TR－SLT－0060

直積混合分布を用いた音響モデルの検討
Acoustic Model using Direct-Product Distributions

武田 晴登　Haruto Takeda
松田 繁樹　Shigeki Matsuda
中村 哲　Satoshi Nakamura

2003 年 10 月 23 日

概要

隠れマルコフモデル(Hidden Markov Model)の状態出力確率分布として直積混合分を用いた新しい音響モデルに関する定式化と評価実験について報告する．音声認識に広く用いられている HMM は，パラメータ数を増加させることによりモデルを精密化することができる．しかし，パラメータ数の増加によって個々のパラメータを推定するために用いられる学習データ量が減少するため，いわゆる過学習によって音声認識性能が低下する．本報告では，特徴量ベクトルの個々の成分における混合分布の直積を状態出力確率分布として用いる新しい音響モデルについて検討を行う．このように表現された分布を本報告では「直積混合分布」と呼ぶこととする．直積混合分布を用いることにより，従来の混合分布と比べて，パラメータ数が等しかったとしても，より広い音響空間を覆うことができると考えられる．従って，少ないパラメータ数で大量の分布を推定することができ，従来よりもロバストな音声認識が可能となることが期待される．本報告では，EM(Expectation Maximization)アルゴリズムに基づく，直積混合分布パラメータの推定方法及び，直積混合分布の構造である，個々の特徴量に割り当てられる混合数の決定方法について述べる．切り出し音素の識別の評価実験では，直積混合分布を用いた場合と，従来の混合分布を用いた場合の認識精度は同程度であった．

（株）国際電気通信基礎技術研究所
音声言語コミュニケーション研究所
〒619－0288「けいはんな学研都市」光台二丁目 2 番地 2　TEL：0774－95－1301

Advanced Telecommunication Research Institute International
Spoken Language Translation Research Laboratories
2-2-2 Hikaridai "Keihanna Science City" 619-0288,Japan
Telephone:+81-774-95-1301
Fax　:+81-774-95-1308

目 次

第1章　はじめに

現在の音声認識技術は，クリーンな音声に対して90%以上の認識性能が達成されている．しかしながら，実際に用いることのできるタスクの大きさは現在でも極めて限定されている．講演音声の自動書き起こしや，自動車内雑音下で使用されるカーナビゲーションシステムでは，話者や発話環境に依存しない頑健な音声認識が求められている．音声認識の研究は，音響モデルや言語モデルなどの認識モデルの根幹を成す部分の研究や，対雑音，話者適応，などの実環境で使用されるモデル適応についての研究など，各々の分野において研究が盛んに進められている．

現在の音声認識システムでは，大量の学習データを用いて推定された統計的な確率モデルを用いる手法が一般的である．このような確率モデルとして隠れマルコフモデル (HMM: hidden Markov model) が広く用いられている．確率モデルは，自由パラメータ数を増加させることによって，学習データの分布形状をより精密に表現することができる．しかし，パラメータ数の増加によって個々のパラメータを推定するために用いられる学習データ量が減少するため，いわゆる過学習と呼ばれる状態に陥る可能性がある．つまり，過度に精密な記述能力を持つ確率モデルは，学習データに過度に適応化され，本来音声が持つ統計的性質からは遠ざかると考えられる．また，パラメータ数が増えることの実装上の問題点として，計算に使用する記憶容量が増大することや，計算量が増加することが挙げられる．

音素の前後関係を考慮した音素環境依存モデルは，個々の環境依存音素毎に別々のHMMによってモデル化される．しかし，日本語で用いられる環境依存音素数は数千に及ぶため，個々の環境依存音素毎にHMMを準備した場合，自由パラメータは膨大な数となる．そこで，個々のHMMの状態間で，類似した分布を持つHMM状態同士を共有化することにより，モデル全体のパラメータ数を削減するための手法が提案されている．ML-SSS(Maximum Likelihood Successive State Splitting)法 [1] は，状態共有構造を自動的に推定する手法である．また，モデル全体のパラメータ数を自動決定するための手法として，MDL 基準を用いる手法 [2] が提案されている．

本報告では，特徴量ベクトルの個々の成分における混合分布の直積を状態出力確率分布として用いる新しい音響モデルについて検討を行う．このように表現された分布を「直積混合分布」と呼ぶこととする．直積混合分布を用いることによっ

て，従来の混合分布と等しいパラメータ数を持っていたとしても，より広い音響空間を覆うことができると考えられる．従って，少ないパラメータ数で大量の分布を推定することができ，従来よりもロバストな音声認識が可能となることが期待される．

特徴量ベクトルの個々成分の振る舞いに着目した研究としては，個々の成分の状態遷移がお互いに非同期なタイミングで発生する非同期遷移型HMM [3] や，HMMのモデルパラメータを4つの階層に分類し，各々の階層でパラメータ共有構造を決定する手法 [4] が提案されている．前者の研究は，個々の特徴量の状態遷移の非同期性を表現するための研究である．また後者の共有化手法は，モデルを記憶するための容量の観点からの研究である．従って，本報告で扱うような，より広い音響空間を表現するための研究とは目的が異なる．

本報告では，EM(Expectation Maximization) アルゴリズムに基づく，直積混合分布パラメータの推定方法及び，直積混合分布の構造である，個々の特徴量に割り当てられる混合数の決定方法について述べる．2章で直積混合分布の構造及び，モデルパラメータの推定方法ついて述べる．3章では，提案する確率分布のパラメータ推定の方法について述べる．4章では提案手法を実装し，切り出し音素分類実験を行う．

第2章　直積混合分布

本章では，音響モデルの音声特徴量ベクトルの確率分布としての直積混合分布について述べる．2.1節では，直積混合分布と従来の混合分布を比較する．2.2節では，音声特徴量の確率分布として使用する意義に述べ，2.3節で定式化を行う．

2.1　直積混合分布の導入

D次元の特徴量ベクトル$\boldsymbol{x} = (x^{(1)}, \cdots, x^{(D)})$の出現確率$p(\boldsymbol{x})$は，それぞれの成分の同時確率$p(x^{(1)}, \cdots, x^{(D)})$として与えられる．各多次元正規分布をベクトルの各成分に混合正規分布を使用しその直積を用いた確率分布を，「直積混合分布」と呼ぶこととする．従来のベクトルの確率分布として，多次元混合正規分布 (GMM: Gaussian Mixture Model) と比較した場合，提案する確率分布は，

A. 同じパラメータ（平均，分散）を用いて，より広い空間を覆える

B. 少ないパラメータで効率的に特徴量空間を覆える

という特徴を持つ．直観的な理解を示すため，以下，2次元の特徴量空間について具体例を示す．

A. について，2つの混合分布から成るGMMついて考察する．2つの分布は各々に2次元の平均ベクトルを持つので，図2.1の左に示すように各成分の周辺分布としては2+2=4の合計4つの分布を持つことになる．逆に，これらの各次元の周辺分布がその次元での混合分布であり，その直積としてベクトルの確率分布が与えられるとすると，図2.1の右に示すように特徴量空間内に2×2=4の合計4つの分布が表される．各成分の分布に混合分布の直積を用いた場合は，特徴量ベクトルの確率分布の混合を考えた場合に対して増えた2つの分布を，本報告では「直積混合分布」と呼ぶ．ここに示すように，各成分の確率分布の直積で表した直積混合分布を含む分布は，より広い特徴領空間の領域を覆うことになる．

B. を理解するために，2次元の特徴量で，図2.2に示されるように，ある成分（第2成分）の周辺分布の2つの分布の中心が近接している場合を考える．ベクトル確率分布の混合分布で覆われる空間と，各成分の混合分布の直積によって覆わ

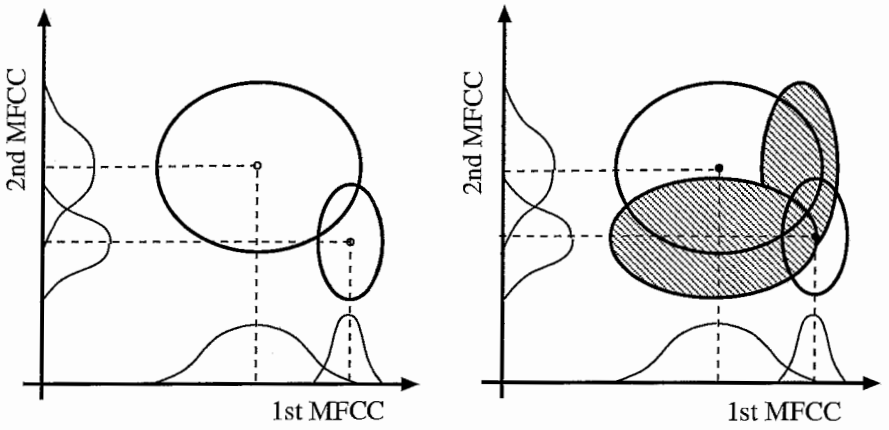

図 2.1: 同じ平均，分散のパラメータを用いた場合，ベクトルの確率分布の混合(左）より各成分の確率分布の混合の直積(右）の方がより広い空間を覆うことができる．

れる空間は，図2.2で見られるようにほぼ等しい．しかし，パラメータ数はより少ないパラメータ数になっている．例えばこのモデルの平均値として記憶すべき倍精度実数の個数について考えると，左では2×2 （2次元の平均ベクトルが2つ）で4であるが，右では2+1（第1成分の中心と第2成分の中心）で3である．

このように定義される直積混合分布を含んだ確率分布は，各成分の直積であるので，従来の多次元正規分布では特徴量の相関の情報が反映されない[1]．そこで，我々はこれらの相関も考慮に入れるために，図に示すように直積混合分布を含んだ特徴量空間内の確率分布の混合分布を考える．

2.2 音声特徴量ベクトルの分布としての直積混合分布

音声認識の音響モデルにおいて，音声特徴量ベクトルが従う確率分布として直積混合分布を使用する場合に期待されることを以下に挙げる．

- **詳細な形状を効率的に表現する確率分布**
  特徴量の成分のレベルでの混合分布と，ベクトルレベルでの混合分布を組み合わせることにより，より詳細な分布形状を表現することができるので，認識精度が向上が期待される．またに，基本分布では各成分ごとに異なる分布数を用いるので，各成分ごとの適した分布形状を表現するように，統計をも

[1] $P(X_1, X_2) = p(X_1)P(X_2)$ のように2変数の確率変数の同時確率が各確率変数の積であることは各確率変数が独立であることを意味する

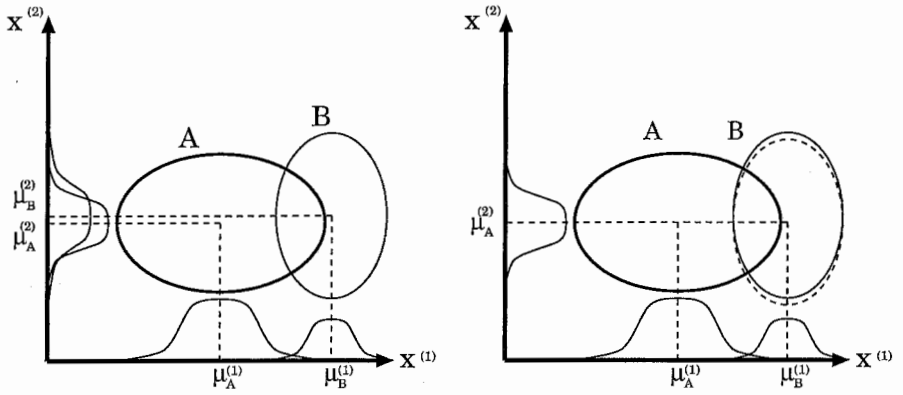

図 2.2: 2つの分布のパラメータが近い場合 (左) は，それらのパラメータを一つの値で代表させた分布を用いた場合 (右) と分布形状はほぼ等しい．

とにバランス良くパラメータ数を配置することにより効率良く分布形状を表現することが可能になると考えられ，少ないパラメータ数で従来手法と同じ確率分布を表現できる可能性もある．

- **学習アルゴリズム**
  次章で述べるように，確実に尤度を上昇させる学習アルゴリズムが存在する．従来と同じ Viterbi Training もしくは Baum-Welch Algorithm の枠組で学習することが可能である．

- **見えないデータの分布をカバー**
  認識では常に未知のデータを対象とするため，学習データには十分現れなかった振舞をするサンプルを扱わなくてはならに場面がある．このように学習データの統計では「見えない」サンプルも，直積混合分布によってカバーされるかも知れない．「見えない」データも既に得られた分布からわずかにずれた位置に存在すると予想されるので，これらのサンプルの分布が直積によって現れる直積混合分布によって覆われる可能性は十分考えられるからである．

しかし，以下の様にマイナス要因となり得る内容もこのモデルには含まれている．

- **分布関数の平滑化**
  直積を用いることにより広い空間を覆うので，分布の確率密度の値は相対的に低くなり，全体として平らな分布になる可能性がある．一般に，広く覆うことと分布形状を鋭くすることはトレードオフの関係にある．このため，認

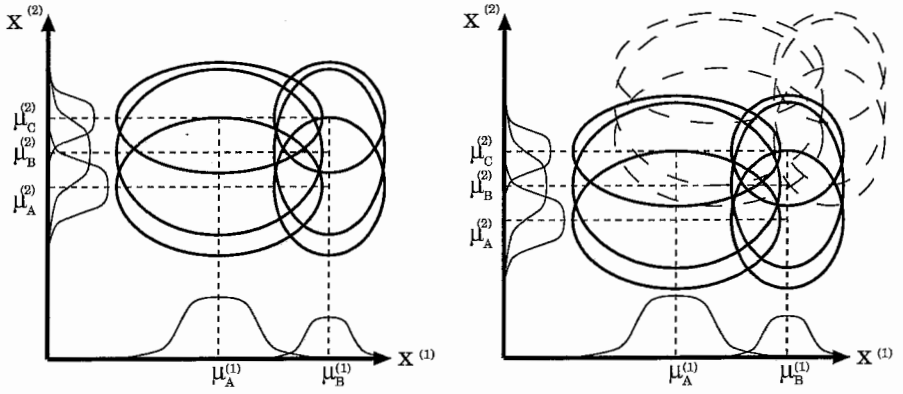

図 2.3: 直積混合分布を含んだ特徴量空間内の確率分布を「基本分布」（左図）とし，基本分布の混合分布を用いて特徴量の各成分の相関を反映させた分布にする

識過程で比較する各仮説の尤度の差が小くなり認識精度の劣化につながる可能性がある．

- **局所最適解への収束**
  特徴量レベルと各成分レベルでの混合分布を用いるために非常に自由度の大きい分布であり，パラメー タ空間が広くなる．本報告で述べる学習アルゴリズムは尤度の上昇のみを約束するものであるため，局所最適解に収束していまい，適切なパラメータ値が得られない可能性がある．

2.3 直積混合分布の定式化

音響モデルの HMM において音声特徴量ベクトルが従う確率分布について，従来の分布と提案モデルの分布を定式化を行う．本報告で使用する数式の記号を表 2.1 に示す．

現在一般的に用いられている特徴量ベクトルの GMM では，共分散行列が対角行列である対角共分散を使用している．平均 $\boldsymbol{\mu} = (\mu^{(1)}, \cdots, \mu^{(D)})$，対角共分散 $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_D)$ の多次元正規分布 $N(\boldsymbol{\mu}, \Sigma)$ の確率密度 $p(\boldsymbol{x})$ は，各成分の確率密度 $p_d \sim N(\mu^{(d)}, (\sigma^2)^{(d)})$ の積と等価になる．

$$p(\boldsymbol{x}) = \prod_{d=1}^{D} p_d(x^{(d)} | \mu^{(d)}, (\sigma^2)^{(d)})$$

これらの混合分布であるベクトル $\boldsymbol{x}$ の確率分布は

表 2.1: 本報告の数式表現で用いる記号

$\boldsymbol{x}_t$	時刻 t の音声特徴量ベクトル
$x_t^{(d)}$	時刻 t の音声特徴量ベクトル $\boldsymbol{x}_t$ の第 d 成分
T	音声特徴量ベクトルの時系列のサンプル数 （フレーム数)
$t = 1, \cdots, T$	音声特徴量ベクトルの時間インデックス
M	基本分布（ベクトルの確率分布) の数
D	音声特徴量の次元
C(d)	ベクトルの各次元での 1 次元の確率分布の混合数
m = 1,$\cdots$,M	基本分布（ベクトルの確率分布) のインデックス
d = 1,$\cdots$,D	音声特徴量の次元のインデックス
c = 1,$\cdots$,C(d)	ベクトルの各次元での混合のインデックス
λ_m	第 m 基本分布（ベクトルの確率分布）m への分岐確率
w_{mdc}	第 m 基本分布の特徴量ベクトルの第 d 成分の第 c_d 分布への分岐確率
μ_{mdc}	第 m 基本分布の次元 d での混合 c の分布の平均
σ_{mdc}	第 m 基本分布の次元 d での混合 c の分布の分散
$p_m(\boldsymbol{x}) \sim N_D(\boldsymbol{\mu}, \Sigma)$	第 m 基本分布の特徴量ベクトル $\boldsymbol{x}$ の確率密度関数
$p_{mdc}(x^{(d)}) \sim N(\mu_{mdc}, \sigma^2_{mdc})$	第 m 基本分布の特徴量ベクトル $\boldsymbol{x}$ の第 d 成分の第 c 分布の確率密度関数
$N(\mu, \sigma^2)$	平均 μ, 分散 σ^2 の正規分布 (確率密度関数は $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$)
$N_D(\boldsymbol{\mu}, \Sigma)$	平均ベクトル $\boldsymbol{\mu}$, 共分散行列 Σ の多次元正規分布 (D 次元)

$$p(x^{(1)},\cdots,x^{(D)}) = \sum_{m=1}^{M} \lambda_m \prod_{d=1}^{D} p_{m,d}(x^{(d)}) \tag{2.1}$$

と表される．ここで，λ_m はベクトルの混合分布における各多次元正規分布への分岐確率を表し，$\sum_{m=1}^{M} \lambda_m = 1$ を満たす正数である．この分布が持つ一次元の正規分布の数は MD となる．

これに対し，我々は特徴量の各成分に混合分布を用いた確率分布を考える．確率分布は各成分に混合正規分布を用い，その直積を「基本分布」と呼ぶ．基本分布は以下の式で表される．

$$p_m(\boldsymbol{x}) = p_m(x^{(1)},\cdots,x^{(D)}) = \prod_{d=1}^{D} \sum_{c=1}^{C(d)} w_{mdc} p_{mdc}(x^{(d)})$$

基本分布は，混合分布の直積でクロスタームが表す直積混合分布を含んでいる．この基本分布の混合分布は，次式で表される．

$$p(x^{(1)},\cdots,x^{(D)}) = \sum_{m=1}^{M} \lambda_m p_m(\boldsymbol{x}) = \sum_{m=1}^{M} \lambda_m \prod_{d=1}^{D} \sum_{c=1}^{C(d)} w_{m,d,c} p_{m,d,c}(x^{(d)}) \tag{2.2}$$

$w_{m,d,c}$ は各次元の確率分布の分岐確率を表す．一次元の正規分布の数は $M\sum_{d=1}^{D} C(d)$ となる．式 (2.2) において，全ての成分について $C(d) = 1$，従って $w_{mdc} = 1$ とすると，式 (2.1) に一致する．GMM では，分岐確率が λ_i のみであるが，直積混合分布を含む分布の混合分布では，基本分布の分岐確率 λ_i と各成分での混合分布の分岐確率 $w_{m,d,c}$ の 2 つのレベルでの分岐確率がある．式 (2.2) で表される直積混合分布を含む確率分布において，各成分の混合数 $C(d)$ を 1 とした場合は，式 (2.1) で表される GMM と等価になる[2]．

[2]GMM と等価であることとは，即ち，直積混合分布が表れない場合に相当する．

第3章　直積混合分布を用いた音響モデルの学習方法

本章では，直積混合分布を用いた音響モデルの学習方法について述べる．この音響モデルは基本的に混合分布であるので，まず3.1節でに混合正規分布のパラメータ推定方法について述べる．続く3.2節で，直積混合分布を含む確率分布のパラメータ推定方法を述べる．3.3節で音響モデルの学習手順を述べる．

3.1　混合正規分布のパラメータ推定

式(2.1)で表されるGMMのパラメータは，分岐確率λ_iと各分布の平均μと分散σである．これらのパラメータ値を推定は，単一正規分布を初期分布として，

1. パラメータ値更新のくり返しによるパラメータ推定
2. 分布の分割による混合数の増加

を目的の混合数に達するまで繰り返す．

パラメータ値更新のくり返しによるパラメータ推定

混合分布のパラメータ推定には，EMアルゴリズムを用いて最尤推定を行うのが一般的である．EMアルゴリズムによって導かれるパラメータの更新式は，

$$\hat{\mu} = \frac{\sum_{t=1}^{N} \boldsymbol{x}_t \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)} \tag{3.1}$$

$$\left(\hat{\Sigma}\right)_{ij} = \frac{\sum_{t=1}^{N} \boldsymbol{x}_t \Sigma^{-1} \boldsymbol{x}_t \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)} - \hat{\mu}^2 \tag{3.2}$$

$$\tag{3.3}$$

であり学習データに対する尤度の増加を保証する．混合正規分布場では各分布からの尤度の寄与は，各分布の尤度の重み付の和で与えられるが，これは，一つの

サンプルが各分布の度の程度寄与するかを確率的に考え，混合正規分布の複数の確率分布の重ね合わせは，一つのサンプルが複数の分布の属する確率を $\gamma_t(i)$) を用いて与えている．EM アルゴリズムによる導出はよく知られており，3.2.2 節における議論で $C(d) = 1$(基本分布全ての成分の混合数が 1) とおいた場合に対応するので，ここでは省略する．

分布の分割による混合数の増加

分布数を増やす過程では，混合分布の各分布を 2 分割するのが一般的である．分割されてできる 2 つの分布は，重みを 2 等分し，新しい平均 $\boldsymbol{\mu}'$ を

$$\boldsymbol{\mu}' = \boldsymbol{\mu} \pm \alpha\sigma$$

のように，分割前の分布の平均 $\boldsymbol{\mu}$ を標準偏差のオーダーでずらす方法などがある．($\alpha < 1$)

3.2 直積混合分布のパラメータ推定

直積混合分布を含んだ確率分布も，GMM と同様にパラメータ値の更新と分布の分割をくり返し分布を推定する．

3.2.1 パラメータ更新式

特徴量ベクトルの分岐確率 $\hat{\lambda}$, ベクトルの各次元の分岐確率 $\hat{w}_{mdc}$, 各自元の各分布の平均 $\hat{\mu}_{mdc}$, 分散 $\hat{\sigma}_{mdc}$ を

$$\hat{\lambda}_m = \frac{\sum_t \gamma_t(m)}{\sum_t 1} \tag{3.4}$$

$$\hat{w}_{mdc} = \frac{\sum_t \gamma_t(m)\gamma'_t m, d, c}{\sum_t \gamma_t(m)} \tag{3.5}$$

$$\hat{\mu}_{mdc} = \frac{\sum_t \gamma_t(m)\gamma'_t(m,d,c)x_t^{(d)}}{\sum_t \gamma_t(m)\gamma'_t(m,d,c)} \tag{3.6}$$

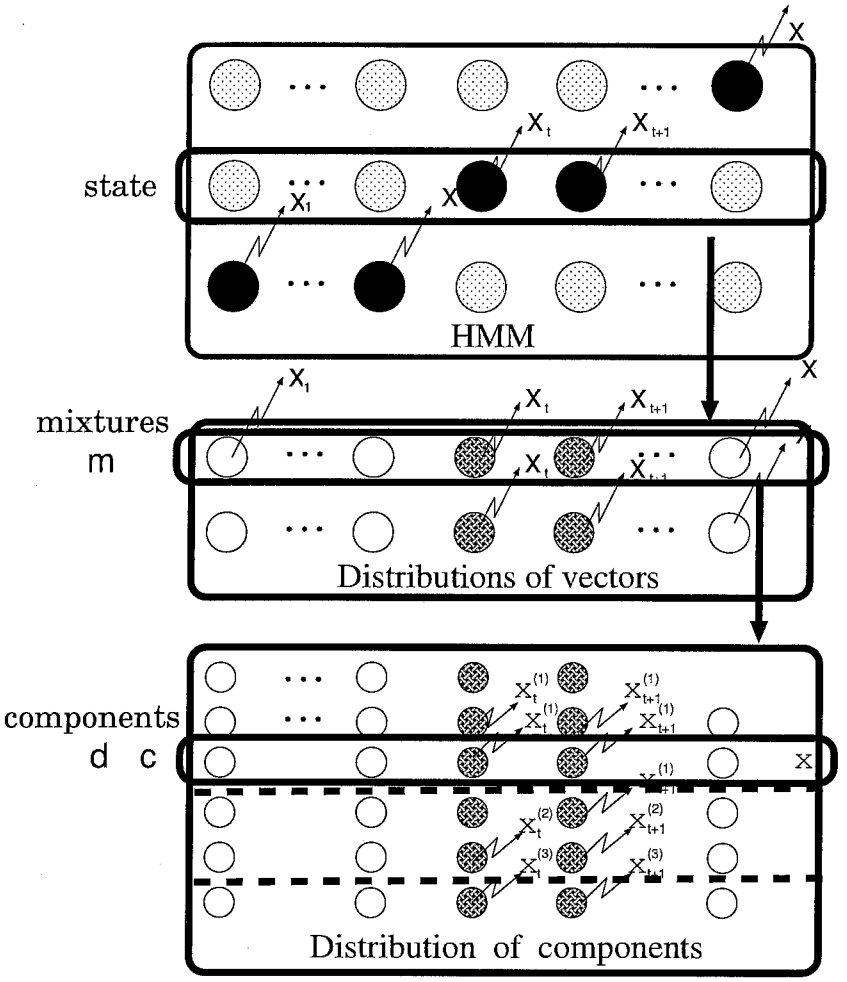

図 3.1: 階層的な十分統計量を用いたパラメータ更新を行う

$$\hat{\sigma}_{mdc} = \frac{\sum_t \gamma_t(m)\gamma'_t(m,d,c)x_t^{(d)2}}{\sum_t \gamma_t(m)\gamma'_t(m,d,c)} - \hat{\mu}^2_{mdc} \tag{3.7}$$

として推定する．ここで，$\gamma_t(m)$ は，更新前のモデルでサンプル $\boldsymbol{x}_t$ が基本分布 m に属する確率であり，$\gamma_t(m,d,c)$ はそのサンプルの第 d 成分が成分レベルの混合分布 c に属する確率である．これらの更新式は次の節で導くが，推定式そのものは十分統計量がモデルの階層性を反映したものとなっており，直観的に理解しやすい．

3.2.2 EMアルゴリズムによるパラメータ更新式の導出

最尤法とは，確率モデルのパラメータ $\boldsymbol{\theta}$ を学習用のサンプルデータ $X = \{\boldsymbol{x}_t\}_{t=1}^{T} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T\}$ に対して尤度 $L(X)$ を最大にするパラメータ $\hat{\boldsymbol{\theta}}$ を推定する方法である．EMアルゴリズムでは，尤度を増加させるパラメータ更新をくり返すことにより最尤推定値を得ようとする方法であり，必ずしも最尤推定値が得られるとは限らないが，学習データに対する尤度の増加は保証されている．

直積混合分布を含んだ確率分布の場合は，パラメータ $\boldsymbol{\theta} = \{\lambda_m, w_{mdc}, \mu_{mdc}, \sigma_{mdc} | m = 1 \cdots M, d = 1 \cdots D, c = 1 \cdots C(d)\}$ のときの学習用データ X に対する尤度は

$$L(X|\boldsymbol{\theta}) = \prod_{t=1}^{T} \underbrace{\left[\sum_{m=1}^{M} \lambda_m \prod_{d=1}^{D} \sum_{c=1}^{C(d)} w_{mdc} p_{mdc}(x_t(d) | \mu_{mdc}, \sigma_{mdc}^2) \right]}_{=P(\boldsymbol{x}_t|\boldsymbol{\theta})} \tag{3.8}$$

と表される．混合分布であるので，各サンプル $\boldsymbol{x}_t$ がどの分布に属しているのか分からないので，パラメータを $\boldsymbol{\theta}$ を $\bar{\boldsymbol{\theta}}$ に更新した時の尤度の対数の増分

$$\log L(X|\hat{\theta}) - \log L(X|\theta) = \log \frac{L(X|\hat{\theta})}{L(X|\theta)} = \sum_{t=1}^{T} \log \frac{P(\boldsymbol{x}_t|\bar{\boldsymbol{\theta}})}{P(\boldsymbol{x}_t|\boldsymbol{\theta})} \tag{3.9}$$

が正となるようにパラメータ更新式を導きたい．各サンプルの基本分布，および各成分でのでの混合分布でそれぞれの分布への帰属についての期待値をとると，

$$\sum_{m=1}^{M} \prod_{d=1}^{D} \sum_{c_d=1}^{C(d)} p(m, c_1, \cdots, c_D | \boldsymbol{x}_t, \boldsymbol{\theta}) \cdot \log \frac{L(X|\hat{\theta})}{L(X|\theta)} \tag{3.10}$$

となる．ここで，期待値をとるとは，

$$\sum_{m=1}^{M} \prod_{d=1}^{D} \sum_{c_d=1}^{C(d)} p(m, c_1, \cdots, c_D | \boldsymbol{x}_t, \boldsymbol{\theta}) = \underbrace{\sum_{m=1}^{M} P(m|\boldsymbol{x}_t, \boldsymbol{\theta})}_{=1} \cdot \prod_{d=1}^{D} \underbrace{\sum_{c=1}^{C(d)} P(m, c | x_t^{(d)})}_{=1} \tag{3.11}$$

による和をとることに相当する．さらに帰属を明らかにしたときの確率を用いるために，

$$P(\boldsymbol{x}_t|\boldsymbol{\theta}) = \frac{P(\boldsymbol{x}_t, m, c_1, \cdots, c_D|\boldsymbol{\theta})}{P(m, c_1, \cdots, c_D|\boldsymbol{x}_t, \boldsymbol{\theta})} \tag{3.12}$$

を式 (3.10) に代入すると，

$$
\begin{aligned}
&= \sum_{t=1}^{T}\sum_{m=1}^{M}\prod_{d=1}^{D}\sum_{c_d=1}^{C(d)} p(m,c_1,\cdots,c_D|\boldsymbol{x}_t,\boldsymbol{\theta})\cdot \\
&\quad \log\frac{P(\boldsymbol{x}_t,m,c_1,\cdots,c_D|\boldsymbol{\theta})/P(m,c_1,\cdots,c_D|\boldsymbol{x}_t,\boldsymbol{\theta})}{P(\boldsymbol{x}_t,m,c_1,\cdots,c_D|\boldsymbol{\theta})/P(m,c_1,\cdots,c_D|\boldsymbol{x}_t,\boldsymbol{\theta})} \\
&= Q(\boldsymbol{\theta},\bar{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta},\boldsymbol{\theta}) + \sum_{t=1}^{T}\sum_{m=1}^{M}\prod_{d=1}^{D}\sum_{c_d=1}^{C(d)} p(m,c_1,\cdots,c_D|\boldsymbol{x}_t,\boldsymbol{\theta})\cdot \\
&\quad \frac{P(m,c_1,\cdots,c_D|\boldsymbol{x}_t,\boldsymbol{\theta})}{P(m,c_1,\cdots,c_D|\boldsymbol{x}_t,\boldsymbol{\theta})}
\end{aligned}
\tag{3.13}
$$

であるが，第3項はJensenの不等式から正であることが保証されているので，第1，2項である$Q(\boldsymbol{\theta},\bar{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta},\boldsymbol{\theta})$を正であれば，パラメータ更新によって尤度を増加させられる．ここで現れた

$$
Q(\boldsymbol{\theta},\bar{\boldsymbol{\theta}}) = \sum_{t=1}^{T}\sum_{m=1}^{M}\prod_{d=1}^{D}\sum_{c_d=1}^{C(d)} p(m,c_1,\cdots,c_D|\boldsymbol{x}_t,\boldsymbol{\theta})\cdot\log\frac{P(\boldsymbol{x}_t,m,c_1,\cdots,c_D|\boldsymbol{\theta})}{P(\boldsymbol{x}_t,m,c_1,\cdots,c_D|\boldsymbol{\theta})} \tag{3.14}
$$

はQ関数である．Q関数の増分が最大になるようにパラメータを更新する．
ここで簡便な表現を行うために，ガンマ変数を以下のように導入する．

$$
\gamma_t(m) = p(m|\boldsymbol{x}_t,\boldsymbol{\theta}) = \frac{\lambda_m p_m(\boldsymbol{x}_t)}{\sum_{m'}\lambda'_m p_{m'}(\boldsymbol{x}_t)} \tag{3.15}
$$

$$
\gamma'_t(m,d,c) = p(m,d,c|\boldsymbol{x}_t,\boldsymbol{\theta}) = \frac{w_{mdc}p_{mdc}(\boldsymbol{x}_t)}{\sum_{c'}w_{mdc'}p_{mdc'}(\boldsymbol{x}_t)} \tag{3.16}
$$

$\gamma_t(m)$は時刻tのサンプル$\boldsymbol{x}_t$がのベクトルの確率分布mに属する確率を表す．$\gamma_t(m,d,c)$は$\boldsymbol{x}_t$がベクトルの確率分布mの第d次元目のベクトル成分の混合分布の第c分布に属する確率を表す．さらに，$\log p(\boldsymbol{x}_t,m,c_1,\cdots,c_{26})$を書き下した

$$
\log\lambda_m + \sum_{d=1}^{D}\left[\log w_{m,d,c} - \frac{1}{2}\log(2\pi\sigma^2_{m,d,c}) - \frac{(x_t^{(d)}-\mu_{m,d,c})^2}{2\sigma^w_{m,d,c}}\right]
$$

を用いてQ関数を書き改めると

$$
\begin{aligned}
Q(\boldsymbol{\theta},\bar{\boldsymbol{\theta}}) &= \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{c_1=1}^{C(m,1)}\cdots\sum_{c_D=1}^{C(m,D)}\gamma_t(m)\gamma_t(m,d,c_1)\cdots\gamma_t(m,d,c_D) \\
&\times\left[\log\lambda_m + \sum_{d=1}^{D}\left[\log w_{m,d,c} - \frac{1}{2}\log(2\pi\sigma^2_{m,d,c}) - \frac{(x_t^{(d)}-\mu_{m,d,c})^2}{2\sigma^w_{m,d,c}}\right]\right]
\end{aligned}
\tag{3.17}
$$

となる．ただし，ベクトル及びその各次元の混合分布の分岐確率は

$$\sum_{m=1}^{M}\lambda_m = 1 \qquad \sum_{c=1}^{C(d)} w_{m,d,c} = 1 \quad (d = 1,\cdots,D) \tag{3.18}$$

を満たす．

式 (3.17) で表される関数を式 (3.18) の束縛条件のもとで最大化するパラメータは Lanrange の未定乗数法を用いて求められる．即ち，最大化する目的関数を

$$L(\bar{\boldsymbol{\theta}}) = Q(\boldsymbol{\theta},\bar{\boldsymbol{\theta}}) + k\left(1-\sum_{m=1}^{M}\lambda_m\right) + \sum_{d=1}^{D}\left\{l_d\left(1-\sum_{c_d=1}^{C(d)} w_{m,d,c}\right)\right\}$$

とし，未定乗数 k,l_d $(d = 1,\cdots,D)$ を用いた項を導入ことにより束縛条件の情報も目的関数に加えて考える．目的関数は，各パラメータに付いて上に凸である[1]ので，極値を求めればよい．$m = 1,\cdots,M$,$d+1,\cdots,D$,$c_d = 1,\cdots,C(d)$ の各パラメータに対し，次が得られる．

$$\frac{\partial}{\partial\bar{\lambda}_m}L(\bar{\boldsymbol{\theta}}) = 0 \text{ より} \qquad \bar{\lambda}_m = \frac{1}{k}\sum_{t=1}^{T}\gamma_t(m) = \frac{\sum_{t=1}^{T}\gamma_t(m)}{\sum_{m'=1}^{M}\sum_{t=1}^{T}\gamma_t(m')}$$

$$\frac{\partial}{\partial\bar{w}_{m,d,c}}L(\bar{\boldsymbol{\theta}}) = 0 \text{ より} \quad \bar{w}_{m,d,c} = \frac{1}{l_d}\sum_{t=1}^{T}\gamma_t(m)\gamma'_t(m,d,c) = \frac{\sum_{t=1}^{T}\gamma_t(m)\gamma'_t(m,d,c)}{\sum_{c'_d=1}^{C(d)}\sum_{t=1}^{T}\gamma_t(m)\gamma'_t(m,d,c_d)}$$

$$\frac{\partial}{\partial\bar{\mu}_{mdc}}L(\bar{\boldsymbol{\theta}}) = 0 \text{ より} \qquad \bar{\mu}_{mdc} = \frac{\sum_{t=1}^{D}\gamma_t(m)\gamma'_t(m,d,c_d)x_t^{(d)}}{\sum_{t=1}^{D}\gamma_t(m)\gamma'_t(m,d,c_d)}$$

$$\frac{\partial}{\partial\bar{\sigma^2}_{mdc}}L(\bar{\boldsymbol{\theta}}) = 0 \text{ より} \qquad \bar{\sigma^2_{mdc}} = \frac{\sum_{t=1}^{D}\gamma_t(m)\gamma'_t(m,d,c_d)(x_t^{(d)}-\bar{\mu_{mdc}})^2}{\sum_{t=1}^{D}\gamma_t(m)\gamma'_t(m,d,c_d)}$$

未定乗数の値は式 (3.18) を用いて定めた．これにより，直積混合分布のパラメータ更新式である式 (3.4)(3.5)(3.6)(3.7) が得られた．

[1] λ_m,$w_{m,d,c}$,$\sigma_{m,d,c}$ は log による凸性，$\mu_{m,d,c}$ については 2 次関数による凸性のため．

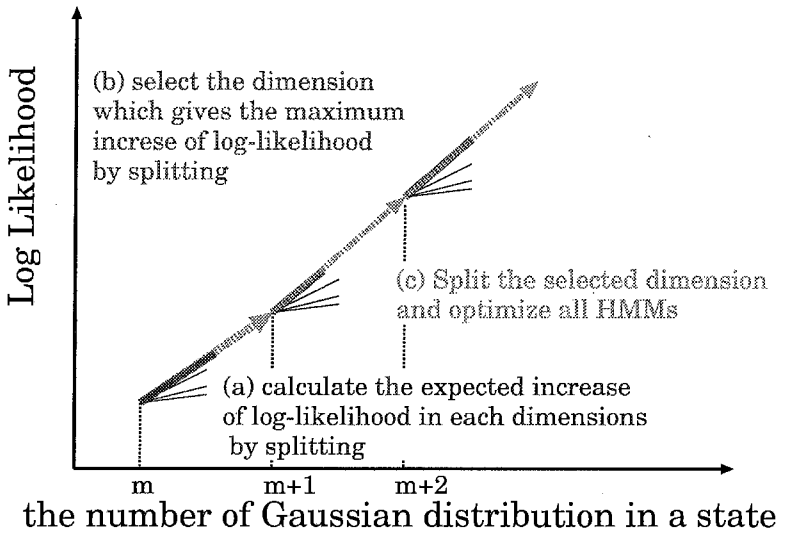

図 3.2: 貪欲なアルゴリズムで混合数を増やす次元を選択し，学習データに対する最尤が上昇する過程のイメージ図

3.2.3 分布の分割方式

分割・最適化の繰り返しにおける各処理で最大限の尤度増加を図るため，混合数を奏化させるための分布の選択には，分割して最適化 (パラメータ学習) した場合に最も尤度の増加が大きい次元を分割すれば良い．このように最大限の目的関数の増加を得るように学習を進めていく方法は，一般に貪欲なアルゴリズムと呼ばれている．貪欲に選択するが，全体での最適性は保証されていない．

このような手順で，

- パラメータ数の追加：貪欲なアルゴリズムによる次元の選択
- パラメータ値の更新：EM アルゴリズムによる最尤推定

をくり返す様子を，図 3.2 に示した．我々の提案手法は，各ステップで最も上昇する選択肢を取ることに当たる．

適切な基本分布の形を表すのに必要な各成分の分布数をどのように決めるかは，今回は複数の分布数のものを試すことにした．尚，各成分の分布数を決定するのに，尤度を用いることは適切でないと思われる．例えば，全ての成分の尤度が一致するときが適切な分布数でるあとすることは，できない．図 3.3 に見られるように，各成分によって収束する尤度が異なるからである．

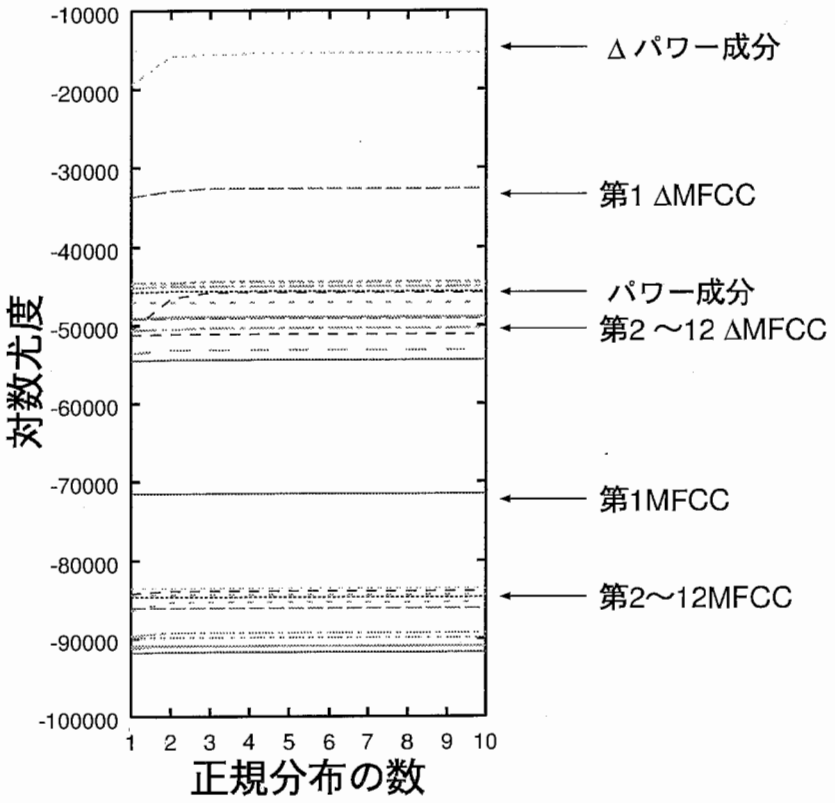

図 3.3: 音素 /a/ の第2状態で各成分に混合分布によって表現したときの尤度: 収束値が各成分で異なるので，単純に尤度から分割停止条件を得ることはできない

3.3　学習手順

音素記号のラベル付けされた音声データに対して，以下の手順で音響モデルを学習する．通常の Viterbi Training と，モデルのパラメータ数の増加をくり返して行く．パラメータ数の増加では，特徴量成分の混合数を増加させるのか，基本分布の混合数を増加させるのかという，2つの可能性があり得る．ここで最適な選択を与えるアルゴリズムについては，今のところ考えていない．

(1) 初期モデルを用いて，Viterbi Aligment により学習データの各フレームと音素の対応を求める．

(2) 各状態に割り当てられた学習データを用いて，式 (3.4)(3.5)(3.6)(3.7) において必要とされる十分統計量を得る．

(3) 各状態のパラメータを更新する．出力確率は式 (3.4)(3.5)(3.6)(3.7) を用い，

遷移確率には通常の最尤推定式 自己遷移確率 $= 1 - 1/N$ (N:滞在回数) を用いる.

(4) 成分の混合数を増やす. または, 基本分布を 2 分割する.

(5) (1) に戻る.

上記の手法はサンプルと各状態への割り当てを行ってから各状態のパラメータを学習しているが, Baum-Welch アルゴリズムを用いて割り当てを確率的に扱い学習することも可能である. ここでは, 式の導出は省略する.

第4章　切り出し音素の識別実験

4.1　評価実験1

4.1.1　実験条件

モデルの学習の条件は以下のとおりである.

- 音響分析
  サンプリング周波数 16kHz, 16bit 量子化, 20ms ハミング窓, フレーム周期 10ms
- 音響特徴ベクトル
  12MFCC + POW + 12 Δ MFCC + Δ POW からなる計26次元（CMN により正規化）
- 学習・評価データ
  ATR より提供される研究用日本語音声データベースの TRA 音素バランス文の話者165名27文のうち，1/4の話者（4で割り切れる番号) を学習データに，他の1/4(4で割って1余る番号) の話者のデータを評価データに使用した.

モデルは以下の手順 (図4.1参照) で作成した.

1. **初期モデル**
   単一の対角分散行列を出力とする HMM を最適化
2. **基本分布の作成**
   貪欲なアルゴリズムにより分割による分布数を増やす成分を選びながらモデルを学習
3. **基本分布の分割**
   成分の分割回数が5,10,15,20,25回であるモデルをそれぞれ基本分布とし，各基本分布を2分割し学習を行う．そして基本分布の2分割と学習を繰り返す.

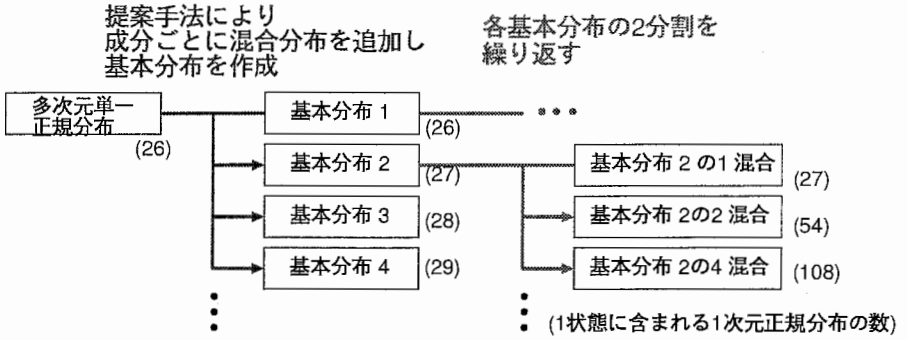

図 4.1: 評価実験 1 のモデル作成手順

こうしてできた各モデルに対して評価データに対する尤度と認識率の評価を行った．音響分析 (特徴量ベクトルを得るまでの過程) には HTK を使用し，続くモデルの学習，評価には，自作のプログラムを使用した．

4.1.2 評価結果

基本分布の作成過程で，貪欲なアルゴリズムよって選択された分割すべき特徴量の次元は，

1. POW, 2. Δ POW, 3. 1st Δ MFCC, 4. 1st MFCC 5. 2nd Δ MFCC, 6. 2nd MFCC, 7. 4th MFCC, 8. 3rd MFCC ...

となり，分割回数とそのときの各成分の分布数を表 4.1 に示す．パワー項，低次の次元から分割され，また全体で同じ数の分布数 (2) になる傾向が観察される．

正規分布の数とモデルの評価データに対する尤度を図 4.2 左に示す．これにより，基本分布の選択により少ない分布数でも学習データに対して高い尤度を実現できている．これは，基本分布が各成分で異なる分布数の混合分布を用いたためで，特徴量の次元ごとに異なる統計的性質を反映しているためと考えられる．

各モデルにおいて認識率は，図 4.2 の右に示すようにどの基本分布を用いても 2 分割を繰り返すことにより 80%程度の性能を実現できている．従来法に及ばない．5 回の分割によって生成した基本分布をもとに基本分割の 2 分割を行った場合が，若干であるが，少ない分布数で高い認識率を出している．

表 4.1: 貪欲なアルゴリズムによって分割する次元を選んだときの特徴量ベクトルの各成分の分布数 (MFCC, Δ MFCC の項は左から順に第 1 次から第 12 次の混合分布の数を表す)

分割回数	MFCC の各成分	Δ MFCC の各成分	POW (Δ)
0	1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1	1 (1)
5	2 1 1 1 1 1 1 1 1 1 1 1	2 2 1 1 1 1 1 1 1 1 1 1	2 (2)
10	2 2 2 2 1 2 1 1 1 1 1 1	2 2 2 1 1 1 1 1 1 1 1 1	2 (2)
15	2 2 2 2 2 2 2 1 1 1 1 1	2 2 2 2 2 1 1 1 1 1 1 2	2 (2)
20	2 2 2 2 2 2 2 2 1 1 1 2	2 2 2 2 2 2 2 1 1 1 1 2	3 (3)
25	2 2 2 2 2 2 2 2 1 2 2 2 2	2 2 2 2 2 2 2 2 2 1 1 1 2	3 (3)

表 4.2: 評価実験 1：音素弁別実験 (正規分布の分布数と音素の認識率)

分布数	認識率	分布数	認識率	分布数	認識率	分布数	認識率	分布数	認識率	分布数	認識率
26	69.8	31	71.1	36	71.3	41	71.5	46	71.5	51	71.5
52	76.0	62	74.6	72	75.1	82	75.0	92	75.1	102	75.3
-	-	124	77.5	144	77.3	164	77.3	184	77.5	204	77.5
-	-	248	79.5	288	79.3	328	79.1	368	79.3	408	79.4
-	-	496	80.1	596	80.1	656	79.9	736	80.0	816	80.0
-	-	992	79.8	1192	79.5	1312	78.9	1472	80.1	1632	78.0

4.2 評価実験 2

評価実験 1 ではベースラインの評価を途中で打ちきり十分でなかったので，以下のようにモデルを作成して音素弁別実験の再実験を行った．ここでは，分割成分の決定は貪欲なアルゴリズムによる決定に従わず，評価実験 1 で選択された成分を用いた．

4.2.1 実験条件

モデルの学習の条件は以下のとおりである．

- 音響分析
  サンプリング周波数 20kHz, 16bit 量子化， 20ms ハミング窓，フレーム周期 10ms
- 音響特徴ベクトル

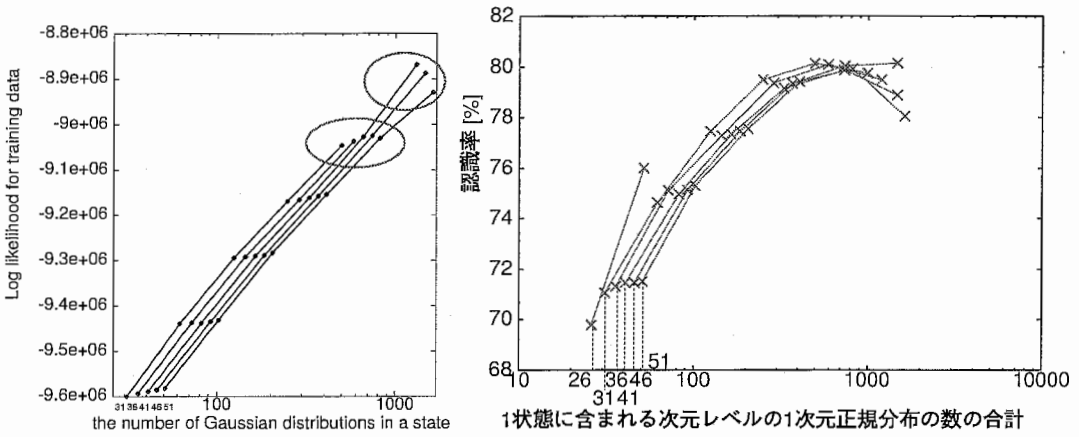

図 4.2: 評価実験 1: 各基本分布をもとに 2 分割と学習を繰り返したモデルの評価データに対する尤度（左）と認識率 (右)

12MFCC + POW + 12 Δ MFCC + Δ POW からなる計 26 次元（CMN により正規化）

- 学習・評価データ
  ATR より提供される研究用日本語音声データベースの A set 音素バランス文のうち，話者 2 名 (5972 発声) を学習データに，話者 1 名 (5972 発声) を評価データに使用した．

モデル作成手順は以下の手順 (図 4.3 参照) で作成した．

1. **初期モデル**
   単一の対角分散行列を出力とする HMM を最適化

2. **基本分布の作成**
   評価実験 1 で選べられた次元の混合数を増やし学習し学習する

3. **基本分布の分割**
   成分の分割回数が 5,10 回であるモデルをそれぞれ基本分布とし，各基本分布を 2 分割し学習を行う．そして基本分布の 2 分割と学習を繰り返す．

こうしてできた各モデルに対して評価データに対する尤度と認識率の評価を行った．

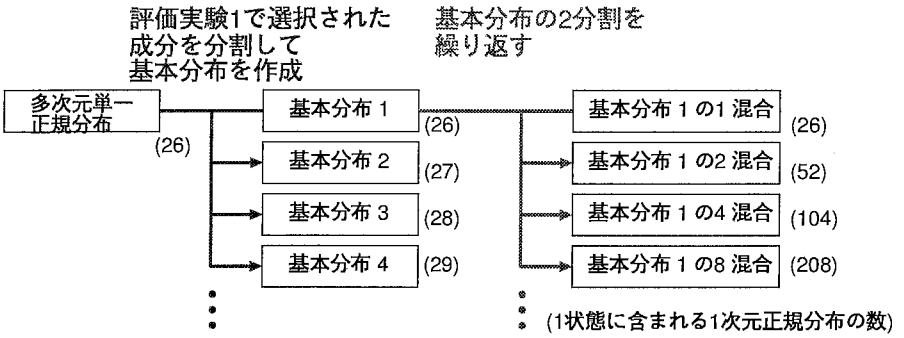

図 4.3: 評価実験 2 のモデル作成手順

表 4.3: 評価実験 2 の認識結果 (正規分布の分布数と音素の認識率)

分布数	認識率	分布数	認識率	分布数	認識率
26	70.5	31	71.4	36	71.4
52	73.4	62	74.3	72	74.2
104	76.9	124	77.8	144	77.8
208	78.8	248	79.1	288	79.2
416	80.7	496	81.0	576	81.1
832	83.0	992	83.3	1152	83.4
1664	84.6	1984	84.4	2304	84.6
3328	85.0	3988	85.1	4608	84.8

4.2.2 評価結果

表 4.3 に示すように，直積混合分布を含んだ確率モデルを用いたモデル (表における第 2，3 列) がが従来の多次元混合正規分布を用いたモデル (表 4.3 における最左列) よりも上回る認識率が得られた．直積混合分布を加えることで，過学習に陥ることなくモデルを詳細化できたことが分る．

4.3 考察

一般的に，モデルのパラメータ数を増やせば，学習データにあるサンプルの分布により近い分布を学習で求めることができるが，パラメータ数を過剰に増やすと学習データの分布に近づき過ぎるため，未知のデータの分布からは却って遠ざかる．これは，評価実験 1 においても見られる．基本分布としては，5 分割 (分布数 31)，10 分割 (分布数 36)，15 分割 (分布数 41)，および 25 分割 (分布数 51) では，

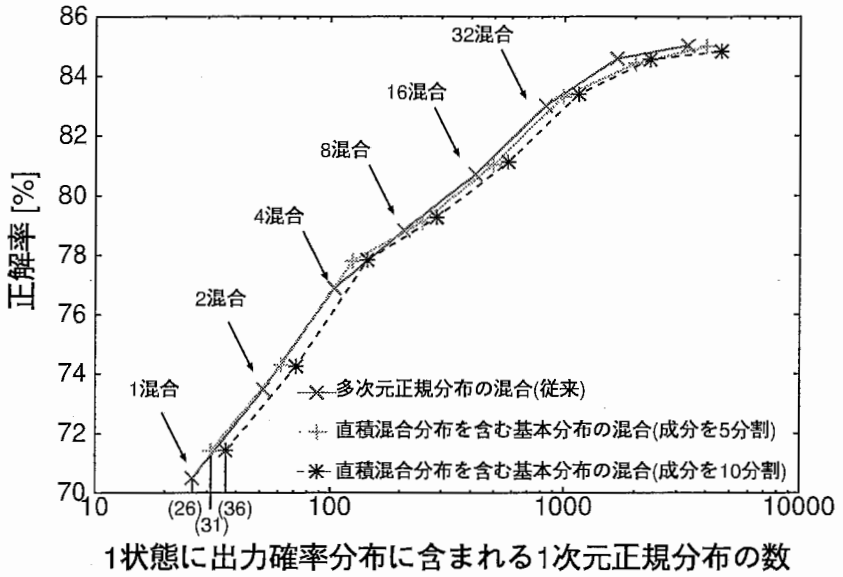

図 4.4: 評価実験 1: 音素弁別の認識率 (右)

各基本分布の 2 分割をくり返した結果認識率の減少という過学習と考えられる傾向が見られた．これに対して，20 分割 (分布数 46) の基本分布から基本分布の 2 分割をくり返した場合は，実験の範囲内では認識率の減少は見られず，混合数 1472 で 80.1%の認識率が得られている．これは，直積混合分布を適切に含んだ基本分布を用いたことにより，過学習に陥らずに有効にパラメータを活用して確率分布を作ることができたためだと考えられる．このことから，直積混合分布を用いることにより，従来の確率分布より，特徴量ベクトルの分布の統計的な性質を反映した詳細なモデルを，過学習に陥らずに作成できたと考えられる．

また，本手法は特徴量ベクトルの時系列を HMM でモデル化することの限界の一例を示していると考えられる．HMM は，特徴量ベクトルの時系列を隠れた状態の時系列との関係をモデル化し，状態遷移により時間発展を表現をし，状態出力により観測される特徴量ベクトルと状態とを関係付けている．複雑な振舞をする特徴量ベクトルの時系列を数個の状態で表現することは，特徴量空間を数個の点 (特徴量ベクトル) で代表させることであり，その点の周囲を覆う出力確率をどのように精密化しても限界があると考えられる．本手法は，状態出力についてモデルの精密化を試み，この評価実験はその限界を示すものと考えることもできる．図 4.4 に示すように，直積混合分布を用いることにより，通常使用される 8, 16 混合の多次元正規分布の混合分布に対しては，僅かながら同程度のパラメータ数で高い音素分別精度を示している．分布数を増やし出力確率分布の形状を精密化し

ても，認識精度向上は僅かであり，図4.2，4.4に見られるように精度には限界があると思われる．

4.4 今後の課題

本研究では基本分布の学習方法として学習データに対する尤度を上昇させる手法を提案したので，今後は最適な混合選択することで基本分布を選択する手法を提案したい．モデルの選択には，情報量基準 (BIC, Bayesian Information Criterion) を用いるアプローチが考えられるが，他にも混合正規分布の形による判定なども可能であると考える．例えば，我々は過剰な分布数を持つ混合正規分布を用いた場合，各分布の平均値が密集する形になることを確認している．この性質を用いれば，例えば，新たに分布を追加した結果，ある2つの分布の平均値が非常に近接した場合は，すでに十分に分布形状を記述していると判断できるであろう．

提案モデルの有効性の検証のため，以下の実験を行い評価するのが好ましい．

- **雑音重畳音声での評価**
  直積混合分布の存在により従来の分布よりも広い空間が覆われているので，観測される特徴量が雑音により多少変動しても認識性能は保たれると考えられる．「見えない」学習サンプルに対する推定を行っていると解釈した場合の利点は，特徴量が変動した場合に活かされることが期待できる．

- **トライフォンモデルでの評価**
  評価実験では，HMMの単位がモノフォンであったので，学習データサンプルのばらつきが大きかったとかが得られる．トライフォン単位のHMMでは，統計的性質がまとまったサンプルでモデルの学習ができるので，従来法との比較が適切にできると思われる．

- **分布形状決定過程の考察**
  直積混合分布を含む特徴量分布は，成分と特徴量ベクトルの2つのレベルでの混合分布構造を持っている．本報告では，始めに成分レベルでの混合分布形状を基本分布として決定し，その後にベクトルレベルでの混合分布を求めた．この方法では，学習データに対する尤度上昇を保証するが，全体での最適性は保証されていない．ベクトルレベルで，異なる構造を持つ基本分布を使用するなど別の学習手順を踏むことで，より適切な分布が得られる可能性がある．

第5章　結論

本報告では，直積混合分布を用いたHMMに対する，(1) モデルの設計，(2) 学習アルゴリズムの構築，(3) モデルの性能評価，について報告した．モデルの学習では，EMアルゴリズムによる導出したパラメータ更新式と貪欲なアルゴリズムを用いる方法を使用した．切り出し音素の識別による評価実験を行い，オープンデータに対するモデルの尤度から，直積混合分布を用いて広い音響空間を効率的に覆えること確認し，80%程度の識別正解率を得た．

謝辞

本報告は，東京大学嵯峨山研究室の学生，武田が実習生としてSTL第1研究室に受け入れて頂いた2003年9月～10月に，松田繁樹氏の指導のもとに行った研究です．研究の機会を与えて下さった中村室長，研究の指導をして下さった松田研究員，並びに快適な研究環境を提供して頂いた研究室の皆様に深く感謝します．

関連図書

[1] Harald Singer, Mari Ostehdorf, "Maximum Likelihood Successive State Splitting," Vol. 43, No. 2, pp. 245–255, 1996.

[2] 篠田　浩一，渡辺　隆夫, "情報量基準を用いた状態クラスタリングによる音響モデルの作成," 情報処理学会，音声言語情報処理研究会研究報告 014-11, 2001.

[3] 松田繁樹, "音声認識における特徴量の非同期性と音素間環境依存性のモデル化に関する研究," 博士論文, 2003.

[4] 高橋敏, 嵯峨山茂樹, "4 階層共有構造の音響モデルによる音声認識," 電子情報通信学会論文誌 D-II, Vol. J82-D-2, No. 3, pp.315–323, Mar 1999.