音声言語情報処理 29－10
(1999. 12. 20)

話者照合における話者モデルの MLLR適応の検討

加納 淳也　　加藤 正治　　伊藤 彰則　　好田 正紀

山形大学工学部
〒992 米沢市城南4丁目3-16
E-mail : kohda@ei5sun.yz.yamagata-u.ac.jp

あらまし　本報告では, 学習データ量に応じた回帰クラスタ設定にMDL基準を用いる方法を提案する. 話者モデルは, 隠れマルコフモデル (HMM : Hidden Markov Model) で表し, 主張話者モデルをMLLR適応により作成する. 回帰クラスを設定するために, 音響的な距離を基準にtop-down clusteringで作成した木構造を用いる. 木構造を用いた回帰クラスタの自動設定には, 次の3通りを実験する. MDL基準を用いる場合, フレーム数を用いる場合, 回帰クラスタを固定する場合. 発声内容指定型話者照合で実験した結果, MDL基準を用いる方法は, クラスタ分割を抑制し, 学習データ量に応じた最適なクラスタ数を選択できる可能性が得られた.

キーワード　話者照合, MLLR適応, 回帰クラスタ, MDL基準

A study on MLLR adapted speaker model for speaker verification

Junya KANOU,　Masaharu KATOH,　Akinori ITO　Masaki KOHDA

Faculty of Engineering, Yamagata University
4-3-16 Johnan,Yonezawa-shi, 992 Japan
E-mail : kohda@ei5sun.yz.yamagata-u.ac.jp

Abstract　In this paper, we propose a method to make automatically the regression cluster corresponding to the amount of adaptation data by MDL criterion. Claimant speaker models are made by MLLR adaptation. To increase the number of regression clusters, we use a tree structure. It is made with top-down clustering based on acoustic distance. The MDL criterion is compared with the frame threshold criterion and fixed regression clusters criterion. In the experiment on the text-prompted speaker verification, MDL criterion becomes the repression of cluster division, and the most suitable number of cluster corresponding to the amount of adaptation data is chosen.

key words　speaker verification, MLLR adaptation, regression cluster, Minimum Description Length criterion

1 はじめに

近年, インターネット等のネットワークの発達により, ネットワークを介したサービス市場が急速に増加しつつある. これに比例して, ユーザに対して高いセキュリティ性を保証する必要も増加しており, 現状の暗証番号のみでは不十分となってきている. 本研究の目的は, 画像・音声等の生体情報を利用した個人認証システムの構築において, 特に音声を用いた高精度の個人認証技術を確立することにある.

音声を用いた個人認証を話者照合という. 話者照合では, 入力音声と同時に自分が誰であるかの ID を入力して, その音声が本当にその ID に対応する人の音声であるか否かを判定する. 入力音声と ID に対応する人の音響モデルとの類似度が, 一定のしきい値よりも大きければ本人の音声であると判定し, それ以外は他人の音声であると判定する. 話者照合では, 登録時に各話者に膨大な量の発声を要求するような方法は現実的ではないので, 小数サンプルからその話者の音響モデルをいかに正確に作成するかが重要である.

発声内容依存型話者照合では, 照合に用いるテキスト (キーワード) をあらかじめ決めておくため, それに固有な音韻情報も利用することができ, 一般的に発声内容独立型話者照合に比べて性能が高い. しかし, 実際にカードや鍵の代わりに, 音声を用いて自動的に話者照合することを考えると, キーワードが固定されている場合, テープレコーダ等によって話者の声を録音し, 照合装置の前で再生すれば, それを受理してしまう危険性がある. 従って, 信頼性の高い話者照合を実現するためには, キーワードを可変に設定することができ, 本人がそのキーワードを正しく発声したときだけ受理するような方式が必要である. これを発声内容指定型話者照合 [1] という.

本報告は, 発声内容指定型話者照合に関して行った研究である. 話者照合システムの構築では, 音響モデルを隠れマルコフモデル (HMM : hidden Markov model) で実現することと, 不特定話者モデルを用いて主張話者モデルのスコアを正規化することを基本とした. 話者適応法による主張話者モデルの学習に関しては, 最尤推定法 (ML : Maximum Likelihood) と最尤線形回帰法 (MLLR : Maximum Likelihood Linear Regression) を用いて. 学習するモデルパラメータの性能比較, 回帰クラスタの単位の性能比較, 回帰クラスタ数の設定法, 等を検討した.

2 話者照合システム

2.1 音声データ

音声データは, 男性 10 名が静かな部屋で発声した 4 桁数字で, 各発声者について学習用 35 組, 評価用 40 組の 4 桁数字を収録する.

2.2 音声分析条件

音声データを 16kHz, 16bit でディジタル化し, フレーム長 32ms, フレーム周期 8ms, ハミング窓, 高域強調を施し, 対数パワーと 12 次元の LPC メルケプストラム, およびそれらの 1 次と 2 次の回帰係数 (計 39 次元) を抽出する. さらにケプストラム平均正規化を行う.

2.3 音響モデル

音素 HMM は, 音素環境独立な 28 種類 (無音を含む) を作成する. 28 種類のうち, 4 桁数字に含まれるのは 17 種類である. 各音素 HMM は 3 状態, 12 混合正規分布, 対角共分散行列をもつ.

背景話者モデルの音素 HMM は, ASJ 音声データベースの男性 20 名が発声した音素バランス 3000 文から作成する. 主張話者用の HMM は背景話者モデルから MLLR 法で話者適応することによって平均ベクトル, 分散を学習する. 分布重みは ML 法で学習する.

2.4 尤度正規化

主張話者モデルによる入力音声の対数尤度を, 背景話者モデルによる入力音声の対数尤度で正規化する.

$$f(s) = g(s) - h \tag{1}$$

ここで, $g(s)$ は主張話者 s のモデルによる対数尤度, h は背景話者モデルによる対数尤度を表す. h は指定された発話の正解音素列に従って背景話者モデルを並べた HMM で求める. 主張話者の受理は, 閾値 θ に対して

$$f(s) > \theta \tag{2}$$

となる時に行う.

2.5 システムの評価

主張話者の発声に対して, それ以外の 9 名を詐称話者とする. システムの評価は, EER(Equal Error Rate : 等誤り率) で行う. 全ての話者に対して共通のしきい値 θ を事後的に設定する.

3 話者モデルのMLLR適応

音声認識に用いられているMLLR法を話者照合に利用する．MLLR法は，学習データから変換行列を求め，背景話者モデルのパラメータ空間から主張話者モデルのパラメータ空間へ一括移動させる方法である．MLLR法では，少量の学習データから変換行列を求めることができ，対応する学習データがないパラメータも適応が可能である．

3.1 平均ベクトル，共分散行列の適応

m 番目の回帰クラスタの r 番目の成分分布の平均ベクトルを $\boldsymbol{\mu}_{m_r}$，共分散行列を $\boldsymbol{\Sigma}_{m_r}$ とする．それらはMLLR法 [2] で次式によって更新される．

$$\hat{\boldsymbol{\mu}}_{m_r} = \hat{\mathbf{A}}_m \boldsymbol{\mu}_{m_r} + \hat{\boldsymbol{\beta}}_m \quad (3)$$

$$\hat{\boldsymbol{\Sigma}}_{m_r} = \mathbf{B}_{m_r}^T \hat{\mathbf{H}}_m \mathbf{B}_{m_r} \quad (4)$$

ここで，$\mathbf{B}_{m_r}$ は $\boldsymbol{\Sigma}_{m_r}^{-1}$ のCholeski factorの逆行列である．

(3), (4) 式の $\hat{\mathbf{A}}_m$, $\hat{\boldsymbol{\beta}}_m$, $\hat{\mathbf{H}}_m$ は m 番目の回帰クラスタの変換行列・ベクトルで，最尤推定によって得られる．

3.2 回帰クラスタの木構造

MLLR法の回帰クラスタ数を増やすことで，クラスタ毎に線形変換を行い，より詳細にパラメータを学習できる．

分布，状態，音素を単位とする回帰クラスタを扱う．2クラスタ間の距離尺度としてBhattacharrya距離を用いる．回帰クラスタの木構造の作成には，top-down clustering[3] をLBG法に基づいて行なう．

3.3 回帰クラスタ数の設定

3.3.1 フレーム数に基づく設定法

フレーム数に基づく設定法では，次の記号を定義して，2通りの設定法を比較する．

- 節点 0, 1, 2 : 回帰クラスタの木構造の節点．但し，1, 2は0の子節点とする．
- N0, N1, N2 : 節点 0, 1, 2 の回帰クラスタに属する学習データのフレーム数 (N0=N1+N2)
- T : フレーム数の閾値
  (節点0ではN0≥Tの条件をすでに満たしているとする)

＜設定法1＞

$(N1 \geq T) \cap (N2 \geq T) \Rightarrow$ 節点 1, 2 の回帰クラスタ
$(N1 < T) \cup (N2 < T) \Rightarrow$ 節点 0 の回帰クラスタ

＜設定法2＞

$(N1 \geq T) \cap (N2 \geq T) \Rightarrow$ 節点 1, 2 の回帰クラスタ
$(N1 \geq T) \cap (N2 < T) \Rightarrow$ 節点 0, 1 の回帰クラスタ
$(N1 < T) \cap (N2 \geq T) \Rightarrow$ 節点 0, 2 の回帰クラスタ
$(N1 < T) \cap (N2 < T) \Rightarrow$ 節点 0 の回帰クラスタ

設定法1では両方の子節点が条件を満たす場合のみ回帰クラスタを分割する．それに対して，設定法2では片方の子節点だけが条件を満たしても回帰クラスタを増やす．条件を満たさない子節点は節点0の回帰クラスタの変換行列を用いる．

3.3.2 MLLR-MDLに基づく設定法

MLLR-MDLに基づく設定法では，MLLR適応後のモデルにMDL基準 [4] を適用してクラスタ分割の判定を行う．つまり，モデルのMLLR適応とクラスタの分割を同時に，MDL基準で評価する．

MDL基準では，モデル i によるデータ x の記述長 $L(i)$ を次式で評価する．

$$L(i) = -\log P_i(x) + \frac{\alpha_i}{2} \log N + \log I \quad (5)$$

但し $P_i(x)$: モデル i によるデータ x の尤度
α_i : モデル i の自由パラメータの個数
N : データ x の長さ (フレーム数)
I : モデルの種類数

$P_i(x)$ はMLLR適応後のモデルを用いる．尤度の計算において，遷移確率は出力確率に比べて影響が小さいと仮定して無視する．モデル i の出力確率が正規分布 $N(x, \mu_i, \Sigma_i)$ で，$\mu_i = K$ 次元平均ベクトル，$\Sigma_i =$ 対角共分散行列，とする．式 (3) の $\hat{\mathbf{A}}_m$ にはフル変換行列，式 (4) の $\hat{\mathbf{H}}_m$ には対角変換行列を用いると，MLLR適応の自由パラメータの個数が $K(K+2)$ であることから，記述長は次式のようになる．

$$L(i) = \frac{1}{2}\Big(K \log 2\pi + \log|\Sigma_i| + K\Big)N + \frac{K(K+2)}{2} \log N + \log I \quad (6)$$

式 (6) の右辺第1項は，データ x の標本平均，標本分散を用いて $P_i(x)$ を計算する場合の式である．MLLR適応後のモデルを用いる場合には，

$$\frac{1}{2}\Big(K \log 2\pi + \log|\Sigma_i| + \sum_{k=1}^{K} \frac{\overline{\sigma}_{ik}^2 + (\mu_{ik} - \overline{\mu}_{ik})^2}{\sigma_{ik}^2}\Big)N$$

ここで　μ_{ik}, σ_{ik}^2 :　$MLLR$ 適応後の平均, 分散
$\overline{\mu}_{ik}, \overline{\sigma}_{ik}^2$:　データ x の標本平均, 標本分散

となるが, ここでは簡単のために, $\mu_{ik} \approx \overline{\mu}_{ik}, \sigma_{ik}^2 \approx \overline{\sigma}_{ik}^2$ と仮定して, 式 (6) を用いる.

いま, 節点 0 を節点 1, 2 にクラスタ分割する場合を考える. 節点 0, 1, 2 の回帰クラスタに属する MLLR 適応後の分布をそれぞれ (μ_{0m}, Σ_{0m}), (μ_{1m}, Σ_{1m}), (μ_{2m}, Σ_{2m}), 分布の占有フレーム数をそれぞれ N_{0m}, N_{1m}, N_{2m} と表す. 分布の占有フレーム数は, MLLR 適応とクラスタ分割の前後で変わらないと仮定して, 背景話者モデルから求めたものを共通に用いる. 節点 0 のモデルによる記述長 $L(0)$ と節点 1, 2 のモデルによる記述長 $L(1,2)$ は,

$$
\begin{aligned}
L(0) &= \frac{1}{2}\sum_{m=1}^{M_0}\Big(K\log 2\pi + \log|\Sigma_{0m}| + K\Big)N_{0m} \\
&\quad + \frac{K(K+2)}{2}\log N_0 + \log I \qquad (7)\\
L(1,2) &= \frac{1}{2}\sum_{m=1}^{M_1}\Big(K\log 2\pi + \log|\Sigma_{1m}| + K\Big)N_{1m} \\
&\quad + \frac{1}{2}\sum_{m=1}^{M_2}\Big(K\log 2\pi + \log|\Sigma_{2m}| + K\Big)N_{2m} \\
&\quad + K(K+2)\log N_0 + \log I \qquad (8)
\end{aligned}
$$

ここで　M_0, M_1, M_2 : 節点 0, 1, 2 の回帰クラスタに属する分布の個数

$$
\begin{aligned}
M_0 &= M_1 + M_2 \\
N_0 &= \sum_{m=1}^{M_0} N_{0m} \\
N_1 &= \sum_{m=1}^{M_1} N_{1m} \\
N_2 &= \sum_{m=1}^{M_2} N_{2m} \\
N_0 &= N_1 + N_2
\end{aligned}
$$

であるので, それらの差は次式のようになる.

$$
\begin{aligned}
\Delta &= L(1,2) - L(0) \\
&= \frac{1}{2}\Big(\sum_{m=1}^{M_1} N_{1m}\log|\Sigma_{1m}| + \sum_{m=1}^{M_2} N_{2m}\log|\Sigma_{2m}| - \\
&\quad \sum_{m=1}^{M_0} N_{0m}\log|\Sigma_{0m}|\Big) + \frac{K(K+2)}{2}\log N_0 \qquad (9)
\end{aligned}
$$

実験では, 上式の代わりに次式を用いる.

$$
\begin{aligned}
\Delta &= \frac{1}{2}\Big(\sum_{m=1}^{M_1} N_{1m}\log|\Sigma_{1m}| + \sum_{m=1}^{M_2} N_{2m}\log|\Sigma_{2m}| - \\
&\quad \sum_{m=1}^{M_0} N_{0m}\log|\Sigma_{0m}|\Big) + c\frac{K(K+2)}{2}\log N_0 \qquad (10)
\end{aligned}
$$

記述長が小さくなるクラスタ分割が望ましいので, 回帰クラスタを下記のルールで設定する.

$$
\begin{aligned}
\Delta < 0 &\quad\Longrightarrow\quad \text{節点 1, 2 の回帰クラスタ} \\
\Delta \geq 0 &\quad\Longrightarrow\quad \text{節点 0 の回帰クラスタ}
\end{aligned}
$$

記述長の差に関する「$\Delta < 0$」の条件は, 対数尤度の増分に関する次式の条件と等価である.

$$
\begin{aligned}
\text{尤度増分} &= -\frac{1}{2}\Big(\sum_{m=1}^{M_1} N_{1m}\log|\Sigma_{1m}| \\
&\quad + \sum_{m=1}^{M_2} N_{2m}\log|\Sigma_{2m}|\Big) \\
&\quad + \frac{1}{2}\sum_{m=1}^{M_0} N_{0m}\log|\Sigma_{0m}| \\
&> c\frac{K(K+2)}{2}\log N_0 \qquad (11)
\end{aligned}
$$

この式からわかるように, 係数 c の値を大きくするとクラスタ分割が抑制される.

4　実験結果と考察

4.1　実験条件

MLLR 法の平均ベクトルの変換行列はフル変換行列, 分散の変換行列は対角変換行列とする. 分布重みは ML 法で推定する. 回帰クラスタの単位は, 音素, 状態, 分布の 3 通りについて行う. 無音の適応は行わない. 主張話者の学習データは, 1, 2, 3, 4, 5, 10, 20, 35 組の計 8 通りについて行う.

4.2　適応モデルパラメータの比較

適応するモデルパラメータの 4 通りの場合について, 学習データ数による EER の変化を図 1 に示す. 図 1 で, 回帰クラスタ数は, 学習データ数によらず 1 としている. 適応するモデルパラメータは, 平均ベクトルに分散, 分布重みを加えた場合に EER が最も小さくなる. 特に, 分散の MLLR 適応の効果が大きい. 分散の ML 適応が少量の学習データでは逆効果になるが, それとは対照的である. 分布重みの ML 適応も一定の効果がある.

4.3　回帰クラスタの単位の比較

回帰クラスタの単位の 3 通りの場合について, 学習データ数による EER の変化を図 2 に示す. 図 2 で, 回帰クラスタ数はフレーム設定法 2 に基づいて設定し, フレーム閾値 1250 としている. 音素, 状態, 分布の各単位で EER に大きな差はない. 以下の実験では, 回帰クラスタは状態単位とする.

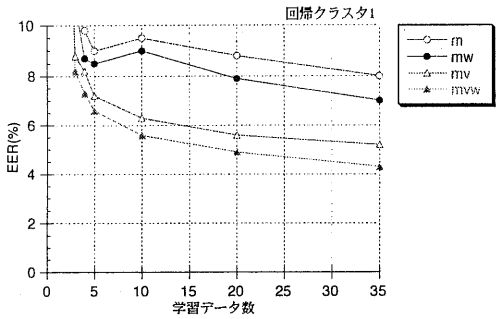

図 1: 適応モデルパラメータの比較 (回帰クラスタ数=1)

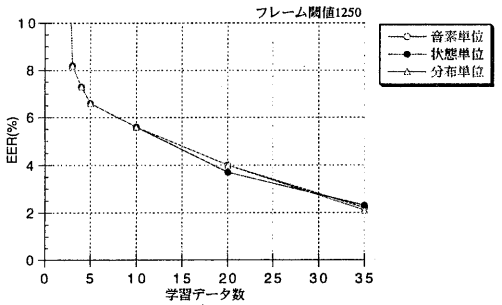

図 2: 回帰クラスタの単位の比較 (フレーム設定法 2・閾値 1250)

4.4 回帰クラスタ数の設定法の検討

4.4.1 フレーム数に基づく設定法

フレーム数に基づく設定法 1, 2 でフレーム閾値を変えた場合, 学習データ数による回帰クラスタ数の変化を図 3, 4, EER の変化を図 5, 6 に示す. これらの結果から, フレーム閾値は設定法 1 では 500, 設定法 2 では 1000 が適当である.

設定法 1 では, 左右の節点が条件を満たさないとクラスタ分割が起こらないので, 回帰クラスタ数は, 学習データが少ない間は抑制されていて, 途中から急に増加する傾向がある. そのため, 回帰クラスタ数が多くなりすぎて, EER が急に悪くなる場合がある. これは, クラスタ当りの学習データが減少し, 変換行列の推定がうまくいかないことによる. それに対して, 設定法 2 では, 回帰クラスタ数が急に増えることはないが, 条件を満たさない片方の節点は, 親節点の変換行列を用いて MLLR 適応を行っているので, EER は設定法 1 より若干劣る場合がある.

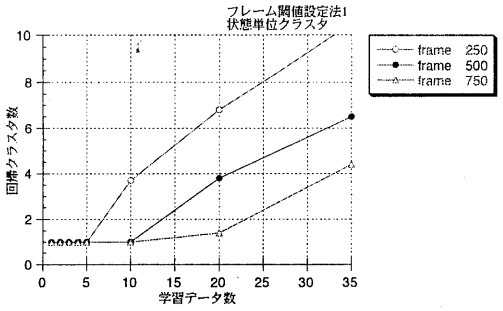

図 3: フレーム設定法 1 での回帰クラスタ数の変化

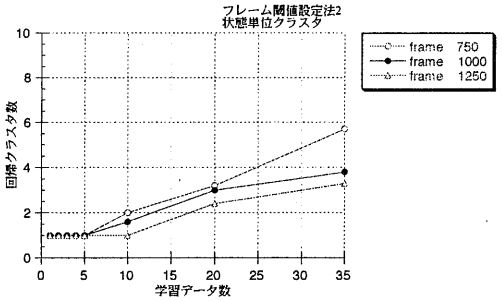

図 4: フレーム設定法 2 での回帰クラスタ数の変化

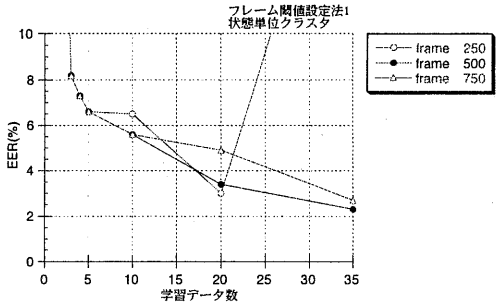

図 5: フレーム設定法 1 での EER の変化

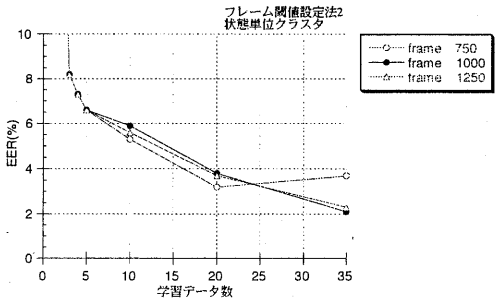

図 6: フレーム設定法 2 での EER の変化

4.4.2 MLLR-MDL に基づく設定法

MLLR-MDL に基づく設定法で係数 c を変えた場合, 学習データ数による回帰クラスタ数の変化を図 7, EER の変化を図 8 に示す. MDL 基準が, クラスタ分割の停止基準となっているので, クラスタを増やしすぎて照合が急に悪くなることはない. しかし, 学習データの量に応じてクラスタを設定するとなると, 係数 c の値が小さいときは, 学習データの多いところで性能が良く, 学習データの少ないところではクラスタを増やしすぎるために悪くなる. 逆に, 係数 c の値が大きいときは, 学習データの多いところでクラスタの分割を抑制しすぎている.

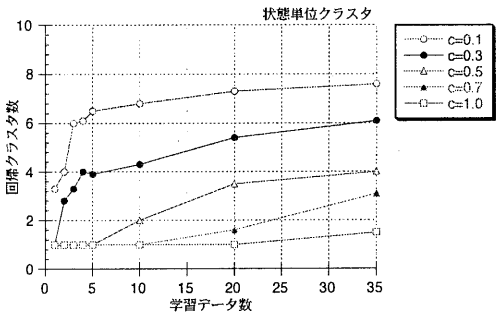

図 7: MLLR-MDL での回帰クラスタ数の変化

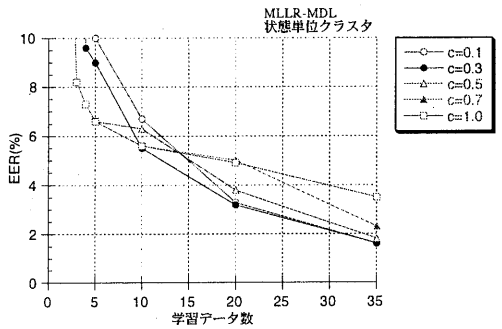

図 8: MLLR-MDL での EER の変化

4.4.3 回帰クラスタ数の設定法の比較

回帰クラスタ数を自動設定する方法の有効性をみるために, 回帰クラスタ 1 または 2 に固定の場合と, 回帰クラスタ自動設定の場合について比較したものを図 9 に示す. 図 9 から, 回帰クラスタを 1, 2 に固定した場合よりも, フレーム閾値や MLLR-MDL の自動設定の方が性能が良い. 回帰クラスタ 1 に固定の場合は学習データの多いところで性能が悪く, 回帰クラスタ 2 に固定の場合は学習データの少ないところで性能が悪い.

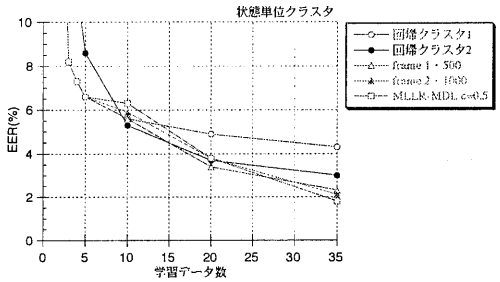

図 9: 回帰クラスタ数の設定法の比較

5 むすび

話者照合における話者モデルの MLLR 適応を検討し, 照合実験を行った結果を報告した. 回帰クラスタの数を適切に増やすことにより照合性能は向上することが分かった. また, MDL 基準を用いることにより, 学習データ量に応じた最適なクラスタ数を選択できる可能性が得られた. 今後は, 他の適応法と組み合わせたり, 時期差に頑健なモデルの作成を検討する予定である.

謝辞

本研究は, NTT データとの共同研究で行われたものである. 有益な議論をしていただいた, NTT データ情報科学研究所の高橋淳一氏, 磯部俊洋氏に深く感謝する.

参考文献

[1] 松井, 古井 : "テキスト指定型話者認識", 信学論 D-II, Vol.J79-D-II, No.5, pp.647-655 (1996-05)

[2] M.J.F.Gales, P.C.Woodland : *"Mean and variance adaptation within the MLLR framework"*, Computer Speech and Language, Vol.10, pp.249-264 (1996)

[3] 小森, 山田, 山本, 小坂, 大洞 : *"Top-Down Clustering* に基づく効率的な *Shared-State Triphone HMM"*, 信学技報, SP95-21, pp.23-30 (1995-06)

[4] 篠田, 渡辺 : "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", 信学技報, SP96-79, pp.9-15 (1996)