

TR-IT-0152

混合分布セグメントモデルを用いた母音識別実験
Experimental Results on Vowel Classification
using Multiple Mixture Stochastic Segment
Models

谷口 泰広
Taniguchi Yasuhiro

深田 俊明
Fukada Toshiaki

1996.2.23

本報告では、確率的セグメントモデルを混合分布化する方法を述べる。混合分布モデルの作成方法として、(1) クラスタリング手法により混合分布を求める方法、(2) EM アルゴリズムを用いたモデルの再推定により混合分布を求める方法、(3) 平均に加えて分散も時間変化の軌跡としてモデル化し、EM アルゴリズムによるモデルの再推定を行なう方法という3つの方法を検討した。TIMIT データベースを用いた母音識別実験を行なった結果、セグメントモデルの混合数を増加させることにより、識別率が向上することが確かめられた。また、上記(3)の分散の時変モデル化は、(1),(2)に比べて高い識別性能が得られ、このモデル化が有効であることが示された。HMM との結果の比較により、一つの音素を1状態で平均軌跡の回帰次数を2次でモデル化したセグメントモデルは、5~7状態のHMMと同程度の識別性能を示すことが分かった。

目次

1	はじめに	1
2	確率的セグメントモデル	3
2.1	セグメントモデル	3
2.2	混合分布セグメントモデル	4
2.2.1	クラスタリングによるモデル化	5
2.2.2	EM アルゴリズムを用いたモデルの再推定	5
2.3	分散の時変モデル化	8
3	母音識別実験	10
3.1	実験条件	10
3.1.1	音声資料	10
3.1.2	特徴パラメータ抽出	10
3.1.3	尤度計算	10
3.2	実験結果及び考察	12
3.2.1	クラスタリングによるモデル化	12
3.2.2	再推定の効果	16
3.2.3	分散の時変モデル化	18
3.2.4	HMM との比較	20
4	まとめ	24
	謝辞	24
	参考文献	25
	付録	27

A	プログラム説明	27
A.1	HCopy	28
A.2	asegm	30
A.3	km-las	31
A.4	allinone	32
A.5	SRest	36
A.6	segclass	37
A.7	HList	37
A.8	lmdl	38
A.9	HRmIO	41

第 1 章

はじめに

現在まで成功を収めている音声認識システムのほとんどは、HMM に基づくものであり、これはこの先もしばらく利用されていくであろうと考えられる。その一方で、近年、HMM の欠点に対処するための新しいモデル化方法も数多く提案されてきている。HMM の制限として次の 2 つのものが挙げられる。

1. 継続時間モデルに弱い
2. 状態系列のもとで観測系列の独立性が仮定されている

HMM の状態継続時間長モデルは幾何学的な分布によって暗黙的に与えられているという第一番目の制限は、継続時間状態分布を明示的に導入することで処理されてきた。第二番目の制限である観測系列の条件付き独立の仮定は、実用的には有益であるが現実的ではないということが広く認識されている。

とりわけ、第二番目の問題に対して、すなわち特徴パラメータの相関を取り入れることを目的としたモデル化の研究が数多くなされており、例えば、HMM の平均値を時間の関数としてモデル化する方法^{[1][2]}、確率的セグメントモデルによる方法^{[3][4]} などがある。前者は、平均値を時間の絶対的な関数として HMM によりモデル化し、文献 [1] で単一分布によりモデル化する効果を、文献 [2] でこれを混合分布とした時の効果が報告されている。文献 [2] の実験では、3 状態の left-to-right HMM において、回帰次数を 0, 1, 2 として比較した結果は報告されているが、HMM の状態数を更に増加させた時の比較は報告されていない。一方、文献 [4] では、平均値を時間の相対的な関数として確率的セグメントモデルによりモデル化している点が文献 [1] と大きく異なる点である。文献 [4] においては、TIMIT データベースを用いた母音識別実験結果が報告されているが、単一混合分布の場合のみであり、これを多混合分布にした場合についての実験は行なわれていない。

本論文は、Gish らの方法によるモデル化を混合分布へ拡張し、母音識別実験を通してその効果を調べることを目的としている。またこのとき、従来広く用いられている HMM の状態数、混合数を多くの組合せにより実験することにより、確率的セグメントモデルと HMM との詳細な性能比較も行なっている。

本論文は4章から構成されており、2章では確率的セグメントモデルを混合分布としてモデル化する3つの方法を述べ、3章では2章で述べた3つの方法とHMMを用いた時の識別実験の結果を述べ、4章でまとめを行なう。

第 2 章

確率的セグメントモデル

2.1 セグメントモデル^{[4][5]}

時刻 t における D 次元の特徴ベクトル系列 (例えばケプストラム) を y_t とすると、長さ L のランダム長観測データ系列 $y_1^L = [y_1, \dots, y_L]$ がラベル a から生成される確率は、次式によって与えられる。

$$P(y_1, \dots, y_L, L|a) = P(y_1, \dots, y_L|L, a)P(L|a) \quad (2.1)$$

ここで、 $P(y_1, \dots, y_L|L, a)$ は出力確率を、 $p(L|a)$ は継続時間長分布を表す。いま出力確率分布が単一ガウス分布で与えられている時、 $P(y_1, \dots, y_L|L, a)$ は、

$$\begin{aligned} P(y_1, \dots, y_L|L, a) &= \prod_{t=1}^L f(y_t) \\ &= \prod_{t=1}^L \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_a|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (y_t - \mu_{at})^T \Sigma_a^{-1} (y_t - \mu_{at}) \right\} \end{aligned} \quad (2.2)$$

により求められる。ここで、 Σ_a はラベル a の $D \times D$ の分散行列を、 μ_{at} は、時刻 t における D 次元の平均ベクトルを表す。ここで、

$$\mu_{at} = z_L^t B_a \quad (2.3)$$

であり、平均軌跡 μ_{at} を R 次で表す場合、 B_a は、

$$B_a = [b_{a0}, b_{a1}, \dots, b_{aR}]^T \quad (2.4)$$

により、また z_L^t は $(R+1)$ 次元のベクトルとして、

$$z_L^t = \begin{cases} [1 \ 0 \ 0 \ \dots \ 0], & t=1 \\ [1 \ \frac{t-1}{L-1} \ (\frac{t-1}{L-1})^2 \ \dots \ (\frac{t-1}{L-1})^R], & 1 < t \leq L \end{cases} \quad (2.5)$$

で与えられる。以後、表記を簡単にするため、 a を除くことにする。

前述の時間伸縮において、ある観測セグメント Y_i に対する B と Σ をパラメータにもつセグメントモデルの平均の最尤推定値は、 $Z_L = [z_L^1, z_L^2, \dots, z_L^L]^T$ として、

$$B_i = Y_i Z_{L_i}^T [Z_{L_i} Z_{L_i}^T]^{-1}, \quad (2.6)$$

となり、分散は、

$$\Sigma_i = \frac{1}{L_i} \sum_{j=1}^{L_i} (y_{i,j} - B_i Z_{L_i})(y_{i,j} - B_i Z_{L_i})^T \quad (2.7)$$

で得られる。ここで、3つの統計量 (B_i, Σ_i, L_i) は、セグメント Y_i をセグメントモデルとして表すための十分な統計量となる。着目しているモデル a に対する全てのセグメントに対してこの統計量を求めると、モデルパラメータの最尤推定値は、

$$\hat{B} = \left[\sum_{i \in \mathcal{I}_a} B_i Z_{L_i} Z_{L_i}^T \right] \left[\sum_{i \in \mathcal{I}_a} Z_{L_i} Z_{L_i}^T \right]^{-1} \quad (2.8)$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{i \in \mathcal{I}_a} L_i \Sigma_i + \frac{1}{N} \sum_{i \in \mathcal{I}_a} L_i (B_i - \hat{B}) Z_{L_i} Z_{L_i}^T (B_i - \hat{B})^T \quad (2.9)$$

で与えられる^[5]。ここで、 $\mathcal{I}_a = \{i : a_i = a\}$ 、 $N = \sum_{i \in \mathcal{I}_a} L_i$ である。

2.2 混合分布セグメントモデル

ここでは、前節での議論を混合分布へ拡張することを考える。すなわち、式(2.2)で表される単一ガウス分布を以下のように混合数 M の多混合分布として表す。

$$\begin{aligned} f(y_t) &= \sum_{k=1}^M w_k f_k(y_t) \\ &= \sum_{k=1}^M w_k \mathcal{N}(y_t, B_k, \Sigma_k) \end{aligned} \quad (2.10)$$

ここで、

$$\mathcal{N}(y_t, B_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (y_t - z_L^t B_k)^T \Sigma_k^{-1} (y_t - z_L^t B_k) \right\} \quad (2.11)$$

であり、重み w_k の総和は

$$\sum_{k=1}^M w_k = 1 \quad (2.12)$$

を満たす。

式(2.10)における B_k, Σ_k, w_k を求める方法として、

- クラスタリング手法に基づくもの
- EM アルゴリズムを用いた再推定に基づくもの

が考えられる。以下、これらの計算方法について述べる。

2.2.1 クラスタリングによるモデル化

混合分布のセグメントモデルを作成する方法の一つとして、セグメントの集合として表される特徴量空間を尤度最大基準に基づいてクラスタリングすることが考えられる。すなわちセグメント集合を、次式の尤度最大化基準による LBG アルゴリズムで M 個にクラスタリングする^[5]。

$$P = -\frac{1}{2}L_i D \log(2\pi) - \frac{1}{2}L_i \log(|\Sigma_c|) - \frac{1}{2} \sum_{t=1}^{L_i} (y_t - \mu_c)^T \Sigma_c^{-1} (y_t - \mu_c) \quad (2.13)$$

ここで、 μ_c はクラスタの平均軌跡ベクトル、 Σ_c はクラスタの共分散行列である。

ここで各クラスタの重み w_n は、

$$w_n = \frac{N_n}{\sum_{i=1}^M N_i}, \quad n = 1, \dots, M \quad (2.14)$$

により求める。ここで、 N_i は、それぞれのクラスタ内の学習セグメント数を表す。

2.2.2 EM アルゴリズムを用いたモデルの再推定

多混合の場合、前節で述べたクラスタリング結果は、モデルパラメータの最尤推定値とはならない。そこで、HMM の場合と同様に、EM アルゴリズムを用いた繰り返し手法に基づいてパラメータを再推定し、最尤推定値を求める。

セグメントモデルにおけるモデルパラメータの再推定を行なうに際し、HMM の場合と同様に次式で与えられる Q 関数を導入する。

$$\begin{aligned} Q(\bar{\Phi}|\Phi) &= E[\log P(y_1^L, L, k|\bar{\Phi})|y_1^L, \Phi] \\ &= \sum_{k=1}^M P(k|y_1^L, \Phi) \log P(y_1^L, L, k|\bar{\Phi}) \\ &= \sum_{k=1}^M \frac{P(k, y_1^L|\Phi)}{P(y_1^L|\Phi)} \log P(y_1^L, L, k|\bar{\Phi}) \end{aligned} \quad (2.15)$$

ここで、 Φ 及び $\bar{\Phi}$ は、それぞれ現在のモデル、再推定後のモデルを表し、 k は混合分布のインデックスを表す¹。ここで、

$$\begin{aligned} \log P(y_1^L, L, k|\bar{\Phi}) &= \log P(y_1^L|\bar{\Phi}) + \log P(L|\bar{\Phi}) + \log P(k|\bar{\Phi}) \\ &= \sum_{t=1}^L \log \bar{f}_k(y_t) + \log P(L|\bar{\Phi}) + \sum_{t=1}^L \log \bar{w}_k \end{aligned} \quad (2.16)$$

である。式(2.16)において、右辺第一項は再推定後のモデルの出力確率の対数尤度を、第二項は再推定後のモデルから継続時間長 L を出力する対数尤度を、第三項は再推定後の k 番目のモデルが出力される対数尤度を表す。ここで、第二項は、着目しているパラメータ B_k, Σ_k, w_k によらない(すなわち再推定の有無によらない)一定の値をとるため、結局、式(2.15)の最大化は、

$$Q_1(\bar{\Phi}|\Phi) = \sum_{k=1}^M \frac{P(k, y_1^L|\Phi)}{P(y_1^L|\Phi)} \log P(y_1^L|\bar{\Phi}) \quad (2.17)$$

¹本論文では、一つの音響単位(例えば音素)に対して、一つのセグメントモデル(すなわち状態数が1となる)でモデル化することを考えているため、通常の HMM における EM アルゴリズムの状態の総和は出てこない。

を B_k, Σ_k について最大化し、

$$Q_2(\bar{\Phi}|\Phi) = \sum_{k=1}^M \frac{P(k, y_1^L|\Phi)}{P(y_1^L|\Phi)} \log P(k|\bar{\Phi}) \quad (2.18)$$

を w_k について最大化することと等価になる。

式(2.17)、(2.18)における $P(k, y_1^L|\Phi)/P(y_1^L|\Phi)$ は、現在のモデルパラメータ Φ を用いて計算できるため、定数として扱うことができる。今、時刻 t におけるこの値を $\gamma_{k,t}$ とすると、これは、前向き変数 α_t 及び後向き変数 β_t を用いて以下のように効率的に計算できる。

$$\gamma_{k,t} = \begin{cases} \frac{\alpha_t \beta_{t+1} w_k f_k(y_{t+1})}{P(y_1^L|\Phi)}, & t = 1, \dots, L-1 \\ \frac{\alpha_T}{P(y_1^L|\Phi)}, & t = L \end{cases} \quad (2.19)$$

前向き変数及び後向き変数は、HMMの場合と同様に以下の再帰的計算により求められる。

$$\alpha_t = \begin{cases} f(y_1), & t = 1 \\ \alpha_{t-1} f(y_t), & t = 2, \dots, L \end{cases} \quad (2.20)$$

$$\beta_t = \begin{cases} 1, & t = L \\ \beta_{t+1} f(y_{t+1}), & t = L-1, \dots, 1 \end{cases} \quad (2.21)$$

平均軌跡の計算方法

まず、平均軌跡の再推定値を求めることを考える。式(2.16)及び式(2.17)より、 m 番目の分布に対する平均軌跡ベクトル $\bar{B}_m = [\bar{b}_{m0}, \bar{b}_{m1}, \dots, \bar{b}_{mR}]^T$ の値は、式(2.17)を \bar{b}_{mr} に関して偏微分したものを零とおき、

$$\frac{\partial Q_1}{\partial \bar{b}_{mr}} = 0 \quad (2.22)$$

を解くことにより得られる。式(2.10)及び式(2.17)より式(2.22)は、

$$\frac{\partial Q_1}{\partial \bar{b}_{mr}} = \sum_{t=1}^L \gamma_{m,t} \frac{\partial \bar{f}_m(y_t)}{\partial \bar{b}_{mr}} \quad (2.23)$$

ここで、式(2.11)及び式(2.3)より、

$$\frac{\partial \bar{f}_m(y_t)}{\partial \bar{b}_{mr}} = \bar{f}_m(y_t) \bar{\Sigma}_m^{-1} (y_t - \bar{\mu}_{m,t}) \left(\frac{t-1}{L-1} \right)^r \quad (2.24)$$

となり、 $\bar{\Sigma}_m^{-1}$ は時刻 t とは独立の定数であることから、

$$\sum_{t=1}^L \gamma_{m,t} \left\{ y_t - \sum_{u=0}^R \bar{b}_{mu} \left(\frac{t-1}{L-1} \right)^u \right\} \left(\frac{t-1}{L-1} \right)^r = 0 \quad (2.25)$$

を求めればよいことになる。式(2.25)を簡単な表現形式で記述すると、

$$\sum_{u=1}^R C(u+r) \bar{b}_{mu} = V(r) \quad (2.26)$$

と表すことができる。ただし、

$$C(l) = \sum_{t=1}^L \gamma_{m,t} \left(\frac{t-1}{L-1} \right)^l, \quad l = 0, \dots, 2R \quad (2.27)$$

$$V(r) = \sum_{t=1}^L \gamma_{m,t} y_t \left(\frac{t-1}{L-1} \right)^r, \quad r = 0, \dots, R \quad (2.28)$$

である。ここで、 $C(l)$ は特徴パラメータの次元 d に寄らないスカラー定数であり、 $V(r)$ は次元依存のベクトル定数であることに注意されたい。結局、式(2.26) を満たす \bar{b}_{mu} は、

$$\begin{bmatrix} C(0) & C(1) & \dots & C(R) \\ C(1) & C(2) & \dots & C(R+1) \\ \vdots & \vdots & \ddots & \vdots \\ C(R) & C(R+1) & \dots & C(2R) \end{bmatrix} \begin{bmatrix} \bar{b}_{m0}^{(d)} \\ \bar{b}_{m1}^{(d)} \\ \vdots \\ \bar{b}_{mR}^{(d)} \end{bmatrix} = \begin{bmatrix} V^{(d)}(0) \\ V^{(d)}(1) \\ \vdots \\ V^{(d)}(R) \end{bmatrix} \quad (2.29)$$

を $d = 1, \dots, D$ について解くことにより得られる。

分散の計算方法

次に、分散の再推定値の計算方法について述べる。平均軌跡の計算と同様に、式(2.16) 及び式(2.17) から、 m 番目の分布に対する分散 $\bar{\Sigma}_m$ の値は、式(2.17) を i 行 j 列の要素 $\bar{\sigma}_{mij}$ について偏微分したものを零とおき、

$$\frac{\partial Q_1}{\partial \bar{\sigma}_{mij}} = 0 \quad (2.30)$$

を解くことにより得られる。式(2.10) 及び式(2.17) より式(2.30) は、

$$\frac{\partial Q_1}{\partial \bar{\sigma}_{mij}} = \sum_{t=1}^L \gamma_{m,t} \frac{\partial \bar{f}_m(y_t)}{\partial \bar{\sigma}_{mij}} = 0 \quad (2.31)$$

となる。ここで、式(2.11) より、

$$\frac{\partial \bar{f}_m(y_t)}{\partial \bar{\sigma}_{mij}} = \bar{f}_m(y_t) \left\{ -\frac{1}{2|\bar{\sigma}_{mij}|} + \frac{(y_t - \bar{\mu}_{m,t})_i (y_t - \bar{\mu}_{m,t})_j}{2\bar{\sigma}_{mij}^2} \right\} \quad (2.32)$$

であるので、式(2.32) を式(2.31) に代入し、両辺に $-2\bar{\sigma}_{mij}^2$ を乗じると、

$$\sum_{t=1}^L \gamma_{m,t} \{ \bar{\sigma}_{mij} - (y_t - \bar{\mu}_{m,t})_i (y_t - \bar{\mu}_{m,t})_j \} = 0 \quad (2.33)$$

が得られる。これより、式(2.33) を満たす $\bar{\Sigma}_m$ は、

$$\bar{\Sigma}_m = \frac{\sum_{t=1}^L \gamma_{m,t} (y_t - \bar{\mu}_{m,t})(y_t - \bar{\mu}_{m,t})^T}{\sum_{t=1}^L \gamma_{m,t}} \quad (2.34)$$

を計算することにより得られる²。

²本来ならば、式(2.34) は式(2.29) から得られた平均軌跡の結果を用いるべきであるが、HMM の場合現在の平均値の結果を用いた場合と大差がないことが確かめられているため^[9]、本論文では現在の平均軌跡の結果を用いて分散の再推定を行なっている。

重みの計算方法

分布 m に対する重み \bar{w}_m は、HMM の場合と同様にして、ラグランジェの未定乗数法を用いれば、式(2.18) から、

$$\bar{w}_m = \frac{\sum_{t=1}^L \gamma_{m,t}}{\sum_{t=1}^L \sum_{k=1}^M \gamma_{k,t}} \quad (2.35)$$

として求めることができる。

2.3 分散の時変モデル化

これまで述べてきたセグメントモデルでは、一つのセグメント内の分散値が一定の値をとっていた。これは、HMM における個々の状態がそれぞれ個別の分散値を保持できることを考えると、分散に着目した場合、HMM の方が観測系列の分散の時間変化の表現能力が高いということの意味している。つまり、HMM の状態数を増加させていった場合、セグメントモデルの認識性能が HMM より劣ってしまう可能性がある。そこで、本節では、これまで分散を時不変の定数として扱っていたものを平均軌跡と同様に、時間の関数としてモデル化することを試みる。分散の時変モデルは、式(2.11) における分散 Σ_k を時間変化の関数 $\Sigma_{k,t}$ として表すことにより実現できる。ここで、

$$\Sigma_{k,t} = z_L^t G_k \quad (2.36)$$

であり、分散軌跡を W 次で表す場合、 G_k は、

$$G_k = [g_{k0}, g_{k1}, \dots, g_{kW}]^T \quad (2.37)$$

により、また z_L^t は式(2.5) と同様に、 $(W+1)$ 次元のベクトルとして、

$$z_L^t = \begin{cases} [1 \ 0 \ 0 \ \dots \ 0], & t=1 \\ [1 \ \frac{t-1}{L-1} \ (\frac{t-1}{L-1})^2 \ \dots \ (\frac{t-1}{L-1})^W], & 1 < t \leq L \end{cases} \quad (2.38)$$

で与えられる。

分散を時変とした場合の EM アルゴリズムにおいても、出力確率を時変の分散を用いて計算するという部分のみが変わるため、平均軌跡、重みの計算は前節と同様に行なうことができる。ただし、分散の計算式が以下のように異なる。

時変分散の計算方法

分散の時変モデル化においても、前節と同様に、式(2.16) 及び式(2.17) から、 m 番目の分布に対する分散 $\bar{\Sigma}_{m,t}$ の値は、式(2.17) を i 行 j 列の要素 $\bar{\sigma}_{m,ij} = [\bar{s}_{m,ij0}, \bar{s}_{m,ij1}, \dots, \bar{s}_{m,ijW}]^T$ の $\bar{s}_{m,ijr}$ について偏微分したものを零とおき、

$$\frac{\partial Q_1}{\partial \bar{s}_{m,ijr}} = \sum_{t=1}^L \gamma_{m,t} \frac{\partial \bar{f}_m(y_t)}{\partial \bar{s}_{m,ijr}} = 0 \quad (2.39)$$

を解くことにより得られる。ただし、ここで、

$$\frac{\partial \bar{f}_m(y_t)}{\partial \bar{s}_{m,ijr}} = \bar{f}_m(y_t) \left\{ -\frac{1}{2|\bar{\sigma}_{m,t,ij}|} + \frac{(y_t - \bar{\mu}_{m,t})_i (y_t - \bar{\mu}_{m,t})_j}{2\bar{\sigma}_{m,t,ij}^2} \right\} \left(\frac{t-1}{L-1} \right)^r \quad (2.40)$$

すなわち、

$$\sum_{t=1}^L \frac{\gamma_{m,t}}{\bar{\sigma}_{m,t,ij}^2} \{ \bar{\sigma}_{m,t,ij} - (y_t - \bar{\mu}_{m,t})_i (y_t - \bar{\mu}_{m,t})_j \} \left(\frac{t-1}{L-1} \right)^r = 0 \quad (2.41)$$

となり、 $\bar{\sigma}_{m,t,ij}$ は時刻 t に依存する値であるため、線形方程式として解くことはできない。そこで、式(2.41)における分母の $\bar{\sigma}_{m,t,ij}$ を現在の分散軌跡 $\sigma_{m,t,ij}$ で代用するという近似手法を導入することを考える。すなわち、式(2.41)を、

$$\sum_{t=1}^L \frac{\gamma_{m,t}}{\sigma_{m,t,ij}^2} \{ \bar{\sigma}_{m,t,ij} - (y_t - \bar{\mu}_{m,t})_i (y_t - \bar{\mu}_{m,t})_j \} \left(\frac{t-1}{L-1} \right)^r = 0 \quad (2.42)$$

として計算する。このとき、前節の平均軌跡の計算方法と同様に線形方程式として解くことができ、

$$\sum_{t=1}^L \gamma_{m,t} \frac{\left(\frac{t-1}{L-1} \right)^r}{\left\{ \sum_{n=0}^W s_{m,ijn} \left(\frac{t-1}{L-1} \right)^n \right\}^2} \left\{ \sum_{u=0}^W \bar{s}_{m,iju} \left(\frac{t-1}{L-1} \right)^u - (y_t - \bar{\mu}_{m,t})_i (y_t - \bar{\mu}_{m,t})_j \right\} = 0 \quad (2.43)$$

を求めればよいことになる。式(2.43)は、

$$\sum_{u=1}^W H(u+r) \bar{s}_{m,iju} = Z(r) \quad (2.44)$$

と表すことができる。ただし、

$$H(l) = \sum_{t=1}^L \gamma_{m,t} \frac{\left(\frac{t-1}{L-1} \right)^l}{\left\{ \sum_{n=1}^W s_{m,ijn} \left(\frac{t-1}{L-1} \right)^n \right\}^2}, \quad l = 0, \dots, 2W \quad (2.45)$$

$$Z(r) = \sum_{t=1}^L \gamma_{m,t} \frac{\left(\frac{t-1}{L-1} \right)^r (y_t - \bar{\mu}_{m,t})_i (y_t - \bar{\mu}_{m,t})_j}{\left\{ \sum_{n=0}^W s_{m,ijn} \left(\frac{t-1}{L-1} \right)^n \right\}^2}, \quad r = 0, \dots, W \quad (2.46)$$

である。ここで、 $H(l)$ 及び $Z(r)$ は共に次元依存のベクトル定数であることに注意されたい。

結局、式(2.44)を満たす $\bar{s}_{m,iju}$ は、

$$\begin{bmatrix} H(0)^{(d)} & H(1)^{(d)} & \dots & H(W)^{(d)} \\ H(1)^{(d)} & H(2)^{(d)} & \dots & H(W+1)^{(d)} \\ \vdots & \vdots & \ddots & \vdots \\ H(W)^{(d)} & H(W+1)^{(d)} & \dots & H(2W)^{(d)} \end{bmatrix} \begin{bmatrix} \bar{s}_{m,ij0}^{(d)} \\ \bar{s}_{m,ij1}^{(d)} \\ \vdots \\ \bar{s}_{m,ijW}^{(d)} \end{bmatrix} = \begin{bmatrix} Z^{(d)}(0) \\ Z^{(d)}(1) \\ \vdots \\ Z^{(d)}(W) \end{bmatrix} \quad (2.47)$$

を $d = 1, \dots, D$ について解くことにより得られる。

第 3 章

母音識別実験

3.1 実験条件

3.1.1 音声資料

音声資料は TIMIT データベースのラベル情報を用いて切り出した 16 母音を用いた。これは、Gish^[4] や Bacchiani^[8] らが、2.1 で述べたセグメントモデルにより、同様の音声資料を用いて認識実験を行っており、比較が行なえるようにするためである。ラベル情報から、45msec 以上の母音を学習 (データベース中の train 全て) 及び識別 (データベース中の test 全て) に用いた。¹

表 3.1 に TIMIT データベースの学習に用いた音素の内訳を示す。また識別に用いた母音データの総数は、14,981 個である。

3.1.2 特徴パラメータ抽出

前述のように切り出されたサンプリング周波数 16KHz の母音波形に対して、分析窓長 25.6msec、フレーム周期 5msec により、分析次数 10 次のメルケプストラム分析 (MFCC)^[9] を行なった。

3.1.3 尤度計算

セグメントの境界は既知であるため、長さ L の識別対象セグメント k に対して、次式の尤度が最大となるモデル \hat{m} を識別結果とする。

$$\hat{m} = \operatorname{argmax}_m P(k|m, L)P(L|m) \quad (3.1)$$

式(3.1) の右辺は、対数をとっても大小関係は変わらないので、実際には、

$$\hat{m} = \operatorname{argmax}_m \{\log P(k|m, L) + \log P(L|m)\} \quad (3.2)$$

を計算した。ここで、式(3.2) の右辺第一項は出力確率の対数尤度であり、式(2.10) により計算され、第二項は学習データの継続時間長分布のヒストグラムから得られる対数尤度である。

¹45msec 以上を用いたのは、HMM との比較を行なう際に、状態数を増加させた場合でもデータ量を同数にするためである。

表 3.1: 学習用音素の内訳と 45msec 以上の割合

ラベル	全体の個数	45msec 以上	割合	ラベル	全体の個数	45msec 以上	割合
※ aa	3064	3054	99.67	※ ae	3997	3995	99.95
※ ah	2306	2276	98.70	※ ao	2940	2927	99.56
※ aw	729	729	100.00	ax	3610	2590	71.75
ax-h	375	99	26.40	axr	3407	3254	95.51
※ ay	2390	2390	100.00	b	2181	32	1.47
bcl	1909	1639	85.86	ch	822	817	99.39
d	3548	376	10.60	dcl	4942	3208	64.91
dh	2826	1179	41.72	dx	2709	342	12.62
※ eh	3853	3813	98.96	el	951	946	99.47
em	124	122	98.39	en	723	684	94.61
eng	38	38	100.00	epi	1464	798	54.51
※ er	2046	2044	99.90	※ ey	2282	2281	99.96
f	2216	2162	97.56	g	2017	487	24.14
gcl	2223	1543	69.41	hh	957	793	82.86
hv	1154	1069	92.63	※ ih	5051	4941	97.82
ix	8642	6521	75.46	※ iy	6953	6865	98.73
jh	1209	996	82.38	k	4874	3093	63.46
kcl	5859	4873	83.17	l	5801	5076	87.50
m	3903	3203	82.07	n	7068	5448	77.08
ng	1330	1160	87.21	nx	971	94	9.68
※ ow	2136	2135	99.95	※ oy	684	684	100.00
p	2588	1468	56.72	pau	952	935	98.21
pcl	2644	2473	93.53	q	3590	2879	80.20
r	6539	5428	83.01	s	7475	7445	99.60
sh	2238	2235	99.87	t	4364	2637	60.43
tcl	6644	4501	67.75	th	751	702	93.48
※ uh	535	508	94.95	utb	9240	9205	99.62
※ uw	555	545	98.20	※ ux	1908	1827	95.75
v	1994	1748	87.66	w	3140	2533	80.67
y	1715	1319	76.91	z	3773	3709	98.30
zh	151	147	97.35				

※は本論文で、識別実験のためのモデル作成に用いた 16 母音を示す。

3.2 実験結果及び考察

3.2.1 クラスタリングによるモデル化

2.2.1 で述べたクラスタリングによる方法で混合分布モデルを作成した場合の識別結果を、表3.2に示す。この実験により混合分布の数を多くすることにより、より高い識別率が得られていることが分かる。次に、継続時間分布の効果を調べるために、継続時間分布を考慮していない識別実験を行なった。これは、式(3.2)の右辺第二項を0としたものである。結果を表3.3に示す。表3.3と表3.2の結果を比較すると、全ての混合数において、継続時間分布を考慮しての方が、より高い識別率が得られていることが分かる。

表 3.2: クラスタリング法の識別結果 (回帰係数 = 0)

音素	サンプル数	1 混合	3 混合	5 混合	7 混合	9 混合
aa	1129	455	469	439	472	482
ae	1407	795	848	859	871	892
ah	869	233	242	239	268	275
ao	1153	758	717	727	729	743
aw	216	57	68	65	69	63
ay	852	251	209	161	162	161
eh	1424	411	436	428	408	422
er	799	591	632	633	630	628
ey	806	411	407	411	392	415
ih	1669	322	382	444	479	447
iy	2680	1735	1674	1684	1700	1738
ow	777	200	223	270	249	254
oy	263	77	24	21	16	28
uh	213	48	58	57	57	55
uw	168	81	79	83	83	81
ux	556	290	323	315	301	303
合計	14981	6715	6791	6836	6886	6987
識別率 (%)		44.8	45.3	45.6	46.0	46.6

表 3.3: 継続時間分布を考慮しない場合の識別結果 (回帰係数 = 0)

音素	サンプル数	1 混合	3 混合	5 混合	7 混合	9 混合
aa	1129	454	468	433	470	476
ae	1407	793	843	855	876	884
ah	869	221	229	221	255	263
ao	1153	753	717	728	728	738
aw	216	55	68	61	69	62
ay	852	249	209	167	164	162
eh	1424	386	404	402	382	388
er	799	590	633	631	633	626
ey	806	405	403	408	386	415
ih	1669	301	352	401	437	412
iy	2680	1729	1665	1673	1687	1724
ow	777	200	214	257	248	252
oy	263	75	22	20	15	22
uh	213	45	55	55	55	51
uw	168	80	78	83	80	81
ux	556	292	324	314	300	305
合計	14981	6628	6684	6709	6785	6861
識別率 (%)		44.2	44.6	44.8	45.3	45.8

回帰係数の効果を調べるために、継続時間分布を考慮した同様の実験を行なった。1次及び2次での識別結果を表3.4と表3.5に示す。ここで、回帰係数0次(表3.2)と、回帰係数1次(表3.4)、回帰係数2次(表3.5)の結果を比べると、全ての混合数において、高次になるほど高い識別率が得られていることが分かる。また、回帰係数を0次から1次へ変化させた場合の識別率の改善率は大きいですが、1次を2次とした時はさほど大きくない。これは、回帰次数を更に大きく(3次以上)していても、識別性能の向上はあまり見込めないことを示唆していると思われる。

表 3.4: クラスタリング法の識別結果 (回帰次数=1)

音素	サンプル数	1 混合	3 混合	5 混合	7 混合	9 混合
aa	1129	463	495	502	538	553
ae	1407	829	893	898	915	924
ah	869	212	241	261	264	264
ao	1153	774	722	753	741	733
aw	216	96	102	98	105	101
ay	852	578	573	589	584	592
eh	1424	443	491	529	490	493
er	799	592	646	644	620	635
ey	806	536	556	556	547	559
ih	1669	428	511	561	551	522
iy	2680	1827	1795	1863	1900	1868
ow	777	306	341	373	362	358
oy	263	145	136	145	142	142
uh	213	60	70	76	77	78
uw	168	92	92	91	84	88
ux	556	320	345	332	322	336
合計	14981	7701	8009	8271	8242	8246
識別率 (%)		51.4	53.5	55.2	55.0	55.0

表 3.5: クラスタリング法の識別結果 (回帰次数=2)

音素	サンプル数	1 混合	3 混合	5 混合	7 混合	9 混合
aa	1129	483	513	539	545	565
ae	1407	836	880	908	918	948
ah	869	225	233	262	275	291
ao	1153	789	728	751	767	764
aw	216	104	118	106	113	113
ay	852	581	597	603	596	597
eh	1424	478	500	555	493	512
er	799	605	641	655	657	654
ey	806	546	545	562	572	565
ih	1669	440	510	562	566	563
iy	2680	1885	1864	1928	1918	1971
ow	777	324	378	389	387	382
oy	263	153	155	170	170	173
uh	213	65	71	63	74	80
uw	168	93	95	95	95	102
ux	556	327	350	338	344	326
合計	14981	7934	8178	8486	8490	8606
識別率 (%)		53.0	54.6	56.6	56.7	57.4

3.2.2 再推定の効果

2.2.2 で述べた、EM アルゴリズムを用いたモデルの再推定を行なった結果を、表 3.6、表 3.7、表 3.8、に示す。ここで、初期値にはクラスタリングによる結果を与えた。

表 3.2 と表 3.6 より、回帰次数が 0 次の場合は全ての混合数について、識別率が向上していることが分かる。しかしながら、表 3.4 と表 3.7 より、回帰次数が 1 次の場合は再推定の効果はあまり見られない。また、表 3.5 と表 3.8 から分かるように、回帰次数が 2 次の場合は、再推定はむしろ逆効果となってしまっている。これは、

- 式(2.34)における $\bar{\mu}_{m,t}$ を現在のモデルの平均軌跡の値を用いたため
- モデル内の最尤推定値は、必ずしもモデル間の識別性能の向上には結び付かないため(これは識別学習により改善される可能性がある)

などが考えられるが、詳細な原因究明は今後の課題である。

表 3.6: 再推定による結果 (回帰次数 = 0)

音素	サンプル数	1 混合 ※	3 混合	5 混合	7 混合	9 混合
aa	1129	455	518	456	447	487
ae	1407	795	907	866	857	848
ah	869	233	259	256	263	276
ao	1153	758	645	706	715	720
aw	216	57	68	66	66	52
ay	852	251	339	345	356	336
eh	1424	411	441	433	415	448
er	799	591	621	636	623	630
ey	806	411	426	428	450	436
ih	1669	322	390	389	411	386
iy	2680	1735	1673	1660	1673	1677
ow	777	200	215	261	235	256
oy	263	77	132	135	147	149
uh	213	48	55	51	56	64
uw	168	81	96	90	89	90
ux	556	290	305	321	317	339
合計	14981	6715	7090	7099	7120	7194
識別率 (%)		44.8	47.3	47.4	47.5	48.0

※ 1 混合の場合はクラスタリングによる結果 (回帰次数 = 0) と同一になる

表 3.7: 再推定による結果 (回帰次数 = 1)

音素	サンプル数	1 混合 ※	3 混合	5 混合	7 混合	9 混合
aa	1129	463	519	509	528	538
ae	1407	829	929	932	880	900
ah	869	212	271	260	284	292
ao	1153	774	704	726	730	734
aw	216	96	111	104	100	108
ay	852	578	572	576	577	576
eh	1424	443	507	517	504	542
er	799	592	637	648	641	645
ey	806	536	544	530	538	544
ih	1669	428	488	516	514	530
iy	2680	1827	1737	1776	1815	1821
ow	777	306	372	376	372	366
oy	263	145	162	183	175	187
uh	213	60	65	74	78	74
uw	168	92	92	95	94	90
ux	556	320	334	332	337	336
合計	14981	7701	8044	8154	8167	8283
識別率 (%)		51.4	53.7	54.4	54.5	55.3

※ 1 混合の場合はクラスタリングによる結果 (回帰次数 = 1) と同一になる

表 3.8: 再推定による結果 (回帰次数 = 2)

音素	サンプル数	1 混合 ※	3 混合	5 混合	7 混合	9 混合
aa	1129	483	536	545	563	579
ae	1407	836	923	893	914	912
ah	869	225	281	283	271	305
ao	1153	789	708	732	749	750
aw	216	104	112	112	111	104
ay	852	581	577	584	587	587
eh	1424	478	499	539	508	530
er	799	605	653	652	662	656
ey	806	546	547	550	562	575
ih	1669	440	513	527	540	530
iy	2680	1885	1834	1878	1911	1911
ow	777	324	384	396	413	408
oy	263	153	181	191	186	182
uh	213	65	74	72	72	74
uw	168	93	86	97	96	95
ux	556	327	341	341	339	341
合計	14981	7934	8249	8392	8484	8539
識別率 (%)		53.0	55.1	56.0	56.6	57.0

※ 1 混合の場合はクラスタリングによる結果 (回帰次数 = 2) と同一になる

3.2.3 分散の時変モデル化

2.3 で論じた分散の時変モデル化についても同様の識別実験を行なった。これまでのモデルでは、分散が時間不変の一定値をとっていたのに対し、この結果は、分散を時間と共に変化する軌跡として、EM アルゴリズムの再推定を行なったものである。ここで、平均及び分散の軌跡の回帰次数は共に 2 としている (本論文では、音素環境非依存の音素モデルを 1 状態のセグメントモデルでモデル化しているため、音素の入り渡り、出渡りにおける分散は、音素の中心付近の分散よりも一般に値が大きくなると予想される。この現象をモデル化するためには、分散の軌跡の回帰次数を最低 2 次とする必要があると考えられるため、ここでは 2 次とした)。また、初期値はクラスタリングによる結果を与えた²。

TIMIT データベースの母音 /aa/(サンプル数 3054 個) について、学習時の繰り返し回数に対する尤度の変化の様子を図 3.1 に示す。図 3.1 において、実線 (下側) は平均の回帰次数を 2 次とし、分散は一定値とした場合の結果 (すなわち 2.2.2 における結果) を示し、破線 (上側) は平均及び分散の軌跡の回帰次数を共に 2 次とした場合の結果 (すなわち 2.3 における結果) である。この結果から、分散の時変モデル化は、

- 式(2.42) による近似で再推定を行なっているが、分散の時不変モデルよりも常に高い尤度が得られていること
- 繰り返し回数に対して尤度が単調増加していること

が分かる。これらのことは、今回用いた 16 母音全てについて確かめられた。

次に、識別結果を表 3.9 に示す。表 3.9 を、表 3.8 の分散を時不変とした再推定による結果 (平均軌跡の回帰次数 = 2) と比較すると、全ての混合数において識別率が向上していることから、分散の時変モデル化は、識別性能の観点からも有効であると考えられる。

²すなわち、式(2.37) における初期値の g_{k1}, g_{k2} は共に 0 とした。

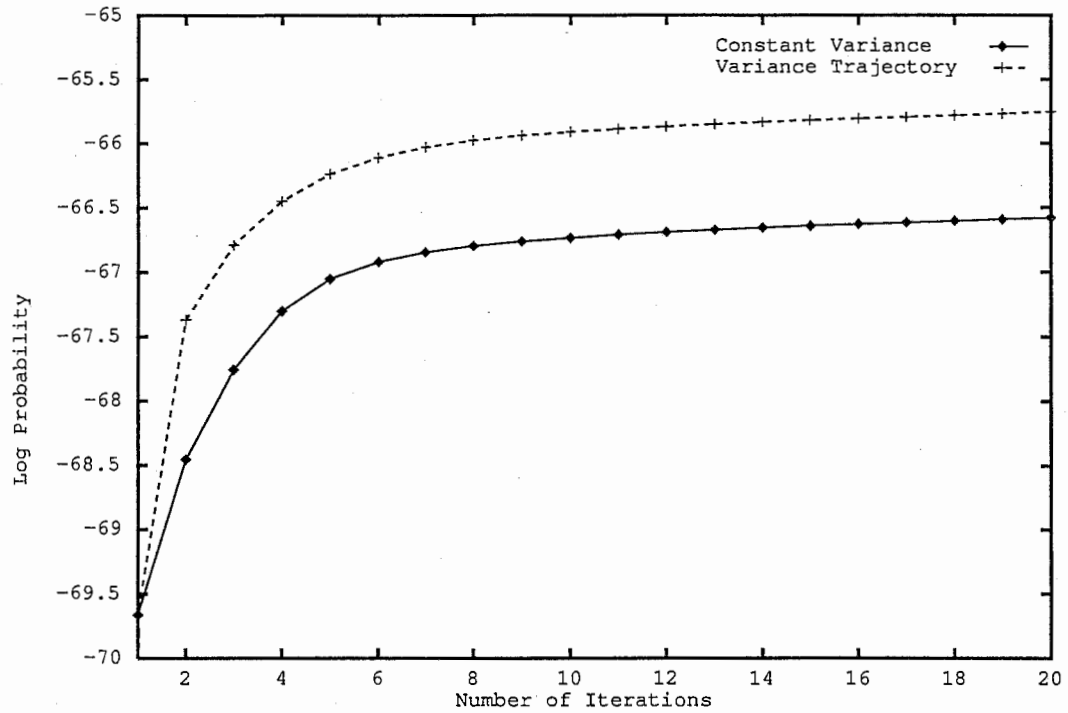


図 3.1: 学習時の尤度の変化の様子

表 3.9: 分散の時変モデル化による識別結果 (平均及び分散の回帰次数 = 2)

音素	サンプル数	1 混合	3 混合	5 混合	7 混合	9 混合
aa	1129	501	525	547	571	587
ae	1407	849	926	937	934	941
ah	869	233	278	291	291	311
ao	1153	781	736	736	760	750
aw	216	106	115	115	115	107
ay	852	580	575	581	583	589
eh	1424	482	503	534	506	522
er	799	619	658	657	665	663
ey	806	540	558	561	575	581
ih	1669	429	512	518	534	541
iy	2680	1930	1857	1902	1916	1941
ow	777	330	390	412	406	425
oy	263	156	179	190	191	191
uh	213	65	67	68	69	74
uw	168	94	85	99	100	94
ux	556	319	343	333	340	335
合計	14981	8014	8307	8481	8556	8652
識別率 (%)		53.5	55.5	56.6	57.1	57.8

3.2.4 HMM との比較

HMM とセグメントモデルの識別結果を比較・検討するために、HMM を用いた同様の識別実験を行なった。学習及び認識には HTK V1.5^[9] を利用した。状態数 1、3、5、7、9 の識別結果をそれぞれ表 3.10、表 3.11、表 3.12、表 3.13、表 3.14 に示す。

これらの結果から、

- いずれの状態数においても混合数 7 で飽和している。
- 状態数の増加による識別率の改善は徐々に小さくなっているが、飽和しているとは言えない。

ということが言える。しかし、これ以上、状態数を増加させると、発話スピードの速い(すなわち 9 フレーム程度の)音声の識別が困難となるため、9 状態 (45msec) 程度までが適切であると考えられる。

次に、これらの結果を再推定を行なったセグメントモデルのもの (表 3.6～表 3.8) と比較すると、

1. 回帰次数が 0 次のセグメントモデルは、すべての混合数において、1 状態の HMM よりも高い性能を示している
2. 回帰次数が 1 次のセグメントモデルは、3～5 状態の HMM と同等の性能を示している
3. 回帰次数が 2 次のセグメントモデルは、5～7 状態の HMM と同等の性能を示している

ということが分かる。1. は、セグメントモデルの継続時間長分布が実データのヒストグラムによって求められている効果であると考えられる。

高次の回帰次数によるセグメントモデルが、7～9 状態の HMM による識別性能より劣っている原因として、

1. 平均や分散の回帰次数が不足している
2. HMM では学習、認識において観測データを状態遷移の結果、非線形伸縮させることができるが、本論文におけるセグメントモデルでは学習、認識がとも線形伸縮で行なわれている

ということが考えられる。1. において、回帰次数を増加させることにより、若干の認識性能の改善は期待されるが、9 状態の HMM までには至らないと予想される。つまり、2. による影響が大きいと考えられる。

2. の問題を解決する一つの方法として、セグメントモデルにおいても、HMM と同様に状態という概念を導入して、例えば、一つの音素を 3 つのセグメントモデルでモデル化するということができる。ここで例えば、回帰次数を 2 次とした 3 状態のセグメントモデルを学習、

認識するためには、最低9フレーム(フレーム周期5msecの場合は45msec)のデータが必要となる。しかしながら、表3.1を見ると、45msec以上のデータがほとんどない音素(例えば、/b/, /d/, /dx/, /nx/ など)も多く見受けられる。このような音素に対しては、

- 状態数を減らす
- 回帰次数を減らす
- ナル遷移を持つセグメントモデルでモデル化する

などにより対処する必要が生じてくる。

2. の問題を解決する別の方法に、学習、識別の前に観測データを各モデルの軌跡に対してダイナミック・プログラミング(DP)により非線形伸縮を行なうということが考えられる。この場合、一つの音素は1状態のセグメントモデルでモデル化できるため、3状態にしたときに生じた問題は回避できる。今後、この方法によるモデル化を検討していきたい。

表 3.10: HMM の識別結果 (状態数=1)

音素	サンプル数	1混合	3混合	5混合	7混合	9混合
aa	1129	454	476	460	464	457
ae	1407	795	882	862	854	846
ah	869	223	262	258	273	260
ao	1153	753	709	720	710	712
aw	216	56	59	73	59	68
ay	852	249	344	341	341	327
eh	1424	387	405	376	400	401
er	799	591	621	624	642	629
ey	806	408	433	434	443	447
ih	1669	302	329	351	367	377
iy	2680	1731	1684	1672	1636	1709
ow	777	201	213	228	235	232
oy	263	77	138	136	150	145
uh	213	45	56	58	56	51
uw	168	81	91	92	95	99
ux	556	291	305	303	321	321
合計	14981	6644	7007	6988	7046	7081
識別率 (%)		44.3	46.8	46.6	47.0	47.3

表 3.11: HMM の識別結果 (状態数=3)

音素	サンプル数	1 混合	3 混合	5 混合	7 混合	9 混合
aa	1129	476	494	505	522	510
ae	1407	774	843	857	838	842
ah	869	208	248	227	260	241
ao	1153	703	698	705	712	702
aw	216	94	92	96	102	99
ay	852	561	550	551	561	579
eh	1424	412	447	463	505	463
er	799	620	624	621	634	651
ey	806	529	524	565	561	536
ih	1669	468	407	472	450	519
iy	2680	1800	1756	1825	1835	1846
ow	777	375	363	401	408	411
oy	263	152	185	196	193	193
uh	213	67	78	70	78	77
uw	168	94	97	93	93	91
ux	556	313	325	334	337	342
合計	14981	7646	7731	7981	8089	8102
識別率 (%)		51.0	51.6	53.3	54.0	54.1

表 3.12: HMM の識別結果 (状態数=5)

音素	サンプル数	1 混合	3 混合	5 混合	7 混合	9 混合
aa	1129	477	527	514	534	559
ae	1407	783	876	851	875	865
ah	869	253	266	274	267	256
ao	1153	738	711	707	705	719
aw	216	106	107	111	109	104
ay	852	590	578	590	604	594
eh	1424	423	475	487	510	486
er	799	626	641	643	648	640
ey	806	572	570	585	575	581
ih	1669	515	480	573	574	600
iy	2680	1789	1898	1886	1912	1929
ow	777	364	414	443	443	447
oy	263	181	197	194	194	196
uh	213	76	89	69	78	78
uw	168	101	96	88	95	92
ux	556	331	335	350	330	351
合計	14981	7925	8260	8365	8453	8497
識別率 (%)		52.9	55.1	55.8	56.4	56.7

表 3.13: HMM の識別結果 (状態数=7)

音素	サンプル数	1 混合	3 混合	5 混合	7 混合	9 混合
aa	1129	467	531	535	539	535
ae	1407	791	872	848	855	859
ah	869	248	267	287	290	294
ao	1153	775	726	705	745	743
aw	216	114	111	112	110	109
ay	852	593	602	604	608	597
eh	1424	476	479	541	543	544
er	799	633	656	660	642	648
ey	806	575	592	614	595	581
ih	1669	504	534	575	646	629
iy	2680	1886	1942	1949	1965	1977
ow	777	394	417	447	453	451
oy	263	194	205	203	205	202
uh	213	87	82	80	73	77
uw	168	98	99	88	96	81
ux	556	330	333	337	344	352
合計	14981	8165	8448	8585	8709	8679
識別率 (%)		54.5	56.4	57.3	58.1	57.9

表 3.14: HMM の識別結果 (状態数=9)

音素	サンプル数	1 混合	3 混合	5 混合	7 混合	9 混合
aa	1129	477	524	555	534	559
ae	1407	832	886	873	875	896
ah	869	266	292	306	323	298
ao	1153	777	738	724	727	747
aw	216	103	122	118	115	111
ay	852	597	595	605	612	606
eh	1424	491	534	533	556	554
er	799	636	665	659	660	647
ey	806	566	603	595	593	592
ih	1669	547	611	644	712	697
iy	2680	1936	1985	1991	2015	2023
ow	777	393	448	450	472	461
oy	263	194	201	206	195	202
uh	213	81	89	81	85	79
uw	168	99	99	91	88	87
ux	556	322	335	339	346	349
合計	14981	8317	8727	8770	8908	8908
識別率 (%)		55.5	58.3	58.5	59.5	59.5

第 4 章

まとめ

本論文では、確率的セグメントモデルの認識性能向上を目的として、混合分布モデルの作成方法を述べた。混合分布モデルの作成方法として、(1) クラスタリング手法により混合分布を求める方法、(2) EM アルゴリズムを用いたモデルの再推定により混合分布を求める方法、(3) 平均に加えて分散も時間変化の軌跡としてモデル化し、EM アルゴリズムによるモデルの再推定を行なう方法の 3 つの手法を述べた。

TIMIT データベースを用いた母音識別実験を行なった結果、セグメントモデルの混合数を増加させることにより、識別率が向上することが確かめられた。また、上記(3)の分散の時変モデル化は、(1),(2)に比べて高い識別性能が得られ、このモデル化が有効であることが示された。HMM との結果の比較により、一つの音素を 1 状態で平均軌跡の回帰次数を 2 次でモデル化したセグメントモデルは、5～7 状態の HMM と同等の性能を示すことが分かった。

本論文では、学習時において、観測データを線形時間伸縮することによりモデル化を行なった。今後は、これを非線形伸縮によるモデル化を行なうことにより、セグメントモデルの認識性能を更に向上させていく予定である。

謝辞

本研究を進めるにあたり、御指導、御支援いただいた K. K. Paliwal 教授、匂坂芳典室長、及び ATR 音声翻訳通信研究所の皆様に深く感謝いたします。さらに実務訓練の機会を与えて下さった豊橋技術科学大学の中川聖一教授及び ATR 音声翻訳通信研究所の山崎泰弘社長に心から感謝いたします。

参考文献

- [1] L. Deng, "A generalized hidden Markov model with state-conditioned trend functions of time for the speech signal," *Signal Processing* **27**, pp.65-78, (1992).
- [2] L. Deng and M. Aksmanovic, "HMMs with mixtures of trend functions for automatic speech recognition," *International Symposium on Speech, Image Processing and Neural Networks*, pp.702-705, (1994).
- [3] M. Ostendorf and S. Roukos, "A stochastic segment model for phoneme-based continuous speech recognition," *IEEE Trans. on Acoust., Speech and Signal Proc.*, vol.37, no.12, pp.1857-1869, (1989).
- [4] H. Gish and K. Ng, "A segmental speech model with applications to word spotting," *Proc. ICASSP-93* pp.447-450, (1993).
- [5] M. Ostendorf, M. Bacchiani, Y. Sagisaka and K. Paliwal, "Speech recognition system design using automatically learned non-uniform segmental units," TR-IT-0147, (1996).
- [6] X. Huang, Y. Ariki and M. Jack, "Hidden Markov models for speech recognition," *Edinburgh information technology series*;7, (1990).
- [7] L. Rabiner and B. Juang, "Fundamentals of speech recognition," PTR Prentice Hall, (1993).
- [8] M. Bacchiani and M. Ostendorf, "Software for design and use of a speech recognition system based on automatically derived non-uniform units," TR-IT-0146, (1996).
- [9] S. Young, P. Woodland and W. Byrne, "HTK : Hidden Markov Model Toolkit V1.5 reference manual," Cambridge University Engineering Department Speech Group and Entropic Research Laboratories Inc., (1993).

付録

A プログラム説明

本実験では、以下のようにプログラムを使った。

- 音声試料の切り出し
 - HCopy
 - * TIMIT データベースより 16 種の母音 (45msec 以上) を取り出す。
- 学習
 - asegm
 - * 切り出された母音区間をセグメントモデルによりモデル化。
 - km-las
 - * セグメントモデルのクラスタリング。
 - allinone (allinone_var)
 - * 1つの母音による全てのクラスタを1つのファイルにまとめる。(allinone_var : 分散の時変モデルに対応。分散の次数を与える。)
 - SRest (SRest_var)
 - * EM アルゴリズムによりモデルの再推定。(SRest_var : 分散の時変モデルに対応)
- 識別
 - segclass (segclass_var)
 - * テストデータの識別 (segclass_var : 分散の時変モデルに対応)
- ファイルの内容を見るもの
 - HList
 - * バイナリのパラメータファイルのダンプ。

- lmdl (lmdl_var)

- * セグメントモデルのパラメータファイルのダンプ。(lmdl_var : 分散の時変モデルに対応)

この章では、本実験に使われたプログラムについて説明する。

A.1 HCopy

HCOPY コマンド^[9] は音声のパラメータファイルの 1 部を取り出すファイルである。このコマンドを用いて TIMIT データベースより、ラベルを参考にして 45msec 以上の 16 母音を取り出す。

図 A.1 にラベルファイルの例として sal.phn の内容を示す。

この時、3 列目が aa, ae, ah, ao, aw, ay, eh, er, ey, ih, iy, ow, oy, uh, uw, ux のいずれかで、

$$\text{int} \left(\frac{(\text{2列目}) - 153}{\frac{1.6}{50}} \right) - \text{int} \left(\frac{(\text{1列目}) - 153}{\frac{1.6}{50}} \right) + 1 \geq 9$$

を満たすサンプルを 45msec(9 フレーム) 以上の 16 母音として取り出した。図 A.1 にラベルファイルの例として sal.phn の内容を示す。

0 3050 utb
3050 4559 sh
4559 5723 ix
5723 6642 hv
6642 8772 eh
8772 9190 dcl
9190 10337 jh
10337 11517 ih
11517 12500 dcl
12500 12640 d
12640 14714 ah
14714 15870 kcl
15870 16334 k
16334 18088 s
18088 20417 ux
20417 21199 q
21199 22560 en
22560 22920 gcl
22920 23271 g
23271 24229 r
24229 25566 ix
25566 27156 s
27156 28064 ix
28064 29660 w
29660 31719 ao
31719 33360 sh
33360 33754 epi
33754 34715 w
34715 36080 ao
36080 36326 dx
36326 37556 axr
37556 39561 ao
39561 40313 l
40313 42059 y
42059 43479 ih
43479 44586 axr
44586 46720 utb

☒ A.1: 'She had your dark suit in greasy wash water all year.' のラベル

以下に、メルケプストラムのパラメータファイル sal.mfc から、ラベルデータ sal.phn を参考にして図 A.1の 8 行目のラベル ih(126 フレームから 140 フレームの 15 フレーム) を取り出し、ih_0001.mfc というファイルに書き込む時の例を示す。

```
% HCopy -s 126 -e 140 -X phn -G HTK sal.mfc ih_0001.mfc
```

以下にオプション群を示す。

オプション		デフォルト
-h	HTK ヘッダの出力をしない	Off
-e N	コピーする終りのサンプル番号	0
-l s	新しいラベルファイル s の出力	Off
-m N	Set margin to N for x/n options	0
-n i [j]	Extract i'th [to j'th] label	Off
-s N	コピーする始めのサンプル番号	0
-x s [n]	Extract [n'th occ of] label s	Off
-F fmt	入力ファイルのフォーマット fmt	HTK
-G fmt	ラベルファイルのフォーマット fmt	HTK
-I s	Set master label file s	none
-L s	ラベルファイルのディレクトリ s	current
-O fmt	出力ファイルのフォーマット fmt	HTK
-P fmt	ラベルファイルのフォーマット fmt	HTK
-S f	スクリプトファイル f	none
-T N	トレースのレベル N	0
-X s	ラベルファイルの拡張子	lab

A.2 asegm

asegm コマンド^[8] は、本来音響的セグメント単位を作成する場合に用いられるプログラムであるが、ここでは HCopy で取り出したファイルをセグメントモデルとするために歪を 10000 とし、1 セグメントのセグメントモデルを作成した。

以下に ih_0001.mfc を、分散が対角行列で、回帰次数が 0 のセグメントモデル ih_0001.mdl.10000 を作る時の例を示す。

```
% asegm -c DIAGONALCOV -d 1 10000 -n 1000 -R 0 -W 0 ih_0001.mfc
```

以下にオプション群を示す。

オプション		デフォルト
-c s	分散行列の型 s	FULLCOV
-d n f1 f2....fn	Set n distortion threshold(s) to f1,...,fn	1 1.0
-m i	最小セグメント長	2 frames
-n i	最大セグメント長	70 frames
-v	エラー出力を行なう	off
-w f	分析窓長 (ms)	25.6
-C s	出力モデルファイルのディレクトリ s	current
-D s	距離測定法 s	MAHALANOBIS
-F fmt	入力ファイルのフォーマット fmt	HTK
-L s	出力ラベルファイルのディレクトリ s	current
-R i	セグメンテーションの回帰次数	0
-S f	スクリプトファイル f	none
-W i	モデルの回帰次数	0
-X s	出力ラベルファイルの拡張子	alab.< dist >
-Z s	出力モデルファイルの拡張子	mdl.< dist >

距離測定法には以下のものがある:

SQ_EUCLID squared Euclidean distance
 MAHALANOBIS Mahalanobis distance

分散行列の型には以下のものがある:

DIAGONALCOV diagonal covariance
 FULLCOV full covariance

A.3 km-las

km-las コマンド^[8] は、セグメントモデルのクラスタリングを行なうものである。

以下に、ih_0001.mdl.10000 ~ ih_0009.mdl.10000 の9個のファイルを、ih_unit1, ih_unit2, ih_unit3 の3つにクラスタリングをおこなう例を示す。但し、分散は対角行列で表される。

```
% km-las -c DIAGONALCOV -C 3 -U ih_unit ih_000[1-9].mdl.10000
```

以下にオプション群を示す。

オプション		デフォルト
-c s	分散行列の型 s	FULLCOV
-d s	距離測定法 s	MAXLIKELIHOOD
-e f	Set splitting parameter to f	0.01
-f f	Set variance floor to f	1e-05
-m i	クラスタの最小フレーム数	25
-t f	Set convergence threshold to f	0.001
-v	エラー出力をだす	off
-B s	コードブックの出力ディレクトリ s	current
-C i	コードブックのサイズ i	1000
-D i	Set codebook export interval to i	10
-I s	Set initial modellist to s	none
-J s	Set initial models directory to s	current
-M i	Set max. nr. of bin. split. iters.	10
-N i	Set max. nr. of K-means iters.	30
-S f	スクリプトファイル f	none
-U s	出力結果のファイル名 s	Unit
-V i	Set start index of new units to i	1

距離測定法には以下のものがある:

SQ_EUCLID	squared Euclidean distance
MAHALANOBIS	Mahalanobis distance
MAXLIKELIHOOD	MaximumLikelihood distance

分散行列の型には以下のものがある:

DIAGONALCOV	diagonal covariance
FULLCOV	full covariance

A.4 allinone

km-las コマンド^[8] でクラスタリングを行なう際、それぞれのクラスタリングの結果がそれぞれ1つのモデルとなってしまふ。allinone は、1つの音素によるクラスタリングの結果を1つにまとめるプログラムである。さらに各クラスタリングの重みも計算し、モデルファイルに加えている。重みは、

$$w_n = \frac{N_n}{\sum_{i=1}^M N_i} \quad n = 1, \dots, M \quad (\text{A.1})$$

により求めている。ここで、 W_n は n 番目の混合分布の重み、 N_n は n 番目の混合分布の時の学習セグメント数、 M は混合分布数を表す。

使用法は、出力ファイル名に続けて、入力ファイル群を並べる。以下に、3つのクラスタ (aa_unit1, aa_unit2, aa_unit3) から3混合分布のモデルファイル (aa) を作る時の例を示す。

```
% allinone aa aa_unit1 aa_unit2 aa_unit3
```

以下に km-las による RegModelFile のフォーマットを図 A.2に、allinone で変換された RegModelFile のフォーマットを図 A.3に示す。

```

typedef      *float      Vector ;
typedef      **float     Matrix ;
typedef enum _CovType{
    DIAGONALCOV,          /* Diagonal covariance */
    FULLCOV              /* Full covariance */
} CovType;
typedef struct{
    int         Dimension;          /* モデルの次元 */
    int         RegressionOrder;    /* 回帰次数 */
    int         NumberOfTrainingSegments; /* 学習セグメント数 */
    long        NumberOfTrainingFrames; /* 学習フレーム数 */
    int         MinDuration;        /* 最小継続時間 */
    int         MaxDuration;        /* 最大継続時間 */
    CovType     CovarianceType;     /* 分散行列の型 */
    int         HeaderSize;        /* ヘッダサイズ */
    Vector       RelFreq;           /* 継続時間分布 */
    Matrix       MeanTrajectory;    /* 平均の軌跡 */
    Matrix       Covariance;       /* 分散 */
}RegModelFile;

```

図 A.2: 従来のフォーマット

```
typedef      *float      Vector ;
typedef      **float     Matrix ;
typedef enum _CovType{
    DIAGONALCOV,          /* Diagonal covariance */
    FULLCOV              /* Full covariance(本実験では未対応)*/
} CovType;
typedef struct{
    int          Dimension;          /* モデルの次元 */
    int          RegressionOrder;    /* 回帰次数 */
    CovType      CovarianceType;    /* 分散行列の型 */
    int          Mixture;           /* 混合分布数 */
    int          HeaderSize;        /* ヘッダサイズ */
    int          *NumberOfTrainingSegments; /* 学習セグメント数 */
    long         *NumberOfTrainingFrames; /* 学習フレーム数 */
    int          *MinDuration;      /* 最小継続時間 */
    int          *MaxDuration;      /* 最大継続時間 */
    Vector       *RelFreq;          /* 継続時間分布 */
    Matrix       *MeanTrajectory;   /* 平均の軌跡 */
    Matrix       *Covariance;       /* 分散 */
    float        *Weight;           /* 重み */
}RegModelFile;
```

図 A.3: 変更後のフォーマット

allinone_var

実験では、分散が時間に対し一定であるプログラムの他に、分散が時間と共に変化するモデルも利用した。これは、再推定を行なうことによって、多次の分散を推定するものである。この場合 allinone_var の実行時に分散の次数が決まる。

分散が時間と共に変化するモデルを利用する時は以下のように入力する。

```
% allinone_var [options] outfile infile1 infile2...
```

以下はオプションの説明である。

オプション		デフォルト
-m i	混合分布数	1
-r i	分散次数	2

これによって作られたモデルを SRest_var で再推定し、分散の時変モデル化に対応した識別プログラム segclass_var で認識する。

A.5 SRest

SRest はセグメントモデルの再推定を行なうプログラムである。このファイルを実行する際には、学習用データファイルの他に、それに対応するラベルファイルも用意する。オプションに続き、再推定したいセグメントモデルファイル名と、学習用データファイル名群を記述する。

```
% SRest [options] SMFile trainFiles...
```

学習用データファイル名群の記述はスクリプトファイルを用いても良い。

以下はオプションの説明である。

オプション		デフォルト
-e f	Set convergence factor epsilon	1.0E-4
-i N	最大再推定回数 N	20
-l s	セグメントのラベル s	None
-o fn	Store new SM def in fn	srcfn
-u mvw	更新の対象 m)eans v)ars w)ghts	mvw
-v f	Set minimum variance to f	0.0
-F fmt	データファイルのフォーマット fmt	HTK
-G fmt	ラベルファイルのフォーマット fmt	HTK
-L s	ラベルファイルのディレクトリ	current
-S f	スクリプトファイル f	none
-T N	トレースのレベル N	0
-X s	ラベルファイルの拡張子	lab

SRest_var

allinone_var の実行時に分散の次数を設定するが、実際は 0 次の係数のみが値を持ち、他は 0 となっている。ここで、SRest_var を実行することで、再推定により高次の係数が求められる。使用法やオプションは SRest コマンドと同じである。

A.6 segclass

これは、テストデータの識別を行なうためのもので、今回の変更では混合分布に対応したフォーマットに対応させた。使用時にはオプションのあとにモデルのリストファイル、その次に認識したいファイル名を続けて記述する。

```
% segclass [options] modellist filenames...
```

認識したいファイル名の記述はスクリプトファイルを用いても良い。

以下はオプションについて説明である。

オプション		デフォルト
-v	エラー出力を行なう	off
-D	継続時間分布を無視	off
-d	継続時間分布をそのまま計算	off
-F fmt	データファイルのフォーマット fmt	HTK
-J s	モデルのディレクトリ s	current
-O s	出力ファイルのディレクトリ	current
-P s	出力ファイルの拡張子	vit
-S f	スクリプトファイル f	none

segclass_var

segclass_var は segclass を分散の時変に対応したものである。使用法やオプションは segclass コマンドと同じものである。

A.7 HList

本実験では、メルケプストラムのバイナリのパラメータを見るために、HList コマンド^[9]を用いた。使う時は、オプションの後に内容の見たいファイル名を続ける。

以下にオプション群を示す。


```

xtaniguc@atrh15<111> HList ih_0001.mfc
File: ih_0001.mfc [HTK]
Num Samples: 15          Sample Period: 5000.0 usecs
Sample Size: 40 bytes    Sample Kind: MFCC_Z
0:  1_ -0.5705  2_  1.6225  3_  1.3316  4_ -0.6165  5_ -0.0350
    6_ -0.8695  7_ -0.6167  8_  0.0047  9_ -0.3329 10_ -0.7977
1:  1_  1.2301  2_  0.5614  3_  1.6652  4_ -0.9981  5_ -0.3782
    6_ -0.9420  7_ -0.8740  8_  0.3201  9_ -0.5800 10_ -0.3341
2:  1_  0.8934  2_  0.0508  3_  1.9718  4_ -1.2630  5_ -0.6739
    6_ -1.2365  7_ -0.8495  8_  0.3846  9_ -0.7853 10_ -0.1550
3:  1_  0.1879  2_ -0.3397  3_  1.9944  4_ -1.3195  5_ -0.9423
    6_ -1.2557  7_ -0.8785  8_  0.4466  9_ -0.8232 10_ -0.0149
4:  1_ -0.0699  2_ -0.5184  3_  1.7592  4_ -1.4059  5_ -1.0452
    6_ -1.1827  7_ -0.7876  8_  0.5486  9_ -0.8618 10_  0.1526
5:  1_ -0.0431  2_ -0.5365  3_  1.5933  4_ -1.4105  5_ -0.9001
    6_ -1.0951  7_ -0.5835  8_  0.5561  9_ -0.7578 10_  0.2081
6:  1_ -0.2346  2_ -0.9262  3_  1.3611  4_ -1.7146  5_ -0.8246
    6_ -1.3303  7_ -0.6172  8_  0.3512  9_ -0.8952 10_ -0.0298
7:  1_ -0.0049  2_ -0.7529  3_  1.6310  4_ -1.5020  5_ -0.3264
    6_ -1.0997  7_ -0.3292  8_  0.5896  9_ -0.6213 10_  0.2146
8:  1_  0.0928  2_ -0.5653  3_  1.6427  4_ -1.3709  5_ -0.1335
    6_ -0.9854  7_ -0.2194  8_  0.7229  9_ -0.5212 10_  0.3547
9:  1_  0.1076  2_ -0.5192  3_  1.6272  4_ -1.4487  5_ -0.0298
    6_ -0.9938  7_ -0.2217  8_  0.7093  9_ -0.4926 10_  0.3618
10: 1_ -0.1758  2_ -0.4966  3_  1.4291  4_ -1.4831  5_ -0.1529
    6_ -1.1475  7_ -0.3280  8_  0.5685  9_ -0.5032 10_  0.1857
11: 1_ -0.4363  2_ -0.3783  3_  1.0826  4_ -1.4964  5_ -0.3581
    6_ -1.2466  7_ -0.5247  8_  0.3933  9_ -0.4721 10_ -0.0258
12: 1_ -0.3172  2_ -0.2206  3_  0.9317  4_ -1.4875  5_ -0.2929
    6_ -1.2205  7_ -0.6793  8_  0.3489  9_ -0.4343 10_ -0.0934
13: 1_ -0.0284  2_ -0.2332  3_  0.7838  4_ -1.3676  5_ -0.3005
    6_ -1.3668  7_ -0.7565  8_  0.3276  9_ -0.5961 10_ -0.1888
14: 1_  0.5632  2_  0.0730  3_  1.0289  4_ -0.8452  5_ -0.1912
    6_ -1.2417  7_ -0.6952  8_  0.3638  9_ -0.5395 10_ -0.1111

```

図 A.4: HList の出力結果

オプション		デフォルト
-e N	End at sample N	0
-f	Use full precision	Off
-h	ヘッダ情報のみ表示	Off
-s N	Start at sample N	0
-w N	表示ライン長 N	80
-F fmt	ファイルのフォーマット fmt	HTK
-S f	スクリプトファイル f	none

以下に HList を使った出力結果を図 A.4に示す。

A.8 lmdl

lmdl は、RegModelFile の内容を見るためのものであるが、今回は、これを混合分布に対応させた。使う時は、オプションの後に内容の見たいファイル名を続ける。

以下にオプション群を示す。

オプション		デフォルト
-F s	見るファイルの型	SegFile
-e N	End at model N	0
-s N	Start at model N	0
-S f	スクリプトファイル f	none

サポートされてるファイルの型は以下の 2 つである。

SegFile (セグメントファイル)
RegModelFile (モデルファイル)

以下に lmdl を使った出力結果を図 A.5に示す。

```

xtaniguc@atrh15<137> lmdl -F RegModelFile model/modells3m.2/aa
FileName          : model/modells3m.2/aa
Regression Order  : 2
Vector Dimension  : 10
Number of TrainingSegments : 901 1143 1010
Number of TrainingFrames : 23750 30808 23429
Minimum Duration  : 9 9 9 frames
Maximum Duration  : 74 70 59 frames
Weight            : 0.2950 0.3743 0.3307
Covariance Type   : DIAGONALCOV
Relative Frequencies1:
 9_ 0.0033 10_ 0.0055 11_ 0.0022 12_ 0.0055 13_ 0.0055
14_ 0.0189 15_ 0.0211 16_ 0.0289 17_ 0.0311 18_ 0.0277
19_ 0.0455 20_ 0.0411 21_ 0.0355 22_ 0.0522 23_ 0.0422
24_ 0.0655 25_ 0.0544 26_ 0.0588 27_ 0.0655 28_ 0.0544
29_ 0.0444 30_ 0.0455 31_ 0.0311 32_ 0.0300 33_ 0.0277
34_ 0.0255 35_ 0.0255 36_ 0.0189 37_ 0.0133 38_ 0.0100
39_ 0.0111 40_ 0.0055 41_ 0.0033 42_ 0.0067 43_ 0.0100
44_ 0.0011 45_ 0.0022 46_ 0.0011 47_ 0.0011 48_ 0.0011
49_ 0.0011 50_ 0.0011 51_ 0.0022 52_ 0.0000 53_ 0.0011
54_ 0.0055 55_ 0.0000 56_ 0.0000 57_ 0.0033 58_ 0.0011
59_ 0.0000 60_ 0.0011 61_ 0.0000 62_ 0.0000 63_ 0.0000
64_ 0.0000 65_ 0.0000 66_ 0.0022 67_ 0.0000 68_ 0.0000
69_ 0.0000 70_ 0.0000 71_ 0.0000 72_ 0.0000 73_ 0.0000
74_ 0.0011
Relative Frequencies2:
 9_ 0.0009 10_ 0.0000 11_ 0.0035 12_ 0.0026 13_ 0.0070
14_ 0.0087 15_ 0.0184 16_ 0.0227 17_ 0.0236 18_ 0.0297
19_ 0.0324 20_ 0.0385 21_ 0.0551 22_ 0.0481 23_ 0.0595
24_ 0.0525 25_ 0.0525 26_ 0.0577 27_ 0.0429 28_ 0.0542
29_ 0.0437 30_ 0.0516 31_ 0.0525 32_ 0.0481 33_ 0.0324
34_ 0.0262 35_ 0.0201 36_ 0.0201 37_ 0.0122 38_ 0.0157
39_ 0.0157 40_ 0.0035 41_ 0.0052 42_ 0.0061 43_ 0.0079
44_ 0.0061 45_ 0.0009 46_ 0.0035 47_ 0.0017 48_ 0.0035
49_ 0.0017 50_ 0.0017 51_ 0.0026 52_ 0.0017 53_ 0.0000
54_ 0.0009 55_ 0.0000 56_ 0.0009 57_ 0.0009 58_ 0.0009
59_ 0.0000 60_ 0.0000 61_ 0.0000 62_ 0.0000 63_ 0.0000
64_ 0.0000 65_ 0.0000 66_ 0.0000 67_ 0.0000 68_ 0.0000
69_ 0.0000 70_ 0.0009
Relative Frequencies3:
 9_ 0.0050 10_ 0.0050 11_ 0.0089 12_ 0.0129 13_ 0.0188
14_ 0.0168 15_ 0.0248 16_ 0.0317 17_ 0.0337 18_ 0.0554
19_ 0.0475 20_ 0.0842 21_ 0.0752 22_ 0.0832 23_ 0.0822
24_ 0.0644 25_ 0.0574 26_ 0.0554 27_ 0.0347 28_ 0.0406
29_ 0.0277 30_ 0.0168 31_ 0.0208 32_ 0.0149 33_ 0.0089
34_ 0.0139 35_ 0.0109 36_ 0.0069 37_ 0.0079 38_ 0.0059
39_ 0.0059 40_ 0.0040 41_ 0.0079 42_ 0.0010 43_ 0.0010
44_ 0.0000 45_ 0.0010 46_ 0.0000 47_ 0.0010 48_ 0.0020
49_ 0.0000 50_ 0.0010 51_ 0.0010 52_ 0.0000 53_ 0.0000
54_ 0.0000 55_ 0.0010 56_ 0.0000 57_ 0.0000 58_ 0.0000
59_ 0.0010
Mean Trajectory1:
1_1 2.21871 1_2 2.43362 1_3 -1.81622 2_1 -0.95122 2_2 -5.96423
2_3 5.64317 3_1 -0.65283 3_2 -3.92007 3_3 3.17836 4_1 -0.38022
4_2 -1.12009 4_3 0.98530 5_1 -0.08238 5_2 2.58497 5_3 -2.34002
6_1 -0.08526 6_2 2.67187 6_3 -2.50565 7_1 -0.21030 7_2 0.64941
7_3 -0.53191 8_1 0.10881 8_2 -1.39895 8_3 1.30527 9_1 0.21335
9_2 1.14463 9_3 -0.99169 10_1 -0.10993 10_2 -0.65533 10_3 0.45296
Mean Trajectory2:
1_1 3.57147 1_2 0.61263 1_3 -1.15791 2_1 -0.08786 2_2 -6.46529
2_3 5.48283 3_1 -1.23227 3_2 -2.72476 3_3 2.74589 4_1 -1.70262
4_2 -1.38867 4_3 2.54578 5_1 -0.38924 5_2 2.28865 5_3 -1.82529
6_1 0.83272 6_2 2.45053 6_3 -3.06803 7_1 0.26101 7_2 0.84679
7_3 -1.02669 8_1 -0.20102 8_2 -1.20603 8_3 1.48921 9_1 -0.00373
9_2 0.32088 9_3 -0.05264 10_1 0.11055 10_2 0.13436 10_3 -0.17542
Mean Trajectory3:
1_1 1.86064 1_2 6.41795 1_3 -4.32233 2_1 -0.89427 2_2 -6.58658
2_3 5.26372 3_1 -0.05576 3_2 -5.17947 3_3 3.18525 4_1 -0.27276
4_2 -2.10072 4_3 2.16935 5_1 -0.39779 5_2 1.80966 5_3 -0.93606
6_1 -0.40046 6_2 3.03153 6_3 -2.81272 7_1 -0.45491 7_2 0.81022
7_3 -0.76498 8_1 0.28630 8_2 0.67434 8_3 -0.24441 9_1 0.19640
9_2 -0.67179 9_3 0.54100 10_1 0.10663 10_2 0.55390 10_3 -0.70979
Covariance1:
1_ 1.59316 2_ 1.25703 3_ 0.90676 4_ 0.56952
5_ 0.66961 6_ 0.55023 7_ 0.49275 8_ 0.30211
9_ 0.27000 10_ 0.25349
Covariance2:
1_ 1.53735 2_ 1.39937 3_ 0.73898 4_ 0.90411
5_ 0.59959 6_ 0.49979 7_ 0.44502 8_ 0.30814
9_ 0.26263 10_ 0.22619
Covariance3:
1_ 1.98102 2_ 1.09123 3_ 1.13608 4_ 0.62831
5_ 0.55942 6_ 0.59452 7_ 0.45914 8_ 0.30700
9_ 0.31208 10_ 0.22168

```

図 A.5: lmdl の出力結果

lmdl_var

lmdl_var は分散時変モデルのパラメータを見るためのものである。

A.9 HRmIO

HRmIO ライブラリでは、セグメントファイルとモデルファイルの2つのフォーマットに対し、入力や出力を行なうものであり、この実験で使ったほとんどのプログラムがこのライブラリを使っている。従来のもの^[8]は、モデルファイルのフォーマットにおいて、混合分布に対応していないので、これを変更し、混合分布に対応したフォーマットに対応させた。ただし、分散は対角行列のみに対応している。