論文 / 著書情報
Article / Book Information

論題(和文)	テキスト指定型話者認識
Title(English)	
著者(和文)	松井知子, 古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌, Vol. J79-D-II, No. 5, pp. 647-656
Citation(English)	, Vol. J79-D-II, No. 5, pp. 647-656
発行日 / Pub. date	1996, 5
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1996 Institute of Electronics, Information and Communication Engineers.

論　文

テキスト指定型話者認識

松井　知子[†]　　古井　貞熙[†]

Text-Prompted Speaker Recognition

Tomoko MATSUI[†] and Sadaoki FURUI[†]

あらまし　本論文では，認識のたびに任意のテキストを指定することが可能な話者認識法，テキスト指定型話者認識法を提案する．この方法では，話者と発声内容の両方の判定を行い，本人が装置側から指定されたテキストを正しく発声したと判定できるときのみ，それを受理する．そのために，各話者ごとに音韻モデル（話者音韻モデル）を用意し，指定したテキストに従って音韻モデルを連結して，そのテキストのモデルを生成する必要がある．しかも，話者認識では，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，限られた量の学習データを用いて，話者音韻モデルを生成する必要がある．ここでは，いくつかの話者音韻モデルを効果的に生成する方法について検討し，不特定話者音韻モデルの音韻依存／独立反復学習による話者適応化に基づく方法が最も有効であることを示す．

キーワード　話者認識，テキスト指定，話者照合，テキスト照合，隠れマルコフモデル，話者音韻モデル

1. まえがき

テキスト依存型話者照合では，あらかじめ照合に用いるテキスト（キーワード）を決めておくため，それに固有な音韻情報も利用することができ，一般的にテキスト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を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にだけ受理するような機構が必要である．

従来，10数字や複数の単語をキーワードとし，照合のたびにそれらのキーワードを新たな順序で並べ換えた系列を指定して，それを話者に発声させるという方法が試みられてきた[1],[2]．しかし，電子化した記憶装置を用いれば，キーワードの任意の系列を容易に再生できるようになると思われ，そのような方法も安全性に欠ける．

† NTTヒューマンインタフェース研究所，武蔵野市
NTT Human Interface Laboratories, Musashino-shi, 180 Japan

本論文では，照合のたびに任意のテキストを指定することが可能な話者認識法，テキスト指定型話者認識法を提案する．この方法では，本人が正しく指定したテキストを発声したか否かを判定するために，学習時に各話者ごとに音韻モデルを生成し，認識時には装置側から指定したテキストに従って，それら音韻モデルを連結して，そのテキストのモデルを生成する．この方法を使えば，極めて高い精度で話者照合ができ，しかも，本人の音声でも，録音音声の再生により異なる内容である場合には，高い割合でそれを棄却することができる．

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，テキスト指定型話者認識法では，少数サンプルからいかに正確に音韻性と話者性をもった音韻モデルを生成するかが重要な課題となる．本論文では，このための方法に重点をおいて検討する．

2. テキスト指定型話者認識法

テキスト指定型話者認識法では，まず各話者の学習データを用いて，各話者の音韻モデル（話者音韻モデル）を生成する．認識時には，システム側から指定した（正書法による）テキストを，無声化・長母音等を考慮して，音韻系列に自動的に変換し，その系列に従っ

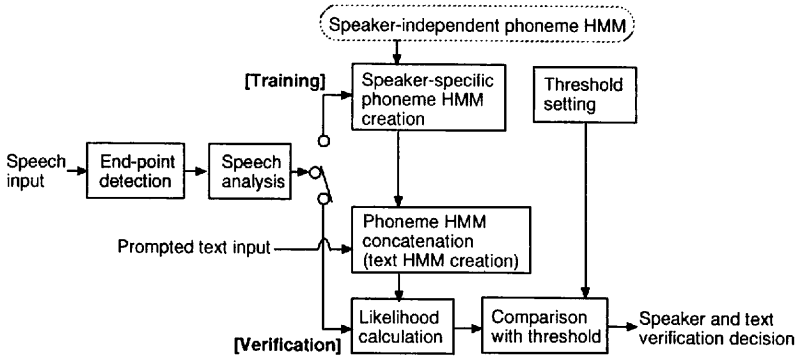

図1 テキスト指定形話者照合システムの構成
Fig. 1 Basic procedure for the text-prompted speaker verification.

て各話者ごとにその話者音韻モデルを連結することにより，そのテキストのモデルを生成する．次いで，入力音声をそのテキストのモデルに与えたときのゆう度を計算し，話者およびテキストの判定を行う．図1にこのシステムの構成を示す．

3. 話者音韻モデルの生成法

話者認識では，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．そのために，限られた量の学習データを用いて，話者性・音韻性をできるだけ正確に表現した話者音韻モデルを生成する必要がある．ここでは，学習データだけから生成する方法（方法1），および学習データ量の不足を補うために不特定話者音韻モデルを利用する方法（方法2，3，4）について検討する．

3.1 学習データだけから生成する方法（方法1）

限られた量の学習データだけを用いて，各音韻ごとにモデルを生成することは難しい．そこで，音韻単位というよりはむしろ，音韻クラスごとにモデルを生成する．

図2にこの方法のブロック図を示す．まず，各話者ごとにその話者のすべての学習データを用いて，音韻独立な話者HMM（1状態64混合ガウス分布の連続型HMM）を作成する．次にこの話者HMMを，その話者の各音韻クラスHMMに共通な初期モデルとし，学習データを再度用いて，そのテキストに合わせて音韻クラスHMMを連結する．連結学習[3]～[5]で各ガウス分布の重みのみを再推定することによって，話者音韻クラスHMMを学習する．音韻クラスの数と，その種類についてはあらかじめ決めておくが，学習データ量の制限から，音韻クラス数をあまり大きくすることはできない．

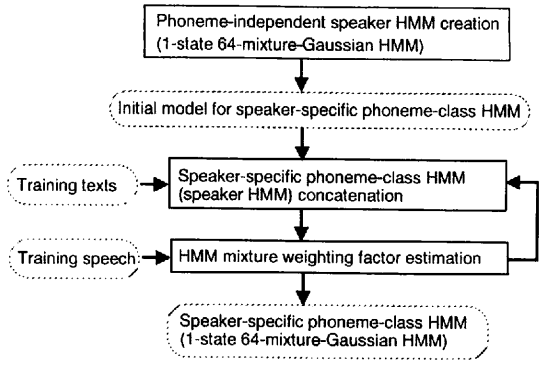

図2 学習データだけから生成する方法（方法1）のブロック図
Fig. 2 Block diagram of the method using only training data (Method 1).

表1 不特定話者音韻モデルを利用する三つの方法の関係
Table 1 Relationship between three methods using speaker-independent phoneme models.

話者性 ＼ 音韻性	なし	あり
なし		不特定話者音韻HMM（方法3、4）
あり	話者HMM（方法2）	話者音韻HMM

3.2 不特定話者音韻モデルを利用する方法

学習データ量の不足を補うために，不特定話者音韻モデルを利用する．ここでは，

方法2： 話者HMMの音韻適応化に基づく方法
方法3： （不特定話者）音韻HMMの話者適応化に基づく方法
方法4： 音韻依存／独立反復学習による方法

の三つの方法について検討する．

表1に示すように，音韻独立な話者HMMはすべての音韻に共通の話者性は有しているが，音韻性は有していない．その反面，不特定話者音韻HMMは音韻性は有しているが，話者性は有していない．音韻性と話者性を併せもつ話者音韻モデルの生成法としては，話者HMMに音韻性を学習させる方法と，不特定話者音韻HMMに話者性を学習させる方法が考えられる．そこで，方法2（話者HMMの音韻適応化に基づく方法）では，話者HMMを基本として，各話者の学習データや不特定話者音韻HMMを用いて，話者HMMに音韻性を学習させ，話者音韻HMMを生成する．方法3（(不特定話者）音韻HMMの話者適応化に基づく方法）では，方法2とは逆に，不特定話者音韻HMMを基本として，各話者の学習データを用いて，不特定話者音韻HMMを各話者の音声に適応化し，話者音韻HMMを生成する．また，方法4（音韻依存／独立反復学習による方法）は，基本的な考え方は方法3と同じであるが，不特定話者音韻HMMを各話者の音声に適応化するときに，音韻依存／独立反復学習を用いる．

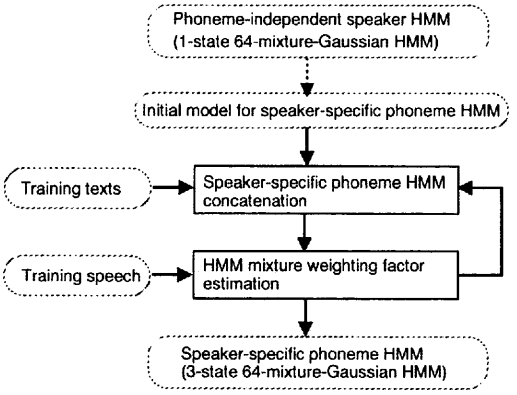

図3　話者HMMの音韻適応化に基づく方法（方法2）のブロック図
Fig. 3　Block diagram of the method based on phoneme-adaptation of speaker HMM (Method 2).

3.2.1　話者HMMの音韻適応化に基づく方法（方法2）

この方法では，音韻独立な話者HMMを基本として，各話者の学習データや不特定話者音韻HMMを用いて，話者HMMに音韻性を学習させ，話者音韻HMMを生成する．

図3，図4にこの方法のブロック図，および初期モデルの構成法を示す．最初に，各話者ごとにその話者の学習データを用いて，音韻独立な話者HMM（1状態64混合ガウス分布の連続型HMM）を作成する．そして，不特定話者音韻HMM（3状態4混合ガウス分布の連続型HMM）の各状態を，その話者HMMを用いて表すことを考える．まず，不特定話者音韻HMMの各状態の混合ガウス分布を満足するようにデータセットを人工的に生成する．生成したデータセットを用いて，話者HMMの各ガウス分布の重みだけを適応化（再推定）する．(不特定話者のデータをそのまま使うことも考えられるが，不特定話者のデータ量が非常に多いことから，適応化にかかる時間も膨大となる．）そして，各不特定話者音韻HMMの各状態に対応して適応化された話者HMMを用いて，話者音韻HMMの初期モデル（3状態64混合ガウス分布の連続型HMM）を生成する．遷移確率は不特定話者音韻HMMのものをそのまま用いる．学習テキストに従って，話者音韻HMMの初期モデルを連結し，その連結HMMに学習音声を与えて，連結学習で再び各ガウス分布の重みだけを再推定し，話者音韻HMM（3状態64混合ガウス分布の

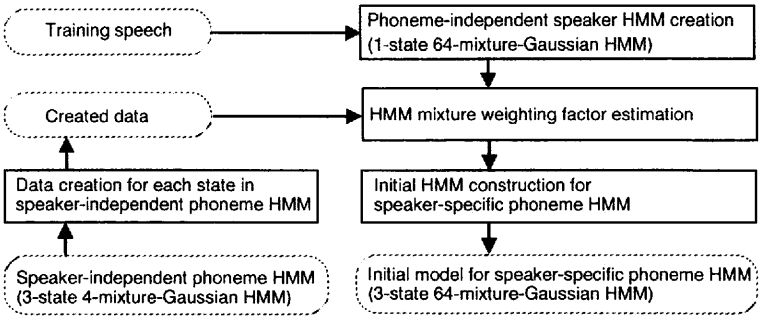

図4　話者HMMの音韻適応化に基づく方法（方法2）における初期モデルの構成法
Fig. 4　Construction of initial models for the method based on phoneme-adaptation of speaker HMM (Method 2).

連続型HMM）を作成する．

以上の手続きにおいて，話者HMMの各ガウス分布の重みだけを適応化するのは，話者性を保持するためである．

3.2.2 音韻HMMの話者適応化に基づく方法（方法3）

この方法では，不特定話者音韻HMMを基本として，各話者の学習データを用いて，不特定話者音韻HMMを各話者の音声に適応化し，話者音韻HMMを生成する．

図5にこの方法のブロック図を示す．不特定話者音韻HMM（3状態4混合ガウス分布の連続型HMM，または3状態256混合ガウス分布の半連続型HMM）を話者音韻HMMの初期モデルとして利用する．学習テキストに従って，不特定話者音韻HMMを連結し，その連結HMMに学習音声を与えて，連結学習で各ガウス分布の平均値と重みを適応化（再推定）し，話者音韻HMM（3状態4混合ガウス分布の連続型HMM，または3状態256混合ガウス分布の半連続型HMM）を作成する．

以上の手続きにおいて，不特定話者音韻HMMの各ガウス分布の分散値を適応化しないのは，各話者の学習データ量が十分でないためである．

3.2.3 音韻依存／独立反復学習による方法（方法4）

不特定話者音韻HMMとして連続型HMMを用いるよりも，半連続型HMMを用いた方が次の2点で優れていると思われる．第1に，半連続型HMMでは，すべての音韻HMMで平均・分散値が結びの状態[6]になっているため，学習データ量が少ないときでもうまくHMMパラメータを推定できる[7]．第2に，音韻共通の混合分布に，テキスト独立型の話者認識[8]で有効に使われる，すべての音韻に共通の話者性を反映させることができる．

また，各話者ごとに，少量の学習データから音韻HMMを生成しようとすると，出現頻度の低い音韻が生じ，そのような音韻のモデルについてはうまく話者の音声に適応化させることは難しい．出現頻度の高い音韻のモデルから，低い音韻のモデルを効果的に推定する方法は今のところない．推定誤りによる性能劣化を避けるために，ここでは，全音韻について平均，分散値に加え，重みも結びの状態にした学習を行うことによって，推定すべきパラメータ数を減らす．これにより，出現頻度の低い音韻のモデルについても，推定誤りを小さくすることを考える．学習データの分布だけを用いるテキスト独立型話者認識[8]がうまくできることから，学習データの分布をその話者の全音韻の分布であると仮定し，それをより正確に表現することが有効であると考える．

この方法では，不特定話者音韻HMMとして半連続型HMMを用い，音韻依存／独立反復学習を通して，不特定話者音韻HMMを各話者の音声に適応化する．

音韻依存／独立反復学習では，図6で示すように，音韻依存／独立学習を一系列で繰り返し行う．音韻依存／独立繰返し学習では，音韻依存学習は従来の音韻モデルの話者適応化に相当し，音韻独立学習は半連続型HMMの混合分布の重みもすべての音韻で結びにし

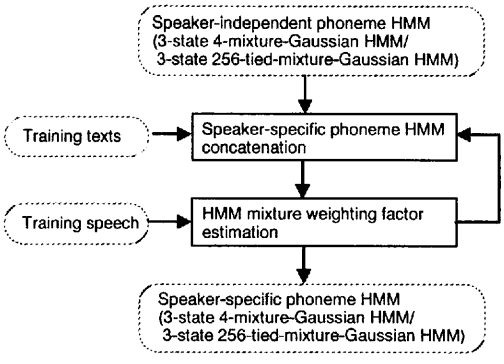

図5 音韻HMMの話者適応化に基づく方法（方法3）のブロック図
Fig. 5 Block diagram of the method based on speaker-adaptation of phoneme HMMs (Method 3).

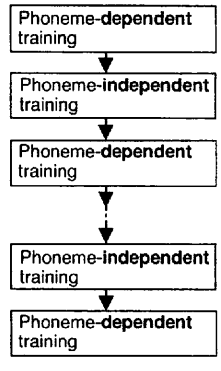

図6 音韻依存／独立反復学習による方法（方法4）のブロック図
Fig. 6 Block diagram of the method using phoneme dependent/independent iterative training (Method 4).

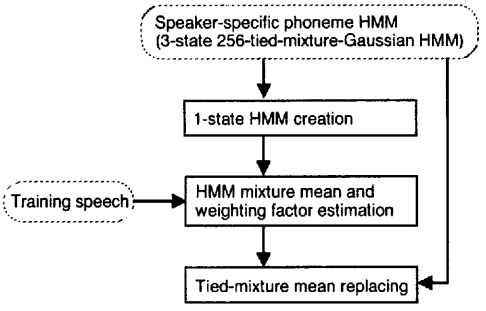

図7　音韻独立学習の手続き
Fig. 7　Phoneme-independent training procedure.

た学習で，すべての音韻に共通な話者性を反映させる．その効果として，出現頻度の少ない音韻のモデルについても，その話者の音声に適応化できる．

音韻依存学習は，**3.2.2**の方法3において用いた学習（図5）であり，まず学習テキストに従い，不特定話者音韻HMM（3状態256混合ガウス分布の半連続型HMM）を連結し，学習音声を与えて，全音韻で結びの平均値，および各音韻ごとの重みを推定する．この学習では，学習データに含まれない音韻のモデルについては，積極的には適応化されない．

音韻独立学習（図7）では，各話者の半連続型HMMの混合分布を用いて，1状態のHMMを生成する．この1状態HMMは，全音韻について重みも結びの状態にしたモデルであると言える．ここでは，重みの初期値は各分布に共通な値に設定した．それに学習音声を与え，平均値と重みを推定する．そして，各話者の混合分布の平均値だけをその1状態HMMの平均値で置き換える．この学習では，出現頻度の少ない音韻のモデルについても，その話者の特徴量分布に適応化される．

4. 実　　験

4.1 概　　要

使用したデータは，男10名，女5名が約10か月にわたる5時期に，発声した文章データである．文章テキストは，ATR連続音声テキスト503文[9]から抜粋した．各話者は，そのテキストを読みあげる形で発声した．なお，録音は雑音のない部屋で行われ，全話者，全時期で同じコンデンサマイクロホンを使用した．

特徴パラメータとして，標本化周波数12kHz，フレーム長32ms，フレーム周期8ms，LPC分析次数16でケプストラムを抽出した．その際，音声パワーは用

表2　学習とテストの音声テキスト
Table 2　Texts for training and testing.

	文章テキスト
学習	1. 飛ぶ自由を得ることは人類の夢だった。 2. 初めてルーブル美術館へ入ったのは14年前のことだ。 3. 自分の実力は自分が一番よく知っているはずだ。 4. これまで少年野球ママさんバレーなど地域スポーツを支え市民に密着してきたのは無数のボランティアだった。 5. ギンザケの卵を輸入してふ化させ海中で育てる養殖も始まっている。 (6). ()の五つのテキストについては、各話者、 (7). 各時期ごとに異なるものを、ATR連続音声 (8). テキスト503文から抜粋して使用した。 (9). (10).
テスト	1. 予防や健康管理リハビリテーションのための制度を充実していく必要があろう。 2. 背の高さは170センチほどで目が大きくやや太っている。 3. 大声を出し過ぎてかすれ声になってしまう。 4. 足し算引き算はできなくても絵は描ける。 5. どの部屋の意匠にも遊び心があふれていて楽しい。

いなかった．表2 に示すように，学習にはある一時期に発声した10文章を用いた．テストでは，その他の2，あるいは4時期に発声した5文章を1文章ずつ用いた．学習で用いた10文章の内の5文章は，そのテキストが全話者，全時期に共通で，残りの5文章のテキストは各話者，各時期ごとに異なる．テストでは，学習に用いた文章とは異なるが，そのテキストは全話者，全時期で同じである．文章の平均時間長は4.2秒であった．実験に用いたテスト音声の総数は，テスト音声として2時期に発声したデータを用いた場合は150文章（15名×2時期×5文章），4時期に発声したデータを用いた場合は300文章（15名×4時期×5文章）である．また，学習データだけから生成する方法では，音韻クラスの種類は25，その他の方法では音韻の種類は41である（表3参照）．なお，不特定話者音韻HMMの作成には，日本音響学会研究用データベースの男30名，女30名の計9,000文章を用い，Baum-Welchアルゴリズムによって HMM パラメータの推定を行った．

評価は，事後確率に基づくゆう度正規化法[10]を適用する／しない場合における1．話者照合誤り率，2．

表3 音韻の種類
Table 3 Phonemes used in the experiments.

音韻クラス数25の場合	(母音) a, i, u, e, o (鼻子音) m, n (流音) r (破裂音) b, d, g, p, t, k (撥音) N (摩擦音) j, z, s, sh, h, f (破擦音) ts, ch (半母音・拗音) y (半母音) w
音韻数41の場合	(母音) a, i, u, e, o (鼻子音) m, n (流音) r (破裂音) b, d, g, p, t, k (撥音) N (摩擦音) j, z, s, sh, h, f (破擦音) ts, ch (半母音) y, w (拗音部) by, gy, my, ny, py, ky, hy, ry (長母音) aa, ii, uu, ee, oo, ei, ou (促音) Q

テキスト照合誤り率，若しくは3．話者およびテキストの照合（話者・テキスト照合）誤り率によって行った．話者およびテキストの判定を行う方法としては，それらを別々に行う個別判定法と，同時に行う一括判定法が考えられるが，1．と2．は前者の判定性能を，3．は後者の判定性能を評価するものである．

1．話者照合では，各話者が正しい文章を発声した場合について，ゆう度の値にしきい値を設けて，そのしきい値よりも大きいゆう度の入力音声は本人の声として受理し，それ以外は棄却する．なお，このしきい値は各話者ごとに，2種類の誤り（詐称者受理率，本人棄却率）が等しくなる値に事後的に設定した．

2．テキスト照合では，本人がシステム側から指定されたテキストとは異なるテキストを発声した場合について，ゆう度の値にしきい値を設けて，そのしきい値よりも大きいゆう度の入力音声は，本人が正しいテキストを発声した音声として受理し，それ以外は棄却する．なお，このしきい値は各話者ごとに，本人が正しいテキストを発声した音声が棄却されることのないように事後的に設定した．

3．話者・テキスト照合では，本人が異なるテキストを発声した音声も，棄却すべき音声として含めて認識対象とした場合について，ゆう度の値にしきい値を設けて，そのしきい値よりも大きいゆう度の入力音声は本人の声として受理し，それ以外は棄却する．なおしきい値は，1．話者照合の場合と同様に事後的に設定した．

4.2 結　　果

図8，図9に，下記の五つの方法によって話者音韻HMMを生成したときの，テスト音声として2時期に発声したデータを用いた場合の話者照合誤り率，テキスト照合誤り率を示す．

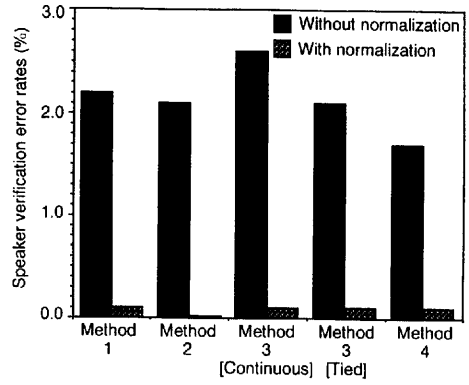

図8 種々の話者音韻HMMの生成法による話者照合誤り率（%）の比較
Fig. 8 Comparison of speaker verification error rates using various methods for making speaker-specific phoneme HMMs.

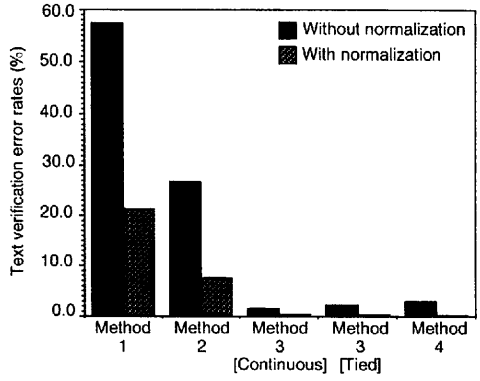

図9 種々の話者音韻HMMの生成法によるテキスト照合誤り率（%）の比較
Fig. 9 Comparison of text verification error rates using various methods for making speaker-specific phoneme HMMs.

（1） 学習データだけから生成する方法（方法1）：1状態64混合ガウス分布の連続型HMM

（2） 話者HMMの音韻適応化に基づく方法（方法2）：3状態64混合ガウス分布の連続型HMM

（3） 音韻HMMの話者適応化に基づく方法（方法3 [Continuous]）：3状態4混合ガウス分布の連続型HMM

（4） 音韻HMMの話者適応化に基づく方法（方法3 [Tied]）：3状態256混合ガウス分布の半連続型HMM

（5） 音韻依存／独立反復学習による方法（方法4）：3状態256混合ガウス分布の半連続型HMM

図8より，ゆう度正規化法を適用した場合，その効果が大きいために，各話者音韻モデルの生成方法による話者照合誤り率にほとんど差は見られないことがわかる．ゆう度正規化法を適用しない場合には，方法4が最も誤り率が小さく，方法3 [Tied]，方法2，方法1，方法3 [Continuous]の順であった．このことは，音韻依存／独立反復学習が，音韻依存学習のみの音韻HMMの話者適応化に基づく方法や他の方法と比べて有効であることを示している．また，方法3 [Tied]の方が方法3 [Continuous] よりも誤り率が小さかったが，**3.2.3** で述べたように，半連続型HMMは少量の学習データから頑健にHMMパラメータが推定でき，音韻共通の混合分布にすべての音韻に共通の話者性を反映させることができるので，連続型HMMよりも話者照合性能が高いと考えられる．更に，方法1や方法2の誤り率は比較的小さかったが，学習データだけから生成する方法や話者HMMの音韻適応化に基づく方法は，すべての音韻に共通の話者性を保持しているため話者照合性能が比較的高いと考えられる．以上より，高い話者照合性能を得るためには，すべての音韻に共通の話者性を表現できるモデルを用いることが重要であると考えられる．

図9より，不特定話者音韻HMMを基本として用い，それを話者適応化して話者音韻HMMを生成する方法，方法3 [Continuous]，方法3 [Tied]，方法4では，テキスト照合誤り率が小さいことがわかる．不特定話者音韻HMMを用いない，学習データだけから生成する方法や話者HMMの音韻適応化に基づく方法では，話者照合性能は比較的高かったが，テキスト照合性能は低い．

図10に，3．音韻HMMの話者適応化に基づく方法（方法3 [Continuous]），4．音韻HMMの話者適応化に基づく方法（方法3 [Tied]），5．音韻依存／独立反復学習による方法（方法4）によって話者音韻HMMを生成したときの，テスト音声として4時期に発声したデータを用いた場合の話者・テキスト照合誤り率を示す．

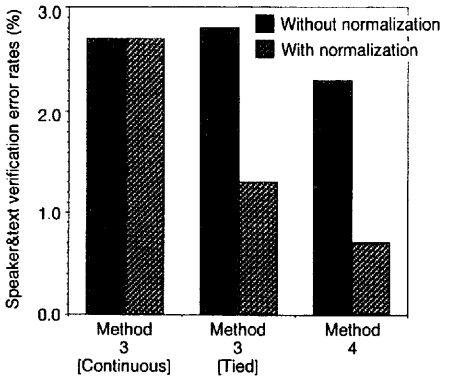

図10 種々の話者音韻HMMの生成法による話者・テキスト照合誤り率（%）の比較

Fig. 10 Comparison of speaker and text verification error rates using various methods for making speaker-specific phoneme HMMs.

図10より，音韻依存／独立反復学習による方法が他の方法に比べて有効であること，特にゆう度正規化法を用いた場合に優れていることを示している．

Viterbiアライメントをとって，各音韻モデルごとにその平均ゆう度を調べた結果，音韻依存／独立反復学習による方法を用いた場合，音韻依存学習のみの音韻HMMの話者適応化に基づく方法に比べて，出現頻度の低い音韻のゆう度が大きくなることが確認できた（**5.4** 参照）．このことは，出現頻度の低い音韻のモデルについても，適応化がうまくできていることを表している．ゆう度正規化法では，申告話者のモデルのゆう度と他の話者のモデルのゆう度との相対値が用いられる．音韻HMMの話者適応化に基づく方法では，上述のように，出現頻度の低い音韻のモデルのゆう度は小さく，また，各音韻の出現頻度は話者によって異なるので，その相対値はテストデータによってばらつく．音韻依存／独立反復学習による方法では，出現頻度の低い音韻のモデルのゆう度もある程度大きいので，その相対値はばらつかず，安定してしきい値が設定できるため，ゆう度正規化法による効果が大きいと思われる．

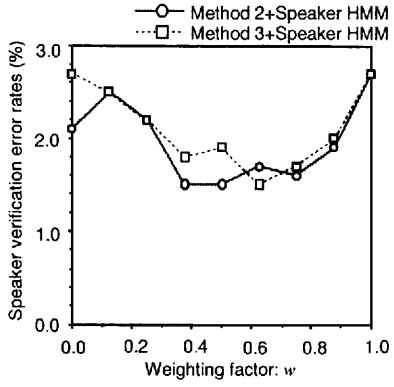

図11 加算重み w の値と話者照合誤り率との関係
Fig. 11 Speaker verification error rates as a function of weighting factor w.

5. 考　　察

5.1 話者HMMとの組合せ

話者HMMの音韻適応化に基づく方法（方法2，3状態64混合ガウス分布の連続型HMM），音韻HMMの話者適応化に基づく方法（方法3，3状態4混合ガウス分布の連続型HMM）において，話者HMMを用いて，話者性を補うことを考える．この方法では，1状態64混合ガウス分布の話者HMMの入力音声に対するゆう度（L_{PI}）を計算し，話者音韻HMMのゆう度（L_{SP}）と次式により重み付き加算したゆう度（L_{sum}）を話者の判定に用いる．

$$L_{sum} = w \times L_{SP} + (1-w) \times L_{PI}$$

$$0 \leqq w \leqq 1$$

なお，加算重み w の値については実験的に求める．

図11に，加算重み w の値を変化させたときの話者照合誤り率を示す．"Method 2+Speaker HMM" は方法2と話者HMMの組合せを，"Method 3+Speaker HMM" は方法3と話者HMMの組合せを表す．この図から，話者性を補うことによって誤り率が小さくなることがわかる．また，誤り率は加算重み w の値にあまり敏感ではなく，ほぼ0.5に設定すればよいことがわかる．

5.2 音韻依存／独立反復学習のバリエーション

音韻依存，独立反復学習の繰返し方には，種々のバリエーションが考えられる．表4に種々のバリエーションによる話者・テキスト照合率を示す．"D" は音韻依存学習を，"I" は音韻独立学習を表す．また，"D^n" は音韻依存学習を n 回繰り返したことを表し，"$(ID)^n$" は音韻依存，独立学習を n 回繰り返したことを表す．"$D(ID)^n$" は図6の繰返しを表す．

表4 音韻依存／独立反復学習の種々のバリエーションによる話者・テキスト照合率（%）（D: 音韻依存学習，I: 音韻独立学習，[]：ゆう度正規化法を適用した場合）
Table 4 Verification error rates (%) according to the types of phoneme-dependent/independent training series. (D: phoneme-dependent training, I: phoneme-independent training, [ ]:with likelihood normalization).

繰返し回数		I				
		0	1	2	3	4
D	1	D 4.7	ID 3.1	-	-	-
	2	D^2 3.2	DID 2.5	$(ID)^2$ 2.7	-	-
	3	D^3 2.8 [1.3]	ID^3 2.7	$D(ID)^2$ 2.3 [0.7]	$(ID)^3$ 2.5 [1.3]	-
	4	D^4 2.5 [1.1]	-	-	$D(ID)^3$ 2.4 [0.6]	$(ID)^4$ 2.4
	5	D^5 2.4 [1.1]	-	-	-	$D(ID)^4$ 2.2 [0.6]

音韻依存学習の繰返し回数が同じ場合，誤り率は音韻独立学習を適用することによって減少する．$D(ID)^4$ のとき，ゆう度正規化後の話者・テキスト照合率は99.4%で最も高かった．

5.3 音韻依存／独立反復学習と音韻依存学習との比較

音韻依存学習の適用回数を変えた場合について，音韻依存／反復学習による方法（方法4）と，音韻依存学習のみの音韻HMMの話者適応化に基づく方法（方法3，3状態256混合ガウス分布の半連続型HMM）の性能を比較した．図12にその結果を示す．この図から，特にゆう度正規化法を用いた場合に音韻依存／独立反復学習による方法の方が有効であることがわかる．

5.4 音韻依存／独立反復学習による話者音韻HMMのゆう度

音韻依存／独立反復学習による方法（方法4）と，音韻依存学習のみの音韻HMMの話者適応化に基づく方法（方法3，3状態256混合ガウス分布の半連続型HMM）について，学習データに含まれる各音韻の出現頻度ごとに，話者音韻HMMの学習データに対する対数ゆう度を調べた．各音韻区間は，Viterbiアライメントによって決定した．なお，対数ゆう度の値は各方法ごとに $N(0,1)$ で正規化した．

表5に，音韻出現頻度に応じて四つにグループ分け

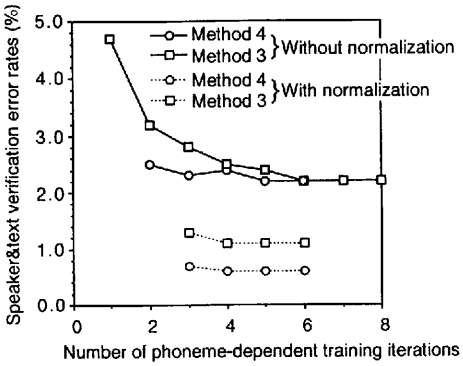

図12 音韻依存／独立反復学習（方法4）と音韻依存学習（方法3）との比較

Fig. 12 Phoneme dependent/independent iterative training (Method 4) vs. phoneme dependent training (Method 3).

表5 正規化対数ゆう度の差（方法4のゆう度 - 方法3のゆう度）

Table 5 Differences between the normalized log-likelihood values for Methods 3 and 4.

時期	音韻出現頻度			
	低い	やや低い	やや高い	高い
T1	0.14	0.00	0.05	-0.07
T2	0.20	-0.04	0.03	-0.05
T3	0.23	-0.04	0.05	-0.06
T4	0.20	0.04	0.00	-0.07
平均	0.19	-0.01	0.03	-0.06

した，各方法による話者音韻HMMの正規化対数ゆう度の平均の差を示す．出現頻度の低い音韻については，音韻依存／独立学習を用いた方がゆう度の値が大きい．このことは，音韻依存／独立学習を用いることにより，出現頻度の低い音韻HMMも話者に適応化されることを示している．

6. むすび

本論文では，認識のたびに任意のテキストを指定することが可能な，テキスト指定型話者認識法を提案した．この方法では，話者性・音韻性をできるだけ正確に表現した話者音韻モデルが必要とされるが，ここでは種々の話者音韻モデルの生成法について検討し，不特定話者音韻HMMの音韻依存／独立反復学習による話者適応化に基づく方法が最も有効であることを示した．また，不特定話者音韻HMMとして連続型HMMを用いるよりも，半連続HMMを用いた方が優れていることや，高い話者照合性能を得るためには，すべての音韻に共通の話者性を表現できるモデルを用いることが重要であることを示した．話者15名，約10か月にわたる5時期のデータを用いて評価した結果，99.4%の話者・テキスト照合率が得られた．

今後は，話者音韻モデル生成法について検討を続けると共に，話者およびテキストの照合において，事前にしきい値を設定する方法についても検討していきたい．また，出現頻度の高い音韻のモデルから，低い音韻のモデルを推定するために，各音韻の特徴量空間の幾何学的な構造について調べていきたい．

謝辞 熱心に討論して頂いた，ヒューマンインタフェース研究所古井特別研究室の方々に感謝致します．

文　献

[1] A. E. Rosenberg and F. K. Soong, "Recent research in automatic speaker recognition," in Advances in Speech Signal Processing, eds. S. Furui and M. M. Sondhi, Marcel Dekker, New York, pp.701–738, 1992.

[2] J. M. Naik, "Speaker verification: a tutorial," IEEE Communications Magazine, vol.28, no.1, pp.42–48, 1990.

[3] K.F. Lee, "Automatic Speech Recognition — The Development of the SPHINX System," Kluwer Academic Publishers, Boston, 1989.

[4] 丸山活輝，花沢利行，川端　豪，鹿野清宏，"HMM音韻連結学習を用いた英単語音声の認識，" 信学技報，SP88-119, 1988.

[5] 南　泰浩，松岡達雄，鹿野清宏，"不特定話者連続音声データベースによる連結学習HMMの評価，" 信学技報，SP91-113, 1992.

[6] 中川聖一，"確率モデルによる音声認識，" 電子情報通信学会，1988.

[7] X. D. Huang, "Phoneme classification using semi-continuous Hidden Markov Models," IEEE Trans. ASSP, vol. 40, no. 5, pp. 1062–1067, 1992.

[8] 松井知子，古井貞熙，"VQひずみ，離散／連続HMMによるテキスト独立形話者認識法の比較検討，" 信学論(A), vol.J77-A, no.4, pp.601–606, 1994.

[9] 桑原尚夫，匂坂芳典，武田一哉，阿部匡伸，"研究用ATR日本語音声データベースの作成（別冊1 連続音声テキスト），" TR-I-0086, 1989.

[10] 松井知子，古井貞熙，"テキスト指定形話者認識のための事後確率に基づくゆう度正規化法，" 日本音響学会秋季研究発表会，1-7-20，山梨，1993.

（平成7年4月5日受付，8月31日再受付）

松井　知子（正員）

昭61東工大・理・情報卒．昭63同大大学院修士課程了．同年NTT（株）入社．以後，NTTヒューマンインタフェース研究所において，話者認識の研究に従事．IEEE，日本音響学会各会員．平5論文賞受賞．

古井　貞熙（正員）

昭43東大・工・計数卒．昭45同大大学院修士課程了．同年NTT電気通信研究所入社．以後，同研究所において，音声認識，話者認識，音声知覚などの研究に従事．昭53～54ベル研究所客員研究員．現在NTTヒューマンインタフェース研究所古井特別研究室長．東京工業大学客員教授．工博．昭50米沢賞，昭63，平5論文賞，平2著述賞，平1科学技術庁長官賞，IEEE ASSP Society Senior Awardなど受賞．著書「ディジタル音声処理」，「Digital Speech Processing, Synthesis, and Recognition」，「音響・音声工学」，編著「Advances in Speech Signal Processing」など．IEEE Fellow．