社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2003－ＳＬＰ－49　(16)
2003／12／18

混合因子分析に基づく話者識別モデルのパラメータ共有構造

山本　啓善†　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　〒464-8603 愛知県名古屋市千種区不老町
E-mail: †{legacy,nanaku,tokuda,kitamura}@ics.nitech.ac.jp, ††miyajima@is.nagoya-u.ac.jp

あらまし　本論文では，話者識別実験において混合因子分析モデルのパラメータ共有構造について検討を行う．また，識別性能を向上させるため，ML 推定により得られたモデルに最小分類誤り学習を適用する．パラメータの共有方法には因子負荷行列を共有する方法や独自因子の分散を共有する方法などが考えられる．我々は，テキスト独立型話者識別実験より，すべての共有方法において対角共分散行列や全共分散行列を用いた GMM に対する優位性を確認した．学習データが少ない場合においても，ほぼ同様の傾向が見られた．また，本実験においては独自因子の分散を共有した場合が，最も有効であり，このとき，対角共分散行列を用いた GMM に対して識別誤りを約 26% 削減することができた．さらに，混合因子分析モデルに対し，最小分類誤り学習を適用することにより，約 3% の識別誤り率の改善が見られた．

キーワード　話者識別，GMM，混合因子分析，パラメータ共有，最小分類誤り学習

Parameter Sharing of Mixtures of Factor Analyzers for Speaker Identification

H. YAMAMOTO†, Y. NANKAKU†, C. MIYAJIMA††, K. TOKUDA†, and T. KITAMURA†

† Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology　Gokiso-cho, Showa-ku, Nagoya, 466-8555, Japan
†† Department of Media Science, Graduate School of Information Science, Nagoya University　Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan
E-mail: †{legacy,nanaku,tokuda,kitamura}@ics.nitech.ac.jp, ††miyajima@is.nagoya-u.ac.jp

Abstract　This paper investigates the parameter tying strategies of mixtures of factor analyzers (MFAs) and discriminative training of MFA based on Maximum Likelihood (ML) solution for speaker identification. The parameters of factor loading matrices or diagonal covariance matrices are shared in different mixtures of MFA. The minimum classification error (MCE) training is applied to the MFA parameters to enhance the discrimination abilities. The results of text-independent speaker identification experiments show that the MFAs outperform the conventional Gaussian mixture models (GMMs) with diagonal or full covariance matrices. Also the same tendency are seen when training data is sparse. MFAs achieve the best performance when sharing the diagonal matrices, resulting in a relative error reduction of 26% over the GMM with diagonal covariance matrices. The recognition performance is further improved by the MCE training with an additional 3% error reduction.

Key words　speaker identification, GMM, mixtures of factor analyzers, parameter tying, minimum classification error training

1. まえがき

テキスト独立型話者識別では一般に，混合ガウスモデル (Mixtures of Gaussian Models, GMMs) [1] が広く用いられている．全共分散行列を用いた GMM の場合，各モデルパラメータの信頼性を保証するためには，十分な学習データが必要になる．そこで分散行列を対角成分のみとした GMM が広く用いられるが，高い認識性能を得るためには，混合数を多くしなければならない．こうした問題を解決するために，近年，混合因子分析 (Mixtures of Factor Analyzers, MFA) [2] が音声認識や話

者認識に適用され，その有効性が示されている [3], [4]．混合因子分析を適用することで分散行列の次元圧縮を行うことができ，少ないパラメータで高い認識率が期待できる．また，パラメータを共有することで，学習データが少ない場合の各モデルパラメータの信頼性を向上させることが可能であると考えられる．

本論文では，話者識別実験において，混合因子分析モデルのパラメータ共有構造について検討を行う．各混合要素間で，因子数は等しいと仮定し，分散行列を構成するパラメータに関して以下の 3 つの共有構造を持った MFA モデルについて，比較検討を行う．

1) パラメータ共有を行わない MFA
2) 独自因子分散を共有した MFA
3) 因子負荷行列を共有した MFA

加えて，MFA で得られた話者モデルに対して，最小分類誤り (Minimum Classification Error, MCE) 学習を適用することにより，話者識別性能のさらなる向上が期待できる．そこで，パラメータ共有を行った MFA における MCE 学習適用の有効性についても，テキスト独立型話者識別実験において評価する．

本稿では，第 2 章で混合因子分析を，第 3 章でその共有構造について，第 4 章で最小分類誤り学習について述べる．第 5 章で話者識別実験を行い，第 6 章でむすびと今後の課題について述べる．

2. 混合因子分析

2.1 因子分析

因子分析 (Factor Analysis, FA) はデータ変数間に潜む因子の存在を仮定し，その因子を通して高次元データの分散構造をモデル化する手法である．つまり FA では，話者データの d 次の音声特徴量 $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^T$ が，それらに共通する $q\,(q < d)$ 次の変数からなる共通因子と，各変数に固有な独自因子により次式で生成されると仮定する．

$$\begin{aligned}\boldsymbol{x} &= \boldsymbol{\mu} + \sum_{i=1}^{q} z_i \boldsymbol{w}_i + \boldsymbol{u} \\ &= \boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{z} + \boldsymbol{u} \end{aligned} \tag{1}$$

ここで，$\boldsymbol{\mu}$ は平均ベクトルを，$\boldsymbol{z}=(z_1, z_2, \ldots, z_q)^T$ は共通因子，$\boldsymbol{u}=(u_1, u_2, \ldots, u_d)^T$ は独自因子，$\boldsymbol{W}=(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q)$，$\boldsymbol{w}_i=(w_{i1}, w_{i2}, \ldots, w_{id})^T$ は音声特徴量に対する共通因子の負荷量を示す．ただし，共通因子 $\boldsymbol{z}$，独自因子 $\boldsymbol{u}$ は，平均零の多次元無相関正規分布を仮定する．すなわち，$p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$，$p(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi})$ とする．ここで，$\boldsymbol{0}$ は零ベクトル，$\boldsymbol{I}$ は q 次元単位行列，$\boldsymbol{\Psi}$ は d 次元対角行列を表す．

隠れ変数 $\boldsymbol{z}$ が与えられたときの音声特徴量 $\boldsymbol{x}$ の条件付き分布は次式で得られる．

$$p(\boldsymbol{x} \mid \boldsymbol{z}) = \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{z}, \boldsymbol{\Psi}) \tag{2}$$

したがって，話者モデルにおける音声特徴量 $\boldsymbol{x}$ の出力確率は，隠れ変数 $\boldsymbol{z}$ に関して積分することにより，次式で与えられる．

$$\begin{aligned} p(\boldsymbol{x}) &= \int p(\boldsymbol{x} \mid \boldsymbol{z}) p(\boldsymbol{z}) d\boldsymbol{z} \\ &= \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{W}\boldsymbol{W}^T + \boldsymbol{\Psi}) \end{aligned} \tag{3}$$

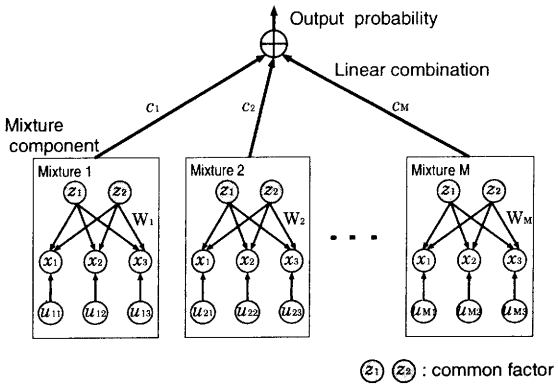

図 1 混合因子分析モデル
Fig. 1 Mixtures of Factor Analyzers

2.2 混合因子分析への拡張

単一の FA モデルは，あるガウス分布に従う単純なデータ空間に対しては，因子数を適切に選択することでうまく働く．しかし実際には，音声データのように複雑なデータ空間をモデル化することが求められる．そこで，図 1 に示すような複数の FA モデルの線形和としての混合因子分析モデル（MFA）が提案され，その有効性が示されている [3], [4]．M 個の FA を考えた場合，音声特徴量 $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T)$ の確率密度関数 $p(\boldsymbol{X})$ は，m 番目のモデルの混合重みを c_m とすると，次式のように表すことができる．

$$p(\boldsymbol{X} \mid \boldsymbol{\Theta}) = \prod_{t=1}^{T} \sum_{m=1}^{M} \int p_m(\boldsymbol{x}_t \mid \boldsymbol{z}) p_m(\boldsymbol{z}) c_m d\boldsymbol{z} \tag{4}$$

ただし，$\boldsymbol{\Theta} = \{P_m, \boldsymbol{\mu}_m, \boldsymbol{W}_m, \boldsymbol{\Psi}_m \mid m = 1, \ldots, M\}$ とする．

3. パラメータ共有構造

ここでは，MFA の共分散行列 $\boldsymbol{\Sigma}_m$ を構成するパラメータ $\boldsymbol{W}_m, \boldsymbol{\Psi}_m$ の共有について検討する．パラメータを共有することにより，学習データ量が少ない場合におけるモデルパラメータの信頼性を向上させることができると考えられる．MFA の構造としてはいくつか考えられる [6] が，本研究では，共通因子数がすべての混合要素で等しいと仮定し，以下の 3 つの共有構造を検討する．

1) Generic MFA: 混合要素間でパラメータ共有を行わない MFA．一般的な MFA．

2) $\boldsymbol{\Psi}$-shared MFA: 因子負荷行列を共有した MFA．つまり，独自因子 $\boldsymbol{u}$ を混合要素によらないセンサーノイズとみなして，$\boldsymbol{\Psi}_1 = \boldsymbol{\Psi}_2 = \cdots = \boldsymbol{\Psi}$ としたもの．

3) $\boldsymbol{W}$-shared MFA: 独自因子の分散を共有した MFA．つまり，各因子の影響量は混合要素間ですべて同じとみなして，$\boldsymbol{W}_1 = \boldsymbol{W}_2 = \cdots = \boldsymbol{W}$ としたもの．

尤度最大化 (Maximum Likelihood, ML) 基準による MFA のパラメータ推定問題は EM アルゴリズムにより解くことができる [2]．以下，それぞれの共有構造における EM ステップを

示す．

3.1 E-step

E-step では，共通因子 z に関する期待値と混合要素 m の事後確率 h_{tm} を計算する．

$$\langle \boldsymbol{z}_{tm} \rangle = E[\boldsymbol{z}|\boldsymbol{x}_t, m] = \boldsymbol{\beta}_m(\boldsymbol{x}_t - \boldsymbol{\mu}_m) \tag{5}$$

$$\begin{aligned}\langle \boldsymbol{z}\boldsymbol{z}_{tm} \rangle &= E[\boldsymbol{z}\boldsymbol{z}^T|\boldsymbol{x}_t, m] \\ &= I - \boldsymbol{\beta}_m \boldsymbol{W}_m + \langle \boldsymbol{z}_{tm} \rangle \langle \boldsymbol{z}_{tm} \rangle^T \end{aligned} \tag{6}$$

$$h_{tm} = \frac{c_m \mathcal{N}(\boldsymbol{x}_t \mid \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)}{\sum_m c_m \mathcal{N}(\boldsymbol{x}_t \mid \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)} \tag{7}$$

ただし．$\boldsymbol{\beta}_m = \boldsymbol{W}_m^T \boldsymbol{\Sigma}_m^{-1}$, $\boldsymbol{\Sigma}_m = \boldsymbol{W}_m \boldsymbol{W}_m^T + \boldsymbol{\Psi}_m$ とする．これは generic MFA の場合である．$\boldsymbol{\Psi}$-shared MFA, $\boldsymbol{W}$-shared MFA についてはそれぞれ $\boldsymbol{\Psi}_m = \boldsymbol{\Psi}, \boldsymbol{W}_m = \boldsymbol{W}$ とすればよい．

3.2 M-step

3 つの共有構造における各モデルパラメータ $\boldsymbol{\mu}$, $\boldsymbol{W}$, $\boldsymbol{\Psi}$, c_m の更新式を以下に示す．

(1) generic MFA

ここでは，更新式を求める際の簡略化のため，$\tilde{\boldsymbol{W}}_m = (\boldsymbol{W}_m \ \boldsymbol{\mu}_m)$, $\tilde{\boldsymbol{z}}^T = (\boldsymbol{z}^T \ 1)$ と定義すると，期待値の部分は

$$\langle \tilde{\boldsymbol{z}}_{tm} \rangle = \begin{pmatrix} \langle \boldsymbol{z}_{tm} \rangle \\ 1 \end{pmatrix}, \langle \tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}_{tm} \rangle = \begin{pmatrix} \langle \boldsymbol{z}\boldsymbol{z}_{tm} \rangle & \langle \boldsymbol{z}_{tm} \rangle \\ \langle \boldsymbol{z}_{tm} \rangle & 1 \end{pmatrix}$$

となり，パラメータ $\tilde{\boldsymbol{W}}'_m, \boldsymbol{\Psi}'_m$ の更新式は以下で与えられる．

$$\tilde{\boldsymbol{W}}'_m = \left(\sum_t h_{tm} \boldsymbol{x}_t \langle \tilde{\boldsymbol{z}}_{tm} \rangle^T \right) \cdot \left(\sum_l h_{lm} \langle \tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}_{lm} \rangle \right)^{-1} \tag{8}$$

$$\boldsymbol{\Psi}'_m = \frac{1}{\sum_t h_{tm}} \mathrm{diag} \left\{ \sum_t h_{tm} \left(\boldsymbol{x}_t - \tilde{\boldsymbol{W}}'_m \langle \tilde{\boldsymbol{z}}_{tm} \rangle \right) \boldsymbol{x}_t^T \right\} \tag{9}$$

ここで，diag(·) は正方行列の対角成分以外を 0 としたものを表す．また，混合重み c_m の更新式は以下で与えられる．

$$c'_m = \frac{1}{T} \sum_{t=1}^{T} h_{tm} \tag{10}$$

(2) $\boldsymbol{\Psi}$-shared MFA

$\boldsymbol{\Psi}$-shared MFA においても，(1)generic MFA と同様な行列操作を行う．すると $\tilde{\boldsymbol{W}}'_m, c'_m$ については式 (8),(10) と同じとなる．混合要素間で共有する $\boldsymbol{\Psi}'$ は以下で与えられる．

$$\boldsymbol{\Psi}' = \frac{1}{T} \mathrm{diag} \left\{ \sum_{t,m} h_{tm} \left(\boldsymbol{x}_t - \tilde{\boldsymbol{W}}'_m \langle \tilde{\boldsymbol{z}}_{tm} \rangle \right) \boldsymbol{x}_t^T \right\} \tag{11}$$

(3) $\boldsymbol{W}$-shared MFA

$\boldsymbol{W}$-shared MFA において，要素間で共有する因子負荷行列 $\boldsymbol{W}'$ は，

$$\begin{aligned}\boldsymbol{W}'_{(k)} = &\left(\sum_{t,m} h_{tm} \boldsymbol{\Psi}^{-1}_{m(k)} (\boldsymbol{x}_t - \boldsymbol{\mu}_m)_{(k)} \langle \boldsymbol{z}_{tm} \rangle^T \right) \\ &\cdot \left(\sum_{t,m} h_{tm} \boldsymbol{\Psi}^{-1}_{m(k)} \langle \boldsymbol{z}\boldsymbol{z}_{tm} \rangle \right)^{-1} \end{aligned} \tag{12}$$

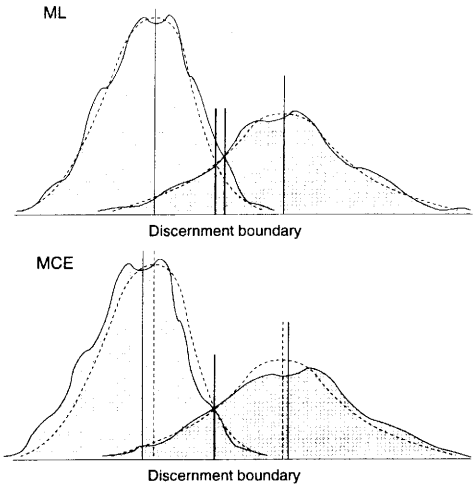

図 2 ML 推定，MCE 学習による識別境界

Fig. 2 Discrnment boundary of ML estimation and MCE training

となる．ただし，k 行目の $\boldsymbol{W}$ を $\boldsymbol{W}_{(k)}$ とする．また，$\boldsymbol{\mu}'_m$, $\boldsymbol{\Psi}'_m$ は以下で与えられる．

$$\boldsymbol{\mu}'_m = \frac{\sum_t h_{tm} (\boldsymbol{x}_t - \boldsymbol{W}' \langle \boldsymbol{z}_{tm} \rangle)}{\sum_t h_{tm}} \tag{13}$$

$$\begin{aligned}\boldsymbol{\Psi}'_m = \frac{1}{\sum_t h_{tm}} diag \sum_t \Big\{ &h_{tm} (\boldsymbol{x}_t - \boldsymbol{\mu}'_m)(\boldsymbol{x}_t - \boldsymbol{\mu}'_m)^T \\ &- h_{tm} \boldsymbol{W}' \left(2 \langle \boldsymbol{z}_{tm} \rangle (\boldsymbol{x}_t - \boldsymbol{\mu}'_m)^T - \langle \boldsymbol{z}\boldsymbol{z}_{tm} \rangle \boldsymbol{W}'^T \right) \Big\} \end{aligned} \tag{14}$$

4. 最小分類誤り学習

図 2 に示すように，ML 基準に基づくモデルパラメータ推定は，各話者の特性を最もよく表現しようとするものであるといえる．しかし，学習したモデルと真のモデル分布とで識別境界がずれてしまう可能性があり，識別誤りを最小にする保証がない．ここでは，ML 法に基づく MFA 話者モデルの識別性能をさらに高めるために，一般的確率的降下 (generalized probabilistic descent, GPD) 法 [5] に基づく MCE 学習を MFA のモデルパラメータに適用する [3]．MCE 学習はモデルの識別境界と学習データの識別境界を合わせ，識別誤りを最小にすることを目的とする．

4.1 損失関数の定義

MCE 学習を行う為に，学習データ $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T)$ と MFA パラメータ $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_S)$ が与えられたとき，誤分類測度を次式で定義する．

$$d_s(\boldsymbol{X}; \boldsymbol{\Theta}) = -g_s(\boldsymbol{X}; \boldsymbol{\Theta}) + \log \left[\frac{1}{S-1} \sum_{y \neq s} \exp \{ g_y(\boldsymbol{X}; \boldsymbol{\Theta}) \eta \} \right]^{\frac{1}{\eta}} \tag{15}$$

ここで，$g_s(\cdot;\cdot)$ は話者 s の平均対数尤度を示す．本論文では，式 (15) の比較操作を制御する η を無限大に設定することにより得られる，以下の誤分類測度を用いる．

$$d_s(\boldsymbol{X};\boldsymbol{\Theta}) \approx -g_s(\boldsymbol{X};\boldsymbol{\Theta}) + \max_{y \neq s} g_y(\boldsymbol{X};\boldsymbol{\Theta}) \tag{16}$$

つまり，式 (16) は，学習データ $\boldsymbol{X}$ が属する話者 s とそれ以外で最も尤度の高い話者（競合話者）y の間で比較するように近似したものである．次に，誤分類測度を 0 − 1 に近似するシグモイド関数を用いて得られる損失を次式により求める．

$$l_s(\boldsymbol{X};\boldsymbol{\theta}) = \left(1 + \exp(-\gamma \cdot d_s)\right)^{-1} \tag{17}$$

ここで，γ はシグモイド関数の傾きを表す．識別学習のゴールは確率的降下規則に基づき損失を最小にすることである．

4.2 MFA パラメータの再学習

MCE 学習におけるパラメータ推定では，パラメータの制約条件（例えば，混合重みはすべて正）などを満たすため，すべての話者モデルのパラメータセット $\boldsymbol{\Theta}$ を新しいモデルパラメータセット $\bar{\boldsymbol{\Theta}}$ に変換しておく．

$$\bar{\boldsymbol{\Theta}} = \{\bar{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2, \ldots, \bar{\boldsymbol{\theta}}_S\} \tag{18}$$

$$\bar{\boldsymbol{\theta}} = \{\bar{c}_m, \bar{\boldsymbol{\mu}}_m, \boldsymbol{W}_m, \bar{\boldsymbol{\Psi}}_m \mid m = 1, 2, \ldots, M\} \tag{19}$$

但し，$\bar{c}_m = \log c_m$, $\bar{\mu}_{mi} = \frac{\mu_{mi}}{\Sigma_{mii}}$, $\bar{\Psi}_{mii} = \log \Psi_{mii}$ とする．その上で，各パラメータの修正量を求める．r 回適用後のパラメータを $\bar{\boldsymbol{\Theta}}(r)$ とすれば，確率的降下規則に基づき以下のように調整が行われる．

$$\bar{\boldsymbol{\Theta}}(r+1) = \bar{\boldsymbol{\Theta}}(r) - \varepsilon_r \nabla l_s(\boldsymbol{X};\bar{\boldsymbol{\theta}}) \tag{20}$$

ここで，ε_r は r 回目の学習係数を示す．

式 (20) の勾配は，

$$\nabla_{\bar{\boldsymbol{\theta}}_y} l_s(\boldsymbol{X};\bar{\boldsymbol{\theta}}) = \frac{\partial l_s}{\partial d_s}\frac{\partial d_s}{\partial g_y} \cdot \nabla_{\bar{\boldsymbol{\theta}}_y} g_y(\boldsymbol{X};\bar{\boldsymbol{\theta}}) \tag{21}$$

となり，$\frac{\partial l_s}{\partial d_s}$, $\frac{\partial d_s}{\partial g_y}$ はそれぞれ，

$$\frac{\partial l_s}{\partial d_s} = \gamma l_s(1 - l_s), \quad \frac{\partial d_s}{\partial g_y} = \begin{cases} -1, & y = s \\ 1, & y \neq s \end{cases} \tag{22}$$

となる．また，式 (21) の $\nabla_{\bar{\boldsymbol{\theta}}_y} g_y(\boldsymbol{X};\bar{\boldsymbol{\theta}})$ は以下となる．

$$\nabla_{\bar{\boldsymbol{\theta}}_y} g_y(\boldsymbol{X};\bar{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{b_y(\boldsymbol{x}_t)} \nabla_{\bar{\boldsymbol{\theta}}_y} b_y(\boldsymbol{x}_t) \tag{23}$$

ここで，$\nabla_{\bar{\boldsymbol{\theta}}} b(\boldsymbol{x}_t)$ はデータ $\boldsymbol{x}_t$ に対する出力確率 $b(\boldsymbol{x}_t)$ をモデル $\bar{\boldsymbol{\theta}}$ の各パラメータでそれぞれ偏微分したものであるので，これを計算することにより，更新量を求めることができる．以下にそれぞれの共有構造における各パラメータの更新量を示す．

(1) generic MFA

$f_m = c_m \mathcal{N}(\boldsymbol{x}_t \mid \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$, $\boldsymbol{\delta}_m = \boldsymbol{\Sigma}_m^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu}_m)$ とおくと，$b(\boldsymbol{x}_t)$ を各パラメータで偏微分したものは次式となる．これを用いて更新する．

$$\frac{\partial b(\boldsymbol{x}_t)}{\partial \bar{c}_m} = f_m, \quad \frac{\partial b(\boldsymbol{x}_t)}{\partial \bar{\mu}_{mi}} = f_m \delta_{mi} \Sigma_{mii} \tag{24}$$

$$\frac{\partial b(\boldsymbol{x}_t)}{\partial W_{mij}} = -f_m \left\{ [\boldsymbol{\Sigma}_m^{-1}\boldsymbol{W}_m]_{ij} - \delta_{mi}[\boldsymbol{\delta}_m^T \boldsymbol{W}_m]_j \right\} \tag{25}$$

$$\frac{\partial b(\boldsymbol{x}_t)}{\partial \bar{\Psi}_{mii}} = -\frac{1}{2} f_m \left\{ \Sigma_{mii}^{-1} - \delta_{mi}^2 \right\} \Psi_{mii} \tag{26}$$

ここで，$[\cdot]_i$ はベクトルの i 番目を，$[\cdot]_{ij}$ は行列の i,j 要素を示す．

(2) $\boldsymbol{\Psi}$-shared MFA

$\boldsymbol{\Psi}$-shared MFA の場合，$\bar{c}_m$, $\bar{\boldsymbol{\mu}}_m$, $\boldsymbol{W}_m$ の勾配は式 (24), (25) と同じである．式 (26) だけが以下のようになる．

$$\frac{\partial b(\boldsymbol{x}_t)}{\partial \bar{\Psi}_{ii}} = \sum_{m=1}^{M} \frac{\partial b(\boldsymbol{x}_t)}{\partial \bar{\Psi}_{mii}} \tag{27}$$

(3) $\boldsymbol{W}$-shared MFA

$\boldsymbol{W}$-shared MFA については，式 (24), (26) は同じである．式 (25) だけが以下のようになる．

$$\frac{\partial b(\boldsymbol{x}_t)}{\partial W_{ij}} = \sum_{m=1}^{M} \frac{\partial b(\boldsymbol{x}_t)}{\partial W_{mij}} \tag{28}$$

5. 話者識別実験

5.1 実験条件

実験データとして，ATR 日本語音声データベースの c-set を利用した．話者は男女各 40 名の計 80 名で，各話者 216 単語を学習に，520 単語を認識に用いた．サンプリング周波数は 10kHz, 分析周期は 10ms とし，音声の特徴ベクトルは 25.6ms 長ブラックマン窓を用いて得られた 0 次を除く 12 次のメルケプストラム係数を用いた．

GMM の初期化には LBG アルゴリズムを用いた．MFA モデルの初期化については，平均と混合重みは LBG アルゴリズムにより得られた値を，因子負荷量は分散を考慮した乱数を，独自因子の分散は全共分散の対角成分を用いた [2]．混合数を $4, 8, 16, 32, 64$ と変化させ，それぞれの混合数において因子数を $2, 4, 6, 8, 10$ と変化させて実験を行った．

5.2 実験結果

本研究では，対角共分散及び全共分散行列を用いた GMM (diag-GMM, full-GMM) と 3 種類の共有構造の MFA (generic MFA, $\boldsymbol{\Psi}$-shared MFA, $\boldsymbol{W}$-shared MFA) についてテキスト独立型話者識別実験を行った．図 3–5 は，それぞれの共有構造を持った MFA と GMM との間で，識別誤り率を比較したものである．グラフは各混合数において因子数を 2〜10 に変化させた結果を示し，横軸はパラメータ数を対数軸上でとってある．図 3 は，generic MFA と diag-GMM, full-GMM とで比較を行ったものである．generic MFA は因子数が少ない場合において，良い認識性能を得ていることが分かる．しかし，識別誤り率が，因子数が増えるに従って full-GMM に近づいて行くことが分かる．これは，初期値の違いはあるにせよ，因子数 $q = 12$ とした場合の generic MFA の分散は，full-GMM のそれとほぼ等しくなるためであると考えられる．図 4 は，$\boldsymbol{\Psi}$-shared MFA の結果を示している．$\boldsymbol{\Psi}$-shared MFA は混合数が多い場合において，

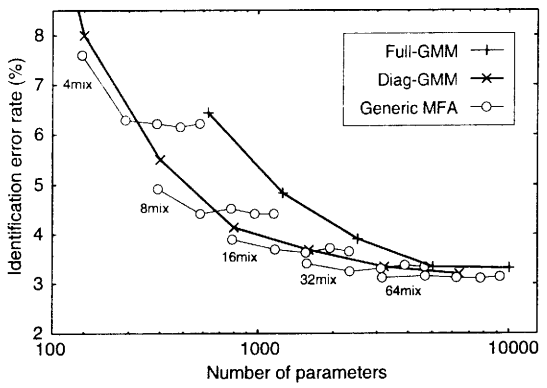

図 3　generic MFA と diag-GMM, full-GMM との識別誤り率の比較

Fig. 3　Comparison between generic MFA and conventional GMMs with diagonal or full covariance matrices.

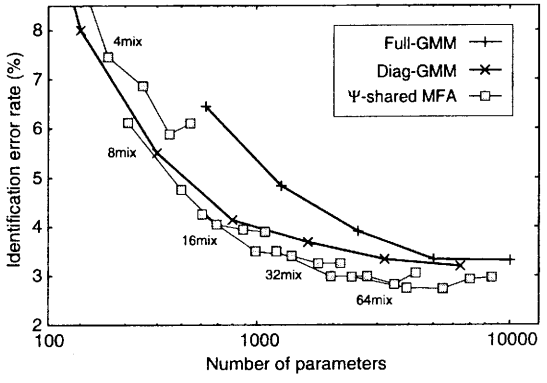

図 4　Ψ-shared MFA と diag-GMM, full-GMM との識別誤り率の比較

Fig. 4　Comparison between Ψ-shared MFA and conventional GMMs with diagonal or full covariance matrices.

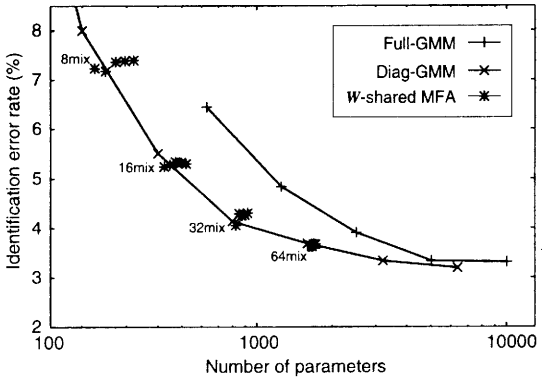

図 5　$\boldsymbol{W}$-shared MFA と diag-GMM, full-GMM との識別誤り率の比較

Fig. 5　Comparison between $\boldsymbol{W}$-shared MFA and conventional GMMs with diagonal or full covariance matrices.

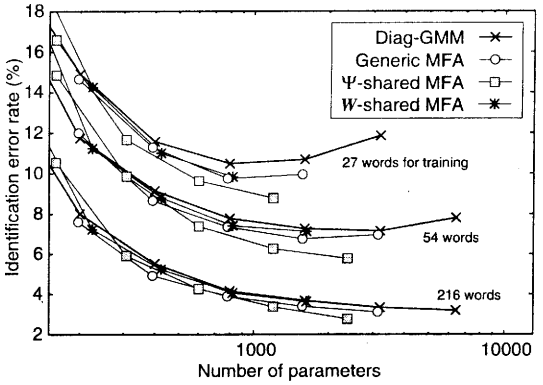

図 6　学習データとして 27 単語（上），54 単語（真中），216 単語（下）を用いた場合の 3 手法の MFA と diag-GMM の混合数の増加における識別誤り率の比較（因子数：q=2）

Fig. 6　Comparison between diag-GMM and three kinds MFA ($q = 2$) with increasing the number of mixtures, using 27 words (upper), 54 words (middle), 216 words (lower) for training.

GMM や generic MFA に対して大きな改善が見られることが分かる．また，図 3 の generic MFA と異なり，因子数に関らず高い認識性能が得られていることが分かる．64 混合において，diag-GMM に対して，19%($q = 2$), 26%($q = 6$) の識別誤り率の改善が得られた．図 5 は，$\boldsymbol{W}$-shared MFA の結果を示している．$\boldsymbol{W}$-shared MFA の識別性能は diag-GMM とほぼ等しい結果となった．これは $\boldsymbol{W}$ を共有することにより，モデル構造が diag-GMM によく似たものとなり，3 つの MFA の中で，パラメータの自由度が最も低くなるためであると考えられる．

図 6 には，因子数が 2 の場合の 3 つの手法の MFA と diag-GMM とで，学習データ数を 27, 54, 216 単語と変えた場合の結果を示す．因子数が 2 の場合でも十分高い認識性能が得られていることが分かる．また，学習データが少ない場合においても，3 つの MFA はすべて，diag-GMM より優れた認識性能を得ているとことが分かる．共有構造の違いで見ると，学習データを変えても，ほぼ同様の傾向が見られ，Ψ-shared MFA が最も良い結果となっていることが分かる．また，学習データが少ない場合においては，$\boldsymbol{W}$-shared MFA の識別性能が generic MFA の識別性能とほぼ等しくなり，各モデルパラメータの低い信頼性をパラメータの共有という手段によって補えているといえる．

さらに，それぞれの共有構造を用いた MFA に対して MCE 学習を適用した．それぞれ因子数が 6 の場合の結果を図 7–9 に示す．すべての共有構造において，特に混合数の少ないモデルにおいて大きな改善が得られたが，全体として見ても識別性能が改善されていることが分かる．因子数を変えた場合にもほぼ同様の傾向が見られた．ただし，$\boldsymbol{W}$-shared MFA では MCE 適用による改善が小さいことが分かる．これは学習係数の与え方が影響している可能性もあり，技術者による経験的な係数の設定が求めら れるという欠点もある．しかしながら，Ψ-shared

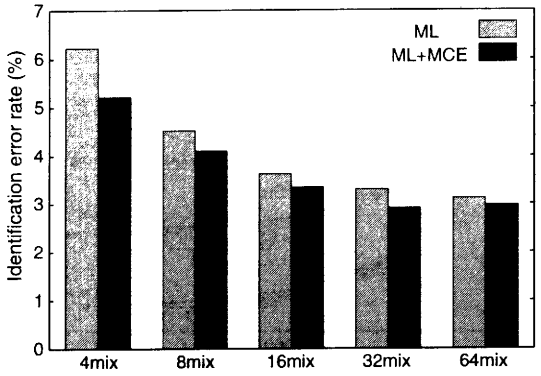

図 7 gneric MFA の MCE 学習適用による識別誤り率の改善（因子数：q=6）

Fig. 7 Comparison of generic MFA before and after MCE training ($q = 6$).

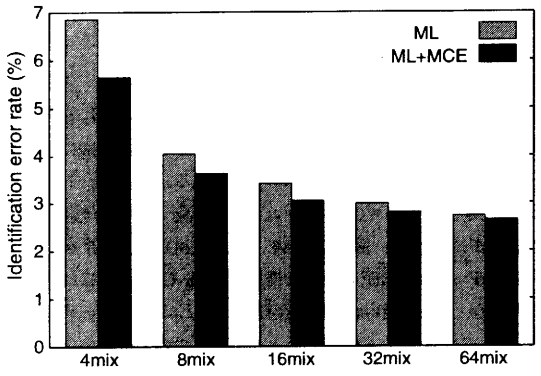

図 8 Ψ-shared MFA の MCE 学習適用による識別誤り率の改善（因子数：q=6）

Fig. 8 Comparison of Ψ-shared MFA before and after MCE training ($q = 6$).

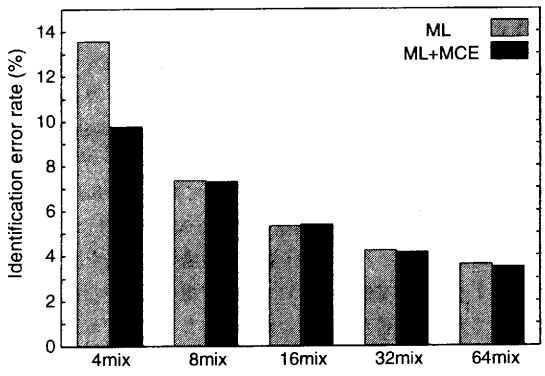

図 9 $\boldsymbol{W}$-shared MFA の MCE 学習適用による識別誤り率の改善（因子数：q=6）

Fig. 9 Comparison of $\boldsymbol{W}$-shared MFA before and after MCE training ($q = 6$).

MFA の 4 混合においては MCE 適用前から 17%の改善がみられ，また，ML 推定で最も誤り率が低かった 64 混合 6 因子においては誤り率 2.73% から 2.65%となり，3%の改善率が得られた．これにより，MFA に基づく話者識別モデルに対する最小分類誤り学習が有効であると言える．

6. むすび

本研究では，MFA に基づく話者識別モデルに対して ML 及び MCE 学習を行い，MFA の共有構造に関する検討を行った．実験より，独自因子の分散 Ψ を共有した場合が最もよい結果となり，対角共分散行列を用いた GMM に対して 26%の識別改善が得られた．また，MFA に基づく話者モデルへの MCE 学習の適用により認識性能のさらなる向上が得られた．

今後の課題として，他の様々な共有構造を持った MFA モデルの話者識別における評価 [6]，また変分ベイズアプローチによる混合数や因子数の自動決定 [7] などが挙げられる．

7. 謝　辞

本研究の一部，中部電力基礎技術研究所研究助成，及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] Z. Ghahramani and G. E. Hinton, "The EM algorithm for mixtures of factor analyzers," *Tech. Rep. Univ. of Toronto*, CRGTR-96-1, May 1996.

[3] L. K. Saul and M. G. Rahim, "Maximum likelihood and minimum classification error factor analysis for automatic speech recognition," *IEEE Trans. on Speech and Audio Processing*, vol.8, no.2, pp.115–125, Mar. 2000.

[4] P. Ding, Y. Liu, and B. Xu, "Factor analyzed Gaussian mixture models for speaker identification," *Proc. of ICSLP-2002*, pp.1341–1344, Sept. 2002.

[5] B.-H. Juang and S Katagiri, "Discriminative learning for minimum error classification," *IEEE Trans. Signal Process.*, vol.40, no.12, pp.3043–3054, Dec. 1992.

[6] A-V. I.Rosti and M.J.F. Gales, "Generalised linear Gaussian models," *Tech. Rep. Cambridge Univ.*, CUED/F-INFENG/TR.420, Nov. 2001.

[7] Z. Ghahramani and M.J. Beal, "Variational inference for Bayesian mixtures of factor analysers," *Neural Information Processing Systems 12*, 1999.