

話者照合における HMM の局所的音響情報に基づく尤度正規化

磯部俊洋 高橋淳一

株式会社 NTT データ 情報科学研究所

〒210-0913 神奈川県川崎市幸区堀川町 66-2 興和川崎西口ビル 9 階

E-mail: isobe@lit.rd.nttdata.co.jp

あらまし HMM に基づく話者照合における新しい尤度正規化法を提案する。提案手法ではコホートモデルを HMM の音素、状態、分布が表す局所的な音響情報に基づいて選択される近傍話者の HMM から合成する。HMM の細部の音響情報を考慮した話者類似性に基づいてコホートモデルを生成することにより、音響空間において本人話者との統計的モデル間距離を小さくすることが可能である。これにより、入力音声の尤度の変動を統計的に吸収できる効果的な正規化処理を実現できる。100 人の電話音声データを用いた話者照合実験において、分布に基づくコホートモデルによる尤度正規化(提案手法)と、従来の話者別にコホートモデルを選択する方法を比較した結果、Equal Error Rate(EER)は 1.30% から 0.70% まで低減し、本手法の有効性を確認した(本人話者 25 人、詐称者 75 人の Open テスト、コホートサイズ 5)。

キーワード 話者照合、尤度正規化、局所的音響情報

A New Normalization Method Using Local Acoustic Information of HMMs for Speaker Verification

Toshihiro Isobe and Jun-ichi Takahashi

Laboratory for Information Technology, NTT DATA CORPORATION

66-2, Horikawa-cho, Saiwai-ku, Kawasaki-shi,

Kanagawa 210-0913 Japan

E-mail: isobe@lit.rd.nttdata.co.jp

Abstract A new cohort normalization method is described for HMM based speaker verification. In the proposed method, cohort models are synthesized on the basis of the similarity of local acoustic features between speakers. The similarity can be determined using acoustic information lying in model components such as phonemes, states, and the Gaussian distributions of HMMs. With the method, the synthesized models can provide an effective normalizing score for various observed measurements because the difference between the individual reference model and the synthesized cohort models is statistically reduced through fine evaluation of acoustic similarity in model structure level. In the experiments using telephone speech of 100 speakers, the Equal Error Rate (EER) can be drastically reduced from 1.30 % (obtained by the conventional speaker-selection based cohort normalization) to 0.70 % (obtained by the proposed method on distribution-based selection) in open test (reference speaker: 25, impostor: 75), in which the other speakers than the reference speaker were used as impostors.

key words Speaker verification, Score normalization, Local acoustic information

1 はじめに

話者照合では、入力音声に対する本人の音響モデルの照合スコアと事前に設定する閾値との大小関係により、受理/棄却の判定を行う。HMM を音響モデルとする場合、照合スコアは入力音声の HMM に対する尤度である。このため、発話内容の違いや音声入力系の影響による尤度変動が照合性能を大きく低下させることが問題となっている。

尤度の影響を低減するために、これまでにいくつかの正規化手法が研究されてきた。多くの手法が各本人話者に対して背景話者を設定し、本人話者の尤度を背景話者の尤度で正規化する方法を基本としている。Higgins 等の方法では、入力音声に対して最大尤度を持つ話者を背景話者としている[1]。Rosenberg 等は、背景話者として、事前に各本人話者に対して設定するコホート話者を用いる方法を提案している[2]。また、Liu 等は多数話者モデルの尤度の平均値と分散値に基づく正規化法を[3]、松井等は、事後確率に基づき、本人話者を含む多数話者モデルの尤度による正規化方法を提案している[4]。いずれの手法も、背景話者の効率的な表現法を主題としており、本人話者との類似性を、入力音声に対する尤度の大きさや話者の音響モデル間の類似度に基づいて評価し、背景話者を設定する。しかし、入力音声全体の尤度や話者モデルを単位として類似度を評価しているため、本人話者と背景話者間の類似性は発話内容の違いによって影響を受けやすい。

本報告では、入力音声の局所において話者間の音響的類似性が異なることに着目し、音素、HMM の状態、分布といった局所的な音響情報に基づくコホート話者の選択法とそれを用いた背景話者の生成方法[5]を提案する。提案手法の特徴を述べると共に、実験による本手法の効果を示す。

2 話者照合における尤度正規化

話者照合において尤度比に基づく話者 I の正規化スコアは、(1)式で表される。

$$L(I|\mathbf{o}) = \log p(\mathbf{o}|\lambda^{(s=I)}) - \log p(\mathbf{o}|\lambda^{(s\neq I)}) \quad (1)$$

$L(I|\mathbf{o})$: 正規化スコア

\mathbf{o} : 音声特徴ベクトル(観測シンボル列)

$\lambda^{(s=I)}$: 話者 I の HMM セット

$\lambda^{(s\neq I)}$: 話者 I 以外の HMM セット

一般に、(1)式の右辺第2項の $\lambda^{(s\neq I)}$ としては、各本人話者 I に対して設定される背景話者モデルが使用される。背景話者モデルは、その生成方法によって大きく2つに分類される。1つは、新規ユーザを登録する場合、システムが保有する既存の話者モデルから新規ユーザの背景話者モデルを選択する方法である。選択された背景話者はコホート話者と呼ばれ、コホート話者による尤度正規化では、(1)式の右辺第2項は(2)式のように表される。

$$((1)式右辺第2項) = \log \left\{ \frac{1}{K} \sum_{k=1}^K p(\mathbf{o}|\lambda^{(c_k(I))}) \right\} \quad (2)$$

K : コホートサイズ

$c_k(I)$: 話者 I の k 番目のコホート話者

各本人話者に対するコホート話者は、システムが保有する話者の中から、音響空間上において本人話者の近傍に位置する話者を選択される。話者照合時には、選択された各コホート話者のモデルが出力する尤度の平均値を用いて尤度正規化を行う。

もう一方の方法では、背景話者モデルは複数の近傍話者の音声データを用いて学習される(以後ブールドモデルと呼ぶ)。Liu、松井等の方法がその代表例である。得られる背景話者モデルは、近傍話者の音響的特徴を正確に表現できる特徴を持つが、新規ユーザの登録の都度、背景話者モデルを再学習する必要がある。

いずれの方法においても、背景話者は本人話者の近傍話者とするが、正規化に用いる尤度計算において、複数の近傍話者の特徴を各々の尤度の和として表現するか、音響空間上で学習したモデルの尤度として表現するかの違いがある。

しかし、いずれの手法も近傍話者を話者単位で選択するため、発話内容によって話者間距離が変動すると考えられ、変動に強い近傍話者を選択しているとは言い難い。Rosenberg 等は、連続数字を用いた話者照合において、数字モデル毎にコホート話者を選択する方法及びブールドモデルを学習する方法の評価を報告しているが[6]、話者間距離は音韻によって異なるため、同様の問題があると考えられる。

3 局所的音響情報に基づくコホート生成法

各話者の音響モデルの局所的な特徴に着目して背景話者を選択し、統計的に正確な近傍話者を設定する手法を提案する。これは、話者間の類似性が入力音声において局所的に異なり得る可能性に着目したものである。局所的な部分としては、音素、HMM の状態、分布が考えられ、それぞれに対して近傍話者を求めて仮想的なコホート話者を生成する。従来方法での背景話者の設計手法——本人の尤度値に最も近い、あるいは、本人の音響モデルに最も類似した他の(複数の)話者を選択——と比較すると、提案手法では、背景話者の設定自由度が増加し、統計的に正確な近傍話者を設定することが可能となる。従って、尤度の変動に頑健な正規化スコアを出力することができる。

提案手法におけるコホートモデル生成の概念図を図1に示す。簡単化のため、ここではそれぞれの話者 (I, A, B, C, D) を1状態3混合分布の HMM で表す。従来の話者を単位とするコホートモデルでは、本人話者 I に対して、近傍話者 A, B, C, D から1人、または、複数の話者のモデルがコホートモデルとして選択される。一方、提案手法における分布を

単位としたコホートモデルでは、仮想的に生成される話者 V のモデルをコホートモデルとする。このコホートモデルは、話者 A の分布 a_3 、話者 C の分布 c_3 、話者 D の分布 d_1 から合成される。それぞれの分布は話者 I のモデルの各分布 I_1, I_2, I_3 に対して最近傍分布となっており、話者の枠を超えて選択される。従って、話者 V のモデルは、詳細な音響情報をもとに生成された、話者 I のモデルの近傍モデルと見なすことができる。本モデルを背景話者モデルとして用いることで、尤度変動に頑健且つ安定した正規化スコアが得られると考えられる。

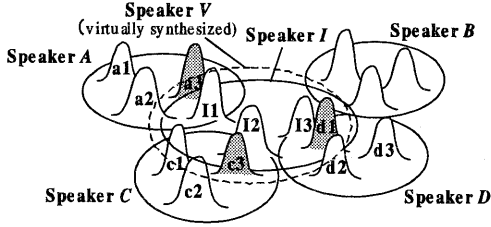


図1. 分布を単位としたコホート生成の概念図

4 提案手法の定式化

4.1 音素を単位とするコホート生成

音素を単位としてコホートモデルを生成する場合は、(1)式の右辺第2項は(3)式で表される。

$$\left\{ \frac{1}{K} \sum_{k=1}^K p(\mathbf{o} | \lambda^{(\hat{c}_k(i))}) \right\} \quad (3)$$

$\hat{c}_k(i)$: 話者 i の k 番目の仮想コホート話者

$\lambda^{(\hat{c}_k(i))}$: 仮想コホート話者 $\hat{c}_k(i)$ の HMM セット

話者 i の HMM セットを

$$\lambda^{(i)} = \{\lambda_1^{(i)}, \lambda_2^{(i)}, \dots, \lambda_P^{(i)}\} \quad (4)$$

$\lambda_p^{(i)}$: 話者 i の p 番目の音素 HMM

P : 音素数

とすると、仮想コホート話者の HMM セットは、(5)式で表される。

$$\lambda^{(\hat{c}_k(i))} = \{\lambda_1^{(\hat{c}_k(i,1))}, \lambda_2^{(\hat{c}_k(i,2))}, \dots, \lambda_P^{(\hat{c}_k(i,P))}\} \quad (5)$$

$c_k(i, p)$: 話者 i の p 番目の音素に対する k 番目のコホート話者

$c_k(i, p)$ は音素を単位としたコホート話者であり、話者 i と他の登録話者の同じ音素に対する音素モデル間距離に基づいて、話者 i に対する近傍話者上位 K 人を選択する。

4.2 状態を単位とするコホート生成

音素 HMM を状態数 S 、混合数 M の left-to-right 型の混合連続型 HMM とすると、 p 番目の音素 HMM λ_p は状態 s における自己ループ遷移確率 $a_{p,s,s}$ 、隣接する状態への遷移確率 $a_{p,s,s+1}$ 、分布重み係数 $w_{p,s,m}$ ($m=1, 2, \dots, M$)、及びガウス分布 $N_{p,s,m}$ ($m=1, 2, \dots, M$) をパラメタセットとして表現できる。状態を単位としたコホート生成では、状態 s におけるパラメタセットを単位として話者 i に対する近傍話者を選択する。仮想コホート話者の HMM セットを(6)式に示す。 $c_k(i, p, s)$ は状態を単位としたコホート話者であり、話者 i と登録話者において同一音素、同一状態での状態間距離に基づいて近傍話者上位 K 人を選択する。

$$\lambda^{(\hat{c}_k(i))} = \{a_{p,s,s}^{(\hat{c}_k(i,p,s))}, a_{p,s,s+1}^{(\hat{c}_k(i,p,s))}, w_{p,s,m}^{(\hat{c}_k(i,p,s))}, N_{p,s,m}^{(\hat{c}_k(i,p,s))}\}_{p=1,2,\dots,P; s=1,2,\dots,S; m=1,2,\dots,M} \quad (6)$$

$c_k(i, p, s)$: 話者 i の p 番目の音素 HMM の状態 s に対する k 番目のコホート話者

$a_{p,s,s}^{(\hat{c}_k(i,p,s))}$: 話者 $c_k(i, p, s)$ の p 番目の音素 HMM の状態 s から s への遷移確率

$w_{p,s,m}^{(\hat{c}_k(i,p,s))}$: 話者 $c_k(i, p, s)$ の p 番目の音素 HMM の状態 s のガウス分布 m の重み係数

$N_{p,s,m}^{(\hat{c}_k(i,p,s))}$: 話者 $c_k(i, p, s)$ の p 番目の音素 HMM の状態 s のガウス分布 m

4.3 分布を単位とするコホート生成

分布を単位として生成するコホートモデルは(7)式で表される。

$$\lambda^{(\hat{c}_k(i))} = \{a_{p,s,s}^{(\hat{c}_k(i))}, a_{p,s,s+1}^{(\hat{c}_k(i))}, w_{p,s,m}^{(\hat{c}_k(i))}, N_{p,s,m}^{(\hat{c}_k(i,p,s,m))}\}_{p=1,2,\dots,P; s=1,2,\dots,S; m=1,2,\dots,M} \quad (7)$$

$$a_{p,s,s}^{(\hat{c}_k(i))} = \frac{\sum_m a_{p,s,s}^{(c_k(i,p,s,m))}}{\sum_{j=0,1}^m \sum_m a_{p,s,s+j}^{(c_k(i,p,s,m))}} \quad (8)$$

$$w_{p,s,m}^{(\hat{c}_k(i))} = w_{p,s,m}^{(I)} \quad (9)$$

$c_k(i, p, s, m)$: 話者 i の p 番目の音素 HMM の状態 s の分布 m に対する k 番目のコホート話者

$w_{p,s,m}^{(I)}$: 本人話者 I の p 番目の音素 HMM の状態 s 、分布 m の重み係数

コホートモデルの各状態の分布は、本人話者 i の各分布に対して他の話者の同じ状態に含まれる分布群から近傍分布として選択する。この時、選択された分布に対する遷移確率も考慮し、各状態において各々の分布に対する遷移確率が均一であると仮定して、遷移確率についても合成する(8式)。重み係数 w については、本人話者モデルの

重み係数を使用する。 $c_k(i, p, s, m)$ は分布に基づいたコホート話者であり、分布毎に選択された近傍話者である。

4.4 音響モデルの距離尺度

提案手法では、音響モデルにおける局所的な音響情報を表すコンポーネント(音素、HMMの状態、分布)を単位として本人話者モデルに対する近傍のコンポーネントを選択する必要がある。分布を単位とする場合には、(10)式に示すように、単一ガウス分布 N_p, N_q 間の Battacharyya 距離を用いた。 μ_p, Σ_p は、それぞれガウス分布 N_p の平均ベクトルと分散行列を表わす。

$$d(N_p, N_q) = \frac{1}{8} (\mu_p - \mu_q)^T \left(\frac{\Sigma_p + \Sigma_q}{2} \right)^{-1} (\mu_p - \mu_q) + \frac{1}{2} \ln \frac{|\Sigma_p + \Sigma_q|/2}{|\Sigma_p|^{1/2} |\Sigma_q|^{1/2}} \quad (10)$$

状態を単位とする場合には、各状態において混合分布から代表分布を生成し、代表分布間での距離を評価した。音素を単位とする場合には、対応する音素 HMM 間における各状態の代表分布間距離の平均値を用いた。また、話者間距離は、各話者の全ての音素モデル間距離の平均値とした。

5 評価実験

5.1 実験条件

音声データは、防音室内でヘッドセットマイク(SHURE SM10)を用いて収録し、疑似口、電話機、公衆回線網を経由し、テレホニーボード(Dialogic DH1.x)で再収録して電話音声に変換したものを使用した。発声内容は4桁連続数字である。実験に用いた音声データの諸元を表1に、分析条件を表2に示す。

表1. 評価用音声データ

セット	話者数	発声数/人	時期差
R	25(男 12, 女 13)	30	無し
E ₀	25(Aと同一)	40	無し
E ₃	25(Aと同一)	40	3ヶ月
E ₆	25(Aと同一)	40	6ヶ月
X	75(男 38, 女 37)	40	無し

(但し、R, E_i(i=0,3,6)は話者同一。RとE_iでは発声内容が異なる。XとE_iの発声内容は同じ。)

表2. 分析条件

量子化	16bit, 8kHz サンプリング
分析窓(窓長)	ハミング窓 (25 msec.)
フレームシフト	10 msec.
プレエンパシス	H(z)=1-0.95z ⁻¹
分析	LPC 分析(12次)
特徴ベクトル	LCP cep(12次)+ Δcep(12次)+ Δpower(1次)

音響モデルは、コンテキスト独立の right-to-left 型音素 HMM であり、状態数3、混合数3の混合連続型とした。話者登録では、最尤推定法を用いて HMM の全てのパラメタを求めた。

5.2 コホート生成単位の違いによる性能比較

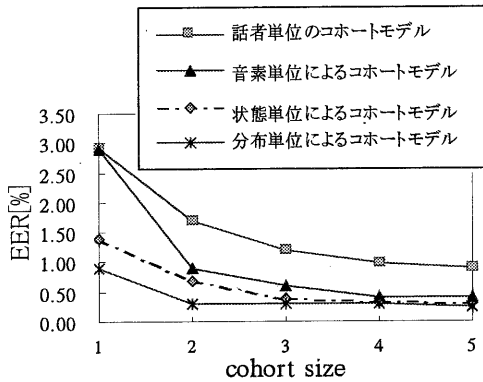
話者照合実験はテキスト指定型で行った。実験では、音素、状態、分布のそれぞれを単位として生成するコホートモデルによる EER と、従来の話者を単位とするコホートモデルによる EER を、表3に示す6種類の項目において比較した。受理/棄却の閾値は全話者共通とし、事後的に設定した。コホートモデルは本人以外の登録話者のモデルから生成した。ここで Closed 実験とは、本人以外の登録話者を詐称者とみなし、詐称者の音響的特徴をコホートモデルに反映する条件での実験である。一方、Open 実験は、詐称者は登録話者以外の話者とし、実際の話者照合システムの使用条件を想定した実験である。

表3. 実験項目

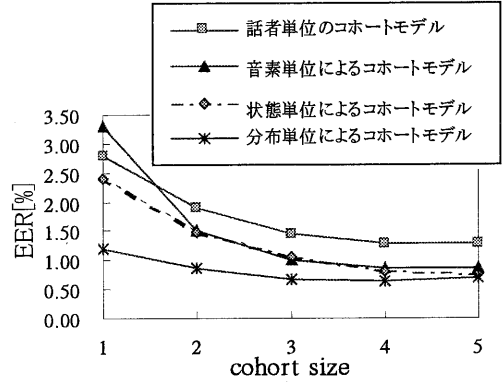
	項目	使用データセット
(a)	Closed 実験 (時期差無し)	登録: R 照合: E ₀
(b)	Closed 実験 (時期差3ヶ月)	登録: R 照合: E ₃
(c)	Closed 実験 (時期差6ヶ月)	登録: R 照合: E ₆
(d)	Open 実験 (時期差無し)	登録: R 照合: X+E ₀ (本人のみ)
(e)	Open 実験 (時期差3ヶ月)	登録: A 照合: X+E ₃ (本人のみ)
(f)	Open 実験 (時期差6ヶ月)	登録: A 照合: X+E ₆ (本人のみ)

話者、音素、HMMの状態、分布のそれぞれのコンポーネントを単位として生成したコホートモデルによる実験結果を図2(a)~(c)(Closed 実験)、図3(d)~(f)(Open 実験)に示す。図中、横軸はコホートサイズを表し、コホート話者数、または仮想的に生成したコホート話者数である。

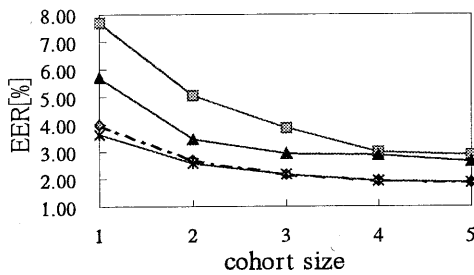
Closed 実験では、全ての場合において、コホートサイズが増加するにつれて EER が低減し、コホートサイズが約4の時に飽和する傾向を得た。また、コホートモデルの生成単位を音素、状態、分布と詳細化するほど、性能が向上する傾向がみられた。分布を単位として生成するコホートモデルを従来の話者単位のコホートモデルと比較すると、時期差無し、コホートサイズ=5の場合、EER の削減率は70.8%(0.89→0.26)であり、提案手法の大きな効果がみられた。時期差6ヶ月の実験においても、前記の削減率は24.4%(4.50→3.40)であり、時期差ありの場合においてもその有効性が認められた。



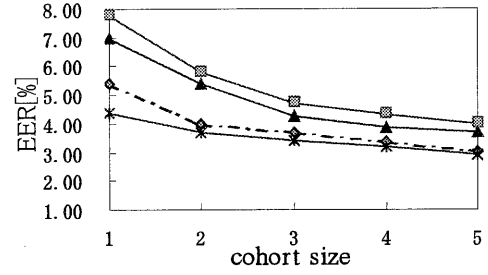
a) 時期差無し



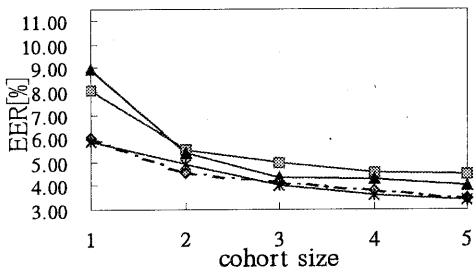
d) 時期差無し



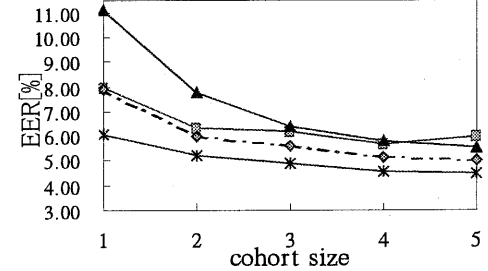
b) 時期差3ヶ月



e) 時期差3ヶ月



c) 時期差6ヶ月



f) 時期差6ヶ月

図2. Closed 実験結果

図3. Open 実験結果

Open 実験においても、Closed 実験と同様の傾向がみられた。コホートサイズ=5の場合、分布を単位とするコホートモデルと話者単位のコホートモデルとの比較では、EER の削減率は、時期差無しで 46.2%(1.30→0.70)、時期差6ヶ月で 25.0%(6.00→4.50)であり、提案手法の有効性が確認された。

5.3 コホートモデルとプールドモデルの比較

第2節で述べたように、尤度正規化用の背景話者モデルの生成は、コホートモデルによる方法と、プールドモデルによる方法が代表的である。実験項目(a)(時期差無し)の Closed 実験)及び(d)(時期差無し)の Open 実験)において、提案手法によるコホートモデルとプールドモデルによる尤

度正規化の比較を行った。閾値は全話者共通とし、事後的に設定した。コホートモデルのコホートサイズは3とし、プールドモデルは、各本人話者に対して話者単位で選択した近傍話者3人の登録用音声データから学習した。各話者のモデルの混合数を3とし、プールドモデルの混合数は12とした。比較結果を表4に示す。(a)、(d)とも、プールドモデルによる正規化法は、話者別のコホートモデルに比べて性能は低かった。

表4. コホートモデル、プールドモデルによる EER

	話者別 コホート	分布別 コホート	プールド モデル
(a)	1.20 %	0.30%	3.00%
(d)	1.44%	0.67%	3.00%

6 考察

コホートモデルを生成するコンポーネントと性能の関係について考察する。実験結果からわかるように、提案手法ではコンポーネントを話者、音素、HMM の状態、分布と詳細化するにつれて、性能を向上できる。これは、コンポーネントが表す音声の局所部分に着目して音響モデルの類似性(音響空間上での距離)を評価して、近傍のコンポーネントを選択でき、詳細なレベルでコホートモデルの生成を制御できるからである。得られた背景話者モデルは統計的に本人話者モデルの最近傍となり、正規化における尤度変動を吸収する働きを考えると考えられる。

各コホートモデルの性能とコホートサイズの関係について考察する。各コホートモデルとも、コホートサイズが増加するにつれて性能が向上するが、これは、近傍コンポーネント数の増加により、コホートモデルの表現がより正確になるためと考えられる。しかし、コホートサイズをさらに増加してもモデルの表現能力はほとんど変わらないため、性能は飽和すると考えられる。また、本実験から、コンポーネントの詳細化と大きなコホートサイズの併用が、コンポーネントの詳細化効果を打ち消すことも明らかになった。これは、コホートサイズが増加することにより、どのような単位でも音響モデルの表現能力が飽和することが原因と思われる。従って、コンポーネントの詳細化は、コホートサイズが小さいところで大きな効果が期待できる。

話者照合における閾値設定のロバスト性について考察する。実験では受理/棄却の閾値は全話者共通とし、事後的に設定したが、実アプリケーションでは、事前に設定する必要がある。話者全体で最適な閾値を共有できれば、新規ユーザに対してもその閾値を使用でき、システム管理を効率的に行える。このような見地から提案方法の有効性を検証するため、EER を求める際の閾値の分散を調査した。これは、閾値の分散が小さければ、話者間で閾値の共通化ができるからである。表5に実験項目(d)において、話者、音素、HMM の状態、分布を単位として生成したコホートモデルによって求めた EER の閾値の平均と分散を示す。詳細なコンポーネントから生成するコホートモデルの方が閾値の分散が小さくなっており、提案手法は閾値の共通化についてロバスト性を向上する効果を持つと考えられる。

表5. 話者個別 EER 基準による閾値の平均・分散

コンポーネント	話者	音素	状態	分布
平均値	29.9	28.0	27.6	10.4
分散	133.5	125.2	108.9	72.6

コホートモデルとプールドモデルの性能比較においては、プールドモデルは従来の話者を単位とするコホートモデルほどの性能を得られなかった。これは、プールドモデルの方法は学習データの発声内容への依存性が強く、テ

キスト固定型には適するが([6]参照)、音韻の任意性が必要なテキスト指定型への適用が困難なためと考えられる。

7 まとめ

HMM の局所的な音響情報に基づいてコホートモデルを設計する方法を提案した。提案手法は、局所的な音響情報として、音素、HMM の状態、分布といったコンポーネント毎に近傍コンポーネントを選択し、コホートモデルを生成するものである。従来の話者を単位として設定するコホートモデルの場合に比べて、話者の特徴を微細なレベルで制御できるため、本人話者モデルに統計的に近接したコホートモデルを生成することができる。したがって、生成されたコホートモデルは尤度の変動を吸収できる正規化処理を可能とする。提案手法により、背景話者をより効率的に表現でき、小さいコホートサイズで照合性能が向上できることを実験的に示した。時期差無し Open 実験では、サイズ5の場合において、分布を単位とする提案方法は話者を単位とする従来方法に比べて、高い EER 誤り削減率 46.2%を達成した。また、話者を単位とする従来方法でのサイズ5による性能は、本手法ではより小さいサイズ1で実現できることを示した。

8 参考文献

- [1] A. Higgins, L. Bahler, and J. Porter, "Speaker Verification Using Randomized Phrase Prompting," *Digital Signal Processing*, vol. 1, pp.89-106, 1991.
- [2] A. E. Rosenberg, J. Delong, C-H. Lee, B-H. Juang, and F.K. Soong, "The Use of Cohort Normalized Scores for Speaker Verification," *Proc. ICSLP 92*, vol. 1, pp.599-602, 1992.
- [3] C-S. Liu, H-C. Wang, and C-H. Lee, "Speaker Verification Using Normalized Log-Likelihood Score," *IEEE Transactions on Speech and Audio Processing*, vol. 4, No. 1, pp.56-60, 1996.
- [4] T. Matsui and S. Furui, "Concatenated Phoneme Models for Text-Variable Speaker Recognition," *Proc. ICASSP 93*, vol. 2, pp.391-394, 1993.
- [5] 磯部, 高橋, "HMM の局所的音響情報を用いたコホート生成に基づく話者照合", 音講論集, pp.115-116(1998-9)
- [6] A. E. Rosenberg S. Parthasarathy, "Speaker Background Models for Connected Digit Password Speaker Verification," *Proc. ICSLP 96*, vol. 1, pp.81-84, 1996.