\boldsymbol{q}}, \boldsymbol{\Sigma}_{\boldsymbol{q}}\right) = K_{\boldsymbol{q}} \cdot \mathcal{N}\left(\boldsymbol{c} \mid \bar{\boldsymbol{c}}_{\boldsymbol{q}}, \boldsymbol{P}_{\boldsymbol{q}}\right)$$



$$\bar{\boldsymbol{c}}_{\boldsymbol{q}} = \boldsymbol{P}_{\boldsymbol{q}} \boldsymbol{r}_{\boldsymbol{q}} \qquad \boldsymbol{R}_{\boldsymbol{q}} = \boldsymbol{W}^{\top} \boldsymbol{\Sigma}_{\boldsymbol{q}}^{-1} \boldsymbol{W} = \boldsymbol{P}_{\boldsymbol{q}}^{-1} \qquad \boldsymbol{r}_{\boldsymbol{q}} = \boldsymbol{W}^{\top} \boldsymbol{\Sigma}_{\boldsymbol{q}}^{-1} \boldsymbol{\mu}_{\boldsymbol{q}}$$



トラジェクトリHMMの定義

静的特徴系列$\boldsymbol{c}$に対するトラジェクトリHMMの尤度

$$P(\boldsymbol{c} \mid \lambda) = \sum_{\text{all } \boldsymbol{q}} P(\boldsymbol{c} \mid \boldsymbol{q}, \lambda) P(\boldsymbol{q} \mid \lambda)$$

$$P(\boldsymbol{c} \mid \boldsymbol{q}, \lambda) = \mathcal{N}\left(\boldsymbol{c} \mid \bar{\boldsymbol{c}}_{\boldsymbol{q}}, \boldsymbol{P}_{\boldsymbol{q}}\right)$$ ：トラジェクトリ尤度

平均ベクトル$\bar{\boldsymbol{c}}_{\boldsymbol{q}}$は滑らかに変化するトラジェクトリ

⇒ 状態内での時間依存性を表現可能

時間方向の共分散行列$\boldsymbol{P}_{\boldsymbol{q}}$が全共分散

⇒ 出力ベクトルの時間依存性を表現可能

トラジェクトリHMMのパラメータ構造 ＝ 通常のHMMのパラメータ構造

尤度の定義が異なる ⇒ 学習アルゴリズムも異なる

発表内容

HMMから導出されるトラジェクトリモデル

1. 静的特徴・動的特徴間の関係をHMMに導入

   ⇒ HMMをトラジェクトリモデルとして再定式化

2. Viterbi学習形式のパラメータ更新式の導出
3. 準最適状態系列選択アルゴリズムの提案
4. 連続音声認識実験による評価

トラジェクトリHMMの学習アルゴリズム

トラジェクトリHMMの$\mathcal{Q}$関数（隠れ変数：状態遷移系列 $\boldsymbol{q}$ ）

$$\mathcal{Q}(\lambda, \lambda') = \sum_{\text{all } \boldsymbol{q}} P(\boldsymbol{q} \mid \boldsymbol{c}, \lambda) \log P(\boldsymbol{c}, \boldsymbol{q} \mid \lambda')$$

全ての状態系列に関する和を計算することは困難

⇒ 一本の状態系列による近似を導入

$$\boldsymbol{q}' = \arg\max_{\boldsymbol{q}} \ \log P\left(\boldsymbol{c} \mid \boldsymbol{q}, \lambda\right) \quad \text{：最適状態系列の探索}$$

$$\lambda' = \arg\max_{\lambda} \ \log P\left(\boldsymbol{c} \mid \boldsymbol{q}', \lambda\right) \quad \text{：パラメータの最適化}$$

トラジェクトリHMMの学習アルゴリズム

対数トラジェクトリ尤度関数

$$\log P(\boldsymbol{c} \mid \boldsymbol{q}, \lambda) = -\frac{1}{2}\Big\{MT\log(2\pi) - \log\left|\boldsymbol{R_q}\right| + \boldsymbol{c}^\top \boldsymbol{P_q}^{-1}\boldsymbol{c} + \boldsymbol{r_q}^\top \boldsymbol{P_q}\boldsymbol{r_q} - 2\boldsymbol{r_q}\boldsymbol{c}^\top\Big\}$$

更新するパラメータ

$$\boldsymbol{m} = \left[\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top, \ldots, \boldsymbol{\mu}_N^\top\right]^\top$$

$$\boldsymbol{\phi} = \left[\boldsymbol{\Sigma}_1^{-1}, \boldsymbol{\Sigma}_2^{-1}, \ldots, \boldsymbol{\Sigma}_N^{-1}\right]^\top$$

$\boldsymbol{m}$ ： 全状態の平均ベクトル

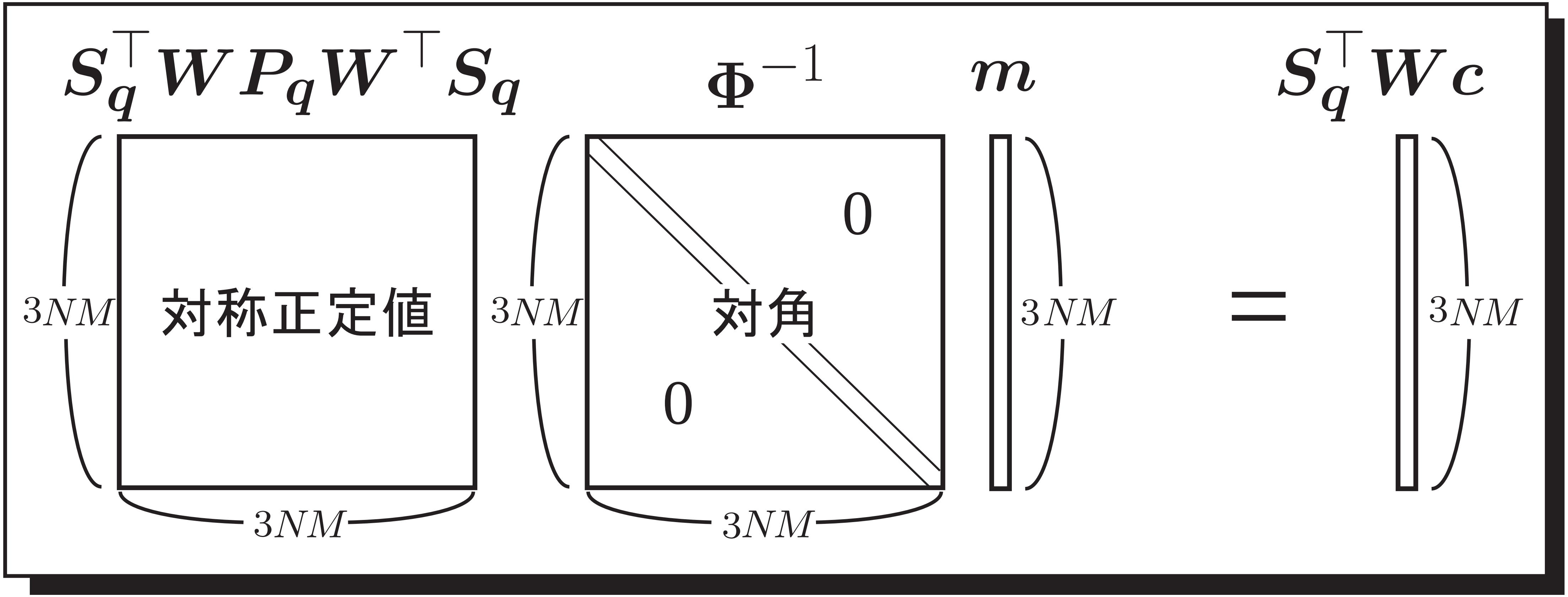
$\boldsymbol{\phi}$ ： 全状態の共分散行列

N ： 総状態数

全モデルのパラメータを同時に更新

パラメータ更新式(平均)

対数トラジェクトリ尤度関数を $\boldsymbol{m}$ で偏微分して0とおくと



式中の各変数の意味は予稿を参照

線形方程式を解くことで $\boldsymbol{m}$ が一意に決定される

パラメータ更新式(分散)

対数トラジェクトリ尤度関数をϕで偏微分

$$\frac{\partial \log P(\boldsymbol{c}, \boldsymbol{q} \mid \lambda)}{\partial \phi} = \frac{1}{2} \boldsymbol{S}_{\boldsymbol{q}}^{\top} \operatorname{diag}^{-1}\Big(\boldsymbol{W} \boldsymbol{P}_{\boldsymbol{q}} \boldsymbol{W}^{\top} - \boldsymbol{W} \boldsymbol{c} \boldsymbol{c}^{\top} \boldsymbol{W}^{\top} + 2 \boldsymbol{\mu}_{\boldsymbol{q}} \boldsymbol{c}^{\top} \boldsymbol{W}^{\top} + \boldsymbol{W} \bar{\boldsymbol{c}}_{\boldsymbol{q}} \bar{\boldsymbol{c}}_{\boldsymbol{q}}^{\top} \boldsymbol{W}^{\top} - 2 \boldsymbol{\mu}_{\boldsymbol{q}} \bar{\boldsymbol{c}}_{\boldsymbol{q}}^{\top} \boldsymbol{W}^{\top} \Big)$$

非線形方程式であるため容易には解けない

⇒ 上式を用いて最急降下法で更新

発表内容

HMMから導出されるトラジェクトリモデル

1. 静的特徴・動的特徴間の関係をHMMに導入

   ⇒ HMMをトラジェクトリモデルとして再定式化

2. Viterbi学習形式のパラメータ更新式の導出
3. 準最適状態系列選択アルゴリズムの提案
4. 連続音声認識実験による評価

トラジェクトリ尤度の計算

$$P(\boldsymbol{c} \mid \boldsymbol{q}, \lambda) = K_{\boldsymbol{q}}^{-1} \cdot P(\boldsymbol{o} \mid \boldsymbol{q}, \lambda)$$

$$K_{\boldsymbol{q}} = \frac{\sqrt{(2\pi)^{MT} \left|\boldsymbol{P}_{\boldsymbol{q}}\right|}}{\sqrt{(2\pi)^{3MT} \left|\boldsymbol{\Sigma}_{\boldsymbol{q}}\right|}} \cdot \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_{\boldsymbol{q}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{q}}^{-1}\boldsymbol{\mu}_{\boldsymbol{q}} - \boldsymbol{r}_{\boldsymbol{q}}^{\top}\boldsymbol{P}_{\boldsymbol{q}}\boldsymbol{r}_{\boldsymbol{q}}\right)\right\}$$

トラジェクトリ尤度 = 正規化定数 $K_{\boldsymbol{q}}$ ×HMMの尤度

トラジェクトリ尤度を直接計算 ⇒ 高次元の線形演算が必要

トラジェクトリ尤度を時間方向に再帰的に計算したい

⇒ $\left|\boldsymbol{P}_q\right|$ と $\boldsymbol{r}_q^{\top}\boldsymbol{P}_q\boldsymbol{r}_q$ を時間方向に再帰的に計算

トラジェクトリ尤度の再帰計算

$$P(\boldsymbol{c} \mid \boldsymbol{q}, \lambda) = \prod_{t=1}^{T} \frac{1}{K_{\boldsymbol{q}_{t+L}}^{(t)}} \cdot P(\boldsymbol{o}_t \mid \boldsymbol{q}_t, \lambda)$$

$$K_{\boldsymbol{q}_{t+L}}^{(t)} = \frac{\sqrt{(2\pi)^M} \left| \boldsymbol{U}_{\boldsymbol{q}_{t+L}}^{(t,t)} \right|^{-1}}{\sqrt{(2\pi)^{3M} \left| \boldsymbol{\Sigma}_{q_t} \right|}} \cdot \exp\left\{ -\frac{1}{2} \left(\boldsymbol{\mu}_{q_t}^{\top} \boldsymbol{\Sigma}_{q_t}^{-1} \boldsymbol{\mu}_{q_t} - \left[\boldsymbol{g}_{\boldsymbol{q}_{t+L}}^{(t)} \right]^{\top} \boldsymbol{g}_{\boldsymbol{q}_{t+L}}^{(t)} \right) \right\}$$

時間方向に再帰的に計算可能！

$\boldsymbol{U}_{\boldsymbol{q}} \Rightarrow \boldsymbol{R}_{\boldsymbol{q}} = \boldsymbol{U}_{\boldsymbol{q}}^{\top} \boldsymbol{U}_{\boldsymbol{q}}$ ：コレスキー分解より得られる上三角行列

$\boldsymbol{g}_{\boldsymbol{q}} \Rightarrow \boldsymbol{R}_{\boldsymbol{q}} \bar{\boldsymbol{c}}_{\boldsymbol{q}} = \boldsymbol{r}_{\boldsymbol{q}} \qquad \boldsymbol{U}_{\boldsymbol{q}}^{\top} \boldsymbol{U}_{\boldsymbol{q}} \bar{\boldsymbol{c}}_{\boldsymbol{q}} = \boldsymbol{r}_{\boldsymbol{q}}$

$\boldsymbol{U}_{\boldsymbol{q}} \bar{\boldsymbol{c}}_{\boldsymbol{q}} = \boldsymbol{g}_{\boldsymbol{q}}$ ：前進代入より得られるベクトル

トラジェクトリHMMのためのViterbiアルゴリズム

トラジェクトリHMM

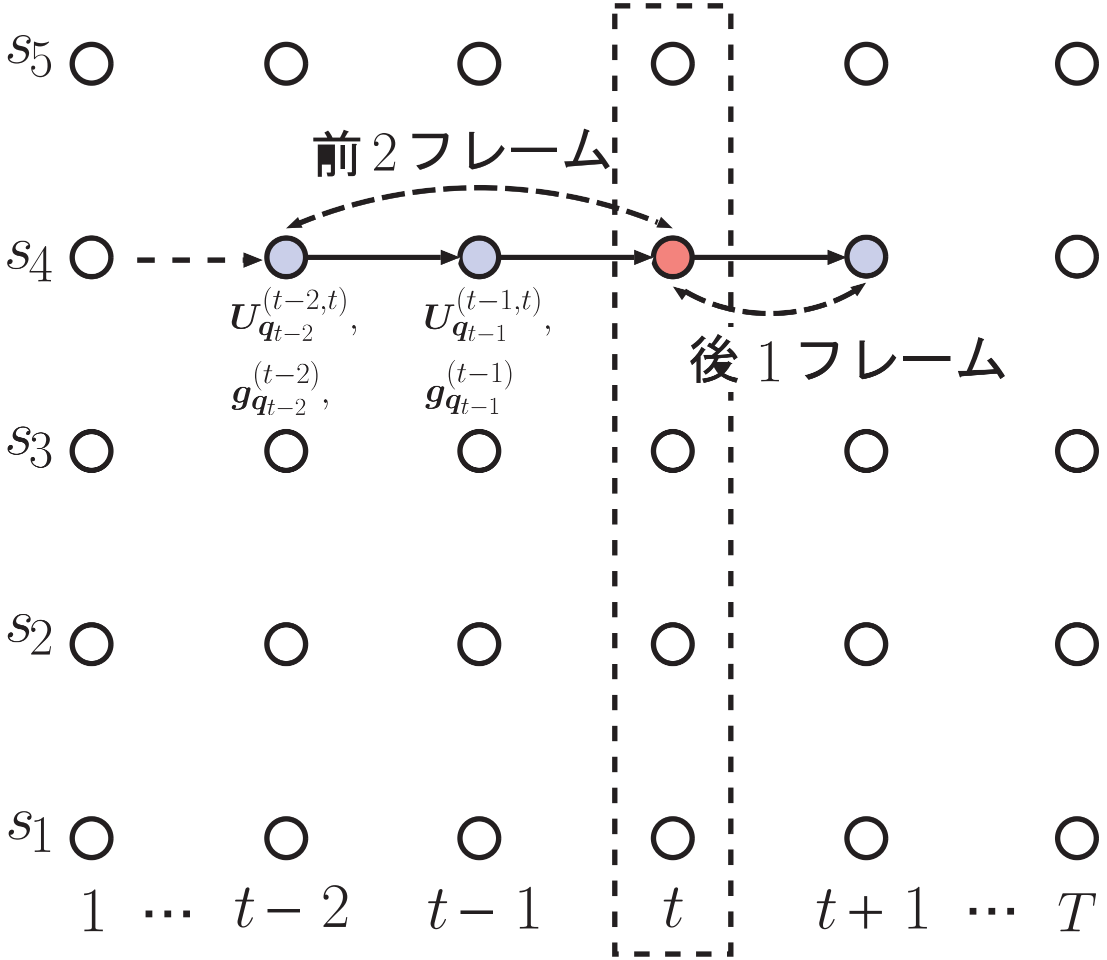
- 各時刻の出力ベクトル間に依存関係(P_q が全共分散)
- マルコフ過程でない(∞重マルコフ過程)
- 最適状態系列の探索 ⇒ 全探索

提案法

- J 重マルコフ過程で近似
- Viterbiアルゴリズムにより準最適状態系列を選択

トラジェクトリHMMのためのViterbiアルゴリズム

（例） 前後1フレームから動的特徴を計算(L=1)
2重マルコフモデルの決定過程で近似(J=2)

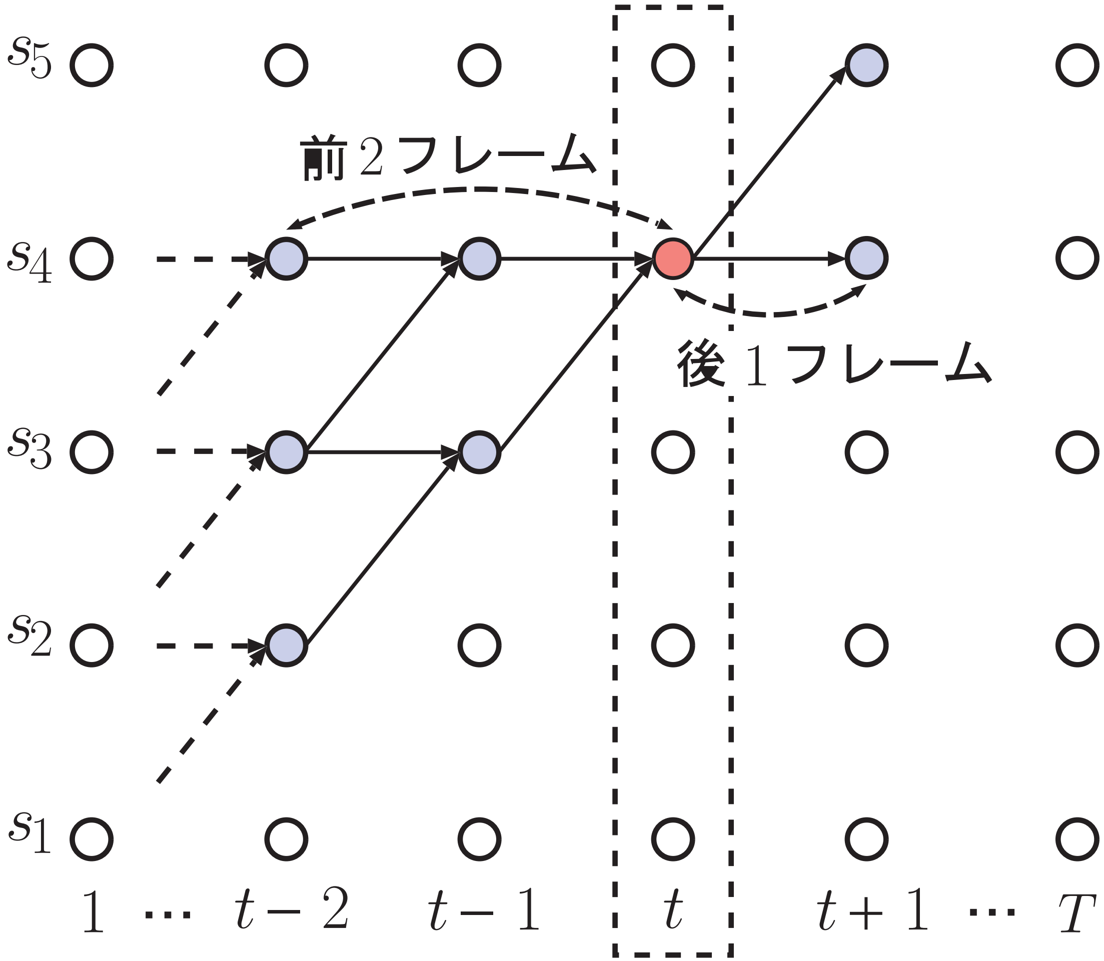


- 2フレーム前までの状態系列は既に決定
- 動的特徴計算の窓が前後1フレームにかかる

⇒ 1フレームの先読みが必要

トラジェクトリHMMのためのViterbiアルゴリズム

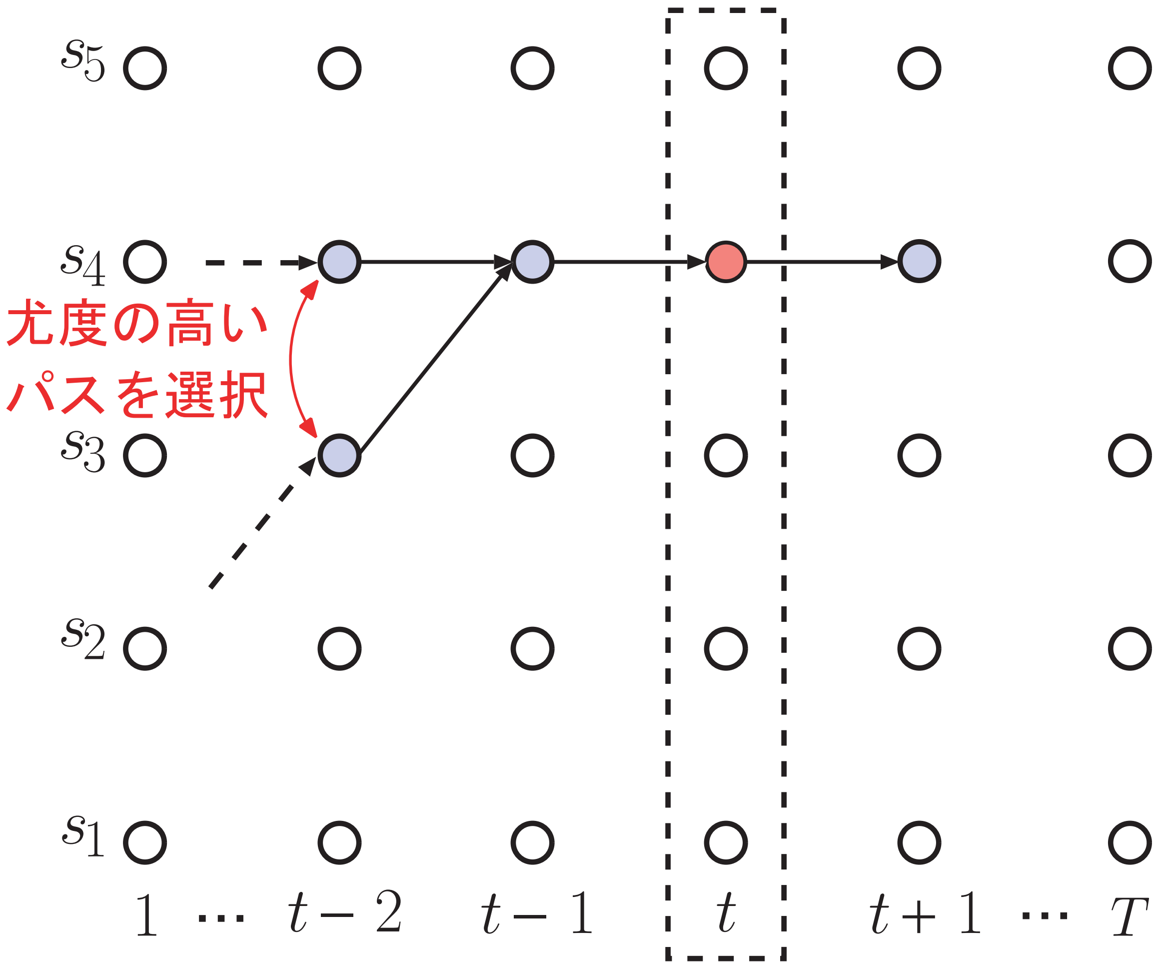
（例）前後1フレームから動的特徴を計算(L=1)
　　　2重マルコフモデルの決定過程で近似(J=2)



- 2^{J+L} 個のパスについて
  トラジェクトリ尤度を計算

トラジェクトリHMMのためのViterbiアルゴリズム

（例） 前後1フレームから動的特徴を計算 (L=1)
2重マルコフモデルの決定過程で近似 (J=2)



Jフレーム遅れて
滞在する状態を決定

⇒ 近傍に与える影響を
考慮して状態を決定

調音結合の影響 ：
100ms～200ms

⇒ Jを10程度とれば十分
（フレーム周期が10msの場合）

発表内容

HMMから導出されるトラジェクトリモデル

1. 静的特徴・動的特徴間の関係をHMMに導入

   → HMMをトラジェクトリモデルとして再定式化

2. Viterbi学習形式のパラメータ更新式の導出
3. 準最適状態系列選択アルゴリズムの提案
4. 連続音声認識実験による評価

実験条件

学習データ	ATR日本語連続音声データベース B-set 男性話者 MHT 503文章中の450文章
評価データ	同一話者 学習データに含まれない53文章
周波数	16 kHz
分析窓	25-ms Blackman窓
フレーム周期	10-ms
分析方法	18次 メルケプストラム分析
動的特徴	前後1フレームから回帰係数として計算
特徴ベクトル	$c(0) \sim c(18), \Delta c(0) \sim \Delta c(18), \Delta\Delta c(0) \sim \Delta\Delta c(18)$
トポロジー	3状態 スキップのないleft-to-right モデル 混合数 1 コンテキスト非依存

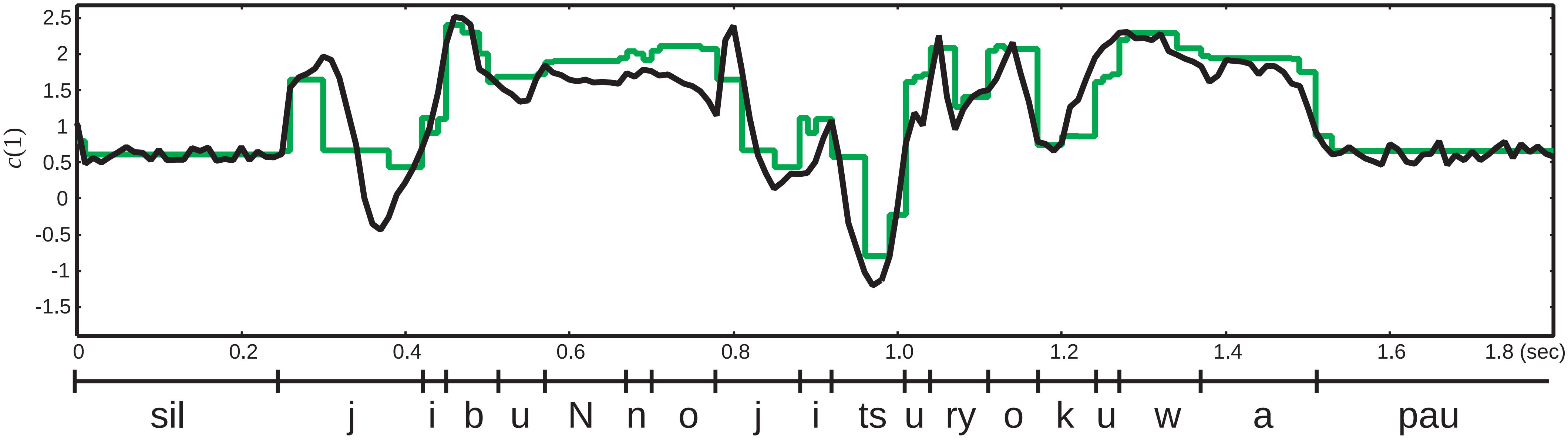
学習結果

学習データ及び評価データに対する
フレーム平均対数トラジェクトリ尤度

	training data	test data
standard-HMM	6.87	6.93
trajectory-HMM	15.0	15.1

導出されたパラメータ更新式により
トラジェクトリ尤度が増加

平均系列の例

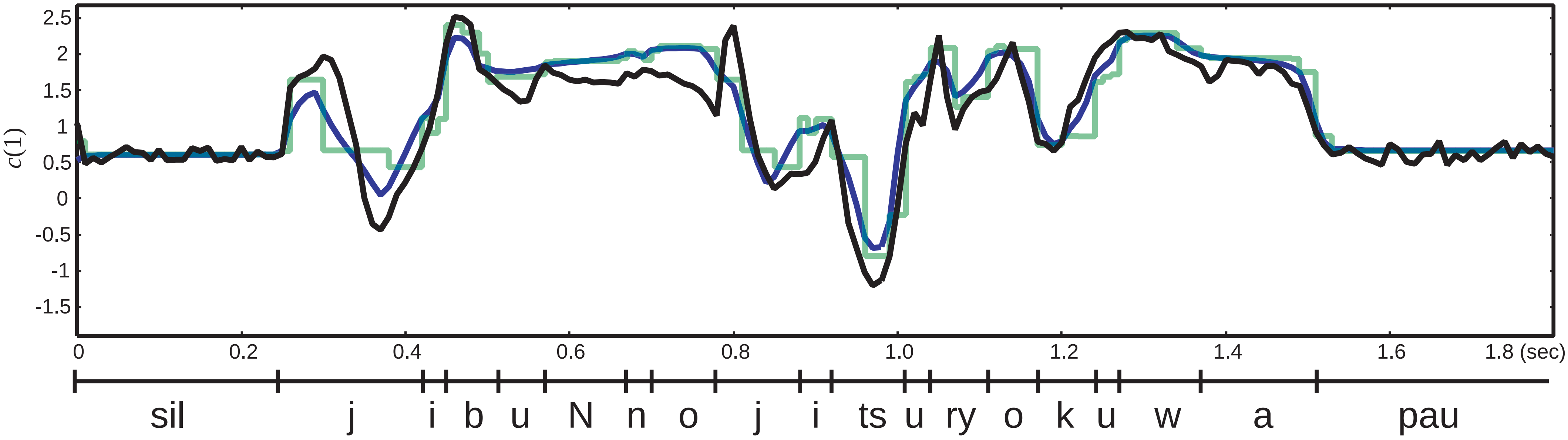


実発声（学習データ）

通常のHMMの平均系列

通常のHMMの平均系列 → 時間的に不連続

平均系列の例



実発声（学習データ）

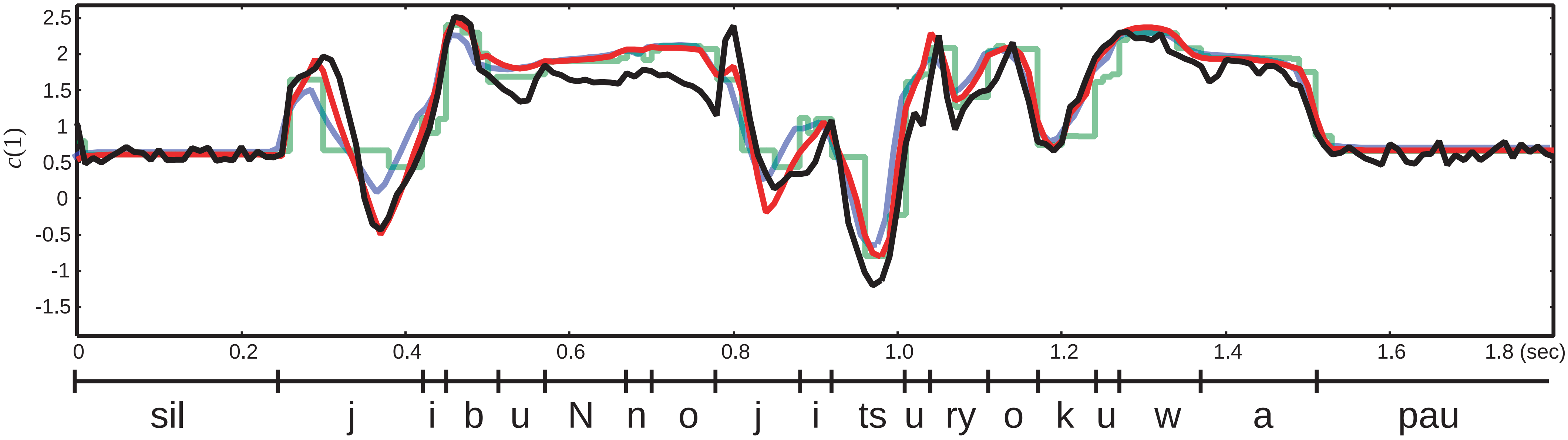
通常のHMMより生成された系列

通常のHMMの平均系列

通常のHMMの平均系列 → 時間的に不連続

通常のHMMより生成された系列 → 時間的に滑らかに変化

平均系列の例



実発声（学習データ）

通常のHMMより生成された系列

通常のHMMの平均系列

トラジェクトリHMMの平均系列

通常のHMMの平均系列 → 時間的に不連続

通常のHMMより生成された系列 → 時間的に滑らかに変化

トラジェクトリHMMの平均系列 → 時間的に滑らかに変化

より実発声に近い平均系列

認識実験結果

通常のHMMを用いて上位200個の仮説を出力

トラジェクトリ尤度でリスコア

	%Err.	#Corr.	#Del.	#Sub.	#Ins.	#Total
standard-HMM	20.1	1855	119	200	119	2174
trajectory-HMM	17.2	1903	118	153	103	2174

誤り削減率　約15%

まとめ

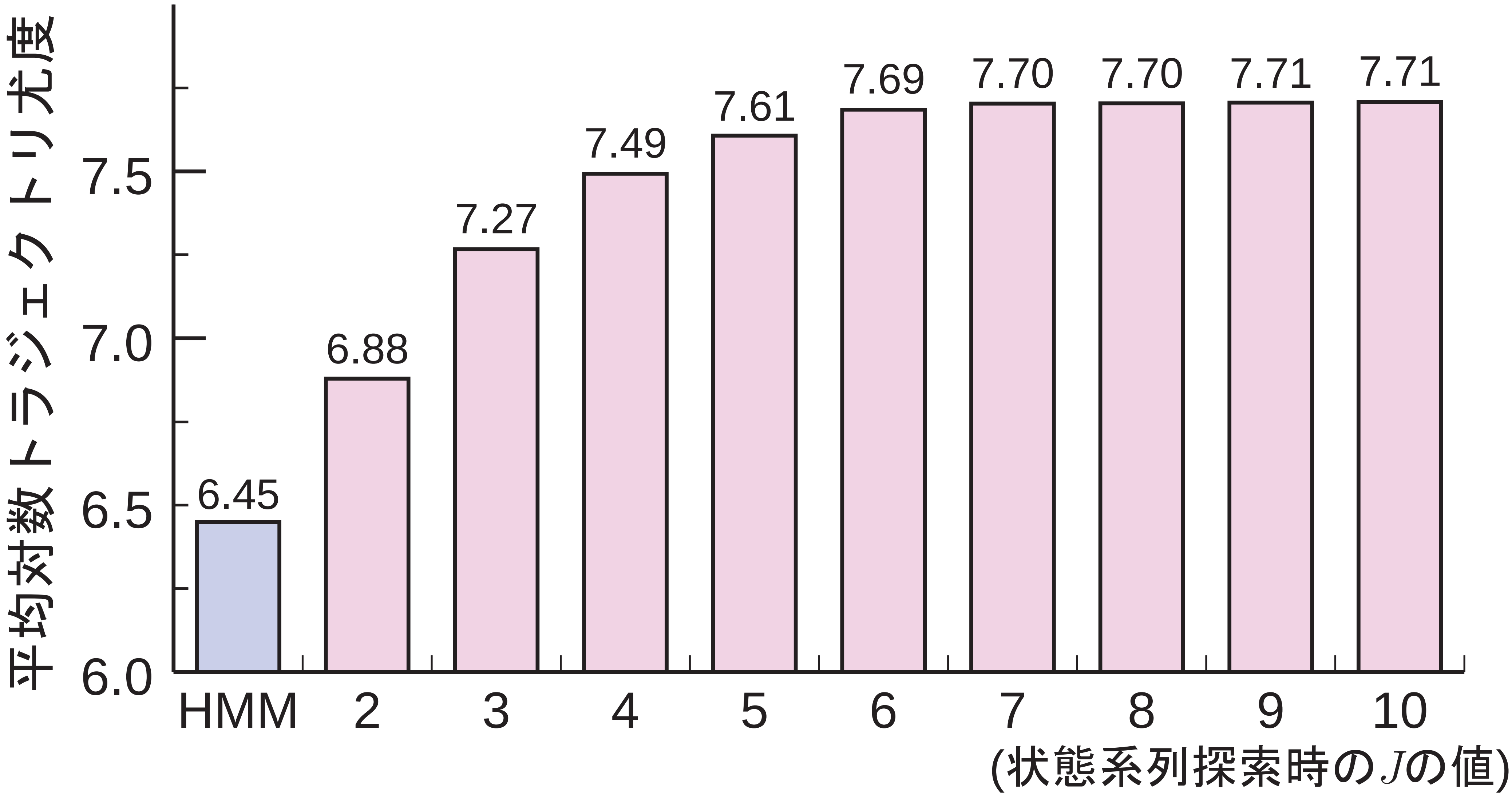
HMMよりトラジェクトリモデルを導出

- 静的特徴と動的特徴間の関係をHMMに導入
  → トラジェクトリモデルとして再定式化
- Viterbi学習形式の学習アルゴリズムを提案
- 連続音声認識実験により有効性を確認

今後の予定

- Baum-Welch（EM）型学習アルゴリズムの導出
- Viterbiデコーダの実装
- より大規模な評価実験
- HMM音声合成への適用

選択された状態系列の尤度



$J\rightarrow$大 ⇒トラジェクトリ尤度が高い状態系列を選択

$J=6$ 程度でほぼ収束