論文 / 著書情報
Article / Book Information

論題(和文)	話者照合におけるモデルとしきい値の更新法
Title(English)	
著者(和文)	松井 知子, 西谷 隆, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌, Vol. J81-D-II, No. 2, pp. 268-276
Citation(English)	, Vol. J81-D-II, No. 2, pp. 268-276
発行日 / Pub. date	1998, 2
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) 1998 Institute of Electronics, Information and Communication Engineers.

論　文

話者照合におけるモデルとしきい値の更新法

松井　知子[†]　　西谷　隆[††]　　古井　貞熙[††]

A Study of Model and a Priori Threshold Updating in Speaker Verification

Tomoko MATSUI[†], Takashi NISHITANI[††], and Sadaoki FURUI[††]

あらまし　本論文では話者照合において，最近に発声した少量の更新用データを用いて，発声変動に対して頑健になるように，話者のモデルを更新する方法，およびその話者のモデルの頑健性を考慮しながら，話者の判定に用いるしきい値を適切に設定する方法について検討する．話者のモデルは，隠れマルコフモデル（hidden Markov model; HMM）で表す．モデルの更新では，話者 HMM のパラメータを，更新用データと現在のパラメータ値から推定する．しきい値の設定では，話者ごとに新しいしきい値を，更新用データから計算した等誤り率（本人棄却率と詐称者受理率が等しくなる値）よりも高めの FA 率を与える値を初期値として，話者 HMM の更新に合わせて，その等誤り率を与える値へ次第に近づくように設定する．本方法を，話者 20 名によるテキスト独立型，およびテキスト指定型話者照合実験で評価した結果，モデルやしきい値を更新しない場合に比べて，平均誤り率はテキスト独立型ではほぼ 4 割，テキスト指定型ではほぼ 8 割に削減された．

キーワード　話者照合，モデルの更新，しきい値，隠れマルコフモデル，テキスト独立型，テキスト指定型

1. まえがき

話者照合では，入力音声と本人のモデルとの類似度（ゆう度）を調べ，一定のしきい値と比較して話者の判定を行う．実システムでは，話者の発声負担も考慮して，各話者の初期モデルは一時期に発声した少量のデータから作成されることが多いが，高い話者照合性能を得るためには，各話者ごとに，過去のデータを蓄積しておき，いろいろな発声変動（主に発声内容や発声時期の違いに起因する）を含んだそのデータを用いて，話者のモデルを再作成することが望ましい．

テンプレートマッチングによる話者照合においては，古井が最近の複数回の発声の重ね合せによって作成したテンプレートによって，話者テンプレートを更新する方法を提案している[1]．この方法では，声が時間の経過につれて変わることから，遠い過去の発声は用いずに，最近の複数回の発声を用いている．HMM による話者照合では，話者 HMM の更新に関する研究例は少ないが，近年 Setlur らが蓄積したデータ量に合わせて，発声変動を表現しやすいようにモデルの構造を複雑化しながら，話者 HMM を再作成する方法について報告した[2]．この方法のように話者 HMM を再作成する方法は効果的である一方，過去のデータを多数蓄積して用いるために，必要とするメモリ量や計算量は大きくなる．本論文では，最近に発声した少量の更新用データを用いて，発声変動に対して頑健になるように，話者 HMM を逐次的に更新する方法について検討する．本方法では，更新用データと現在のパラメータ値を用いて，過去のデータを多数蓄積しておき，話者 HMM を再作成した場合のパラメータ値を近似する．

実システムではまた，話者の判定に用いるしきい値を事前に設定しておく必要がある．しきい値を事前に設定する場合，2 種類の誤り，本人の棄却（false rejection; FR）と詐称者の受理（false acceptance; FA）が考えられる．FR 率は本人のモデルと本人のデータから，FA 率は本人のモデルと本人以外のデータから求めることができる．本論文では，話者 HMM の更新に合わせて，しきい値を FR 率と FA 率が等しい，等誤り率を与える値の近傍に設定する方法について検討する．

学習用データ量が少ない場合や複数の時期に発声した音声データを学習に用いることができない場合には，本人の学習用データとは異なるデータ（オープン

[†] NTT ヒューマンインタフェース研究所，横須賀市
NTT Human Interface Laboratories, Yokosuka-shi, 239-0847 Japan
[††] 東京工業大学，東京都
Tokyo Institute of Technology, Tokyo, 152-0033 Japan

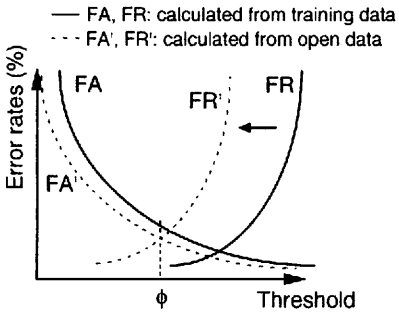

図1 学習用データ，オープンデータから計算したFA，FR率としきい値との関係

Fig. 1 Relationship between threshold and FA and FR rates calculated from training and open data.

データ）のFR率を推定することは非常に難しい．図1に学習用データ，およびオープンデータから計算したFA，FR率としきい値との関係を示す．なお，話者の判定にはHMMのゆう度を用いている．図中の ϕ は，本方法で目標とするしきい値，すなわちオープンデータの等誤り率を与えるしきい値を表す．一般に誤り率が小さいときには特に，学習用データ，オープンデータの誤り率曲線のずれは，FR率の方がFA率よりも大きく，オープンデータのFR率曲線は，学習用データのFR率曲線を基準とした場合に，しきい値が小さくなる方向（左側）へシフトする傾向にある（付録）．このため，本方法での目標のしきい値 ϕ は，学習用データのFA率曲線上では高めのFA率を与える値に相当すると考えられる．そして，この傾向は特に，話者のモデルが発声変動に対して十分に頑健でない場合に著しい．この問題に対処するために，古井は本人のモデルと本人以外のデータとの類似度の分布の平均と標準偏差の値を用いてしきい値を設定する方法，つまり比較的推定精度が良いFA率だけを考慮する方法について報告し，テンプレートマッチングによる話者照合実験においてその有効性を示した[1]．本研究では，更新用データから計算した等誤り率よりも高めのFA率を与えるしきい値を初期値として，話者HMMの更新に合わせて，その等誤り率を与えるしきい値へ次第に近づける方法を考える．

本論文では以下，各話者の初期モデルの作成を「学習」，以後のモデルの更新を「更新」と呼ぶ．

2. モデルの更新

本方法では，話者HMMの更新に必要とするメモリ量や計算量が少なくなるように，更新用データと現在のパラメータ値を用いて，過去のデータを多数蓄積しておき，話者HMMを再作成した場合のパラメータ値を近似する．

時期 U までに発声された音声データをすべて用い，話者HMMを最ゆう推定（Baum-Welchアルゴリズムによる繰返し推定）に基づいて再作成する場合，そのHMM（無相関混合ガウス分布型）の状態 s，混合分布 m の平均ベクトル $\tilde{\mu}_{sm}$，重み係数 $\tilde{w}_{sm}$ は式(1)，式(2)となる．

$$\tilde{\mu}_{sm}=\frac{\sum_{u=1}^{U}\sum_{t=1}^{T_u}c_{smt}^{u}x_t^u}{\sum_{u=1}^{U}\sum_{t=1}^{T_u}c_{smt}^{u}} \tag{1}$$

$$\tilde{w}_{sm}=\frac{\sum_{u=1}^{U}\sum_{t=1}^{T_u}c_{smt}^{u}}{\sum_{m=1}^{M}\sum_{u=1}^{U}\sum_{t=1}^{T_u}c_{smt}^{u}} \tag{2}$$

$$\mu_{sm}^{U}=\frac{\mu_{sm}^{U-1}\sum_{u=1}^{U-1}\sum_{t=1}^{T_u}C_{smt}^{u}+\tilde{\mu}_{sm}^{U}\sum_{t=1}^{T_U}C_{smt}^{U}}{\sum_{u=1}^{U-1}\sum_{t=1}^{T_u}C_{smt}^{u}+\sum_{t=1}^{T_U}C_{smt}^{U}} \tag{3}$$

$$w_{sm}^{U}=\frac{w_{sm}^{U-1}\sum_{m=1}^{M}\sum_{u=1}^{U-1}\sum_{t=1}^{T_u}C_{smt}^{u}+\tilde{w}_{sm}^{U}\sum_{m=1}^{M}\sum_{t=1}^{T_U}C_{smt}^{U}-1}{\sum_{m=1}^{M}\sum_{u=1}^{U-1}\sum_{t=1}^{T_u}C_{smt}^{u}+\sum_{m=1}^{M}\sum_{t=1}^{T_U}C_{smt}^{U}-M} \tag{4}$$

ここで，$u(=\{1,2,\ldots,U\})$ は発声時期を表す識別子であり，T_u は時期 u に発声された音声データの長さを表す．また c_{smt}^{u} はそのHMMにおいて，時期 u の時刻 t に，状態 s，混合分布 m で観測されたベクトル x_t^u が出現する確率を表している．

本方法では，式(1)の平均ベクトル $\tilde{\mu}_{sm}$，式(2)の重み係数 $\tilde{w}_{sm}$ を，最近の時期 U に発声された音声データ（更新用データ）を用いて，それぞれ式(3)の μ_{sm}^{U}，式(4)の w_{sm}^{U} で近似する．C_{smt}^{u} は，時期 $u-1$ までに本方法によって逐次的に更新した話者HMMを初

期 HMM とし，時期 u の更新用データを用いた最ゆう推定による HMM において，観測ベクトル x_t^u が出現する確率を表している．また，$\tilde{\mu}_{sm}^U$，$\tilde{w}_{sm}^U$ は式 (5)，式 (6) で定義され，時期 U に発声された更新用データのみを用いた最ゆう推定による HMM の各混合分布の平均ベクトル，重み係数を表す．なお，式 (4) の分子，分母の第 3 項は実験的に設定した補正項である．

$$\tilde{\mu}_{sm}^U = \frac{\sum_{t=1}^{T_U} C_{smt}^U x_t^U}{\sum_{t=1}^{T_U} C_{smt}^U} \tag{5}$$

$$\tilde{w}_{sm}^U = \frac{\sum_{t=1}^{T_U} C_{smt}^U}{\sum_{m=1}^{M}\sum_{t=1}^{T_U} C_{smt}^U} \tag{6}$$

ここで，$\sum_{u=1}^{U-1}\sum_{t=1}^{T_u} C_{smt}^u$ は，平均ベクトル μ_{sm}^{U-1} を推定するのに用いた，過去（時期 $1 \sim U-1$）のすべての音声データに含まれるフレーム数に相当し，$\sum_{m=1}^{M}\sum_{u=1}^{U-1}\sum_{t=1}^{T_u} C_{smt}^u$ は，重み係数 w_{sm}^{U-1} を推定するのに用いたフレーム数に相当する．従って，μ_{sm}^U，w_{sm}^U は基本的に，もとの μ_{sm}^{U-1}，w_{sm}^{U-1} と更新用データから得られる最ゆう推定値 $\tilde{\mu}_{sm}^U$，$\tilde{w}_{sm}^U$ との，それぞれのパラメータを推定するのに用いたフレーム数による重み付き平均として解釈できる．

式 (3)，式 (4) は（補正項を無視した場合），再帰的に展開していくと，式 (1)，式 (2) の c_{smt}^u を C_{smt}^u で置き換えた形となる．出現確率 C_{smt}^u は，時期 u までに発声された音声データをすべて用いた最ゆう推定による HMM から求まる c_{smt}^u とは異なるが，本方法では両者に大差は生じないと考えて，

$$c_{smt}^u \approx C_{smt}^u \quad \text{for all } u \tag{7}$$

の近似を行っている．

3. しきい値の設定

1. で述べたように，図 1 のオープンデータの FR 率曲線の更新用データのそれに対するシフト幅は，データ量（つまりは話者 HMM の発声変動に対する頑健性）が増えるに従い，減少すると考えられる（付録）．そこで本方法では，しきい値 $\tilde{\phi}$ を，データ量が増加して話者 HMM が発声変動に対して頑健になるにつれて，その更新用データから計算される，高めの FA 率を与える値から，等誤り率を与える値へ次第に近づくように設定する（式 (8)）．更新用データに関する等誤り率を求めるために，FR 率を次のように計算する．まず各話者ごとに，その更新用データに含まれる各サンプルについて，そのサンプルを除いた残りのすべてのサンプルを用いて，その話者のモデルを更新し，そのサンプルとのゆう度を計算する．そして，それらのゆう度から FR 率を計算する．

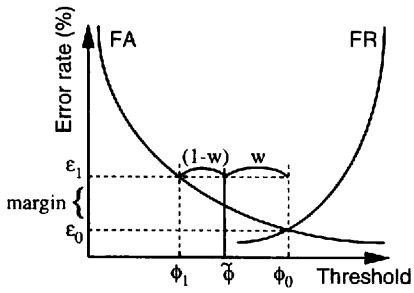

図 2 しきい値の設定法
Fig. 2 Method of updating the a priori threshold.

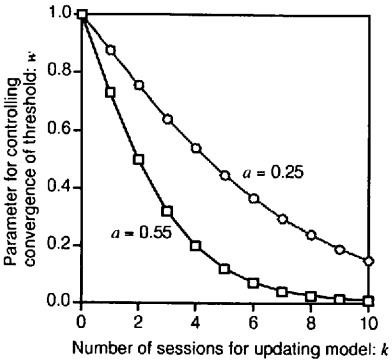

図 3 しきい値が等誤り率の値へ漸近していく度合を制御するパラメータ：w
Fig. 3 Parameter for controlling the convergence of the threshold: w.

$$\tilde{\phi} = w\phi_1 + (1-w)\phi_0 \tag{8}$$

ここで，ϕ_0 は更新用データの等誤り率を与えるしきい値を，ϕ_1 は高めの FA 率を与えるしきい値（実験的に設定）を表す（図 2）．w はオープンデータのためのしきい値が，更新用データの等誤り率の値へ近づいていく度合を制御するパラメータで，滑らかに減衰する関数（例えば，自由パラメータ a を含む式 (9)）で表す（図 3）．k は話者モデルの更新回数（$k = 0, 1, 2, \ldots$）を表す．

$$w = \frac{2}{1 + \exp(a \cdot k)} \tag{9}$$

4. 実験条件

使用したデータは，男性20名が約12か月にわたる5時期（T1～T5）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[3]から抜粋した．男性10名を登録話者，その他の男性10名を詐称者として用いた．特徴パラメータとして，標本化周波数12 kHz，フレーム長32 ms，フレーム周期8 ms，LPC分析次数16でケプストラムを抽出した．学習では時期T1に発声した10文章を用いて，各話者のモデルを生成し，時期T2，T3に発声した文章を用いて，そのモデルを更新した．テストでは，2時期T4，T5に発声した各5文章を1文章ずつ判定に用いた．学習と更新で用いた各時期の10文章のうちの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．テストに用いたテキストは，学習と更新に用いたテキストとは異なるが，全話者，全時期で同じである（表1）．実験に用いたテスト音声の総数は200文章（話者(20)×時期(2)×文章(5)）である．1文章長は平均4.2秒であった．

評価は，テキスト独立型，およびテキスト指定型[4]話者照合実験を通じて行った．本実験ではHMMのゆう度は，事後確率に基づくゆう度正規化法[5]によって正規化した．テキスト独立型においては，平均話者照合誤り率で評価した．登録話者1人当りの照合実験回数は，110回（{登録話者(1)＋詐称者(10)}×時期(2)×文章(5)）であった．各話者のモデルとしては，連続型HMM（1状態，64無相関混合ガウス分布）を用いた．また，最初に話者HMMを生成するときには，Baum-Welchアルゴリズム（推定繰返し回数3）を用いた．テキスト指定型においては，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の話者・テキスト照合誤り率で評価した．登録話者1人当りの照合実験回数は，150回（{登録話者の正しいテキスト(1)＋その登録話者の異なるテキスト(4)＋詐称者(10)}×時期(2)×文章(5)）であった．各話者の声は，各音韻ごとに半連続型HMM（3状態，256混合ガウス分布）を用いてモデル化した．音韻数は41である．また，最初に各話者の音韻HMMを生成するときには，初期モデルとして不特定話者用音韻HMMを用い，その初期モデルをBaum-Welchアルゴリズム（推定繰返し回数3）によってその話者に適応化した．

表1 学習および更新とテストの音声テキスト
Table 1 Transcription of the training/updating and testing data.

	文章テキスト
学習・更新	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). (7). ( )の五つのテキストについては，各話者， (8). 各時期ごとに異なるものを，ATR連続音声 (9). テキスト503文から抜粋して使用した． (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センチほどで日が大きくやや太っている． 3. 大声を出しす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．

なお本実験ではしきい値の設定において，式(8)のϕ_1と式(9)のaは事後的に最適な値となるように設定した．テキスト独立型ではFA率が等誤り率よりも1%ほど高い値と0.25に，テキスト指定型ではFA率が等誤り率よりも6%ほど高い値と0.55に設定した．ϕ_1に関して，テキスト独立型よりもテキスト指定型の方が高めのFA率を与える値に設定することになったのは，テキスト指定型では各話者の声を各音韻ごとにモデル化する必要があるために，推定すべきHMMパラメータ数が多いことに起因すると考えられる．モデルの更新に同じ10文章を用いた場合には，テキスト独立型の方が各パラメータの推定精度が比較的良いために，更新用データのFR率曲線とオープンデータのFR率曲線とのずれが少なくなると考えられる．なお，ϕ_1，aの事前設定は今後の課題である．

5. 結　果

5.1 モデルの更新の効果

ここでは，次の二つの更新の型について実験を行った．

バッチ型：　10文章を一度に用いて更新

逐次型：　1文章ずつ用いて更新

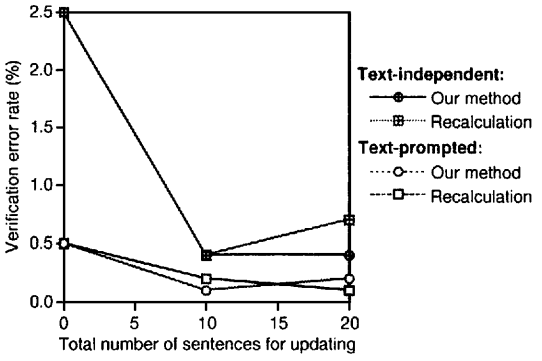

図4 10文章を用いて話者モデルを更新した場合の照合誤り率（バッチ型）
Fig. 4 Verification error rates when updating the reference models using 10 consecutive sentences.

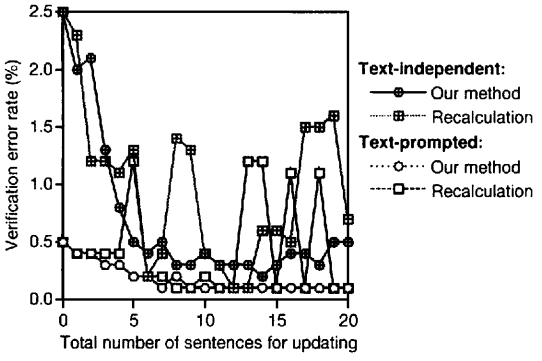

図5 1文章ずつ用いて話者モデルを更新した場合の照合誤り率（逐次型）
Fig. 5 Verification error rates when updating the reference models successively using each sentence.

図4にバッチ型で更新した場合，図5に逐次型で更新した場合のテキスト独立型（text-independent），およびテキスト指定型（text-prompted）話者照合における平均誤り率を示す．なお，話者モデルの更新の効果だけを見るために，しきい値は認識用データから計算したFR率とFA率が等しくなる値に仮に設定できたものとして，その等誤り率で評価した．"Recalculation"は，学習用データや前回までの更新用データを含むすべてのデータを用いて，Baum-Welchアルゴリズム（推定繰返し回数3）によって，話者HMMの各混合分布の平均，分散，重み係数を推定し，話者HMMを再作成する方法を表す．本方法は再作成する方法と比べて，ほぼ同等の性能を示した．本方法によって話者HMMを更新し，更新に計20文章を用いた場合の平均誤り率は，テキスト独立型，テキスト指定型ともに更新しない場合のほぼ2割（テキスト独立型：0.4/2.5，テキスト指定型：0.1/0.5）であった．

表2 再作成する方法で必要とするメモリ量を1とした場合の本方法で必要とする量（概算）
Table 2 Memory size required by our method normalized by that of the recalculation method.

	更新回数 k			
バッチ型	1		2	
	1/2 (= 10/20)		1/3 (= 10/30)	
	更新回数 k			
逐次型	5	10	15	20
	1/15	1/20	1/25	1/30

また，表2に二つの更新の型について，再作成する方法で必要とするメモリ量（≈計算量）を1とした場合の本方法で必要とする量の概算を示す．バッチ型における更新回数 $k = 1, 2$ は，時期T2，T3に発声した各10文章をそれぞれ用いて更新した場合を表し，逐次型における更新回数 $k = 5 \sim 20$ は，時期T2，T3に発声した計20文章を1文章ずつ用いて更新し，計5〜20文章を用いた場合を表す．再作成する方法では必要とするメモリ量は，初期モデルの作成に用いた10文章分に加えて，更新のたびにバッチ型では10文章分ずつ増え，逐次型では1文章分ずつ増えていく．一方，本方法では必要とするメモリ量は，更新用データの分のみである．このことから，いずれの更新の型についても，本方法で必要とするメモリ量は再作成する方法よりも少ないことがわかる．

以上の結果から，本方法による話者モデルの更新は有効であることがわかる．

5.2 しきい値の再設定の効果

10文章を一度に用いて，本方法によって話者モデルを更新した場合（バッチ型）について，しきい値を再設定する効果を調べた．図6に，本方法によってしきい値を再設定した場合のFA，FR率を示す．更新回数 $k = 0, 1, 2$ は，時期T1，T2，T3に発声した各10文章をそれぞれ用いて学習/更新した場合を表す．また，図7に，しきい値を更新用データのFA率が等誤り率よりもテキスト独立型では1%ほど高い値に，テキスト指定型では6%ほど高い値に再設定した場合，図8に，最初しきい値を学習用データのFA率が等誤り率よりもテキスト独立型では1%ほど高い値に，テキスト指定型では6%ほど高い値に設定し，その後は再設定しなかった場合，図9に，しきい値を更新用データ

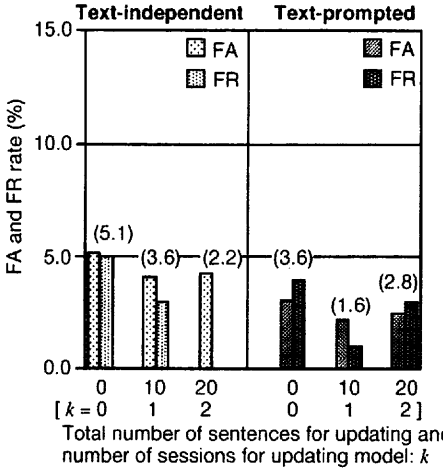

図6 本方法によってしきい値を再設定した場合の FA, FR 率 (( ): FA, FR 率の平均誤り率)

Fig. 6 FA and FR rates when using our method of resetting the a priori threshold.

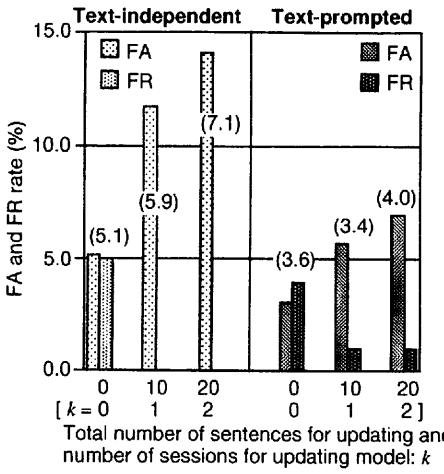

図8 しきい値を再設定しなかった場合の FA, FR 率 (( ): FA, FR 率の平均誤り率)

Fig. 8 FA and FR rates without resetting the a priori threshold.

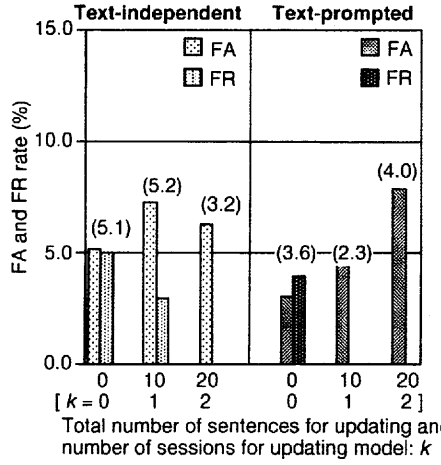

図7 しきい値を更新用データから計算した高めの FA 率を与える値に再設定した場合の FA, FR 率 (( ): FA, FR 率の平均誤り率)

Fig. 7 FA and FR rates when resetting the a priori threshold to where the FA rate is higher than the error rate at the equal error threshold.

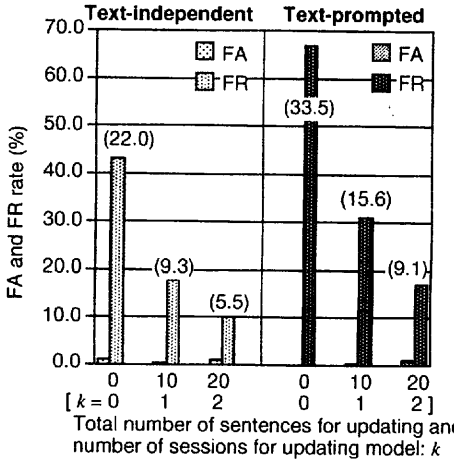

図9 しきい値を更新用データの等誤り率の値に再設定した場合の FA, FR 率 (( ): FA, FR 率の平均誤り率)

Fig. 9 FA and FR rates when resetting the a priori threshold to the equal error threshold.

の等誤り率の値に再設定した場合の FA, FR 率（および，それらの平均誤り率）を示す．

本方法によってしきい値の再設定を行った場合（図 6, $k=2$）の FA, FR 率の平均誤り率は，更新用データの等誤り率の値に設定する場合（図 9, $k=2$）と比べて，テキスト独立型ではほぼ 4 割（2.2/5.5），テキスト指定型ではほぼ 3 割（2.8/9.1）であり，またモデルやしきい値を更新しない場合（図 6, $k=0$）と比べると，テキスト独立型ではほぼ 4 割（2.2/5.1），テキスト指定型ではほぼ 8 割（2.8/3.6）であった．これらの結果は，本方法が有効であることを示している．また，本方法によってしきい値の再設定を行った場合（図 6）と，しきい値を更新用データから計算した高めの FA 率を与える値に再設定した場合（図 7）とを比較すると，本方法の方が FA 率が低かった．これは，

モデルの更新が進むにつれて，オープンデータのためのしきい値を，その更新用データから計算される，高めのFA率を与える値から，等誤り率を与える値へ近づけているためである．更に，しきい値を再設定しなかった場合（図8）では，平均誤り率が高くなった．このことから，モデルの更新に合わせて，しきい値を再設定する必要があることがわかる．図8において，特にFA率が高くなったのは，モデルの更新が進み，話者HMMがいろいろな発声変動に対して頑健になるのに従って，本人のデータに対する本人のモデルのゆう度の値が大きくなるのと同時に，話者HMMが表す特徴量空間が広がることによって，他人のデータに対する本人のモデルのゆう度の値も多少大きくなるためと考えられる．

6. むすび

本論文では，最近に発声された少量の更新用データを用いて，各話者のモデルを更新する方法，および話者のモデルの更新に合わせて，しきい値を設定する方法について報告した．話者モデルの更新に関しては，すべてのデータを蓄積しておき，話者のモデルを最ゆう推定によって再作成する方法と比べ，その近似法の一つである本方法は必要とするメモリ量（≈計算量）が少ないだけでなく，ほぼ同等の性能を示した．本方法によって，話者モデルを更新した場合の話者照合誤り率（しきい値は等誤り率を与える値に事後的に設定）は，テキスト独立型，およびテキスト指定型ともに更新しない場合のほぼ2割であった．モデルの更新に加えて，しきい値の再設定を行った場合の平均誤り率は，更新用データの等誤り率の値に設定する場合と比べて，テキスト独立型ではほぼ3割，テキスト指定型ではほぼ4割に，またモデルやしきい値を更新しない場合と比べると，テキスト独立型ではほぼ4割，テキスト指定型ではほぼ8割に低下した．

今後は，データの時期数を増やして実験を行い，本方法の効果を確かめると共に，話者HMMの各混合分布の分散を頑健に推定する方法についても検討していく．

謝辞 貴重な討論をして頂いた，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] S. Furui, "Cepstral analysis technique for automatic speaker verification," IEEE Trans., vol.ASSP-29, no.2, pp.254–272, 1981.

[2] A. Setlur and T. Jacobs, "Results of a speaker verification service trial using HMM models." Proc. Eurospeech, pp.I-53-56, 1995.

[3] 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I連続音声テキスト)," TR-I-0086, 1989.

[4] 松井知子，古井貞熙，"テキスト指定形話者認識," 信学論(D-II)，vol.J79-D-II, no.5, pp.647–656, May 1996.

[5] T. Matsui and S. Furui, "Likelihood normalization using a phoneme- and speaker-independent model for speaker verification," Speech Communication, vol.17, no.1-2, pp.109–116, 1995.

付録

学習用データ，オープンデータのFA，FR率曲線

図A·1に，テキスト独立型話者照合実験（4.）において，学習用データとして時期T1に発声した10文章を，オープンデータとして時期T2～T4に発声した各5文章を用いたときの話者ごとのFA，FR率曲線を示す．実線は学習用データから計算したFA，FR率曲線を表し，FA率を求めるときには本人を除く全登録話者（9名）の学習用データを用いた．点線は各時期のオープンデータから計算した各FA，FR率曲線を表し，FA率を求めるときには話者についてもオープンとなるように，詐称者として4.の登録話者とは異なる話者10名を用いた．特に誤り率が小さいとき（誤り率が25%以下）には，学習用データ，オープンデータの誤り率曲線のずれは，FR率の方がFA率よりも大きく，オープンデータのFR率曲線は学習用データのFR率曲線を基準とした場合に，しきい値が小さくなる方向（左側）へシフトする傾向があることがわかる．また，各時期のオープンデータの誤り率曲線のばらつきも，FR率の方がFA率よりも大きく，FR率の方が時期差による発声変動の影響を大きく受けることがわかる．

図A·2に，テキスト独立型話者照合実験（4.）において，3.のしきい値が等誤り率の値へ漸近していく度合を制御するパラメータwに関して，時期T2，T3に発声した各10文章を用いてバッチ型で更新した場合の各更新時期，各話者ごとの最適値を示す．点線はその最小2乗回帰直線を表す．更新のたびに，wの最適値は小さくなる傾向にあった．このことから，オープンデータのFR率曲線の更新用データのそれに対するシフト幅は，話者HMMが更新され，その発声変動に対する頑健性が増えるに従って，減少する傾向にあることがわかる．

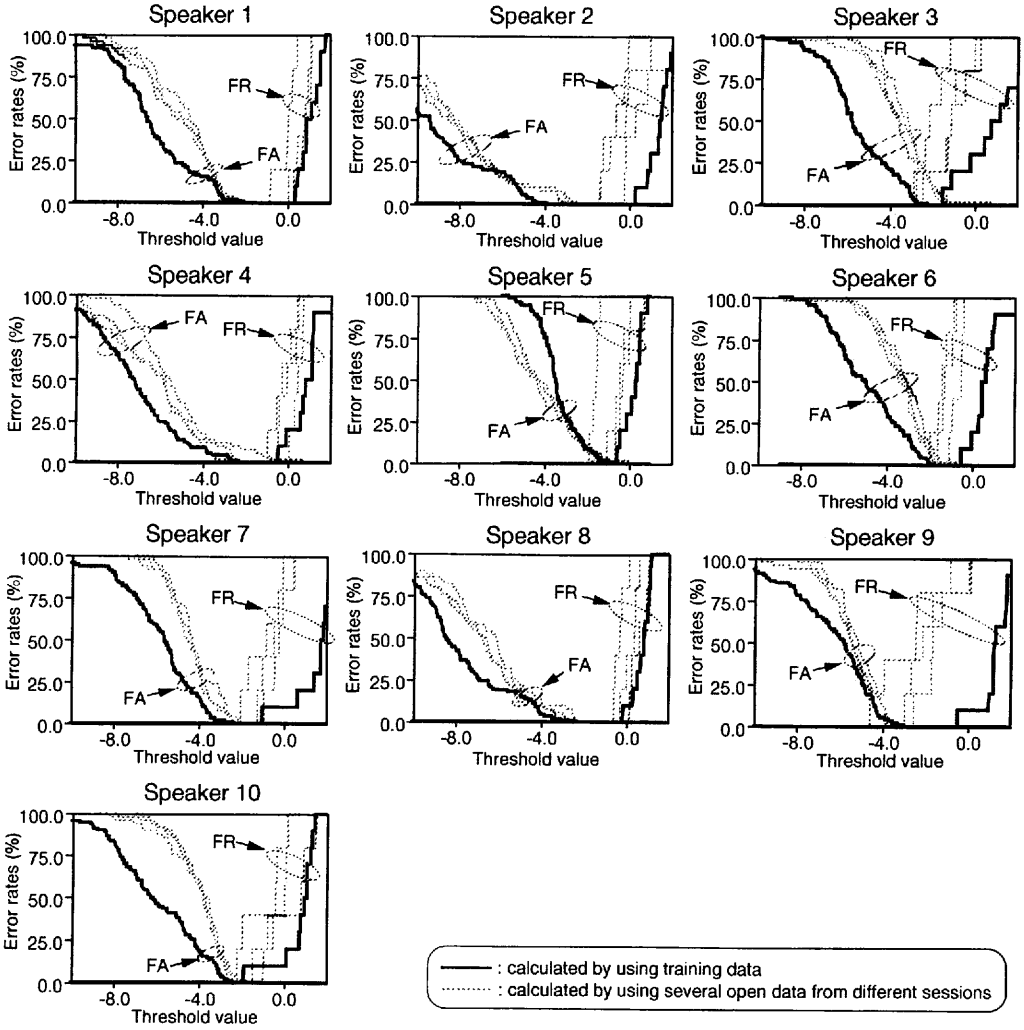

図 A·1 話者ごとの学習用データ，オープンデータから計算した FA，FR 率
Fig. A·1 FA and FR rates calculated by using training and open data for each speaker.

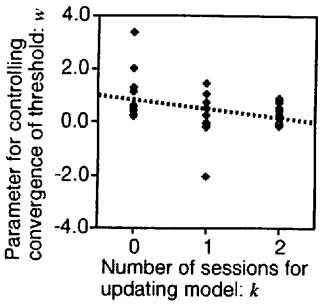

図 A·2 話者ごとのしきい値が等誤り率の値へ漸近していく度合を制御するパラメータ w の最適値
Fig. A·2 Optimal values of parameter w for each speaker. (w controls the convergence of the threshold.)

(平成 9 年 3 月 17 日受付，8 月 7 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．工博．IEEE，日本音響学会各会員．平5論文賞受賞．

西谷　隆

平8東工大・工・電電卒．同大大学院修士課程在学中．

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を経て，現在東京工業大学教授．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，米国音響学会Fellow．