

DAEM アルゴリズムの話者・音声認識における有効性の検討

板谷 洋平[†] 全 炳河[†] 南角 吉彦[†] 宮島千代美^{††} 徳田 恵一[†]
北村 正[†]

[†] 名古屋工業大学 〒466-8555 愛知県名古屋市昭和区御器所町

^{††} 名古屋大学 〒466-8603 愛知県名古屋市千種区不老町

E-mail: †{yoheir32,zen,nankaku,tokuda,kitamura}@ics.nitech.ac.jp, ††miyajima@is.nagoya-u.ac.jp

あらまし GMM (Gaussian Mixture Model) や HMM (Hidden Markov Model) のパラメータ推定には、広く EM (Expectation Maximization) アルゴリズムが用いられる。しかし、EM アルゴリズムは、推定結果が初期値設定に依存してしまうという、局所最適性の問題を有する。この問題に対処するために DAEM (Deterministic Annealing Expectation Maximization) アルゴリズムが提案された。本稿では、この DAEM アルゴリズムを GMM、及び音素境界情報が得られない場合の HMM のパラメータ推定 (フラットスタート) にそれぞれ適用し、GMM を用いた話者認識、及び HMM を用いた連続音声認識における有効性に関する検討を行う。

キーワード EM アルゴリズム, DAEM アルゴリズム, 話者認識, 連続音声認識

Deterministic Annealing EM algorithm for speaker and speech recognition

Y. ITAYA[†], H. ZEN[†], Y. NANKAKU[†], C. MIYAJIMA^{††}, K. TOKUDA[†], and T. KITAMURA[†]

[†] Department of Computer Science and Engineering, Nagoya Institute of Technology Gokiso-cho, Showa-ku, Nagoya, 466-8555, Japan

^{††} Department of Media Science, Nagoya University Furo-cho, Chikusa-ku, Nagoya, 466-8603, Japan

E-mail: †{yoheir32,zen,nankaku,tokuda,kitamura}@ics.nitech.ac.jp, ††miyajima@is.nagoya-u.ac.jp

Abstract This paper investigates the effectiveness of a DAEM (Deterministic Annealing Expectation Maximization) algorithm for speaker and speech recognition. The EM (Expectation Maximization) algorithm is widely used for parameter estimation of statistical models. However, the EM algorithm has the problem of initialization dependence. To relax this problem, the DAEM algorithm has been proposed. In this paper, we apply the DAEM algorithm to estimate acoustic models for speaker recognition and continuous speech recognition.

Key words EM algorithm, DAEM algorithm, speaker recognition, continuous speech recognition

1. ま え が き

GMM (Gaussian Mixture Model) 及び、HMM (Hidden Markov Model) に基づく音声認識では、パラメータ推定に、EM (Expectation Maximization) アルゴリズム [4] を用いた尤度最大化基準に基づく枠組みが用いられている。不完全データから最尤推定値を求める EM アルゴリズムは、多くの場合、実現容易なアルゴリズムが導出されることから、様々な統計モデルの最尤推定に広く用いられている。しかし、基本的には勾配型のアルゴリズムであるため、局所最適性の問題を持ち、推定パラメータの初期値の設定に推定結果が左右され、設定した初期値によっては、十分な認識性能を得ることができない場合

がある。

この問題に対し、最大エントロピー原理と統計力学のアナロジーを用いて確定的アニーリング EM (Deterministic Annealing Expectation Maximization: DAEM) アルゴリズム [5] が提案され、その有効性が報告されている。DAEM アルゴリズムは、尤度関数の最大化問題を、自由エネルギー関数の最小化問題として再定式化したものであり、温度パラメータ β を用いてアニーリング過程を制御することで、初期値依存の問題に対応するものである。

本稿では、まず GMM を用いた話者認識における DAEM アルゴリズムの有効性を示す。次に連続音声認識のための HMM の学習における DAEM アルゴリズムの有効性に関しての検討

を行う。温度パラメータ β が 0 に近いとき、HMM の状態系列の事後確率は一様となり、 β の値を増加させながら連結学習を繰り返していくことにより、音素境界情報が与えられない条件下での学習（フラットスタート）でも高い推定精度を得ることが期待される。本研究では、特にこのような音素の境界情報が得られない場合の HMM 学習における、DAEM アルゴリズムの有効性を示す。

以下、2. で DAEM アルゴリズムについて、3. で GMM を用いた話者認識実験について、4. で HMM を用いた連続音声認識実験について、最後に 5. で結論を述べる。

2. DAEM アルゴリズム

2.1 DAEM アルゴリズムの導出

DAEM アルゴリズムは、尤度関数の最大化問題を、“温度” に依存する、自由エネルギー関数の最小化問題として再定式化し、この“温度”を用いてアニーリング過程を制御することにより EM アルゴリズムの局所最適性の問題に対処するものである。

EM アルゴリズムは、E ステップ、M ステップを逐次実行していくことにより、対数尤度関数を最大化する未知パラメータ Λ の最尤推定値を求めて行くアルゴリズムである。対数尤度関数 L は次式で表すことができる。

$$L(\Lambda; \mathbf{O}) = \log \int p(\mathbf{O}, \mathbf{q}; \Lambda) d\mathbf{q} \quad (1)$$

ここで、 $\mathbf{O} = (\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T)$ は観測ベクトル、 $\mathbf{q} = (q_1, q_2, \dots, q_T)$ は隠れ変数で、HMM においては状態系列を表し、 Λ はモデルパラメータである。EM アルゴリズムは、式 (1) から導出される Q 関数を最大化するモデルパラメータ Λ を計算する。EM アルゴリズムにおける Q 関数は、次式で表すことができる。

$$Q(\Lambda, \Lambda') = \int p(\mathbf{q} | \mathbf{O}, \Lambda) \log p(\mathbf{O}, \mathbf{q} | \Lambda') d\mathbf{q} \quad (2)$$

また、 Q 関数における事後確率 $p(\mathbf{q} | \mathbf{O}, \Lambda)$ は、ベイズ則から、次式のように与えられる。

$$p(\mathbf{q} | \mathbf{O}, \Lambda) = \frac{p(\mathbf{O}, \mathbf{q} | \Lambda)}{\int p(\mathbf{O}, \mathbf{q} | \Lambda) d\mathbf{q}} \quad (3)$$

ここで、推定過程の初期段階におけるパラメータ推定値は信頼が低いために、事後確率も信頼が低いことになる。そこで DAEM アルゴリズムでは、新たな事後確率密度 f を導入し、この f を用いて式 (2) の Q 関数の代わりに、次に示す U 関数を考える。

$$U = E_f\{L(\Lambda; \mathbf{O}) | \mathbf{O}; \Lambda\} \\ = \int f(\mathbf{q} | \mathbf{O}, \Lambda) \log p(\mathbf{O}, \mathbf{q} | \Lambda) d\mathbf{q} \quad (4)$$

ここで、 $f \equiv p$ のとき、 $U \equiv Q$ となる。

次に、事後確率 f を最大エントロピー原理 [8] を用いて求める。具体的には、次式に示すような最大化問題を解くことになる。

$$S + \beta U \rightarrow \max_f$$

$$\text{制約条件: } \int f d\mathbf{q} = 1 \quad (5)$$

ここで、 S はエントロピーで次式で定義される。

$$S = - \int f(\mathbf{q} | \mathbf{O}; \Lambda) \log f(\mathbf{q} | \mathbf{O}; \Lambda) d\mathbf{q} \quad (6)$$

また、 β は U と S の最大化のバランスを定めるための定数であり、式 (5) の最大化問題は、次式で示す f を変数とする汎関数 $J[f]$ の極値問題と等しくなる。

$$J[f] = S + \beta U + \lambda \int f d\mathbf{q} \quad (7)$$

上式において、変数 λ はラグランジュ乗数である。式 (7) において f に関して変分をとることにより、次式が得られる。

$$\delta J = \int \{[-1 - \log f(\mathbf{q} | \mathbf{O}; \Lambda) \\ + \beta \log p(\mathbf{O}, \mathbf{q}; \Lambda) + \lambda]\} \delta f d\mathbf{q} \quad (8)$$

ここで、任意の δf に対して $\delta J = 0$ が成立するには、式 (8) における [] 内が恒等的に $= 0$ となる必要があることから、次式を得る。

$$f(\mathbf{q} | \mathbf{O}, \Lambda) = \exp\{\beta \log p(\mathbf{O}, \mathbf{q}; \Lambda) + \lambda - 1\} \quad (9)$$

式 (9) を式 (5) に代入して λ を消去すると、次式の Gibbs 分布を得る。

$$f(\mathbf{q} | \mathbf{O}; \Lambda) = \frac{1}{Z} \exp\{-\beta(-\log p(\mathbf{O}, \mathbf{q}; \Lambda))\} \quad (10)$$

式 (10) における Z は分配関数と呼ばれ、次式で与えられる。

$$Z = \int \exp\{-\beta(-\log p(\mathbf{O}, \mathbf{q}; \Lambda))\} d\mathbf{q} \quad (11)$$

ここで $1/\beta$ は“温度”に相当し、 β は逆温度を表す温度パラメータである。統計力学のアナロジーと式 (11) の分配関数を用いて、次式で表す、温度に依存する自由エネルギー関数が得られる。

$$F_\beta(\Lambda; \mathbf{O}) = -\frac{1}{\beta} \log Z \\ = -\frac{1}{\beta} \log \int p(\mathbf{O}, \mathbf{q}; \Lambda)^\beta d\mathbf{q} \quad (12)$$

DAEM アルゴリズムでは、式 (12) の自由エネルギー関数から導出される、次式で表される Q_β 関数を最大化することにより、尤度 $p(\mathbf{O} | \Lambda)$ を極大に導く。

$$Q_\beta(\Lambda, \Lambda') = \int f(\mathbf{q} | \mathbf{O}, \Lambda) \log p(\mathbf{O}, \mathbf{q} | \Lambda') d\mathbf{q} \quad (13)$$

また、式 (13) における事後確率^(注1) $f(\mathbf{q} | \mathbf{O}, \Lambda)$ は、式 (10) を単純化することにより、次式のように与えられる。

(注1)：正確には関数 f は EM アルゴリズムにおける事後確率とは異なるものであるが、本稿では以降、事後確率と呼ぶことにする。

$$f(q|\mathbf{O}, \Lambda) = \frac{p(\mathbf{O}, q|\Lambda)^\beta}{\int p(\mathbf{O}, q|\Lambda)^\beta dq} \quad (14)$$

ここで、式(3)と式(14)を比較すると、DAEMアルゴリズムにおける事後確率は、EMアルゴリズムの事後確率の分母と分子をそれぞれ β 乗するだけで得ることができることがわかる。つまり、式(14)において、 $\beta = 1$ のとき、 $f(q|\mathbf{O}, \Lambda) = p(q|\mathbf{O}, \Lambda)$ となり、式(13)は式(2)のEMアルゴリズムにおける Q 関数と等しくなる。

先に述べたように、推定過程の初期段階では、パラメータ推定値の信頼性が低いために、その推定値から計算される事後確率も信頼が低く、初期値の影響を受けやすい。そこで、初期値の影響を弱めるために、事後確率 $f(q|\mathbf{O}, \Lambda)$ が一様分布となるように β を小さく、つまり、温度($1/\beta$)を高くした状態でEMステップを繰り返しパラメータの推定を行う。 β が十分に小さい時、式(12)の自由エネルギー関数 F_β における極小値は唯一となるため、その最小値を与える最適なパラメータ Λ が初期値にかかわらずに求めることができる。次に β の値をやや大きくし、温度を少しだけ下げて同様のEMステップを繰り返す。この操作を $\beta = 1$ になるまで繰り返す。これにより、 F_β は尤度関数の大域的構造を徐々に詳細にしつつ、各温度における最適解を逐次求めて行くために、局所最適性の問題を軽減することができる。温度パラメータ β は1に近くなれば、従来のEMアルゴリズムに近づき、 $\beta = 1$ でDAEMアルゴリズムはEMアルゴリズムの処理と完全に一致する。

EMアルゴリズムは、DAEMアルゴリズムにおいて、温度パラメータ β をいきなり0から1にあげてEMステップを行うことに対応するため、乱数等で初期値を与えた場合、初期値の影響を大きく受けてしまうのに対し、DAEMアルゴリズムでは、 β を徐々に大きくしながらEMステップを繰り返して尤度関数の形状を変化させながら推定を行うため、初期値の影響に依存しにくく、つまり初期化を考慮せずにより推定結果を得ることができる、という利点がある。

2.2 GMM への適用

混合数 M のGMM λ を例に考えると、EMアルゴリズムにおける事後確率は、時刻 t の観測ベクトル \mathbf{o}_t と、ガウス分布 $\mathcal{N}(\mathbf{o}_t | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$ で表される混合要素 m により、

$$p(m | \mathbf{o}_t, \lambda) = \frac{w_m \mathcal{N}(\mathbf{o}_t | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)}{\sum_{m=1}^M w_m \mathcal{N}(\mathbf{o}_t | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)} \quad (15)$$

と表される。これをDAEMアルゴリズムへ移行させるには、事後確率を

$$f(m | \mathbf{o}_t, \lambda) = \frac{\{w_m \mathcal{N}(\mathbf{o}_t | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)\}^\beta}{\sum_{m=1}^M \{w_m \mathcal{N}(\mathbf{o}_t | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)\}^\beta} \quad (16)$$

のように変更し、さらに β に関するアニーリングのループを付加すればよい。

2.3 HMM への適用

DAEMアルゴリズムをHMMに適用した場合、式(14)にお

ける $p(\mathbf{O}, q|\Lambda)^\beta$ は次式のように書き換えることができる。

$$\begin{aligned} p(\mathbf{O}, q|\Lambda)^\beta &= p(\mathbf{O}|q, \Lambda)^\beta p(q|\Lambda)^\beta \\ &= \prod_t p(\mathbf{o}_t|q_t, \Lambda)^\beta \prod_t p(q_t|q_{t-1}, \Lambda)^\beta \end{aligned} \quad (17)$$

ここで、 $p(\mathbf{o}_t|q_t, \Lambda)$ と $p(q_t|q_{t-1}, \Lambda)$ は、それぞれHMMにおける出力確率と遷移確率を表す。よって、 $f(q|\mathbf{O}, \Lambda)$ の計算は $p(q|\mathbf{O}, \Lambda)$ と同じ形をしており、 $p(\mathbf{o}_t|q_t, \Lambda)^\beta$ と $p(q_t|q_{t-1}, \Lambda)^\beta$ をそれぞれ出力確率と遷移確率に置き換えて、Forward-Backwardアルゴリズムを実行することにより、計算可能である。

HMMのパラメータ推定にDAEMアルゴリズムを用いた場合、 β が0に近い小さな値のとき、全ての状態系列の事後確率は一様となる。つまり、音素境界情報が得られない状況と等しくなるが、先にも述べたように、DAEMアルゴリズムでは初期化は考慮しないでよいために、このような場合でもパラメータ推定が可能である。 β の値を徐々に大きくして繰り返し状態系列を学習して行くことで、一様だった状態系列が徐々に明確になっていき、 $\beta = 1$ となったとき、初期値に依存しにくい状態系列を得ることができると考えられる。

3. GMM を用いた話者認識実験

3.1 実験条件

ATR日本語音声データベースのcセットに含まれる男女各40名計80名の単語音声を使用し、テキスト独立型の話者識別実験を行った。各話者について、音韻バランス単語216単語を学習に、重要単語520単語(計41600単語)を評価に用いた。サンプリング周波数は10kHzとし、フレーム周期10ms、窓幅25.6msのBlackman窓で切り出したフレームに対して12次のメルケプストラム分析[9]を行い、得られたメルケプストラム係数の0次を除いたものを特徴量として使用した。

3.2 モデルの学習条件

各話者は混合数4, 8, 16, 32, 64の対角共分散を持ったGMMでモデル化した。DAEMアルゴリズムで推定する前のGMMパラメータの初期値は、混合重みについては全混合等しい値とし、平均ベクトルの各成分には平均0分散1の正規分布に従う乱数を与え、対角共分散の各要素には同じく平均0分散1の正規分布に従う乱数の絶対値を与えた。

DAEMにおける温度パラメータ β を0に近い値から1に単調増加させる方法はいくつか考えられるが、ここでは、第 t 回目の更新における温度パラメータの値 $\beta(t)$ を、次の非線形関数に基づいて設定した。

$$\beta(t) = \sqrt{\frac{t}{T}} \quad (18)$$

ここで、 T は温度パラメータの総更新回数であり、3.3.1の実験以外では、 $T = 20$ とした。つまり本実験においては、図1のような非線形な無理関数に従い、温度パラメータを更新した。また、温度パラメータを更新するごとに、温度パラメータの値を固定した状態でEMステップを繰り返すが、このステップ数については、1, 5, 10, 20回の4通りについて行った。この場

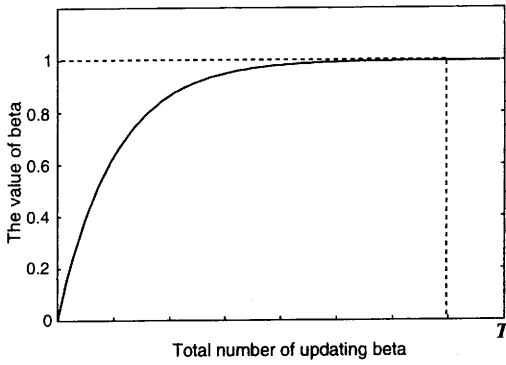


図1 温度パラメータの更新関数

Fig. 1 The updating function of the temperature parameter

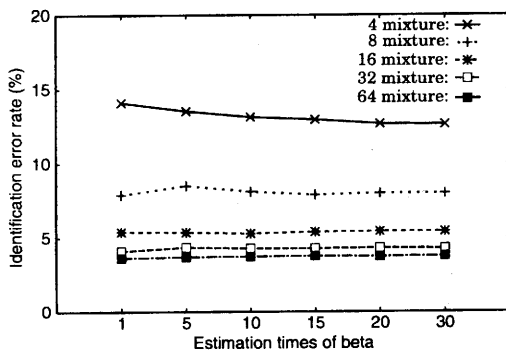


図2 温度パラメータの更新回数の違いに関する検討

Fig. 2 Comparison of the difference of the temperature parameter

合, EM ステップはそれぞれ計 20, 100, 200, 400 回行うことになる。ここで, EM ステップの回数は温度ごとに变えることも可能であるが, 本実験では, どの温度についても同じ回数だけ繰り返すこととした。

3.3 実験結果

3.3.1 温度パラメータ β の繰り返し回数の検討

本実験では, DAEM アルゴリズムの温度パラメータ β の更新回数を変更して, 認識率の比較を行った。 β の更新回数は 1, 5, 10, 15, 20, 30 回とした。 β の 1 回の更新の間の EM ステップの繰り返し回数は 20 回と固定した。 識別結果を図 2 に示す。 図のように β の更新回数に比例して認識率が向上するとは限らなかった。 しかし, 混合数 4 などの小さな混合数では β の更新回数を増やしていくと, 認識誤り率が低くなる傾向がみられた。

3.3.2 各温度における繰り返し回数の検討

図 3 に, DAEM アルゴリズムにより学習した GMM による識別結果を示す。 結果より, 各温度における EM ステップの回数は, 少な過ぎても多過ぎても良くないことがわかる。 今回の場合, 1 回や 20 回に比べて, 5 回, 10 回の方が結果が良く, 特に大きな混合数においては, 20 回の結果が 1 回のみ結果よりも悪くなっている。 このことから, DAEM アルゴリズムによる効果を得るためには, 各温度における EM ステップの回数

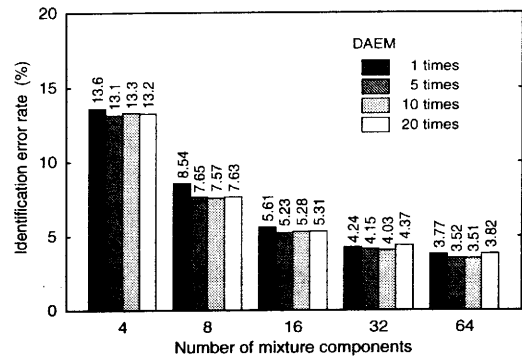


図3 各温度における繰り返し回数の違いに関する比較

Fig. 3 Comparison about the difference of the total number of EM-step in each temperature

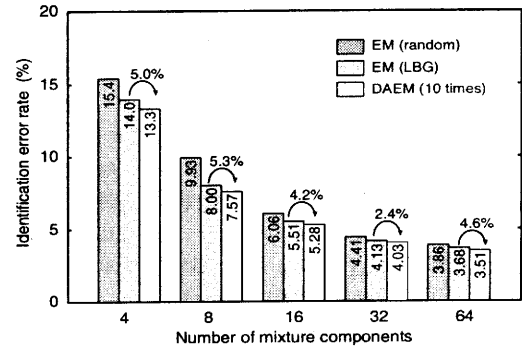


図4 乱数や LBG で初期化した場合の EM との比較

Fig. 4 Comparison with EM algorithm with random initialization and LBG initialization

は, 適度な値に制限する方が良いと言える。

3.3.3 EM アルゴリズムとの比較

続いて, DAEM アルゴリズムと EM アルゴリズムの比較実験を行った。 結果を図 4 に示す。 図中の DAEM の結果は, 各温度における繰り返し回数が 10 回の場合の結果である。 今回は, DAEM アルゴリズムの比較対象として, GMM のパラメータを乱数で初期化した場合の EM アルゴリズム, および LBG アルゴリズム [10] により作成したコードブックに基づいて初期値を与えた EM アルゴリズムについて実験を行った。 但し, 乱数による初期化では, DAEM アルゴリズムの初期化と同様に, 混合重みは全混合同じとし, 平均と分散の初期値には, 平均 0 分散 1 の正規分布に従う乱数とその絶対値を用いた。 また, LBG のコードブックを用いた場合は, 混合重みの初期値をコードブックの各セントロイドに属する学習データ量の割合とし, 平均ベクトルはコードベクトル, 対角共分散は各セントロイドに属する学習データの分散を与えて初期化した。 乱数, LBG どちらの場合も, 初期化の後, EM ステップを 100 回繰り返した。 なお, EM ステップの回数が DAEM アルゴリズムの場合と同じ 200 回の場合についても同様に実験を行ったが, 識別結果に大きな変化は見られなかった。

図4より、乱数により初期化したEMアルゴリズムでは、他の2つに比べて全体的に誤り率が高くなっている。一方、LBGで初期化した場合については、乱数の場合に比べ、大きく誤り率が減少している。これについては、VQで得られた結果に基づいて、初期値のガウス分布を学習データの分布に合わせて配置していることから、初期値の影響緩和されたためと考えられる。さらに、DAEMアルゴリズムに基づいて学習したモデルでは、LBGで初期化した場合よりも良い結果が得られた。また、同じ乱数で初期化したEMアルゴリズムの結果と比較して、全ての混合数で改善がみられ、混合数8で最大5.3%の改善率が得られた。これらの結果から、DAEMアルゴリズムは話者認識のためのモデル学習において、初期値依存性の問題の解決策として有効であることがわかる。

4. HMMを用いた連続音声認識実験

4.1 HMMの学習方法

本研究では、DAEMアルゴリズムの初期値依存の問題に対する有効性を示すため、連続音声認識実験において、以下の3種類の学習法によるHMMを比較した。

Method-1: 音素境界ラベルを用いて、セグメンタルk-meansアルゴリズムで初期化、及びEMアルゴリズムを用いて再推定した後、連結学習を行う。(音素境界付き学習)

Method-2: 音素境界の情報を用いず、HMMの全ての状態に等しい初期値(平均、分散)を与えるフラットスタートにより初期化を行い、(フラットスタート)EMアルゴリズムを用いて連結学習を行う。

Method-3: Method-2の連結学習にDAEMアルゴリズムを適用する。(DAEMアルゴリズムによるフラットスタート)

4.2 実験条件

特定話者、不特定話者モデルを用いてそれぞれ認識実験を行った。特定話者モデルでは、ATR日本語音声データベースbセットの503文章のうち、学習に450文章、評価データに学習データに含まれない53文章を用いた。不特定話者モデルにおける学習データは、ASJ-PBの150文章64名分及び、ASJ-JNASの150文章200名分を使用した。評価データはIPA98テストセットを用いた。音声の分析条件として、サンプリング周波数は16kHzとし、フレーム周期10ms、窓幅25msのBlackman窓で切り出したフレームに対して12次メルケプストラム分析を行い、得られたメルケプストラム係数の0次を除いたもの、及びその動的特徴を特徴量として使用した。また、各HMMにおいて1, 2, 4, 8混合、それぞれmonophone-HMMとtriphone-HMMでモデル化した。triphoneは、決定木に基づく状態クラスタリング[11]により、パラメータの共有を行った。

Method-1, Method-2のHMMでは連結学習の繰り返し回数は10回とした。Method-3において、DAEMアルゴリズムにおける温度パラメータ β を0に近い値から1に単調増加させる方法はいくつか考えられるが、ここでは、 β の更新方法は先に示したGMMの場合と同様、式(18)の更新式に従った。ただし、 $T=10$ とし、各温度におけるEMステップの繰返し回数を5回とした。従って、Method-3の連結学習における総

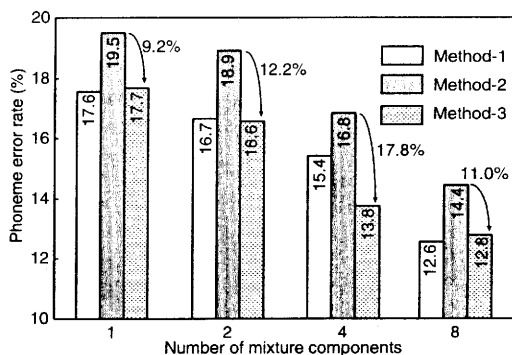


図5 特定話者モデルにおける比較 (monophone)

Fig. 5 Results of speaker-dependent model (monophone)

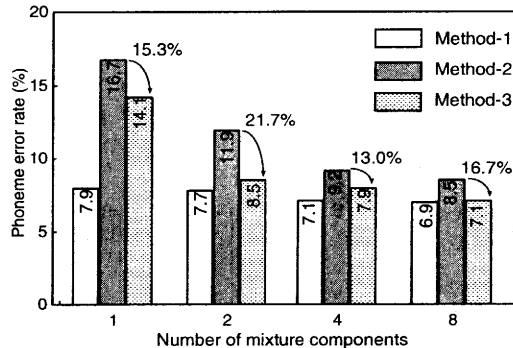


図6 特定話者モデルにおける比較 (triphone)

Fig. 6 Results of speaker-dependent model (triphone)

EMステップ数は100回となる。

4.3 実験結果

4.3.1 特定話者モデルにおける比較

図5に特定話者 monophone モデルを用いた場合の実験結果を示す。Method-2のフラットスタートで初期化を行ったモデルは、EMアルゴリズムの局所最適性の問題から、音素境界情報を用いて初期化したMethod-1と比べ、認識誤りが大きくなっていることがわかる。一方、DAEMアルゴリズムを用いたMethod-3では、同じフラットスタートで初期化したMethod-2と比較して、全ての混合数において低い誤り率となり、混合数4で最大17.8%の改善がみられ、初期値の影響を緩和できていると考えられる。また、音素境界情報を用いて初期化したMethod-1とはほぼ同等の認識性能が得られた。

図6に triphone モデルによる実験結果を示す。パラメータ数を増やした triphone においても、Method-2のモデルに対し、DAEMアルゴリズムを適用したMethod-3は、低い認識誤り率となり、混合数2で最大21.7%の改善がみられた。また、混合数1以外では、monophoneの場合と同様、音素境界を用いて初期化したMethod-1とはほぼ同等の認識性能が得られた。

4.3.2 不特定話者モデルにおける比較

次に、不特定話者 monophone モデルを用いた実験結果を図

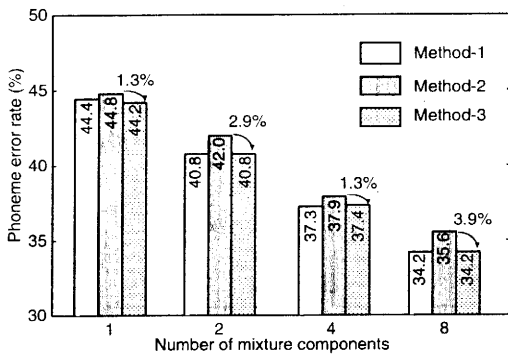


図7 不特定話者モデルにおける比較 (monophone)

Fig. 7 Results of speaker-independent model (monophone)

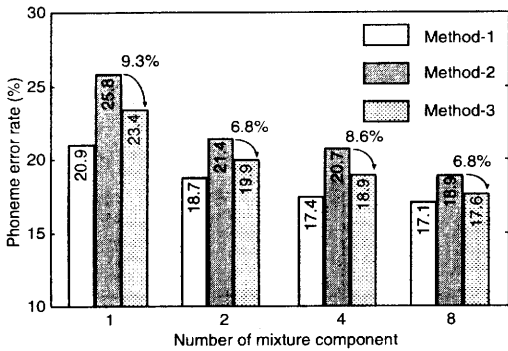


図8 不特定話者モデルにおける比較 (triphone)

Fig. 8 Results of speaker-independent model (triphone)

7に示す。特定話者の monophone での結果と比べ、認識率の差は小さいものの、この場合も Method-3 のモデルが Method-2 のモデルよりもよい結果となり、混合数 8 で、最大 3.9% の改善率を得ることができた。

図 8 に、triphone モデルによる実験結果を示す。こちらも特定話者 triphone を用いた場合と同様の傾向が見られ、最大 9.3% の誤り削減率が得られた。

これらの結果より、DAEM アルゴリズムにおいて、温度を徐々に下げながら学習を行うことにより、初期値の影響を緩和することができ、音素境界情報を用いない場合にも、音素境界情報を用いてあらかじめ整備された初期値を与えられたモデルとはほぼ同等の認識性能が得られることが確認された。

5. むすび

本稿では、GMM を用いた話者認識及び、HMM を用いた連続音声認識において、DAEM アルゴリズムの有効性に関する検討を行った。DAEM アルゴリズムに基づいて温度パラメータによるアンニリング制御を行いながら学習を繰り返して行くことにより、EM アルゴリズムの局所最適性の問題に対処することができ、話者認識及び、連続音声認識における有効性を確認することができた。

今後の課題として、学習データの量を変更しての実験、温度

パラメータ β の更新方法の検討、さらに、コンテキストクラスタリングにおける決定木の大きさや分割基準等に関する検討、等が挙げられる。

6. 謝 辞

本研究の一部は、中部電力基礎技術研究所研究助成、および科学研究費補助金若手研究 (B) No.14780274 による。

文 献

- [1] D. A. Reynolds and R. C. Rose : "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Trans. on Speech and Audio Processing, vol.3, no.1, pp.72-83, Jan. 1995.
- [2] X. D. Huang, Y. Ariki, M. A. Jack : "Hidden Markov models for speech recognition," EDINBURGH UNIVERSITY, pp.119-125, 1990.
- [3] 今井聖 : "音声認識 (情報・電子入門シリーズ 16)", 共立出版, pp.90-96, 1995.
- [4] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Roy. Statist. Soc., vol.39, pp.1-38, 1977.
- [5] 上田修功, 中野良平, "確定的アンニリング EM アルゴリズム", 電子情報通信学会論文誌, J80-D-II, no.1, pp.267-276, Jan. 1997.
- [6] N. Ueda and R. Nakano, "Mixture density estimation via EM algorithm with deterministic annealing," Proc. IEEE Neural Networks for Signal Processing 4, pp.60-77, 1994.
- [7] N. Ueda and R. Nakano, "Deterministic annealing variant of the EM algorithm," G. Tesaro et al. eds, Advances in Neural Information Processing Systems 7(NIPS), MIT press, Cambridge MA, pp.69-77, 1995.
- [8] K. Rose, E. Gurewitz, and G.C. Fox, "Vector quantization by deterministic annealing," IEEE Trans. Information Theory, vol.38, no.4, pp.1249-1257, 1992.
- [9] 徳田 恵一, 小林 隆夫, 深田 俊明, 斎藤 博徳, 今井 聖 : "メルケプストラムをパラメータとする音声のスペクトル推定", 信学論 (A), J74-A, no.8, pp.1240-1248, Aug. 1991.
- [10] Y. Linde, A. Buzo and R. M. Gray : "An algorithm for vector quantization design," IEEE Trans. on Communications, vol.28, no.1, pp.84-95, Jan. 1980.
- [11] S.J.J. Odell, "The use of context in large vocabulary speech recognition," PhD dissertation, Cambridge University, 1995.
- [12] 板谷洋平, 宮島千代美, 徳田恵一, 北村正, "DAEM アルゴリズムを用いた話者認識のための GMM 学習法", 日本音響学会春季研究発表会講演論文集, 3-3-6, pp.115-116, Mar. 2003.
- [13] 板谷洋平, 全 炳河, 南角 吉彦, 宮島千代美, 徳田恵一, 北村正, "DAEM アルゴリズムに基づく HMM を用いた連続音声認識", 日本音響学会秋季研究発表会講演論文集, 2-6-4, pp.67-68, Sep. 2003.