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

2007－ＳＬＰ－69（40）
2007／12／21

音素クラスHMMを使用した話者ベクトルに基づく話者識別法の検討

赤津 達也† 加藤 正治† 小坂 哲夫† 好田 正紀†

† 山形大学大学院理工学研究科
〒992-8510 山形県米沢市城南 4-3-16
E-mail: tatsuya17akatsu@yahoo.co.jp, katoh@yz.yamagata-u.ac.jp,
tkosaka@yz.yamagata-u.ac.jp, kohda@yz.yamagata-u.ac.jp

あらまし 本研究では，音素モデルを用いた話者ベクトルに基づくテキスト独立型話者識別について述べる．本話者識別システムはアンカーモデルに基づいており，識別対象話者の発声とアンカーモデル間の尤度からなる話者ベクトルによって，各々の話者が話者空間に配置されている．識別対象話者の音響モデルを必要としないという利点があり，1発話程度の極めて少量の登録用発声で話者識別が可能となる．欠点として従来法では識別性能が低いという問題点があったが，アンカーモデルに従来用いられている混合ガウス分布モデル (GMM) ではなく，音素 HMM を使用することで性能改善が得られている [1]．本研究では，音素をクラスタリングした音素クラス HMM を用いることで更なる性能の向上を図る．音素クラス HMM の対数尤度の計算には，音素認識器を使用する．30 名の日本語話者識別タスクで本手法の評価を行った．実験では，平均 5.5sec の極く短い発話を識別対象話者の登録用データとして使用した．結果として音素決定木に基づいてクラスタリングした 15 音素クラスの HMM を用いたとき，35 音素 HMM ベースのアンカーモデルと比較して 17.1%の相対的改善が得られた．

キーワード 話者認識，話者識別，隠れマルコフモデル (HMM)，混合ガウス分布モデル (GMM)，音素クラス HMM

An investigation on the speaker vector-based speaker identification method with phonetic-class HMMs

Tatsuya AKATSU†, Masaharu KATOH†, Tetsuo KOSAKA†, and Masaki KOHDA†

† Graduate School of Science and Engineering, Yamagata University
Jonan 4-3-16, Yonezawa-city, Yamagata, 992-8510 Japan
E-mail: tatsuya17akatsu@yahoo.co.jp, katoh@yz.yamagata-u.ac.jp,
tkosaka@yz.yamagata-u.ac.jp, kohda@yz.yamagata-u.ac.jp

Abstract This paper presents a phonetic based approach for speaker identification performed in text-independent mode. The identification system is based on the technique of anchor models, where the location of each speaker is represented by the speaker vector. The vector consists of the set of the likelihood between a target utterance and the anchor models. In order to improve the identification performance, phonetic-class HMMs are used instead of phoneme HMM scheme as anchor models. This approach utilizes a phonetic speech recognizer to calculate the log-likelihood with phonetic-class HMMs. The proposed method was evaluated on Japanese speaker identification task with 30 speakers. It showed that the proposed method achieved 17.1% relative improvement over the phoneme HMM-based system.

Key words Speaker recognition, speaker identification, hidden Markov model(HMM), Gaussian mixture model(GMM), phonetic-class HMM

1. はじめに

本稿ではアンカーモデルに基づいた，音素クラス HMM を使用する話者識別について述べる．アンカーモデルシステムはすでに話者インデキシングのために文献 [2] で提案されている．また，話者識別 [3] および話者照合 [4] で既に使用されている．アンカーモデルに基づく話者識別の基本的な考え方は，識別対象話者以外の多数の参照話者のモデル（アンカーモデル）を用い

て, 入力話者と多数話者の相対的位置関係を識別に用いるということである. この方法において, 各々の話者の特徴は話者ベクトルによって表される. 話者ベクトルは, 識別対象話者の発声と多数のアンカーモデル間の尤度から求められる. それは, 識別対象話者の発声の話者空間への写像と考えることが出来る. アンカーモデルは GMM や HMM などの確率モデルで表現されるが, アンカーモデルの話者セットは識別対象話者を含まない. 一般的話者識別手法では, 確率モデルを学習するため識別対象話者の音声が事前にある程度の量必要であった. 一方アンカーモデルに基づく手法では, 登録音声として 1 発話程度の極く少量の音声があれば識別可能である. これにより, ユーザがモデルを学習するために繰り返し発声を行う手間を省くことができる. よって, 初回の音声登録時も経時変化に対応するための再登録時も, ユーザに対する負担が少ない.

しかし従来のアンカーモデルに基づく方法では, 話者識別の性能が不十分という問題があった. 例えば[3] では, アンカーモデルに 16 混合の GMM を利用し, 次元数が最高で 500, 識別対象話者数 50 名よる話者識別タスクで, 識別率 76.6%と報告されている. この方法では音素コンテキストの情報を無視している. また, 音素クラスの種類によっては高い話者性を有している可能性も考えられる. そこで, アンカーモデルに GMM ではなく音素 HMM を使用したところ性能の改善が得られた. 3 状態 10 混合 HMM, 次元数 1000, 識別対象話者数 30 名で識別率が 94.21%まで向上している [1].

本研究の目標は, 音素をクラスタリングして音素クラス HMM を作成し, これをアンカーモデルとして使用することによって本識別法の性能を向上させることである. 文献[1] では音素として 35 種類の音素を設定しているが, このクラス分類が話者識別に最適であるかの検討は行っていない. 音素によっては高い話者性を示す母音や鼻音のようなクラスもある一方, 話者識別に悪影響を及ぼす音素が存在する可能性も考えられる. また, 音素のクラス数を細かく設定した場合, 音素でアライメントをとった際の精度が低下している可能性もある. そこで, クラスタリングを行い音素のクラス数を種々に設定し, 話者識別における最適なクラス数の検討を行った. 識別対象話者の発声とアンカーモデル間の尤度計算は, 音素対文法などの言語的拘束を用いた HMM ベース音素認識器により行われる. 本提案手法を評価するために, GMM ベースシステム, 音素 HMM システムと音素クラス HMM システムをアンカーモデルフレームワークで比較する.

以下では, 関連研究を紹介する. 近年, 音素ベースの話者識別手法がいくつか提案されている. Hebert らは, 音素クラスの木構造に基づいた話者照合法を提案した [6]. この論文では, 音素クラスに基づいたシステムは, 従来法である GMM アプローチよりも優れた性能が得られたことを示している. Park らは, 各話者について音素クラス GMM を使用する話者識別手法を提案した [7]. 識別対象話者のためにモデルが必要という点で, この 2 つの方法は本提案手法とは異なる. Andrews らは, 音響特徴ベクトルに基づいた方法のかわりに, 音素列にのみ基づいた話者認識システムを開発した [8]. この方法において, 評価話者モデルは音響モデルではなく n-phone 頻度数を使用して生成される点が本手法とは異なる.

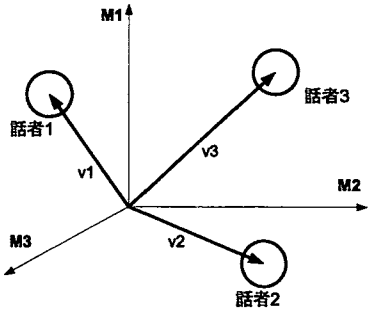

図 1 話者ベクトルの概念図
Fig. 1 A conception diagram of speaker vector.

本稿は以下のように構成される. 2 章では話者識別の方法について記述する. 3 章ではタスクおよびデータセットについて記述する. また, 実験条件についても示す. 4 章では実験結果および考察を記述する. 最後に, 5 章では結論および検討課題を示す.

2. 音素モデルを用いた話者識別

2.1 アンカーモデルを用いた話者空間の構成

アンカーモデルに基づいたシステムでは, 識別対象話者の音声はアンカーモデルの尤度による話者ベクトルによって特徴付けられる. j 番目の発話の話者ベクトル V_j は以下のように求められる.

$$V_j = \begin{bmatrix} \frac{p(s_j|M_1)-\mu_j}{\sigma_j} \\ \frac{p(s_j|M_2)-\mu_j}{\sigma_j} \\ \vdots \\ \frac{p(s_j|M_N)-\mu_j}{\sigma_j} \end{bmatrix} \quad (1)$$

$$\mu_j = \frac{1}{N}\sum_{n=1}^{N} p(s_j|M_n) \quad (2)$$

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(p(s_j|M_n)-\mu_j)^2} \quad (3)$$

ここで s_j は j 番目の発話の入力特徴時系列全体を表わし, $p(s_j|M_n)$ はアンカーモデル M_n における s_j の対数尤度を表わす. ベクトルは発話間のスコア変動を抑えるために平均 0, 分散 1 に正規化される [9]. s_j を発声する識別対象話者はアンカーモデルとして利用されている N 人の話者には含まれない. 本手法では, 入力音声から話者ベクトル V_j を生成し, 識別対象話者の登録音声の話者ベクトルとのユークリッド距離を計算して, 距離が最短のものを入力音声の話者であると識別する. 図 1 に話者ベクトル空間の 3 次元での概念図を示す. 各軸は式 (1) で求められた話者ベクトルの要素を示す. この図において, 登録話者と各軸を構成する話者は異なることに留意する必要がある.

登録話者の発話は上記の方法により話者ベクトル空間上に写像され，入力話者のベクトルと距離計算が行われる．

従来の GMM に基づく一般的な話者識別手法では，識別対象話者の話者モデルを作成する必要があり，学習用の発声が 10 文程度は必要であった．提案手法では識別対象話者のためにモデルを学習する必要がないので，登録ベクトル生成には 1 発声程度あれば良い．

2.2 アンカーモデルの音素表現

本手法では式 (1) で現れる尤度 $p(s_j|M_n)$ を計算するために，音素認識器を用いる．話者 n における対数尤度は，話者依存音素 HMM の認識器によって得られる．本研究では音素のベースの種類数は 35 とした．音素クラス HMM のクラス分けについては，2.3 節に詳述する．認識器は未知発話をデコードするので，テキスト独立型の話者識別が可能である．音素認識では，音素対文法などの言語的拘束を用いてデコードを行い，このときの音響尤度を話者ベクトルの計算に利用する．デコード時にビームから落ちる場合もあるが，この場合はベクトル中の最小尤度と置き換える操作を行う．

図 2 は，音素決定木によりクラスタリングした 15 音素クラスの HMM を用いた場合の，$N = 1000$ 次元で構成された話者ベクトルの一例を示している．x 軸は話者 (F.AIFU：女性) の 26 番目の入力発話の話者ベクトルの値を表し，y 軸は同じ話者あるいは異なる話者の 27 番目の発話の値を表す．散布図の各点が個々のアンカーモデルを表わす．この場合 1000 次元のため 1000 個の点がプロットされている．1000 個の点の x 座標の値が F.AIFU の 26 番目の入力発話から計算された話者ベクトルの各次元に対応する値そのものとなる．もし，x 軸に対応する入力発話と y 軸に対応する入力発話の，話者空間上の位置が完全に一致すれば，45 度の直線上に 1000 点がすべて乗り，両軸の相関は 1.0 となる．すなわち相関が高ければ両発話の位置が近く，低ければ遠いことを示す．上の図が同じ話者で発話内容が異なる場合，下の図は異なる話者 (M.YUIT：男性) である．発話内容が異なるが，話者が同じであれば 2 つの値のセットは高い相関を示す．一方異なる話者では相関が低い．これは，発話内容が異なっても，発声話者が同じなら，話者空間上で近い位置に写像されることを意味する．よってテキスト独立型話者識別が本手法で実行可能であることを示唆している．また図では男女のアンカーモデルを区別してプロットしているが，性別差のためその分布形状が大きく異なるのが分かる．上図は縦軸，横軸とも女声のため，女声のアンカーモデルが高い尤度を示すが，男声は低い尤度を示す．一方下図では，横軸が女声，縦軸が男声のため，男女のアンカーモデルの分布が二分化される傾向を示す．このように，男女の差は個人性に大きな影響を与える．

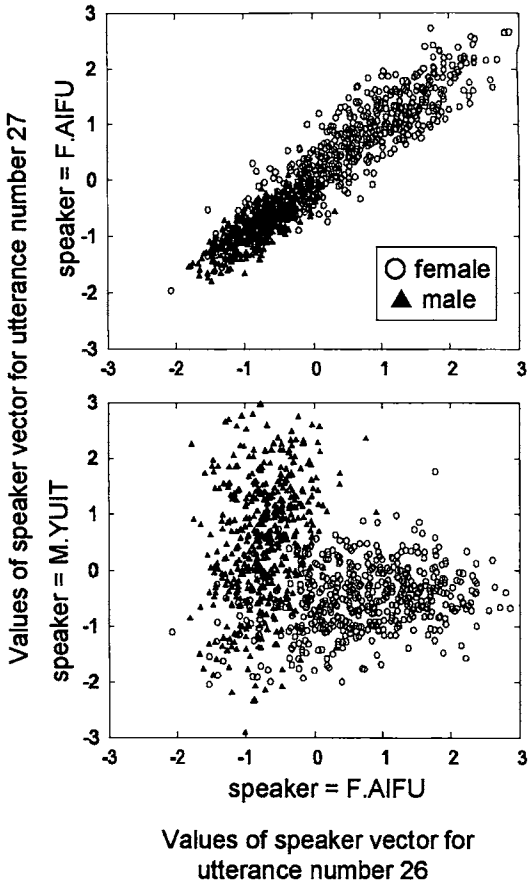

図 2 話者ベクトルの値の例
Fig. 2 Example of values of speaker vector.

2.3 音素クラス HMM

文献 [1] では，音素 HMM は音素 35 種類 {a,i,u,e,o,aa,ii,uu,ee,oo,ei,ou,w,xy,y,r,h,f,z,j,s,sh,ts,ch,p,t,k,b,d,g,m,n,N,cl,sil} について音響モデルを作成していた．本稿では 35 種類の音素をクラスタリングして半数以下のクラスに分けて音素クラス HMM とし，これを用いてアンカーモデルを作成する．本研究で検討した音素クラスは，音声学的知識を利用して人手により分類したクラスと，MLLR 適応などに用いられる音素決定木を利用して自動クラスタリングした場合の 2 種類である．音声学的知識による音素クラスとしては 15 クラスを設定した．音素決定木によるクラスタリングでは 10, 15, 17 の 3 種類について検討した．クラスタリングの詳細を表 1, 表 2 に示す．

2.4 音素構造 GMM

音素クラス HMM をアンカーモデルとした話者ベクトルとの比較のために，音素 HMM, GMM のほかに音素構造 GMM [10] を使用したアンカーモデルも扱う．音素構造 GMM は，モデルが 1 状態で表現される点は一般の GMM と同じであるが，ガウス分布が音素クラス別に学習される点が異なる．文献 [1] では，音素構造 GMM は音素 HMM と GMM の中間の性能を示している．音素構造 GMM は GMM と比較すると状態遷移構造は等しいが pdf が異なる．また音素 HMM と比較すると pdf は等しいが，状態遷移構造は異なる．このことは，pdf を音素ごとに学習すること，および HMM の構造を用いることが共に効果的であることを示唆している．

音素構造 GMM も本研究と同様に，音素コンテキストの情報を，話者識別の際積極的に利用するために提案された手法であ

表 1 音声学的知識による音素クラス
Table 1 Definition of phonetic-class based on knowledge.

有声破裂音	b,d,g	有声摩擦音	z,j
無声破裂音	p,k,t	無声摩擦音	s,sh,h,f
母音 1	a,aa	鼻音	N,n,m
母音 2	i,ii	半母音	w,y
母音 3	u,uu	流音	r
母音 4	e,ee,ei	拗音	xy
母音 5	o,oo,ou	無声破擦音	ch,ts
無音	sil,cl		

表 2 音素木に基づくクラスタリングによる音素クラス
Table 2 Definition of phonetic-class based on tree clustering.

10 クラス	15 クラス	17 クラス
a,aa	a,aa	a,aa
u,uu,ou	u	u
o,oo	uu,ou	uu,ou
i,ii,ei	o,oo	o,oo
e,ee	i,ii,ei	i,ii,ei
m,n,N	e,ee	e,ee
w	N	N
y,xy,r,z,j,b,d,g	m,n	m,n
p,t,k	w	w
ch,cl,f,h,s,sh,sil,ts	y,xy,r	y,xy,r
	z,j	z,j
	b,d,g	b,d,g
	p,t,k	p,t
	ch,f,h,s,sh,ts	k
	sil,cl	ch,f,s,sh,ts
		h
		sil,cl

る．文献[10]で，Faltlhauser らによって提案された手法では，まず各話者ごと個々の音素クラスの GMM を学習し，音素クラス別のガウス分布を得る．それらのガウス分布に対し混合重みを与え，1 状態に合成することにより GMM を作成する．本研究では，音素 HMM に含まれるガウス分布に混合重みを与えることにより，1 状態に合成して音素構造 GMM を作成し，比較対象とする．

3. 実験条件

評価には，スピーチコーパスとして ATR SDB-I を使用する[5]．このコーパスは多数話者の読み上げ音声と対話音声からなり，多数話者による音響的変動をカバーしている．アンカーモデルの学習用に，744 名の男性と 1,288 名の女性からなる計 2,032 名の話者によって発声される，音素バランス音声データを使用する．発話の総数は 51,131 である．各アンカーモデルの学習データは，ほぼ 25 発話であるが，話者によって若干のばらつきがある (23〜30 発話)．話者空間の最大次元数は $N = 2032$ となるが，文献[1]で次元数の検討を行ったところ，数百次元で性能は飽和している．そこで本稿では次元数 $N = 1000$ で検討を行う．このときのアンカーモデルの学習データ数は，男性 500 名，女性 500 名の計 1,000 名，発話総数 25,173 である．個々の話者の発声量が少ないため音素 HMM および音素クラス HMM の学習には，MAP 推定を使用した．学習に当たっては，まず 2,032 名の全データを用い ML 推定で不特定話者 HMM を作成する．次にこのモデルを初期モデルとして 1,000 名の各話者のデータを用いて MAP 推定を行う．学習回数はいずれも 5 回である．また MAP 推定の初期パラメータと最尤推定値のバランスを表わす τ の値は 10.0 とした．評価データセットは音素バランス音声データのうち学習に用いられる 2,032 名とは異なる 30 名の話者 (男女各 15 名) からなり，それぞれ 25 の発声データを持っている．評価セットの平均発話長は約 5.5sec である．

表 3 分析条件
Table 3 Analysis Conditions.

フロントエンド	ETSI Advanced front-end (AFE-WI008) Blind Equalization なし
標本化	16kHz
量子化	16bit
周期	10msec
フレーム長	25msec
分析窓	ハミング窓
高域強調	$1 - 0.9z^{-1}$
分析	MFCC(1-12 次), 対数パワー +Δ +ΔΔ (計 39 次元)

表 3 に分析条件を示す．本研究は雑音状況下での話者識別法にも応用するため[11]，分析には雑音に頑健なアルゴリズムが使用されている ETSI advanced front-end (AFE-WI008) を使用する[12]．このフロントエンドは雑音対策として，加算性雑音にはウィナーフィルタによる雑音除去法を，乗算性雑音には blind equalization を用いている．予備実験の結果から，blind equalizaton が話者識別の性能に悪影響を与えることが示されたため，本研究では blind equalizaton 処理は省略する．

音素 HMM および音素クラス HMM を用いた場合の尤度計算には音素認識器を用いる．音素認識器は，時間同期ビームサーチを用いた one-pass アルゴリズムに基づくものである．言語モデルとしては音素対文法を用いる．

本研究では，登録用発声として 1 発話を用いるが，発話長は平均 5.5sec と極めて短い．このため登録発話の内容に識別性能が影響を受ける可能性があるため，以下の方法で，登録発声内容の違いによる識別性能の変動を平均化して評価する．

- 各評価話者の 25 発声のうち 24 発声を評価用，残り 1 発声を登録用として使用する
- 25 の異なる登録について実験し，その平均を識別率とする

以上により，一人当たりの評価データ数は $24 \times 25 = 600$ サンプルとなる．

4. 実験結果および考察

音素クラス HMM を使用した話者識別について，他のアンカーモデルを用いた話者識別法と性能を比較し検討・考察を行

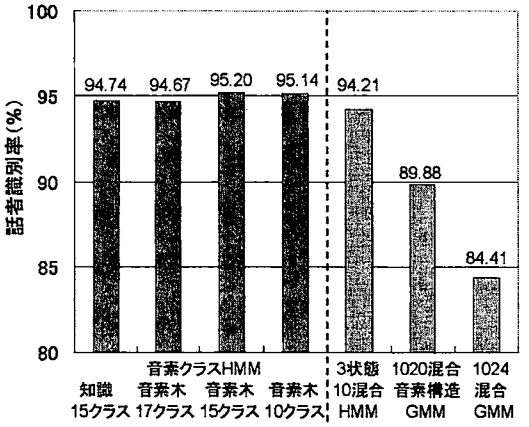

図 3 アンカーモデルの違いによる性能比較

Fig. 3 Performance comparison by the difference in anchor model.

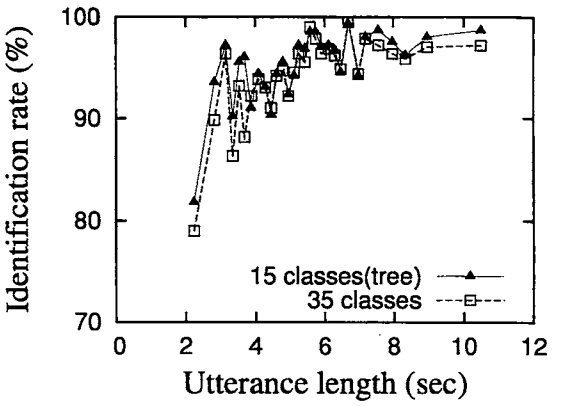

図 4 登録音声の時間長と識別率の関係

Fig. 4 Results with various length of reference utterances.

う．pdf 数を比較データと合わせるために，音素クラス HMM の混合数は以下の通りにする．

- 音声学的知識による音素 15 クラス：音素 15 クラス× 3 状態× 23 混合=1035
- 音素木に基づくクラスタリングによる音素 17 クラス：音素 17 クラス× 3 状態× 20 混合=1020
- 音素木に基づくクラスタリングによる音素 15 クラス：音素 15 クラス× 3 状態× 23 混合=1035
- 音素木に基づくクラスタリングによる音素 10 クラス：音素 10 クラス× 3 状態× 34 混合=1020

音素クラス HMM と音素 HMM, GMM，さらに音素構造 GMM との比較検討を行う．音素構造 GMM は，pdf の総数を考慮して，3 状態 10 混合 HMM に含まれるガウス分布を使用して作成する．以下に音素構造 GMM の作成法を記述する．

step1 3 状態 10 混合 HMM の学習モデルから，無音モデルを除く 34 音素についてガウス分布を取り出す．(pdf 総数： 3 状態× 10 混合× 34 音素 = 1020)．

step2 上記ガウス分布に混合重みを与え 1 状態とし，音素構造 GMM を得る．

step3 ML により混合重みのみを再学習する．

音素クラス HMM と HMM, GMM，および音素構造 GMM の性能比較を図 3 に示す．この実験では音素 HMM デコード用の文法として音素対文法を使用した．次元数は $N = 1000$ で，アンカーモデルは音素クラス HMM の 4 タイプ (音声学的知識による音素 15 クラス，および音素木に基づくクラスタリングによる 10, 15, 17 の音素クラス) とその他の 3 タイプ (3 状態 10 混合 HMM, 1020 混合音素構造 GMM, 1024 混合 GMM) について比較を行っている．モデルの pdf の総数はほぼ同数である．

比較すると，音素クラス HMM を使用した 4 パターンすべてについて他のアンカーモデルを用いたときより性能が向上している．本識別法において，音素クラスの使用は性能向上に有効であることが示された．最良の結果として，95.20%の話者識別率が音素木に基づいたクラスタリングによる 15 音素クラスの

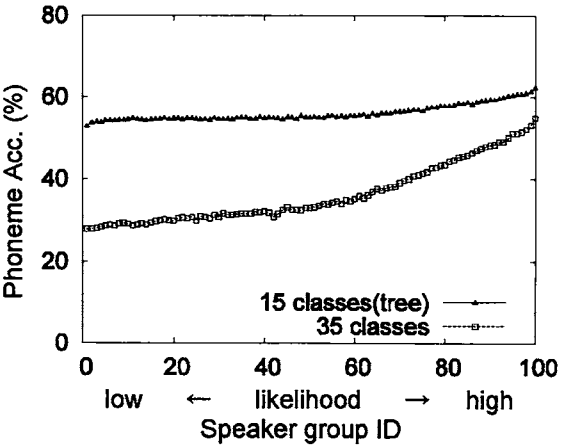

図 5 音素クラス HMM による音素認識率の差

Fig. 5 Phoneme recognition rate with two phonetic-class HMM.

3 状態 23 混合 HMM で得られている．これは GMM ベースのアンカーモデルシステムと比較して，69.2%の相対的改善となる．また 35 音素の 3 状態 10 混合 HMM と比較して 17.1%の相対的改善となる．

本実験では，識別対象話者の登録発話の時間長は平均 5.5sec と短いが，登録発話の時間長と識別性能の関連を，さらに詳細に検討した．評価話者 30 人で 3 状態 23 混合 15 音素クラス (音素木) の HMM を用いた実験について，登録発話の時間長ごとに識別率を求めた．比較として，3 状態 10 混合 HMM での時間長ごとの識別率もあわせて載せる．実験では 30 名の登録発話を 25 の異なる登録について実験するため，$30 \times 25 = 750$ 通りの識別率が得られる．この識別率を登録音声の時間長順に並べ替え，25 通りの識別率ごと平均を出しプロットした図を，図 4 に示す．時間長が 6sec を下回ると徐々に識別率が低下するが，2.2sec でも 80%以上の識別率が得られていることが分かる．

アンカーモデルを音素 HMM から音素クラス HMM にする

ことで性能が向上したが，この要因として音素アライメントの精度が改善した可能性が考えられる．本識別手法では，特定話者の音響モデルで異なる話者の音声をデコードしている．このため，デコード結果の音素認識率自体は極めて低いと予想される．しかし音素クラス HMM は音素 HMM と比べクラス数が半数以下となっており，音素クラス認識率が向上していると予想される．そこで音素クラス認識率について調査を行った．評価データには 3 状態 23 混合 15 音素クラス (音素木) の HMM を使用する．比較データとして 35 音素の 3 状態 10 混合 HMM の音素認識率についても調査した．この結果を図 5 に示す．話者 1000 人を尤度順に 10 人ごとのグループに分け，音素認識精度で評価を行なった．この結果，15 音素クラス (音素木) の HMM は 35 音素の HMM に比べ，良好な音素クラス認識率が得られていることが分かった．いずれのグループにおいても 15 音素クラス (音素木) の HMM の性能が上回っており，音素認識精度の向上が，話者識別性能の向上に寄与していると考えられる．

5. 結　論

本研究では話者ベクトルを使用した話者識別法について，アンカーモデルに音素クラス HMM を用いたときの検討を行った．アンカーモデルとして GMM の代わりに音素 HMM を使用することで識別性能改善が得られているが [1]，音素クラス HMM とすることで更なる性能の向上を図った．本提案手法は日本語話者識別タスクで評価を行った．実験では，平均 5.5sec の極く短い発話を識別対象話者の登録用データとして使用した．比較実験の結果，音素クラス HMM をアンカーモデルに使用したすべてのパターンにおいて，音素 HMM や GMM を用いたものよりも性能が向上した．このことから，音素クラスの使用は本識別法において性能の向上に有効であることが示された．また音素クラスの設定法としては知識によるものよりも，音素決定木に基づくクラスタリングが良いことが分かった．最良の結果として，95.20%の識別率が 15 音素クラス (音素木)3 状態 23 混合 HMM システムの 30 名の話者識別タスクで得られた．これは GMM ベースの話者ベクトル法と比べると，69.2%の相対的改善となる．また，35 音素 HMM ベースシステムと比較して 17.1%の相対的改善となる．

本研究では，全音素クラスが識別に使用されている．しかし，ある音素クラスが他のものより高い話者特徴を有している可能性がある [13]．よって話者識別における音素クラスの影響を検討する．さらに計算量削減のために，アンカー話者選択法についても検討を行う予定である．また雑音への耐性についても，検討を行う [11]．

文　献

[1] 赤津達也, 加藤正治, 小坂哲夫, 好田正紀, "音素モデルを用いた話者ベクトルに基づく話者識別の検討", 信学技報, SP2006-101, pp.95-99, 2006.

[2] D.Sturim, D.Reynolds, E.Singer, and J.Campbell, "Speaker indexing in large audio databases using anchor models," in *ICASSP01*, vol.1, pp.429-432, 2001.

[3] Yassine Mami and Delphine Charlet, "Speaker identification by anchor models with pca/lda post-processing," in *ICASSP03*, vol.1, pp.180-183, 2003.

[4] Delphine Charlet, Mikael Collet, Yassine Mami and Frederic Bimbot, "Probabilistic anchor models approach for speaker verification," in *INTERSPEECH05*, pp.2005-2008, 2005.

[5] A.Nakamura et al., "Japanese speech databases for robust speech recognition," in *ICSLP96*, pp.2199-2202, 1996.

[6] M.Hebert and L.P.Heck, "Phonetic class-based speaker verification," in *EUROSPEECH03*, pp.1665-1668, 2003.

[7] A.Park and T.J.Hazen, "Asr dependent techniques for speaker identification," in *ICSLP02*, pp.1337-1340, 2002.

[8] M.A.Kohler, W.D.Andrews and J.P.Campbell, "Phonetic speaker recognition," in *EUROSPEECH01*, pp.149-153, 2001.

[9] 秋田祐哉, 河原達也, "多数話者モデルを用いた討論音声の教師なし話者インデキシング", 信学論, Vol. J87-D-II, No. 2, pp. 495-503, 2004.

[10] R.Faltlhauser and G.Ruske, "Improving Speaker Recognition Performance Using Phonetically Structured Gaussian Mixture Models," in *EUROSPEECH01*, pp.751-754, 2001.

[11] 後藤佑樹, 赤津達也, 加藤正治, 小坂哲夫, 好田正紀, "話者ベクトルによる雑音下話者識別の検討", 信学技報, SP2007-18, pp. 61-66, 2007.

[12] ETSI ES 202 050 V1.1.1, "Stq; distributed speech recognition; advanced front-end feature extraction algorithm; compression algorithms," *ETSI standard*, 2002.

[13] M. A. Fattah, F. Ren and S. Kuroiwa, "Effects of Phoneme Type and Frequency on Distributed Speaker Identification and Verification," in *IEICE Trans. Inf & Syst.*, vol.E89-D, No.5, pp. 1712-1719, 2006.