

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[†]

[†] 名古屋工業大学 大学院 工学研究科 情報工学専攻
〒 466-8555 名古屋市 昭和区 御器所町
E-mail: †{zen,tokuda,kitamura}@ics.nitech.ac.jp

あらまし 本報告では、静的特徴・動的特徴間の関係を明示的に導入することにより、HMM から導出されるトラジェクトリモデルを提案する。音声認識において、音響モデルに HMM、特徴ベクトルに MFCC 及びその動的特徴量を用いた枠組みが広く利用されている。一般に動的特徴量は、前後数フレームの静的特徴量から回帰係数として計算されることが多い。よって静的特徴量・動的特徴量間には、線形変換で表現できる確定的な関係が存在するが、これまでの枠組みではこの関係を無視してモデル化を行っていた。このため、実際の静的特徴量・動的特徴量間の関係と、出力確率分布の静的特徴量・動的特徴量に関するパラメータ間の関係との間に矛盾が生じていた。本報告では、HMM に静的特徴量・動的特徴量間の関係を明示的に導入することにより、HMM はトラジェクトリモデルとして再定式化されることを示す。また、得られたトラジェクトリモデルの Viterbi 学習アルゴリズムを導出する。特定話者連続音声認識実験の結果、学習したトラジェクトリモデルは通常の HMM と比較して、音素誤り率を約 15%削減できた。キーワード 隠れマルコフモデル、トラジェクトリモデル、動的特徴量、HMM からのパラメータ生成

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

Heiga ZEN[†], Keiichi TOKUDA[†], and Tadashi KITAMURA[†]

[†] Department of Computer Science and Engineering, Graduate School of Engineering,
Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, 466-8555 Japan
E-mail: †{zen,tokuda,kitamura}@ics.nitech.ac.jp

Abstract In this paper, we propose a trajectory model derived from HMM with explicit relationship between static and dynamic features. In many speech recognition systems, hidden Markov model and MFCC and its dynamic features are often used. Generally, dynamic features are computed as regression coefficients from static features at its neighboring frames. Therefore, there is a deterministic relationship between static and dynamic features. However, HMM with static and dynamic features allows inconsistent statistics between the model parameters for static and dynamic features. In this paper, we derive a trajectory model from HMM by imposing the relationship between static and dynamic features. Furthermore, we also derive the Viterbi-type training algorithm for proposed trajectory model. In speaker-dependent continuous speech recognition experiment, proposed trajectory model achieved about 15% error reduction over HMM.

Key words hidden Markov model, trajectory model, dynamic feature, parameter generation from HMMs

1. はじめに

計算機の高速度化、大規模音声コーパスの整備などに伴い、大規模連続音声認識に関する研究が盛んに行われている。音声認識における代表的な枠組みとして、音響モデルに統計モデルの一種で

ある隠れマルコフモデル (Hidden Markov Model; HMM)、特徴ベクトルに MFCC とその動的特徴量 (Δ MFCC, Δ^2 MFCC) を用いる枠組みがある。HMM は、学習データに基づきパラメータを推定する実現容易なアルゴリズムが存在する、トポロジーを認識対象に応じて設計できる、現実的な計算量で学習・

である．ここで， $L_-^{(0)} = L_+^{(0)} = 0$ ， $w^{(0)}(0) = 1$ である．以上の関係を用いることで，式 (4) は以下のように変形できる．

$$P(\mathbf{W}c | q, \lambda) = \mathcal{N}(\mathbf{W}c | \boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q) \quad (14)$$

$$= K_q \cdot \mathcal{N}(c | \bar{c}_q, P_q) \quad (15)$$

但し，

$$\mathbf{R}_q \bar{c}_q = r_q \quad (16)$$

$$\mathbf{R}_q = \mathbf{W}^\top \boldsymbol{\Sigma}_q^{-1} \mathbf{W} = P_q^{-1} \quad (17)$$

$$r_q = \mathbf{W}^\top \boldsymbol{\Sigma}_q^{-1} \boldsymbol{\mu}_q \quad (18)$$

$$K_q = \frac{\sqrt{(2\pi)^{MT} |P_q|}}{\sqrt{(2\pi)^{DMT} |\boldsymbol{\Sigma}_q|}} \cdot \exp \left\{ -\frac{1}{2} \left(\boldsymbol{\mu}_q^\top \boldsymbol{\Sigma}_q^{-1} \boldsymbol{\mu}_q - r_q^\top P_q r_q \right) \right\} \quad (19)$$

である．式 (15) において正規化係数 K_q を除くことにより，以下のようにトラジェクトリ HMM を定義する．

$$P(c | \lambda) = \sum_{\text{all } q} P(c | q, \lambda) P(q | \lambda), \quad (20)$$

$$P(c | q, \lambda) = \mathcal{N}(c | \bar{c}_q, P_q). \quad (21)$$

以後，式 (21) で計算される尤度をトラジェクトリ尤度と呼ぶ．興味深いことに，系列 \bar{c}_q は HMM からの音声パラメータ生成アルゴリズム（文献 [26] の case 1）より得られるトラジェクトリと等しい．つまり， \bar{c}_q を静的特徴ベクトル系列 c の平均ベクトル系列， P_q を時間方向の共分散行列とみなせば，通常の連続分布型 HMM をトラジェクトリモデルに変換できる．

トラジェクトリ HMM の平均ベクトル系列 \bar{c}_q は，通常の HMM の平均ベクトル系列のように階段状に変化するのではなく，滑らかに変化する系列となるため，状態内で統計量が一定であるという HMM の制約を避けることができる．また，時間方向の共分散行列 P_q は一般に全共分散行列となるため，全時刻の観測ベクトルと全時刻において滞在している状態は依存関係を有している．このため，独立性の仮定も避けることができる．また，トラジェクトリ HMM のモデルパラメータ構造は HMM と完全に一致しており，パラメータ数が増加することはない．

提案するトラジェクトリ HMM は，HMM-trajectory 法 [12–14] と近い．HMM-trajectory 法では，HMM からの音声パラメータ生成アルゴリズム [26] より得られた系列を平均ベクトル系列とし，平均ベクトル系列と学習データ系列との差分を用いて分散のみを再学習している．このとき，時間方向の共分散行列は対角である．これに対して，トラジェクトリ HMM は次章で述べるアルゴリズムにより平均及び分散を更新する．また，時間方向の共分散行列は全共分散である．

ここで，トラジェクトリ HMM を違う観点から分析してみる．簡単化のため， $M = 1$ として記述する．トラジェクトリ HMM の時間方向の共分散行列 P_q の逆行列（precision matrix と呼ばれる） R_q は，式 (11)，(12)，(17) より以下のように変形できる．

$$\begin{aligned} \mathbf{R}_q &= \mathbf{W}^\top \boldsymbol{\Sigma}_q^{-1} \mathbf{W} \\ &= \sum_{t=1}^T \left\{ \frac{1}{\Sigma_{q_t}^{(1)}} \cdot \left[\mathbf{w}_t^{(0)} \right]^\top \mathbf{w}_t^{(0)} \right. \\ &\quad \left. + \sum_{d=1}^{D-1} \frac{1}{\Delta^{(d)} \Sigma_{q_t}^{(1)}} \cdot \left[\mathbf{w}_t^{(d)} \right]^\top \mathbf{w}_t^{(d)} \right\} \quad (22) \end{aligned}$$

式 (22) は， \mathbf{R}_q に対する拡張最尤線形変換（Extended Maximum Likelihood Linear Transform; EMLLT）[27] つまり，時間方向の全共分散行列 P_q の逆行列 \mathbf{R}_q （大きさ $MT \times MT$ ）を，基底行列 \mathbf{W} （大きさ $DMT \times MT$ ）を用いて高次の空間（大きさ $DMT \times DMT$ ）に射影し，対角行列 $\boldsymbol{\Sigma}_q$ （大きさ $DMT \times DMT$ ）でモデル化している．EMLLT の一般形では基底行列，対角行列の両方が推定すべきパラメータであるが，トラジェクトリ HMM では基底行列は窓行列 \mathbf{W} で与えられるため，基底行列を固定して対角行列を推定するような形式の EMLLT とみなせる．推定すべきパラメータである $\boldsymbol{\Sigma}_q$ の要素は，状態遷移により決定し，同一の状態に滞在する時刻では，同じパラメータが共有されている．このため，推定すべきパラメータの数は非常に少ない．このようにしてトラジェクトリ HMM では，通常の HMM と同じパラメータ数で時間方向の全共分散行列 P_q をモデル化していると考えられる．

3. 学習アルゴリズム

本節では，前節で定義したトラジェクトリ HMM のモデルパラメータの更新式を導出する．まず， \mathcal{Q} 関数を以下のように定義する．

$$\mathcal{Q}(\lambda, \lambda') = \sum_{\text{all } q} P(q | c, \lambda) \log P(c, q | \lambda') \quad (23)$$

式 (23) を計算するには，全ての状態系列について式 (21) を計算する必要があり，容易ではない．そこで，HMM の Viterbi 学習と同様の，単一の状態系列による近似を行う．その結果学習の手順は，1) 状態系列の決定，2) パラメータ更新，の二段階に分かれる．状態系列の決定法については次節で述べ，本節ではパラメータ更新について述べる．パラメータ更新問題は，状態系列 q が与えられたとき対数トラジェクトリ尤度

$$\begin{aligned} \log P(c | q, \lambda) &= -\frac{1}{2} \left\{ MT \log(2\pi) - \log |\mathbf{R}_q| \right. \\ &\quad \left. + c^\top P_q^{-1} c + r_q^\top P_q r_q - 2r_q c^\top \right\} \quad (24) \end{aligned}$$

を最大にするパラメータを求めることに帰着される．ここで推定すべきパラメータは， m 及び ϕ とする．但し，

$$m = \left[\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top, \dots, \boldsymbol{\mu}_N^\top \right]^\top \quad (25)$$

$$\phi = \left[\boldsymbol{\Sigma}_1^{-1}, \boldsymbol{\Sigma}_2^{-1}, \dots, \boldsymbol{\Sigma}_N^{-1} \right]^\top \quad (26)$$

であり， N は学習データ中に出現する HMM の総分布数である．式 (21) を最大にする m は，以下に示す関係，

$$\boldsymbol{\Phi}^{-1} = \text{diag}(\phi) \quad (27)$$

$$\boldsymbol{\mu}_q = S_q m \quad (28)$$

$$\boldsymbol{\Sigma}_q^{-1} = \text{diag}(S_q \phi) \quad (29)$$

を利用すれば，式 (24) を m について偏微分し 0 とおくことに

より得られる線形方程式

$$S_q^\top W P_q W^\top S_q \Phi^{-1} m = S_q^\top W c \quad (30)$$

によって定められる。但し、 S_q (大きさ $DMT \times DMN$) は状態系列 q によその要素が 0 または 1 に定まる行列である。式 (30) の左辺の係数行列の要素数は $DMN \times DMN$ となり、数千から数万に及ぶが、市販されている計算機で十分計算できる。

次に、式 (24) を ϕ について偏微分すれば、

$$\begin{aligned} \frac{\partial \log P(c, q | \lambda)}{\partial \phi} &= \frac{1}{2} S_q^\top \text{diag}^{-1} \left(W P_q W^\top \right. \\ &\quad \left. - W c c^\top W^\top + 2 \mu_q c^\top W^\top \right. \\ &\quad \left. + W \bar{c}_q \bar{c}_q^\top W^\top - 2 \mu_q \bar{c}_q^\top W^\top \right) \quad (31) \end{aligned}$$

を得る。式 (31) を 0 とおいた式を ϕ に関して直接解くことは容易でないが、勾配法を用いて ϕ の値を更新できる^(注1)。

4. 状態系列の探索

4.1 尤度の再帰計算

トラジェクトリ尤度を式 (4) を用いて計算するには、高次の線形演算 (逆行列、行列式など) が必要となる。これを避けるため、本節ではトラジェクトリ尤度を時間方向に再帰的に計算するアルゴリズムについて述べる。

式 (4) より、トラジェクトリ尤度は以下ようになる。

$$P(c | q, \lambda) = K_q^{-1} \cdot P(o | q, \lambda) \quad (32)$$

式 (19) において、 $|\Sigma_q|$ と $\mu_q^\top \Sigma_q \mu_q$ は、 Σ_q が対角共分散行列であるため各時刻独立に計算できるが、 $|P_q|$ と $r_q^\top P_q r_q$ は、 P_q が全共分散であるため独立に計算できない。しかし、 P_q の特別な性質を利用することで、 K_q を時間方向に再帰的に計算できる。

式 (13) より、

$$L = \max \left\{ L_-^{(1)}, L_+^{(1)}, \dots, L_-^{(D-1)}, L_+^{(D-1)} \right\} \quad (33)$$

とすれば、 R_q は $(4L+1)$ 重対角対称正定値行列となる。ここで、 R_q をコレスキー分解すると、

$$R_q = U_q^\top U_q, \quad (34)$$

となる。但し、 U_q は $(2L+1)$ 重上三角行列である。式 (34) より、 $|P_q|$ は以下のように変形できる。

$$|P_q| = |R_q|^{-1} = |U_q|^{-2} = \prod_{t=1}^T \left| U_{q_{t+L}}^{(t,t)} \right|^{-2} \quad (35)$$

ここで、 $q_{t+L} = \{q_1, \dots, q_{t+L}\}$ である。 $U_{q_{t+L}}^{(t,t)}$ は時刻 1 から $t+L$ に滞在する状態にのみ依存するため、 $|P_q|$ は時間方向に再帰的に計算できる。

次に、式 (16)、(17)、(34) より、 $r_q^\top P_q r_q$ は以下のように変形できる。

$$r_q^\top P_q r_q = r_q^\top P_q^\top R_q P_q r_q = \bar{c}_q^\top U_q^\top U_q \bar{c}_q \quad (36)$$

$$= g_q^\top g_q \quad (g = U_q \bar{c}_q = U_q^{-1} r_q) \quad (37)$$

$$= \sum_{t=1}^T \left[g_{q_{t+L}}^{(t)} \right]^\top \cdot g_{q_{t+L}}^{(t)} \quad (38)$$

但し、

$$g_q = U_q^{-1} r_q = U_q \bar{c}_q \quad (39)$$

である。 $g_{q_{t+L}}^{(t)}$ は時刻 1 から $t+L$ に滞在する状態にのみ依存することから、 $r_q^\top P_q r_q$ も時間方向に再帰的に計算できる。よって、トラジェクトリ尤度は以下のように再帰的に計算できる。

$$P(c | q, \lambda) = \prod_{t=1}^T \frac{1}{K_{q_{t+L}}^{(t)}} \cdot P(o_t | q_t, \lambda), \quad (40)$$

但し、

$$\begin{aligned} K_{q_{t+L}}^{(t)} &= \frac{\sqrt{(2\pi)^M} \left| U_{q_{t+L}}^{(t,t)} \right|^{-1}}{\sqrt{(2\pi)^{3M} |\Sigma_{q_t}|}} \\ &\quad \cdot \exp \left\{ -\frac{1}{2} \left(\mu_{q_t}^\top \Sigma_{q_t}^{-1} \mu_{q_t} - \left[g_{q_{t+L}}^{(t)} \right]^\top g_{q_{t+L}}^{(t)} \right) \right\}. \quad (41) \end{aligned}$$

である。

4.2 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 は時間方向の共分散行列 P_q が全共分散であるため、動的計画法によって最適状態系列を得ることが困難である。そこで本節では、“準最適” な状態系列を探索するアルゴリズムについて述べる。このアルゴリズムは、トラジェクトリ尤度に基づき時刻 t において状態 q_{t-J} を決定するアルゴリズムとなっており、 J フレーム遅れて状態を決定する Viterbi アルゴリズムとみなせる。

提案する Viterbi アルゴリズムを以下に示す。

0) 初期化: $t = 1; \forall q_{1-J}^{1+L}$

$$\delta_1 \left(q_{1-J}^{1+L} \right) = \pi_{q_1} b \left(q_{1+L} \right)$$

$$\psi_1 \left(q_{1-J}^{1+L} \right) = 0$$

1) 繰り返し: $t = 2, \dots, T; \forall q_{t-J}^{t+L}$

$$\delta_t \left(q_{t-J}^{t+L} \right) = \max_{q_{t-J-1}} \left[\delta_{t-1} \left(q_{t-J-1}^{t+L-1} \right) a_{q_{t-1}, q_t} \right] b \left(q_{t+L} \right)$$

$$\psi_t \left(q_{t-J}^{t+L} \right) = \arg \max_{q_{t-J-1}} \left[\delta_{t-1} \left(q_{t-J-1}^{t+L-1} \right) a_{q_{t-1}, q_t} \right]$$

2) 終了:

$$\hat{P} = \max_{q_{T-J}^{T+L}} \left[\delta_T \left(q_{T-J}^{T+L} \right) \right]$$

$$\hat{q}_{T-J}^{T+L} = \arg \max_{q_{T-J}^{T+L}} \left[\delta_T \left(q_{T-J}^{T+L} \right) \right]$$

$$= \{ \hat{q}_{T-J}, \dots, \hat{q}_{T+L} \}$$

(注1): 実装では、分散の値は常に正の値であるという制約が必要であるため、対数領域において更新している。

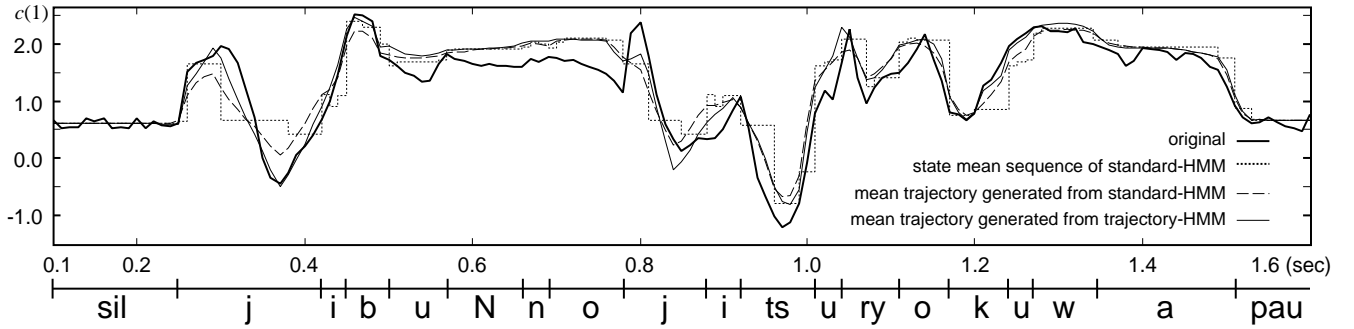


図1 実発声と平均系列の例(太線:実発声の系列,点線:通常のHMMの平均系列,破線:通常のHMMより生成された平均系列,細線:再学習したトラジェクトリHMMより生成された平均系列)

Fig.1 Generated trajectories \bar{c}_q for the trajectory HMM (thin line) and the standard-HMM (broken line), the state mean sequence μ_q (dotted line), and one of training data c (thick line).

3) バックトラック: $t = T, \dots, J + 2$

$$\hat{q}_{t-J-1} = \psi_t(\hat{q}_{t-J}^{t+L})$$

$$\hat{q}_{t-J-1}^{t+L-1} = \{\hat{q}_{t-J-1}, \dots, \hat{q}_{t+L-1}\}$$

ここで, π_{q_1} は初期状態確率, a_{q_{t-1}, q_t} は状態遷移確率, q_{t-J}^{t+L} は時刻 $t - J$ から $t + L$ までに滞在した状態系列であり,

$$q_{t-J}^{t+L} = \{q_{t-J}, \dots, q_{t+L}\} \quad (42)$$

である. また $b(q_{t+L})$ は,

$$b(q_{t+L}) = K_{q_{t+L}}^{(t)} \cdot P(o_t | q_t, \lambda) \quad (43)$$

で与えられる.

ここで, $J = T$ であれば, 提案した Viterbi アルゴリズムにより最適な状態系列を得ることができる. J の値が大きいくほどより尤度の高い状態系列を得ることができるが, 非常に大きな記憶容量と計算量が必要になる. このため, J の値はその性能と計算量のバランスを考慮して決定する必要がある. 音声のモデル化においては, 調音結合が影響を及ぼす時間が 100ms から 200ms と考えられることから, 10ms のフレーム周期のとき J が 10 以上であればほぼ最適な状態系列を選択できると思われる.

5. 実験

5.1 実験条件

トラジェクトリ HMM とその学習アルゴリズムを評価するため, 特定話者連続音声認識実験を行った. ATR 日本語音声データベース b-set, 話者 MHT の 503 文章中 450 文章を学習に用い, 学習データに含まれない 53 文章を評価に用いた. 音声の分析条件は, サンプリング周波数 16kHz, フレーム長 25ms (ブラックマン窓), フレーム周期 10ms であり, 18 次のメルケプストラム分析 [28] を用いて特徴抽出を行った. 特徴ベクトルとしては 0~18 次のメルケプストラム, Δ メルケプストラム, Δ^2 メルケプストラムから成る 57 次元を用いた. 動的特徴量は前後 1 フレームからの回帰係数として計算した. モデルのトポロジーは 3 状態の left-to-right モデル (スキップなし) で, 各状態の出力確率分布は対角共分散単一ガウス分布とした.

トラジェクトリ HMM の学習は以下のように行った.

表1 通常の HMM とトラジェクトリ HMM の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

Table 1 Average log “trajectory likelihood” of standard-HMM and trajectory-HMM for training data and test data

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

(1) 通常の HMM を学習し, これを初期モデルとする.

(2) 4 節の Viterbi アルゴリズムを用いて状態系列 q を決定する.

(3) 得られた q を用いて, 式 (30), (31) に基づき m と ϕ を更新する^(注2).

初期モデルとした HMM は, HTK [29] を用いて学習したコンテキスト非依存 HMM である. 繰り返し学習は行っていない.

認識には, 音響尤度のリスクを用いた. 初期モデルとした通常の HMM を用いて, 日本語の音素接続の制約を考慮した音素ネットワーク^(注3)上で, HTK [29] の Viterbi デコーダにより上位 200 個の音素仮説を出力した. 各仮説に対して, 4 節の Viterbi アルゴリズムを用いて状態系列 q を探索することにより, トラジェクトリ尤度を再評価した.

学習・認識における状態系列探索時の Viterbi アルゴリズムは, $J = 4$ として探索した.

5.2 実験結果

表1は, 初期モデルとした通常の HMM 及び学習したトラジェクトリ HMM の学習データ, 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である. 導出した更新式によりトラジェクトリ HMM を学習することにより, 学習データ及び評価データに対する尤度が向上していることが確認できる.

図1は学習データ中の一発話に対する, 平均系列の例である. 通常の HMM の平均系列は状態内で一定であるため, 状態の切り替わりの部分で平均系列が不連続に変化している. これに対して, 通常の HMM をトラジェクトリ HMM とみなして得られる平均トラジェクトリ系列, 及び提案法で学習したトラジェクトリ HMM の平均トラジェクトリ系列は, 状態内及び状態間

(注2): ϕ の更新には最急降下法を用いた.

(注3): 言語モデルの影響を除くため, 音素 bi-gram 等は用いていない.

表 2 実験結果

Table 2 recognition experimental results

	%Err.	#Corr.	#Del.	#Sub.	#Ins.	#Total
baseline	20.1%	1855	119	200	119	2174
proposed	17.2%	1903	118	153	103	2174

で連続的に変化していることが確認できる。加えて、提案法により学習したトラジェクトリ HMM の平均トラジェクトリ系列は、通常の HMM をトラジェクトリ HMM とみなして得られる平均トラジェクトリ系列と比較して、より実発声のトラジェクトリ系列に近付いていることが確認できる。また、音素/j/は図 1 中に 2 度出現しているが、通常の HMM の平均系列は近傍の音素に何ら影響を受けないため、どちらの場合も同じ値である。これに対してトラジェクトリ HMM では、近傍の音素によりその統計量が動的に変化していることが確認できる。これよりトラジェクトリ HMM は、調音結合などによるスペクトルの変化を自然にモデル化できる可能性がある。

表 2 に実験結果を示す。認識実験の結果、提案したトラジェクトリ HMM を用いてリスクアすることで、通常の HMM を用いた場合と比較して、約 15%の誤り削減率が得られた。

6. まとめ

本稿では、静的及び動的特徴量間の明示的な関係を導入することで、HMM をトラジェクトリモデルとして再定義した。また、得られたトラジェクトリモデルの Viterbi 学習型式の学習アルゴリズムを導出し、連続音声のモデル化に適用した。連続音声認識実験の結果、約 15%の誤り削減率が得られた。

今後の課題として、Baum-Welch (EM) 型式の学習アルゴリズムの導出、Viterbi デコーダの実装、より規模の大きい評価実験、HMM 音声合成システムへの適用があげられる。

謝辞 有意義な討論をして下さった東京工業大学小林隆夫教授、益子貴史助手、名古屋工業大学南角吉彦氏に感謝致します。

文 献

- [1] M. Ostendorf, V. Digalakis, and O.A. Kimball, "From HMMs to segment models," *IEEE Trans. on Speech & Audio Process.*, vol.4, no.5, pp.360-378, 1996.
- [2] 中川 聖一, "音声認識研究の動向", 信学論 (D-II), vol.83-D-II, no.2, pp.433-457, 2000.
- [3] 中川 聖一, "音声認識において HMM とトライグラムを超えるもの", 人工知能学会誌, no.17, pp.35-40, 2002.
- [4] L. Deng, M. Aksmanovic, X. Sun, and J. Wu, "Speech recognition using hidden Markov models with polynomial regression functions as nonstationary states," *IEEE Trans. on Speech & Audio Process.*, vol.2, no.4, pp.507-520, 1994.
- [5] L. Deng and M. Aksmanovic, "Speaker-independent phonetic classification using hidden Markov models with mixture of trend functions," *IEEE Trans. Speech & Audio Process.*, vol.5, no.4, pp.319-324, 1997.
- [6] H. Gish and K. Ng, "Parametric trajectory models for speech recognition," *Proc of ICSLP'96*, vol.I, pp.466-469, 1996.
- [7] H.B. Richards and J.S. Bridle, "The HDM : a segmental hidden dynamic model of coarticulation", *Proc. of ICASSP'99*, vol.I, pp.357-360, 1999.
- [8] J.S. Bridle, L. Deng, J. Picone, H.B. Richards, J. Ma, T. Kamm, M. Shuster, S. Pike, and R. Regan. "An investigation of segmental hidden dynamic models of speech coarticulation for automatic speech recognition," In CLSP/JHU

- Summer Workshop on Language Engineering, 1998.
- [9] J. Ma, "Spontaneous speech recognition using statistical dynamic models for the vocal tract resonance dynamics," Ph.D thesis, University of Waterloo, Canada, 2000.
- [10] J.-L. Zhou, F. Seide, and L. Deng, "Coarticulation modeling by embedding a target-directed hidden trajectory model into HMM - model and training," *Proc. of ICASSP2003*, vol.I, pp.744-747, 2003.
- [11] L. Deng, I. Bazzi, and A. Acero, "Tracking vocal tract resonances using an analytical nonlinear predictor and a target-guided temporal constraint," *Proc. of Eurospeech2003*, pp.73-76, 2003.
- [12] Y. Minami, E. McDermott, A. Nakamura, and S. Katagiri, "A recognition method with parametric trajectory synthesized using direct relations between static and dynamic feature vector time series," *Proc. of ICASSP2002*, vol.I, pp.957-960, 2002.
- [13] Y. Minami, E. McDermott, A. Nakamura, and S. Katagiri, "Recognition method with parametric trajectory generated from mixture distribution HMMs," *Proc. of ICASSP2003*, vol.I, pp.124-127, 2003.
- [14] 南 泰浩, "音声生成モデルを考慮した音声認識" 日本音響学会誌, vol.59, no.11, pp.682-687, 2003.
- [15] P.F. Brown, "The acoustic-modelling problem in automatic speech recognition," *Technical Report of IBM T.J. Watson Research Centre*, 1987.
- [16] C.J. Wellekens, "Explicit correlation in hidden Markov model for speech recognition," *Proc. of ICASSP'87*, pp.383-386, 1987.
- [17] 高橋 敏, 松岡 達雄, 南 泰浩, 鹿野 清宏, "フレーム間相関を利用した音韻 HMM による音声認識", 信学論 (A), vol.J77-A, no.2, pp.153-161, 1994.
- [18] K.K. Paliwal, "Use of temporal correlation between successive frames in hidden Markov model based Speech recognizer," *Proc. of ICASSP'93*, pp.215-218, 1993.
- [19] G. Qing, Z. Fang, W. Jian, and W. Wenhui, "A new method used in HMM for modeling frame correlation," *Proc of ICASSP'99*, vol.I, pp.169-172, 1999.
- [20] 古山 純子, 小林 哲則, "部分隠れマルコフモデルによる単語音声認識", 信学論 (D-II), J83-D-II, no.11, pp.2379-2387, 2000.
- [21] 小川 哲司, 小林 哲則, "部分隠れマルコフモデルによる連続音声認識", 信学技報, SP2002-40, 2002.
- [22] M.J.F. Gales and S.J. Young, "The theory of segmental hidden Markov models," *Technical Report of CUED/F-INFENG*, no.133, Cambridge University, UK, 1993.
- [23] M. Ostendorf and S. Roukos, "A stochastic segment model for phoneme-based continuous speech recognition," *IEEE Trans. Acoust., Speech & Signal Process.*, vol.37, no.12, pp.1857-1869, 1989.
- [24] G. Zweig, "Speech recognition using dynamic Bayesian networks," Ph.D. thesis, University of California, Berkeley, 1998.
- [25] S. Furui, "Speaker independent isolated word recognition using dynamic features of speech spectrum," *IEEE Trans. Acoust., Speech, & Signal Process.*, vol.34, pp.52-59, 1986.
- [26] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, and T. Kitamura, "Speech parameter generation algorithms for HMM-based speech synthesis," *Proc. of ICASSP2000*, vol.3, pp.1315-1318, 2000.
- [27] P.A. Olsen and R.A. Gopinath, "Modeling inverse covariance matrices by basis expansion," *Proc of ICASSP2002*, vol.I, pp.945-948, 2002.
- [28] T. Fukada, K. Tokuda, T. Kobayashi and S. Imai, "An adaptive algorithm for mel-cepstral analysis of speech," *Proc. of ICASSP'92*, vol.I, pp.137-140, 1992.
- [29] <http://htk.eng.cam.ac.uk/>.