論文 / 著書情報
Article / Book Information

論題(和文)	話者認識技術
Title(English)	
著者(和文)	松井 知子, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	NTT R&D, Vol. 43, No. 10, pp. 1101-1108
Citation(English)	, Vol. 43, No. 10, pp. 1101-1108
発行日 / Pub. date	1994,

話者認識技術

Speaker Recognition Technology

松井知子*　古井貞熙*
Tomoko MATSUI　Sadaoki FURUI

あらまし

話者認識は，発声者がだれであるかを自動的に判定する技術である．本論文では，3種類の話者認識法，すなわち，あらかじめ決められたキーワードを用いるテキスト依存形，任意の言葉を発声してよいテキスト独立形，装置を用いるたびに装置側から新しい言葉を指定するテキスト指定形について，それらの基本的な方法を解説する．テキスト指定形には，テープレコーダなどによって装置がだまされてしまうという問題を回避できる特徴がある．最近，米国のAT&TやTI社（Sprint社に技術提供）で，話者認識システムのフィールドテストや実用化が開始され，TI社のシステムについてはすでに広く用いられている．しかし，まだ多くの技術的な課題が未解決のまま残されている．本論文ではその中から，NTTが最近取り組んだ，特徴量の組合せ，平均量子化歪みの計算法，尤度正規化に関する課題について述べる．

Abstract

Speaker recognition is the process of automatically recognizing who is speaking. This paper introduces three speaker recognition methods : a text-dependent method that uses predetermined keywords, a text-independent method that does not rely on a specific text being spoken, and a text-prompted method in which the system prompts each user with a new key sentence every time the system is used. The text-prompted recognition system accepts the input utterance only when it decides that the registered speaker has uttered the prompted sentence. A recorded voice can thus be correctly rejected. Recently, AT&T and TI (with Sprint) started field tests and real application of speaker recognition technology, and Sprint's Voice Phone Card is already used by many customers. However, there still remain many technical problems to be solved. Among them, this paper introduces several issues that have recently been investigated at NTT Laboratories : the combination of multiple speech features, calculation of reliable average distance or distortion, and likelihood normalization.

*NTTヒューマンインタフェース研究所　NTT Human Interface Laboratories

まえがき

話者認識とは，コンピュータなどにより，発声者がだれであるかを自動的に判定することをいう．これが実現できれば，車や建物の入口の鍵，電話を用いたバンキングや買い物サービス，情報提供サービス，コンピュータのリモートアクセス，クレジットカードコールなどにおいて，音声による本人確認ができるようになる．

話者認識技術は，**図1**で示すように話者識別と話者照合に分類することができる[1]．話者識別では，入力された音声が登録話者中のだれであるかを判定する．話者照合では，ユーザは音声を入力するとともに，自分はだれであるかを申告する．そして，その音声が本当にその申告された話者の声であるかどうかを判定する．現在考えられている話者認識の応用例のほとんどは，話者照合に該当する．上述の車や建物の入口の鍵，バンキングサービスにおける本人確認なども，話者照合の応用例である．このため，本論文では話者照合に重点を置き，その具体的な方法，システムの基本的な構成，それを実現するための技術的な課題，実用例，今後の課題などについて述べる．

話者照合方法

2.1 概　　要

話者照合方法は，**表**に示すようにテキスト（発声内容）依存形，独立形，指定形の3つに分類することができる．テキスト依存形では，ユーザは“ひらけごま”のような決まった言葉（キーワード）を発声する．テキスト独立形では，ユーザは任意の言葉を発声してよい．テキスト依存形では，キーワードに含まれる固有の話者情報も利用することができるので，テキスト独立形に比べて一般的に性能が高い．言い換えると，同じ認識性能を得るためには，テキスト独立形の方が，長い音声を必要とする．いずれにしても，これらの2つの方法では，テープレコーダなどによってユーザの声を録音し，認識装置の前で再生すれば，装置がそれを受理してしまう危険性がある．

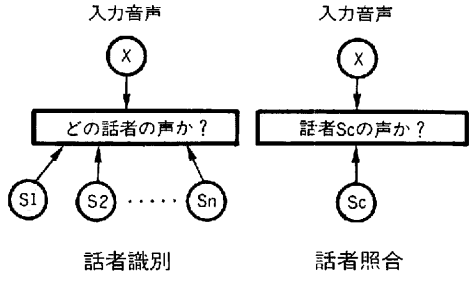

図1　話者識別と話者照合

表　話者照合方法の分類

方　法	ユーザの発声内容	特徴と問題点
テキスト依存形	「ひらけごま」など，決まった言葉（キーワード）	認識性能が比較的高い 録音した音声で騙されてしまう
テキスト独立形	任意の言葉	比較的長い音声が必要 録音した音声で騙されてしまう
テキスト指定形	認識のたびに，装置から新たに指定される言葉	録音した音声で騙されない システム構成がやや複雑になる

そこで近年著者らは，認識のたびに，任意のテキストを装置側から指定することが可能な，テキスト指定形話者照合方法を提案した[2]．この方法では，本人が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．このため，テープレコーダなどによる悪用を防ぐことができる．

以下の節では，これらの3つの方法の基本的な構成について説明する．

2.2 テキスト依存形

図2に，基本的なテキスト依存形話者照合システムの構成を示す[3]．まず学習では，話者ごとに決められたキーワードを数回発声してもらう．各発声について音声区間を検出し，音声分析して，特徴量（ケプストラム，Δケプストラムなど[3]）ベクトル（特徴ベクトル）の時系列を得る．そして，DPマッチング[4]を用いて，時間軸を合わせた後にそれらの時系列を平均化し，その話者の標準パタンを作成する．

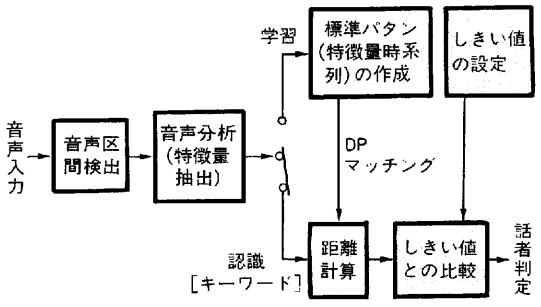

図2　テキスト依存形話者照合システムの構成

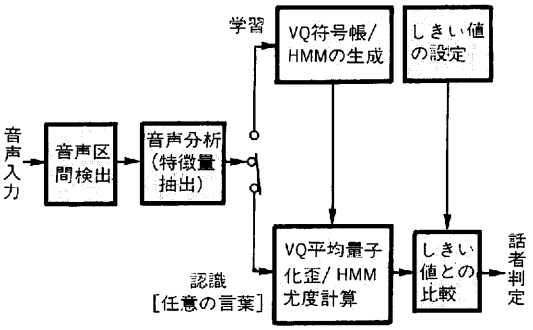

図3　テキスト独立形話者照合システムの構成

認識では，学習と同じように，入力音声から特徴ベクトルの時系列を抽出する．DPマッチングによって，その時系列と本人の標準パタンとの時間軸整合後の距離を計算し，その値を一定のしきい値と比較する．その値が，しきい値よりも大きければ本人の声として受理し，小さければ他人の声として棄却する．

最近では，特徴ベクトルの時系列をそのまま標準パタンとして用いる代りに，各話者ごとにそれを隠れマルコフモデル（HMM；hidden Markov model）[4]に変換して，蓄えておく方法も用いられている．入力音声が各話者のHMMから生成された尤度を計算し，その値を前述の方法と同様にしきい値と比較して，話者照合の判定を行う．

このようにテキスト依存形では，キーワードに含まれる音韻系列固有の特徴ベクトル系列として，各話者の標準パタンあるいはモデルが表される．このため，比較的短い音声でも，安定した話者の特徴を得ることができる．また，学習に必要なデータ量も，キーワードが固定されているために比較的少なくてすむ．実際の電話音声を用いた実験で，高い認識精度が達成されている[2]．実用的には，話者ごとにキーワードを変えることによって，その違いも話者の判定に用いることができるので，一層の精度向上が期待できる（もちろん，キーワードが盗まれないことが前提）．

2.3　テキスト独立形

テキスト独立形ではテキストが可変であるために，テキスト依存形のように，特定のテキストに対応した特徴ベクトル時系列として，各話者の標準パタンを表すことができない．**図3**に基本的なテキスト独立形話者照合システムの構成を示す[5]．学習では，種々のテキスト（音韻）に対する各話者の音声の特徴を得るために，話者ごとに10文章程度の発声をしてもらう．その音声区間を検出し，特徴ベクトル（ケプストラム）の時系列を抽出する．そして，その特徴ベクトル時系列を，**ベクトル量子化**†(VQ；Vector Quantization)[6][7]によってパタン化するか，**エルゴード形**†のHMM[8][9]によってモデル化する．VQによる方法では，時間軸を無視してひとまとめにした特徴ベクトル集合をクラスタ化し，各クラスタの重心（セントロイド）を話者特徴（符号帳；サイズは256等）の要素として蓄える．HMMによる方法では，時間変動を，状態（1～5状態等）間の遷移と特徴ベクトルの出力確率（密度）によって表し，各状態における出力確率は混合ガウス分布（64混合ガウス分布等）で近似する．

認識では，学習と同じように，入力音声から特徴ベクトル時系列を抽出する．VQによる方法では，その時系列を本人の符号帳でベクトル量子化し，その時系列全体にわたる平均量子化歪みを求め，その値をしきい値と比較する．HMMによる方法では，本人のHMMからその時系列が出現する尤度（確率）を計算し，その

†論文中のゴシック文字は，本論文末に解説があります．

値をしきい値と比較する．この比較の結果に基づいて，話者照合の判定を行う．

2.4 テキスト指定形

図4に基本的なテキスト指定形話者照合システムの構成を示す[2]．テキスト指定形では，本人が正しく指定されたテキストを発声したか否かを判定するために，本人がそのテキストを発声した場合の音声のモデルを生成する必要がある．システムの側から任意のテキストが指定できるためには，各話者ごとにすべての基本的な音韻のモデルを学習しておき，それを接続して，テキストどおり発声した音声のモデルを生成すればよい．このため，まず学習では，各話者ごとに10文章程度の発声をしてもらう．各発声から特徴ベクトル（ケプストラム）の時系列を抽出した後，各音韻（/a/や/i/などの言葉の短い単位）のHMM（半連続形，3状態，256混合ガウス分布）を作成する．

その際，各音韻モデルにいかに正確に音韻性・話者性を持たせるかが重要な課題である．10文章程度の学習データだけでは，各音韻モデル(HMM)のパラメータを推定するには十分ではない．しかし，話者認識では各話者に多量の発声を要求するような方法は現実的ではないので，これ以上文章の数を増やすのは適当ではない．そこで，多数話者のデータ（話者60名，計9,000文章)から生成した不特定話者用音韻モデルを初期モデル（たね）とし，各話者の学習データを用いて，その初期モデルをその話者に適応化する．こうして，その話者の音韻モデルを生成する．

図5にその処理の流れを示す．学習データのテキストは既知なので，そのテキストに従って不特定話者音韻モデルを接続して，そのテキストのモデル(テキストモデル)を生成する．それに音声を与えて，**Baum-Welchアルゴリズム**[3]により，テキストモデル中のパラメータ（混合ガウス分布の平均値と重み係数）を推定する．混合ガウス分布の分散の値は，少ない学習データ量から推定することは難しいので，不特定話者用音韻モデルの値をそのまま用いる．その後，そのテキストモデルを各音韻モデルに分ける．以上を全学習データについて何度か繰り返し，各音韻モデルのパラメータを話者に適応化する．

著者らは，上記の適応化法と，音韻を明示的に与えずに半連続形HMMの基底の混合正規分布を適応化する音韻独立学習を，交互に繰り返す学習法によって，不特定話者音韻モデルを各話者に適応化する方法を提案した[10]．この方法を用いれば，学習データに少ししか含まれていない音韻のモデルについても，その話者の声に適応化できる．

認識では，システム側から指定したテキストに従って，本人の音韻モデルを連結し，そのテキストのモデルを作る．そして，入力音声から得られた特徴ベクトルの時系列が，そのテキストモデルから出現する尤度を計算し，その値をしきい値と比較して，話者の判定を行う．話者15名の音声データで評価した結果，前記の適応

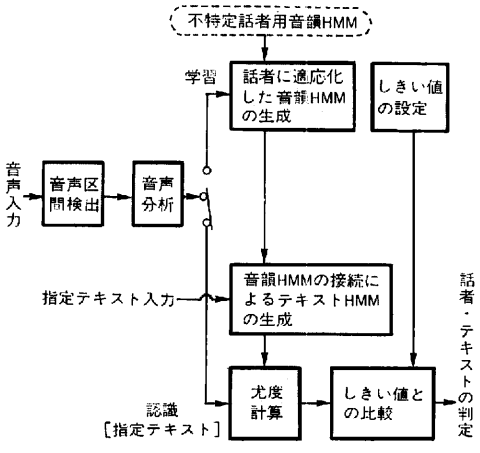

図4 テキスト指定形話者照合システムの構成

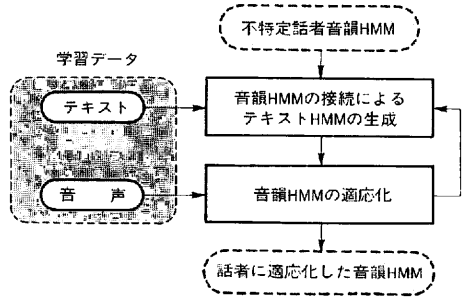

図5 話者に適応化した音韻モデルの作成法

化法により，99.3%の話者照合率が得られることが確認された．

3 技術的な課題

3.1 概　　要

話者認識技術には，いろいろな研究課題がある．例えば，話者照合システムにとって，声が時間の経過につれて変わるという自然の摂理は重大な問題であり，各話者の標準パタンやモデルを定期的に更新していく仕組みを考える必要がある．また，話者を判定するときのしきい値について，事前に最適な値に設定する方法に関する研究[2]は少なく，今後も検討していくことが必要である．さらに，少なくとも日本では，電話を通した場合，雑音のある環境，多数話者の場合など，実環境に則した話者認識性能の評価がまだ十分に行われていない．以下の節では，特徴量の組合せ，平均量子化歪みの計算法，尤度正規化に関する課題などを紹介するとともに，それらに対する著者らのアプローチについて述べる．

3.2 特徴量の組合せ

一般に，音声から抽出される特徴量には，声道の共振・反共振特性を表したスペクトル包絡に関するものと，音源（声帯の振動など）を表したスペクトル微細構造に関するものがある．前者の代表的な特徴量としてはケプストラム，その**線形回帰係数**†のΔケプストラム，後者の代表的な特徴量としてはピッチ（基本周波数，声の高さ），その線形回帰係数の**Δピッチ**†がある．話者認識では，2章でも紹介したように，スペクトル包絡に関する特徴量，特にケプストラムを用いることが多い．これは，ケプストラムを用いれば，高い話者認識性能が得られることが実験的に分かっており，また，その係数は安定して抽出することができるからである．それに比べ，ピッチは有声音区間からのみ抽出され，無声音区間では定義できない．有声音区間でも自動的に正確にピッチを抽出することは難しい．また，ピッチは話者による違いも大きいが，声の高さをいろいろ変えられるように，話者内でのばらつきも大きいため，その扱いが難しい．

しかしながら，人間は声の音色，高さ，大きさなど，様々な情報から個人性を知覚しており，テキストや発声時期の違いにも対応できる．このような観点から，複数の特徴量を組み合わせて用いた方が，安定した認識ができると思われる．

そこで著者らは，信頼度の高い区間のみのピッチ情報を，ケプストラムとうまく組み合わせる方法を試み，それによって，認識性能が向上することを示した[11]．**図6**に，そのケプストラムとピッチを用いた，VQによるテキスト独立形話者照合システムの構成を示す．この方法では，各話者ごとに**有声音/無声音**†区間で別々に符号帳を用意する．有声音区間の特徴ベクトルの要素は，ケプストラム，Δケプストラム，ピッチ，Δピッチであり，無声音区間（ピッチが抽出できない区間）の特徴ベクトルの要素は，ケプストラム，Δケプストラムである．入力音声に対

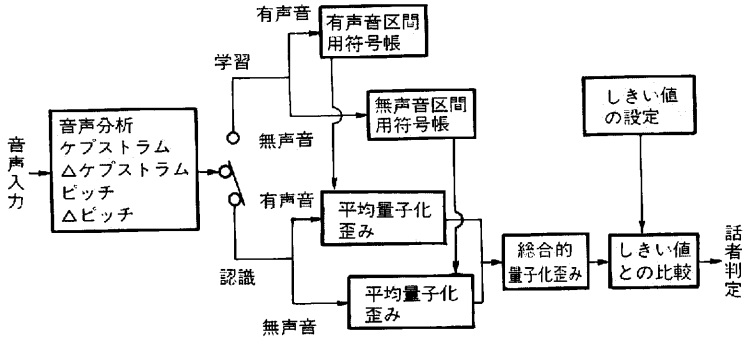

図6　音源-声道特徴を用いたテキスト独立形話者照合システムの構成

して，有声音/無声音の判定を行った後，それぞれに対応する符号帳を用いてベクトル量子化を行い，最後に両者の歪みを加え合わせて最終的な判定を行う．話者9名の音声データを用いて，この方法の評価を行った結果，98.7%の話者照合率が得られることが確認された．

一般に，特徴量は，静的な特徴量（ケプストラム，ピッチなど）と，動的な特徴量（Δケプストラム，Δピッチなど）に分けて考えることができる．Δケプストラムは，テキスト依存形の方法においては，ケプストラムと組み合わせて用いると，認識性能が向上することが確かめられているが[3]，HMMによるテキスト独立形の方法においては，その組合せ効果が認められていない．今後は，動的な特徴量について更に検討していく必要があろう．

3.3 平均量子化歪みの計算法

テキストが違うと，それを発声した音声から抽出される特徴ベクトルの分布も異なる．テキスト独立形では，認識時に任意のテキストの発声を許すので，そのテキストに学習データにはない音韻や言葉が含まれることがある．そのような部分に相当する特徴ベクトルは，本人であってもその学習データの分布から離れている可能性があり，入力音声と本人の符号帳との平均量子化歪み，あるいはHMMとの尤度を求めるときのノイズになる．また，上でも述べたように，同じ人の同じ言葉でも，発声時期が異なるとその物理的特性は変化する．

そのために，著者らはVQによる方法において，入力音声の中から，それと学習データに共通に含まれている言葉の部分や，体調や時期によって変化しにくい部分を自動的に選択して用いる方法を提案した[11]．話者9名の音声データで評価した結果，この方法を用いれば，すべての音声区間を用いる場合に比べて，平均話者照合誤り率がほぼ半分になることが確認された．

このアプローチは，テキスト独立形の話者認識に限らず，予測できない変動を許容しなければならないパタン認識の問題に，広く適用できると考えられる．

3.4 尤度正規化

テキストや発声時期の違いによって，入力音声と本人の符号帳/HMMとの平均量子化歪み/尤度の値が大きくばらつき，話者の判定のためのしきい値を安定して設定することが難しいという問題がある．このため著者らは，HMMによる方法において，次のような事後確率$P(s_i|x)$を用いることによって，尤度の値を効果的に正規化する方法を示した[12]．

$$P(s_c|x)=\frac{P(x|s_c)\times P(s_c)}{\sum_i\{P(x|s_i)\times P(s_i)\}}\cong\frac{P(x|s_c)}{\sum_i P(x|s_i)}$$

ここで，xは入力音声，s_iは話者の識別子，特に，s_cは本人であると申告された話者を示す．$P(s_i)$は，話者s_iの出現確率であるが，各話者共通であると仮定し，キャンセルする．$P(x|s_i)$は，話者s_iのHMMに対する入力音声xの尤度である．この式を右辺のとおりに計算すると，$\Sigma P(x|s_i)$を求めるために，全登録話者のモデルについての累積尤度を計算する必要があり，その計算量は膨大となる．そのため，著者らは全登録話者の声を併せ持つ単一の音韻・話者独立モデルを作成し，この累積尤度をそのモデルに対する尤度で近似した．

音韻・話者独立モデルの作成法は次のとおりである．まず，登録話者とは異なる多数の話者の音声データを用いて，不特定話者の特徴量空間を広く覆うように多数話者モデル（256混合ガウス分布）を作成する．ここでは男30名，女30名が発声した，9,000文の音声データを用いた．次に，各登録話者ごとに学習データを用いて，多数話者モデルの混合ガウス分布の重みだけをBaum-Welchアルゴリズムで推定する．その後，全登録話者について各混合ガウス分布ごとに重みを平均し，音韻・話者独立モデルを生成する．新しい話者を登録する場合は，その新しい話者の学習データを用いて，多数話者モデルの混合ガウス分布の重みを推定する．新しい話者も含めて全登録話者について，各混合ガウス分布ごとに重みを平均するだけで，音韻・話者

独立モデルを更新することができる．この更新に必要とされる計算量は極めて少ない．話者15名の音声データで評価した結果，この正規化法を用いれば，平均話者照合誤り率が，正規化前のほぼ半分になることが確認された．

3.5 データベース

話者認識に関する研究を継続して行っている主な機関は，米国ではAT&T，TI社，ITT社BBN，英国ではSwansea大学，フランスではTelecom-Paris，日本ではNTTなどであるが，これまで各機関では，独自のデータベースを用いて，それぞれの方法を評価してきた．しかし，各方法を比較するためには，共通のデータベースを使用する必要がある．また，大規模なデータベースを単独の機関で構築するとなると，その機関に多大な負担が強いられる．このため，現在，各機関の合意の下に，共通データベースを作成するためのプロジェクトが検討されている．

4 実用化の例

4.1 Sprint社

米国では，クレジットカード式のテレホンカードが用いられており，その盗難などによる悪用が社会的な問題となっている．米国Sprint社では，TI社の技術を用いて，テレホンカード（Voice Phone Card）に話者照合機能を付加し，セキュリティ性の向上を図るサービスを1994年1月より開始した．話者照合機能自体は極めて高い認識性能を有しているわけではないが，本人がカードを紛失する確率，それを悪用される確率，話者照合の確率を掛け合わせれば，非常に高いセキュリティ性が得られるというのがセールスポイントである．具体的には，ガイドメッセージに従い，社会保障番号（social security number；米国では1人ひとりにこの番号が与えられている）10桁を発声する．うまく本人確認されれば，例えば“call Miyawaki”（Miyawakiさんの名前はあらかじめ登録しておく）と発声すると，Miyawakiさんに電話が掛かる．この方法は，発声する言葉が決まっているので，テキスト依存形の方法と言える．すでに10万人のユーザによって用いられていると言われている．

4.2 AT&T

米国AT&Tでも1993年より，テレホンカードや通常のクレジットカードに話者照合機能を付加するためのフィールドテストを行っている．ユーザはあらかじめ，10数字あるいは7種類程度の短い言葉を発声し，カードにそれぞれの言葉に対応したユーザの声の特徴がパタンとして蓄えられる．装置には，入力された音声とカードに蓄えられたパタンとの照合を取る機能が取り付けられている．照合時には，装置側でその言葉の中から複数をランダムに選び，ユーザに提示する．ユーザがその言葉を順に正しく発声したときのみ，本人であると確認される．この方法は，発声する言葉の種類は決まっているが，毎回ランダムに言葉が選ばれて，装置側から指定されるという点で，テキスト依存形と指定形の中間的な方法であると言える．

5 あとがき

本論文では，話者照合の基本的方法，特に，テキスト依存形，独立形，指定形について解説し，技術的な課題，応用例について述べた．現在のところ，高品質マイク，静かな環境で録音したデータに関しては，ケプストラムを特徴量として用いた場合，発声時期が異なっても，数十名の話者について，98.6%のテキスト独立形話者照合率，98.9%のテキスト指定形話者照合率が得られている．今後は実際の電話を通した場合，雑音のある環境など，実環境に則した評価を行っていく必要がある．話者照合では本人か他人かの2値の判定を行うので，理論的には照合率は登録話者の数には依存しないが，実際に極めて多数の多様な話者を含む場合についての実験を行う必要がある．

文　献

(1) 古井：話者認識，テレビジョン学会誌，47，No.12，

pp.1600-1603, 1993.
(2) 松井・古井：連結音韻モデルによる話者認識，信学技報，SP92-128，1993.
(3) S.Furui：Cepstral analysis technique for automatic speaker verification, IEEE Trans. Acoust., Speech, Signal Processing, ASSP-29, No.2, pp.254-272, 1981.
(4) 古井：音響・音声工学，近代科学社，東京，1992.
(5) 松井・古井：VQ，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，J77-A，No.4，pp.601-606，1994.
(6) K.P.Li and E.H.Wrench Jr.：An approach to text-independent speaker recognition with short utterances, Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Boston, MA, 12.9, pp.555-558, 1983.
(7) F.K.Soong and A.E.Rosenberg：On the use of instantaneous and transitional spectral information in speaker recognition, Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Tokyo, Japan, 17.5, pp.877-880, 1986.
(8) A.B.Poritz：Linear predictive hidden Markov models and the speech signal, Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Paris, S11.5, pp.1291-1294, 1982.
(9) M.Savic and S.K.Gupta：Variable parameter speaker verification system based on hidden Markov modeling, Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Albuquerque, S5.7, pp.281-284, 1990.
(10) 松井・古井：テキスト指定形話者認識のための話者適応化法，日本音響学会春季研究発表会，千葉，3-7-5，1994.
(11) 松井・古井：音源・声道特徴を用いたテキスト独立形話者認識，信学論，J75-A，No.4，pp.703-709，1992.
(12) 松井・古井：音韻・話者独立モデルによる話者照合尤度の正規化，電子情報通信学会技報，SP94-22，1994.

■ 用語解説 ■

ベクトル量子化

特徴量をパラメータごとに（スカラー）量子化するのではなく，複数のパラメータの組（ベクトル）をまとめて，1つの符号で表現する方法．この方法は，情報速度・歪み理論に基づく情報理論的な帯域圧縮の限界に漸近的に近づく方法として，広く用いられている．

エルゴード形

HMMの構造として，すべての状態相互間の遷移を許した対称形のモデル．

Baum-Welchアルゴリズム

HMMのパラメータ（初期確率，遷移確率，出現確率）推定において，学習用音声に対するHMMの尤度が最大とするように推定するアルゴリズム．直接解析的に推定することはできないので，EM（expectation-maximization）アルゴリズムと呼ばれる方法を用いる．

線形回帰係数

あるパラメータ列を，2乗誤差が最小になるように直線で近似して，その傾きを表した係数．

Δピッチ

ピッチの時間パタンの線形回帰係数．

有声音/無声音

声帯振動を伴う音を総称して有声音，伴わない音を総称して無声音と呼ぶ．有声音には母音と有声子音，無声音には無声子音が含まれる．有声音にはピッチがあるが，無声音にはピッチがない．

松井　知子

ヒューマンインタフェース研究所古井特別研究室研究主任
昭和63年入社，以後，ヒューマンインタフェース研究所にて，話者認識の基礎研究に従事．
昭和61年東京工業大学理学部情報科学科卒業，63年同大学院修士課程修了．
IEEE・日本音響学会・電子情報通信学会会員．
平成5年度電子情報通信学会論文賞・学術奨励賞受賞．

古井　貞熙

ヒューマンインタフェース研究所古井特別研究室長
昭和45年入社，主として音声認識，話者認識，音声知覚などの基礎研究に従事．53～54年米国ベル研究所客員研究員，平成1年ヒューマンインタフェース研究所音声情報研究部長，3年より現職，6年より東京工業大学客員教授．
昭和43年東京大学工学部計数工学科卒業，45年同大学院工学系研究科修士課程修了，53年工学博士（東京大学）．
IEEE Fellow・ESCA・電子情報通信学会・日本音響学会会員．
昭和50年電子情報通信学会業績賞，63年および平成5年同論文賞，平成2年同著述賞，昭和60年および62年日本音響学会論文賞，平成元年科学技術庁長官賞，同年IEEE ASSP Society Senior Award各受賞，1993年IEEE Signal Processing Society Distinguished Lecturer．