論文 / 著書情報
Article / Book Information

論題(和文)	話者認識技術の高度化を目指して
Title(English)	
著者(和文)	古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1993年秋季講演論文集, , , pp. 611-612
Citation(English)	, , , pp. 611-612
発行日 / Pub. date	1993, 10

スペシャル・セッション〔音質・声質をめぐって〕（招待講演）

1—7—10

話者認識技術の高度化を目指して*

○古井　貞熙（NTTヒューマンインタフェース研究所）

1．話者認識とは

音声に含まれる個人性情報を用いて、誰の声であるかを自動的に判定することを、話者認識(speaker recognition)という1)2)3)4)。これが実現できれば、電話によるバンキング、買い物、情報案内サービス、音声メール、コンピュータのリモートアクセスなどにおいて、音声によって誰であるかが確認できるので、極めて便利になると期待されている。

話者認識の形態は、話者識別(speaker identification)と話者照合(speaker verification)に分けられる。話者識別では、入力音声が、あらかじめ登録されている誰の声であるかを判定する。一方話者照合では、入力音声と同時に自分が誰であるかを名乗り、その音声が本当にその話者の声であるか否かが判定される。音声を本人確認の鍵として用いる応用のほとんどは、話者照合に該当する。話者識別の性能は、登録話者の数が大きくなると低下するが、話者照合の性能は、登録話者の数がある程度以上大きければ、その数には依存しない。

話者認識の方法はさらに、あらかじめ決まっている言葉（キーワード）を発声する発声内容依存（限定）型と、任意の言葉を発声してよい発声内容独立型に分類できる。前者の場合は、各話者のキーワードの標準パターン（あるいはモデル）と入力音声との時間軸を整合させて、全体としての類似の度合いを調べる方法をとることが多い。この際、音韻に依存した個人性情報を直接的に用いることができるので、一般に発声内容独立型の方法よりも比較的高い認識性能を得ることができる。しかし、応用によっては、決まった言葉を用いるのが難しい場合もある。また、人は発声内容にかかわりなく、話者を認識することができる。このため、発声内容独立型やそれを基本とする方法が、最近活発に研究されている。

さらに発声内容依存型の方法には、本人のキーワードの音声を録音して装置の前で再生すれば、本人になりすまして機械を容易にだますことができるという問題もある。このため、我々は最近、後に紹介する発声内容指定型の話者認識法を提案した。

ここでは主に、発声内容独立型と発声内容指定型の話者認識技術の動向と今後の課題について述べる。

2．話者認識に用いる特徴パラメータと正規化法

音声に含まれる情報は、音源に関連したピッチ（基本周波数）、発声レベルなどの情報と、声道に関連したスペクトル包絡情報に分けることができる。話者認識には、この内の後者、あるいは両者が組み合わせて用いられる。スペクトル包絡を表現する方法としては、音声認識の場合と同様に、ケプストラム（対数スペクトルの逆フーリエ変換）とその線形回帰係数（Δケプストラム）が用いられることが多い。

これらの特徴パラメータは、言葉（音韻）が異なれば、勿論大きく変化する。同じ人の同じ言葉でも、指紋と違って、発声のたびに変化する。数か月も経過したり、異なるマイクロホンや電話機、伝送系を通したりすると大きく変化することもある。周囲雑音の付加によっても変化する。このため、これらの変化を吸収するようなパラメータ変換や、統計的処理、パターン認識処理などを施すことが必要である。音声認識の場合は、認識を行なう直前に声の個人差に対する学習や適応化を行なうことが可能であることが多いが、話者認識の場合は、登録の時期と認識の時期は異なるのが普通である。さらに個人性情報は、音韻性情報よりも一段階細かい情報であるから、特徴パラメータの種々の変化の影響を受けやすく、それへの対処すなわち正規化法は、極めて重要な課題である。

現在用いられているパラメータレベルでの正規化法としては、平均ケプストラムを差し引く方法がある1)。パターン認識レベルでの最も効果的な正規化方法は、尤度比を用いる方法5)と、事後確率を用いる方法6)である。いずれの場合も正規化のための話者セットを用意しておいて、それらとの尺度値を差し引く（あるいは割り算をする）ことによって正規化を行なう。これらの正規化用話者セットの設定法と、その用い方は、重要な課題の一つである。

3．発声内容独立型話者認識

発声内容独立型の方法には、発声内容（音韻）

* Toward the advanced speaker recognition technology by Sadaoki FURUI (NTT Human Interface Laboratories)

による変動と話者による変動を分離する方法と、両者を一括して取り扱う方法が考えられる。任意の発声内容の音声から両者の変動を分離し、音韻に依存した個人性情報を直接的に抽出するためには、音声認識を行なって、各音韻の区間をセグメンテーションしなければならない。不特定話者の任意語彙の音声認識系を用いてこれをある程度行なうことも可能であるが、構成が複雑になるのと、認識誤りに対する対処の問題などがある。そこで、現在試みられている方法では、音韻と話者による変動を一括して取り扱う方法が中心である。

これらの方法はさらに、表1に示すように、音韻に依存した個人性情報を利用する方法と、利用しない方法に分けることができる。後者の比較的古典的な方法に、長時間平均スペクトルのように、比較的長時間（数10秒程度）の音声にわたってパラメータを平均化することによって、発声内容の変動の影響をキャンセルする方法がある。この方法には、長時間の音声が必要で、しかも平均化によって多くの情報が失われてしまうために、認識性能に限界がある。

音韻に依存した情報を利用しない新しい方法として、最近、パラメータの動的特徴に含まれる個人性特徴を、ケプストラムの多次元自己回帰（MAR; Multivariate Auto-Regressive）モデルを用いて表現する方法が提案されている7)。この方法を用いた実験の結果、類似度尺度を適切に選択すれば、かなり高い認識性能が得られることが確かめられている8)。

間接的ではあるが、音韻に依存した個人性情報を用いることができる有効な方法として、ベクトル量子化（VQ）による方法とHMM（hidden Markov model; 隠れマルコフモデル）による方法がある。その構成の例を図1に示す9)10)。いずれの場合も、まず各登録話者について、任意の文章を発声した学習音声から、スペクトル包絡を表す特徴ベクトル、あるいはそれにピッチを組み合わせたベクトルを短時間ごとに抽出する。VQによる場合は、それらをクラスタ化して、その話者固有の符号帳を作成する。

表1　発声内容独立型話者認識法の分類

音韻別個人性情報の利用	方法の例
な　し	長時間平均スペクトル（パラメータ） 多次元自己回帰（MAR）モデル
あ　り	ベクトル量子化 マトリックス量子化 HMM ニューラルネット

HMMによる場合は、符号帳の代わりに、各登録話者の学習用音声からHMMを作成する。各話者の特徴は、モデルの違い、具体的には状態遷移確率と、各状態におけるパラメータの出現確率分布（混合ガウス分布）の違いによって表現される。ピッチは有声音区間からのみ抽出されるので、それを用いる場合は図1に示すように、各話者について、有声音区間用と無声音区間用の符号帳あるいはHMMを別々に作成する。

認識時には、VQによる場合は、入力音声を各登録話者の符号帳でベクトル量子化し、入力音声全体にわたる平均量子化歪みを求める。HMMによる場合は、入力音声を各登録話者のモデルに与えたときの、音声区間全体の平均尤度を求める。それらの値によって、話者識別あるいは話者照合の判定を行なう。

実験の結果9)10)、VQによる方法とHMMによる方法では、ほぼ同程度の認識性能が得られること、後者の場合、HMMの状態間の遷移情報は個人差の表現にはほとんど寄与しないため、状態数の増加と状態内の混合分布数の増加は、同程度の認識性能向上効果があり、性能はほぼ両者の積によって決まることがわかった。また、いずれの方法の場合も、歪みあるいは尤度を計算する際に、入力音声のすべての部分を用いずに、符号帳あるいはモデルとの重なりのある部分のみを用いた方が（DIM法）、学習音声と入力音声に含まれる音韻の違い、時期による音

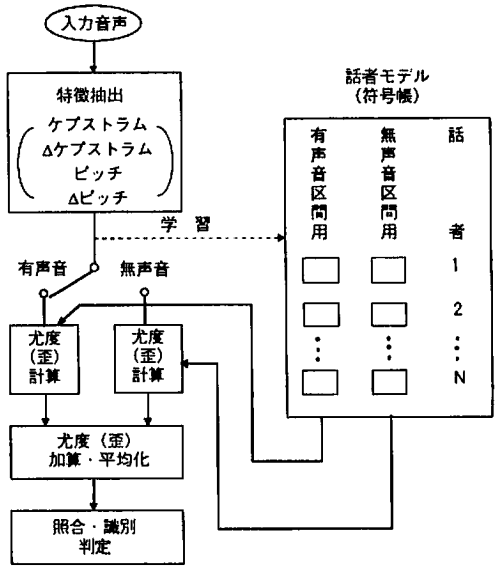

図1　発声内容独立型話者認識システムの構成