

## 特徴量間で状態遷移が非同期な HMM

松田 繁樹 中井 満 下平 博 嵯峨山 茂樹

北陸先端科学技術大学院大学 情報科学研究科

〒 923-1292 石川県能美郡辰口町旭台 1-1

URL <http://www-ks.jaist.ac.jp/index-j.html>

あらまし 音声特徴量間の時間構造に着目した非同期遷移型 HMM(AT-HMM)を提案する。一般に複数の音声特徴量(たとえばケプストラムの各係数)は必ずしも同期的に時間変化しない。従来の HMM ではこれらの特徴ベクトルとしてまとめ、状態遷移により同期的に取り扱ってきた。我々は、時間変化が同期する特徴量に対しては同時に状態遷移し、同期しない特徴量は別々に状態遷移するモデルの方が、より効果的に表現できると考えた。本報告では AT-HMM の多種類の構造を考察し、特に特徴量間で完全に非同期で、音素境界では同期するタイプの AT-HMM について実験評価したところ、従来の HMM と比較して音素認識で約 20% の誤り削減率、孤立単語認識で 40% 以上の誤り削減率が得られた。

キーワード ●音声認識 ●隠れマルコフモデル ●隠れマルコフ網 ●非同期遷移型 HMM

## Asynchronous-Transition Hidden Markov Model

Shigeki MATSUDA Mitsuru NAKAI

Hiroshi SHIMODAIRA Shigeki SAGAYAMA

School of Information Science

Japan Advanced Institute of Science and Technology, HOKURIKU

1-1 Asahidai, Tatsunokuchi, Ishikawa 923-1292

URL <http://www-ks.jaist.ac.jp/>

**Abstract** A new class of hidden Markov model (HMM) called asynchronous-transition HMM (AT-HMM) is proposed to model the asynchronous temporal structure of acoustic feature sequences (e.g., cepstrum coefficients). Temporally changing patterns of acoustic features are not necessarily synchronized with each other, while conventional HMMs treat feature vectors, i.e., synchronized feature sequences. More efficient representation is expected in synchronized and asynchronized state transition modeling depending on synchrony among the features. Various types of AT-HMM are discussed according to optional constraints on state transition of the models. In this report, we focus on a particular AT-HMM type that synchronizes the features at phoneme boundaries. Speaker-dependent speech recognition experiments demonstrated error reduction rates of approximately 20% and more than 40% in isolated word recognition and in phoneme recognition, respectively, compared with conventional HMMs.

**Key words** ●speech recognition ●hidden Markov model ●hidden Markov network ●asynchronous transition HMM

# 1 まえがき

隠れマルコフモデル (HMM) は、状態遷移確率と出力確率分布から構成されている。これまで行われてきた HMM のモデル化能力向上のための研究は、主に出力確率分布に対する検討であった。出力確率分布の表現法として、複数のガウス分布の混合を用いる連続混合出力分布型 HMM[1] や各特徴の離散分布の混合分布を用いる離散混合出力分布型 HMM[2] などが提案されている。しかし、一般に出力分布の形状を精密に表現するためには、多くの自由パラメータが必要となる。HMM は統計的モデルであるため、自由パラメータ数の増大は統計的信頼性の低下に繋がる。そこで、類似した特性を持つパラメータを共有化することにより自由パラメータを削減する手法 [3-7] が提案されている。

HMM は、全ての特徴量が時間方向に同一のタイミングで状態遷移することを前提としているモデルである。各特徴量間の時間変化が非同期な信号は、特徴量毎に別々に状態遷移し、時間変化が同期している信号に対しては、同時に状態遷移した方がより効果的に特徴ベクトルの時系列信号を表現できると考えられる。本報告では、特徴量間の同期非同期構造をモデル化する、非同期遷移型 HMM (Asynchronous-Transition HMM: AT-HMM) を新たに提案する。また、音素モデル内では非同期、音素間では同期を取る部分非同期な AT-HMM の検討を行ない、特定話者音素認識実験及び孤立単語認識実験により音声認識性能を評価する。

特徴量間の非同期性をモデル化する AT-HMM とは目的も着目している点も異なるが、特徴量毎の HMM を用いる手法としてはマルチバンド (複合周波数帯域型) 音声認識 [8-10] が提案されている。この手法は、各周波数バンド毎に認識を行なうことにより、ノイズの混じったバンドの影響を除外することを狙った手法である。

第 2 章では、AT-HMM の全体の枠組と音素モデルの生成法を説明する。第 3 章では、AT-HMM の認識性能を確認するため、特定話者音素認識実験及び、孤立単語認識実験を行なう。第 4 章は、まとめである。

## 2 非同期遷移型 HMM

### 2.1 時間構造に着目した新しい HMM

従来の HMM は、特徴ベクトルの時系列信号を、定常信号源を持つ状態のマルコフ過程としてモデル化している。これは、全ての特徴量が時間方向に同一のタイミングで状態遷移するモデルとして

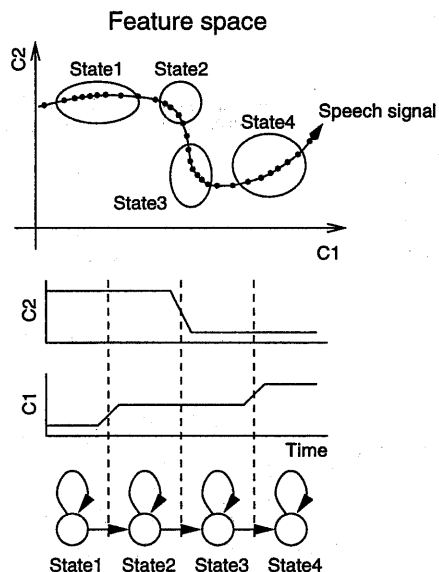


図 1: 従来型 HMM による音声特徴時系列のモデル化

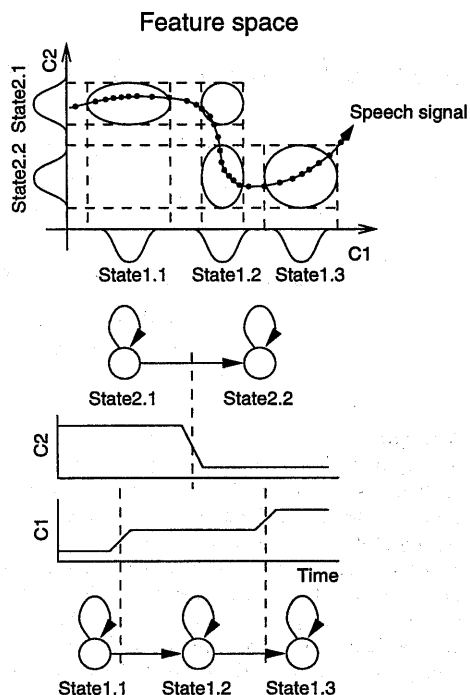


図 2: 特徴量毎の HMM による音声特徴時系列のモデル化

見ることができる。しかし、音声認識に一般に用いられる特徴ベクトル (LPC ケプストラム係数など) の時系列信号を観察すると、各特徴量の時間変化は必ずしも同じタイミングで起こっている訳ではない。各特徴量間の時間変化が非同期な信号は特徴量毎に別々に状態遷移し、時間変化が同期している信号に対しては同時に状態遷移した方が、より効果的に特徴ベクトルの時系列信号を表現できると考えられる。

従来型 HMM を用いて 2 次元の時系列信号を対角共分散行列を持つガウス分布でモデル化した例を図 1 に、特徴量毎にスカラー出力型 HMM を用いてモデル化した例を図 2 に示す。図 1 の従来型 HMM によりモデル化した例では、モデルの各状態は、平均ベクトル、分散ベクトル、自己遷移確率を持つので、状態当りの自由パラメータ数は 5 である。4 つの状態でモデル化しているため、全体として 20 の自由パラメータを持つ。図 2 の特徴量毎にスカラー出力型 HMM によりモデル化した例では、各状態は、平均値、分散値、自己遷移確率を持つので自由パラメータ数は 3 である。5 つの状態でモデル化しているため、全体として 15 の自由パラメータを持つ。このように、特徴ベクトルの各特徴量の時間変化が互いに非同期な信号に対しては、特徴量毎にモデル化した方が、従来型 HMM によるモデル化より少ない自由パラメータ数で時系列信号を表現することができると考えられる。

本報告では、このような特徴量間の時間構造に着目した新しい HMM を非同期遷移型 HMM (AT-HMM) として提案する。

### 2.1.1 特徴量間の同期/非同期性の分類

各特徴量間の同期非同期性を表現する種々の方法を図 3 に示す。

- (A) 完全同期 (Synchronous) HMM  
全ての特徴量の状態 (分布) が同期して遷移するタイプで、従来型の HMM はこれに当たる。
- (B) 部分非同期 (Partially asynchronous) HMM  
同期点と非同期点を混在させる方法である。混在方法により次のように大きく 3 つに分類できる。
  - (B-1) 特徴グルーピング (Feature grouping)  
時間変化が類似している特徴量をグループ化し、グループ内では同期、グループ間では非同期に遷移するタイプである。
  - (B-2) モデル同期 (Model synchronous)  
音素、単語などのモデル境界で同期し、モデル内部では非同期に行なう手法である。
  - (B-3) 遷移部分共有 (Transition tying)  
以上に述べた二種の同期の導入原理とは異なる、より一般的に同期と非同期が混在するタイプである。同期拘束を持つ方がモデルとして効率的である状態遷移の組合せは同期し、同期拘束を持たない方が効率的である組合せには非同期であるように同期拘束を用いることにより、効率的なモデル化を行うものである。
- (C) 完全非同期 (Fully asynchronous) HMM  
各特徴量間の時間変化は完全に無相関と考え、特徴量毎に別々に遷移するタイプである。

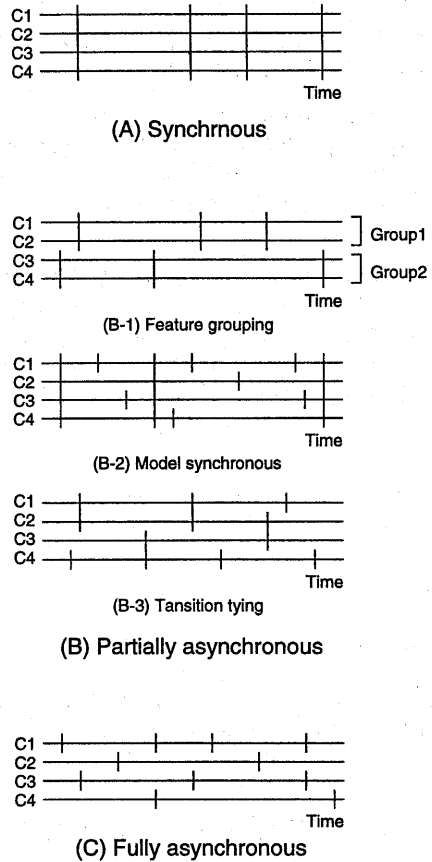


図 3: 特徴量間の非同期性のモデル化。縦線は状態の境界、横軸は時間、C1 から C4 は特徴量を表している

### 2.1.2 非同期な状態遷移の順序制約の分類

各特徴量の状態遷移の順序制約をどのように設けるか、によって別種の分類ができる。

- (I) 順序付き (Ordered) 非同期遷移 HMM  
各特徴量の状態遷移に順序関係を持たせる手法である。図4の非同期性を持つ信号は、 $a \rightarrow b \rightarrow c \rightarrow d$  と状態遷移が起こり、それ以外の順序での状態遷移は起こらない。
- (II) 順序無し (Non-ordered) 非同期遷移 HMM  
状態遷移に順序関係は無く、他の特徴量の状態遷移とは無関係である。図4の信号は、第1特徴量は  $a \rightarrow c$ 、第2特徴量は  $b$ 、第3特徴量は  $d$ 、それぞれ別々に状態遷移が起こる。

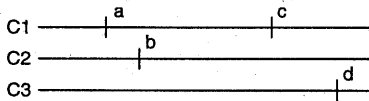


図4: 順序制約あり/なしの非同期状態遷移

このように AT-HMM は、特徴量間の同期非同期性と状態遷移の順序関係を組み合わせることにより様々なモデル化が可能である。

## 2.2 非同期遷移型 HMM の実現法

### 2.2.1 順序付きの非同期遷移型 HMM

状態遷移に対して、順序付きの同期非同期性を持たせた順序付き AT-HMM の実現法を説明する。これは、各特徴量の状態遷移に対して順序関係の拘束を付けたモデルである。図5に順序付きの AT-HMM の構造を示す。

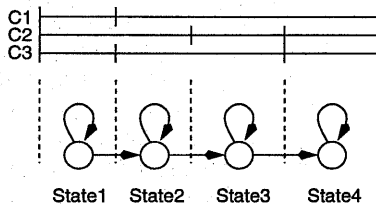


図5: 順序付きの非同期遷移型 HMM の構造

順序付き AT-HMM は従来型 HMM の特徴パラメータレベルの共有として表現される。図5の HMM は、状態2, 3, 4の第1特徴量に対応する

出力確率分布が共有されている。同様に、状態1, 2の第2特徴量、状態2, 3の第3特徴量も共有されている。この結果、最初に第1特徴量と第3特徴量に関して状態遷移が同期して起こった後に、第2特徴量のみに関して状態遷移し、ついで第2特徴量と第3特徴量に関して状態遷移が同期して起こるといったような時間構造を表現することができる。

このタイプは、従来型 HMM において隣合った状態間で特徴パラメータを共有する (tying) ことにより実現でき、従来型 HMM の枠組で特徴ベクトルの同期非同期性を表現することができる。そのため、既存の HMM に対する様々な手法や音声認識システムへの応用がしやすいことから、我々は有望な手法と考えており、現在研究中である。また、これは時間方向の共有構造という新しいモデル構造の概念につながるものである。

### 2.2.2 順序制約無しの非同期遷移型 HMM

各特徴量間の状態遷移に対して順序制約の無い AT-HMM を実現する手法について説明する。特に、部分非同期 (B) の AT-HMM におけるモデル同期 (B-2) の実現法を説明する。モデル同期は、モデル内では非同期、モデル間では同期で状態遷移する手法である。モデル内の順序無しの非同期性を表現するには、各特徴量をスカラー出力型 HMM を用いてモデル化し、それぞれの HMM の尤度を別々に計算することにより実現できる。図7にスカラー出力型 HMM を用いた AT-HMM を示す。モデル間の同期は、同期を取るモデルを連結して大きな単位の AT-HMM を作成し、2段 DP 法を基礎とした手法を用いることにより各モデルの境界を求めることができる。単語/aki/に対する音素モデルの連結例を図6に示す。

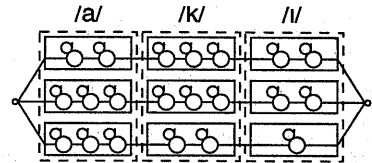


図6: 単語/aki/に対する非同期遷移型 HMM のモデル連結例

連結したモデルの尤度計算は、従来型 HMM では単純にモデルを連結し、全体を Forward アルゴリズムや Viterbi アルゴリズムを用いて計算すれば良い。しかし、AT-HMM はモデル境界を Viterbi 計算で求める事はできないため、大量の計算時間が掛かる。

スカラー出力型 HMM を用いた AT-HMM のモデルパラメータ  $\lambda$  に対する観測ベクトルの時系列信号  $O$  の観測確率  $p(O|\lambda)$  は、各特徴量に対応するスカラー出力型 HMM の観測確率の積として式 (1) により計算される。ここで、 $D$  は特徴ベクトルの次元数、 $\lambda_d$  は  $d$  番目の特徴量に対応するスカラー出力型 HMM のモデルパラメータ、 $O_d$  は観測ベクトルの  $d$  番目の特徴量に対応する 1 次元の時系列信号である。

$$p(O|\lambda) = \prod_{d=1}^D p(O_d|\lambda_d) \quad (1)$$

$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_D\}$$

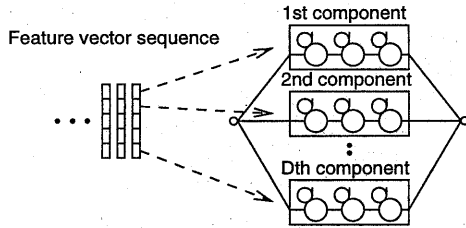


図 7: スカラー出力型 HMM を用いた非同期遷移型 HMM の構造

本報告では AT-HMM の効果を調べるため、従来型 HMM の対極に位置する順序無し完全非同期 (C) な AT-HMM の検討を行なう。しかし、音素単位の単語音声認識では、同一時刻に異なる音素の特徴量が出力されるような HMM は考えにくい。最低限の拘束条件として部分非同期 (B) の AT-HMM におけるモデル同期 (B-2) の検討を中心に行なう。

### 3 音素モデル生成法

本報告では、順序制約無し AT-HMM によるモデル同期 (B-2) の検討を行なう。モデル同期では、モデル内部は順序無し非同期遷移なので特徴量毎にスカラー出力型 HMM を用いて実現することができる。本章では、特徴ベクトルの非同期性を効果的に表現するための、順序無し AT-HMM の音素モデル生成法の説明を行なう。

#### 3.1 特徴量別逐次状態分割法による音素モデル生成

スカラー出力型 HMM による順序制約無し AT-HMM の音素モデル生成法を説明する。特徴ベクトルの各特徴量の時間変化が互いに非同期で異なる特性を持つ時、各特徴量に対応するスカラー出力型 HMM は異なる状態数と状態共有構造を持つと考えられる。しかし、各特徴量に対して最適な状態数と状態共有構造を解析的に求めることは一般に困難である。そこで各特徴量に対応するスカラー出力型 HMM の状態に対して逐次状態分割法 (SSS)[4-5] を基礎とした手法により、時間方向又は環境方向に状態を分割し、モデルを段階的に精密化する、特徴量別逐次状態分割法 (Feature-wise Successive State Splitting: FWSSS) を提案する。FWSSS 法の処理の流れを図 8 に示す。

- (1) AT-HMM の各特徴量に対応するスカラー出力型 HMM の状態全てに対して分割ゲインを計算する。
- (2) 最も分割ゲインの大きな状態を時間方向または、音素環境方向に分割を行なう。
- (3) 分割の影響を受けた状態を再学習する。

この処理を繰り返し行なうことにより、モデル尤度を徐々に上昇させる手法である。本報告で提案する FWSSS 法の分割ゲインの計算は MLSSS 法 [5] を一部変更した式を用いている。FWSSS 法における状態の分割ゲインの計算式を式 (2) に示す。

$$G(s, q_1, q_2) = -\frac{1}{D} N_2(s, s) \log a_{ss} + 0.5 N_1(s) \log \sigma(s) + \frac{1}{D} \sum_{s'=q_0 q_1} \sum_{s''=q_0 q_1} N_2(s', s'') \log a_{s's''} - \sum_{s'=q_0 q_1} 0.5 N_1(s') \log \sigma(s') \quad (2)$$

$$N_1(s') = \sum_t \gamma_t(s')$$

$$N_2(s', s'') = \sum_t \xi_t(s', s'') \quad (3)$$

ここで、 $D$  は特徴ベクトルの次元数、 $s$  は分割ゲインを計算するスカラー出力型 HMM の状態、 $q_1$  と  $q_2$  は状態  $s$  の分割後の状態、 $\gamma_t(s)$  は時刻  $t$  に状態  $s$  に存在する確率、 $\xi_t(s, s')$  は時刻  $t$  に状態  $s$  から状態  $s'$  へ遷移する確率、 $\sigma(s)$  は状態  $s$  の分散値である。右辺第 1 項、第 3 項の係数  $1/D$  の意味は以下の通りである。

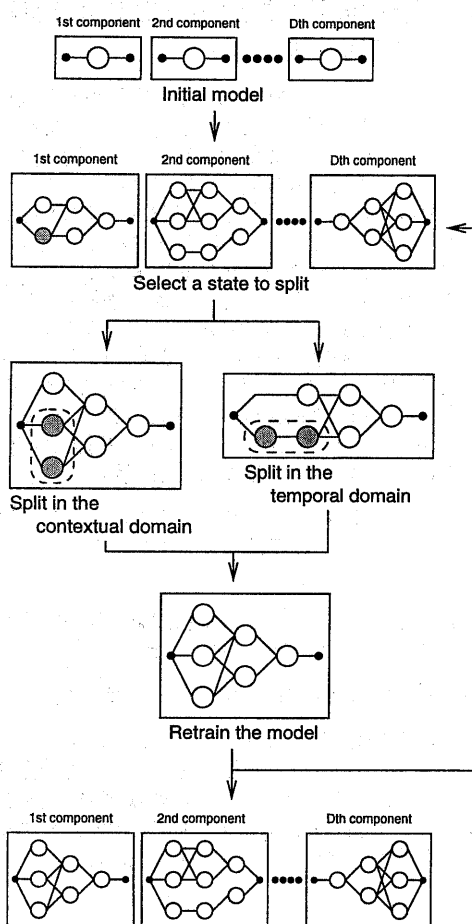


図 8: 特微量別逐次状態分割法 (FWSSS) の処理の流れ

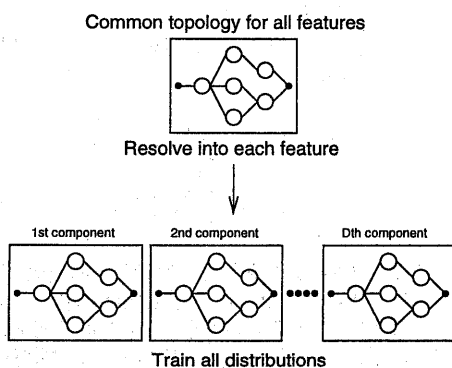


図 9: 次元分解逐次状態分割法 (DRSSS) の処理の流れ

HMM のモデル尤度は、状態遷移確率と状態出力確率の 2 つの確率成分により構成されている。本報告で提案した AT-HMM は状態遷移分の確率が従来の HMM と比較して次元数倍多くなっている。このため、状態遷移確率の影響が大きくなり時間方向分割が行なわれにくい。そこで FWSSS 法では、AT-HMM の状態遷移確率と状態出力確率の割合を従来の HMM と揃えるため、状態遷移確率分の分割ゲインを  $1/D$  として計算を行なう。

この FWSSS 法により、AT-HMM の各特微量に対応するスカラー出力型 HMM は適当な状態数を持つ状態共有構造を持つこととなり、各特微量の動きを効果的にモデル化できると考えられる。

### 3.2 次元分解法による音素モデル生成

SSS 法はトポロジー生成と学習を従来型 HMM (同期型) を対象として行なうのに対し、FWSSS 法は非同期な AT-HMM を対象としている。この 2 つの手法の中間に位置する手法として、トポロジー生成を同期型として行ない学習を非同期型として行なう手法が考えられる。

トポロジー生成を同期型、学習を非同期型で行なう方法を次元分解 SSS 法 (Dimension-Resolved SSS: DRSSS) と呼び、比較実験として行なう。DRSSS 法の処理の流れを図 9 に示す。

- (1) 従来型 HMM 用の音素モデルを SSS 法を用いて生成する。
- (2) 生成された HMnet のトポロジー (状態共有構造) を各特微量に対応するスカラー出力型 HMM に割り当てる。
- (3) それぞれのスカラー出力型 HMM を再学習する。

DRSSS 法により生成された AT-HMM の各特微量に対応するスカラー出力型 HMM の状態数と状態共有構造は同じである。参考に、SSS 法と本報告で提案した DRSSS 法、FWSSS 法の対応関係を表 1 に示す。

表 1: SSS 法, DRSSS 法, FWSSS 法の対応関係

	SSS 法	DRSSS 法	FWSSS 法
トポロジー生成	同期	同期	非同期
学習	同期	非同期	非同期

## 4 認識実験

AT-HMM によるモデル化の効果を確認するために、特定話者音素認識実験及び、特定話者孤立

表 2: 音素認識実験結果. MLSSS, DRSSS, FWSSS 法により生成した音素モデルのパラメータ数に対する誤り認識率と誤り削減率

method	# parameters	% error	% reduction
MLSSS	10600	6.59	0.00
DRSSS	10608	8.86	-34.4
FWSSS	10608	6.15	6.78
MLSSS	21200	4.29	0.00
DRSSS	21216	5.35	-24.7
FWSSS	21216	3.53	17.6
MLSSS	31800	3.76	0.00
DRSSS	31824	4.97	-32.2
FWSSS	31824	3.00	20.3
MLSSS	42400	3.58	0.00
DRSSS	42432	4.28	-19.6
FWSSS	42432	2.83	20.9

単語認識実験を行なった。実験データには、ATR 研究用日本語音声データベース A-set 中、男性 2 話者 (mht, mau) と女性 2 話者 (fms, ffs) の重要語 5240 単語と音韻バランス 216 単語を用いた。音素ラベルは /N, a, b, ch, d, e, f, g, h, i, j, k, m, n, o, p, q, r, s, sh, t, ts, u, w, y, z/ の計 26 音素と無音を用いた。サンプリング周波数 20kHz のデータをハミング窓 25ms, フレーム周期 8ms で分析した。特徴パラメータは、対数パワー, 12 次メルケプストラム, Δ対数パワー, 12 次Δメルケプストラムの計 26 次元を使用した。

#### 4.1 特定話者音素認識実験

特定話者音素認識実験により、FWSSS 法及び DRSSS 法を用いて生成した音素モデルの音素認識性能を評価する。従来型 HMM の音素認識性能と比較するために、MLSSS 法を用いて生成した音素モデルの音素認識実験を合わせて行なった。初期 HMnet は各音素 1 状態を用い、時間方向分割を最大 3 状態に制限した。学習用データは、重要語 5240 単語中奇数番目の 2620 単語と音韻バランス 216 単語を用いた。評価用データには、偶数番目の 2620 単語を用いた。音素サンプルは、ラベルファイルを基に切り出し、前後の音素環境を既知とした。モデル尤度は Forward アルゴリズムを用いた。また、パラメータ数は 10600(従来 HMM で 200 状態), 21200(400 状態), 31800(600 状態), 42400(800 状態) で認識実験を行なった。

各パラメータ数に対する誤り認識率と、MLSSS 法を基準とした時の誤り削減率を表 2 に示す。FWSSS 法はパラメータ数約 10600 では同等以上、パラメータ数 21200 以上では約 20% 以上の誤り削減率が得られた。DRSSS 法では、大きく認識率が低下している。

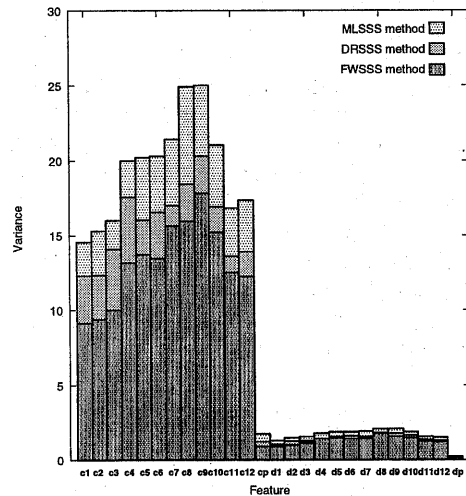


図 10: MLSSS, DRSSS, FWSSS 法により生成した音素モデルの、各特徴量に対応する分散値の平均

表 3: 孤立単語認識実験結果. MLSSS, FWSSS 法により生成した音素モデルのパラメータ数に対する誤り認識率と誤り削減率

method	# parameters	% error	% reduction
MLSSS	10600	5.36	0.00
FWSSS	10608	2.86	46.6
MLSSS	21200	3.17	0.00
FWSSS	21216	1.68	47.0

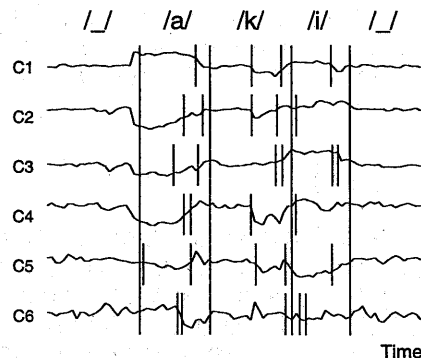


図 11: 単語/aki/に対して得られたメルケプストラム係数第 1 次から 6 次までの AT-HMM による状態セグメンテーションの例

各手法のモデル化の効果を確認するため、生成された音素モデルの各特徴量に対応する分散値の平均を図10に示す。MLSSS法よりも、FWSSS法により生成した音素モデルは分散値の平均は小さくなっており、特徴ベクトルの時系列信号の分布表現効率が向上したと考えられる。しかし、DRSSS法で生成した音素モデルは分散値の平均が小さくも関わらず、音素認識性能が低下している。この原因については検討中である。パラメータ数42400までのモデルの生成時間は、MLSSS法が408分であるのに対してFWSSS法では679分と約1.7倍程度多くかかる。使用計算機は、Alpha21264/500MHzである。

## 4.2 特定話者孤立単語認識実験

特定話者孤立単語認識実験により、FWSSS法を用いて生成した音素モデルの音素認識性能を評価した。従来のHMMの単語認識性能と比較するために、MLSSS法を用いて生成した音素モデルの単語認識実験を合わせて行なった。音素モデルは、特定話者音素認識実験で生成したモデルを用いた。単語辞書は重要語5240単語中偶数番目の2620単語を用い、その内の1310単語の認識実験を行なった。モデル尤度はForwardアルゴリズムを用いて計算し、音素境界での同期については2段DP法を基礎とする手法を用いて尤度計算を行った。パラメータ数は10600, 21200について認識実験を行なった。

各パラメータ数に対する誤り認識率と誤り削減率を表3に示す。FWSSS法はMLSSS法により生成された音素モデルに対し40%以上の誤り削減率が得られた。参考のため、単語/aki/に対して得られた第1次から6次までのメルケプストラム成分の状態に対応するViterbiセグメンテーションを図11に示す。音素内では非同期、音素間では同期して認識している様子が分かる。また、各特徴量に対応するスカラー出力型HMMの状態数も異なっている。

## 5 むすび

本報告では、音声特徴量時系列を効果的にモデル化するために、非同期遷移型HMM(AT-HMM)を提案した。AT-HMMの音素モデル生成法として、逐次状態分割法を基礎として各特徴量に対応するスカラー出力型HMMの状態を徐々に分割する特徴量別逐次状態分割法(FWSSS)を提案した。また、モデルを連結して尤度を計算する手法について論じた。AT-HMMの音声認識性能を調べるため、日本語全音素による特定話者音素認識実験と孤立単語認識実験を行なった。音素認識実験に

において、パラメータ数が21200(従来型HMMで400状態)以上の場合、従来型HMMの音素モデルに対し約20%の誤り削減率が得られた。孤立単語認識実験では、40%以上の誤り削減率が得られた。以上により、特徴量間で非同期な遷移を持つような構造のHMMが、音声認識性能向上の一手法としての可能性があることを示した。

今後はモデル同期(B-2)なAT-HMMに加えて、他の特徴グルーピング(B-1)や遷移部分共有(B-3)の実現法について検討を行なう。また特に、順序制約付きのAT-HMMのモデル生成法の検討と音声認識性能評価を行う予定である。

## 参考文献

- [1] B.H. Juang: "Maximum-likelihood estimation for mixture multivariate stochastic observations of Markov chains," AT&T Tech. J., 64, 6, pp.1234-1249, 1985.
- [2] 高橋 敏, 嵯峨山 茂樹: "離散混合出力分布型HMM," 音響講論, 2-3-2, pp.51-52, 1996.
- [3] 高橋 敏, 嵯峨山 茂樹: "4階層共有構造の音素HMM," 信学技報, SP94-73, 1994-12.
- [4] 鷹見 淳一, 嵯峨山 茂樹: "逐次状態分割法による隠れマルコフ網の自動生成," 信学論(D-II), J76-D-II, 10, pp.1255-2164, 1993-10.
- [5] M. Ostendorf, H. Singer: "HMM topology design using maximum likelihood successive state splitting," Computer Speech and Language, 11(1), pp.17-41, 1997.
- [6] X.D. Huang, K.F. Lee, H.W. Hon, M.Y. Hwang: "Improved Acoustic Modeling with the SPHINX Speech Recognition System," Proc. ICASSP91, pp.345-348, 1991.
- [7] M.Y. Hwang, X.D. Huang: "Subphonetic Modeling with Markov States - Senon," Proc. ICASSP92, pp.33-36, 1992.
- [8] H. Bourlard, S. Dupont: "A new ASR approach based on independent processing and recombination of partial frequency bands," Proc. ICSLP96, pp.426-429, 1996.
- [9] S. Tibrewala, H. Hermansky: "Sub-band based recognition of noisy speech," Proc. ICASSP97, pp.1255-1258, 1997.
- [10] 大川 茂樹, E. Bocchieri, A. Potamianos: "部分周波数帯域の特徴量の融合による雑音環境下音声認識," 音響講論, 3-6-2, pp.85-86, 1998.