

【技術分類】 6 - 2 - 3 音声 / 照合・判定技術 / 話者照合方式

【 F I 】 G10L3/00

【技術名称】 6 - 2 - 3 - 1 テキスト依存方式

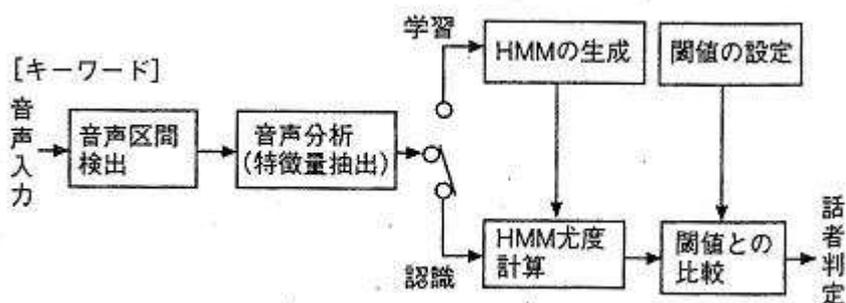
【技術内容】

あらかじめ決めておいたテキスト(パスワード)を発声させ、話者を照合する方法である。基本的なシステム構成を図に示す。学習では、話者ごとに決められたキーワードを数回発声させる。各音声について音声区間を検出し、音声分析して、特徴ベクトル(ケプストラム、ケプストラムなど)の時系列を得る。その時系列を、HMMによりモデル化する。認識では、学習と同様に、入力音声から特徴ベクトルの時系列を抽出する。本人のHMMからその時系列が出現する確率(尤度)を計算し、閾値と比較する。

このようにテキスト依存型では、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、各話者のモデルが表わされる。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、比較的少なくてすむ。

【図】

テキスト依存型話者照合システム



出典:「HMMによる話者認識」,「電子情報通信学会技術研究報告 音声 SP95-111 pp.17-24」,1996年1月発行、松井知子著、電子情報通信学会発行、18頁 図1 テキスト依存型話者照合システムの構成

【出典 / 参考資料】

「HMMによる話者認識」,「電子情報通信学会技術研究報告 音声 SP95-111 pp.17-24」,1996年1月発行、松井知子著、電子情報通信学会発行

【技術分類】 6 - 2 - 3 - 2 音声 / 照合・判定技術 / 話者照合方式 / テキスト独立方式

【 F I 】 G10L3/00

【技術名称】 6 - 2 - 3 - 2 - 1 ベクトル量子化歪方式

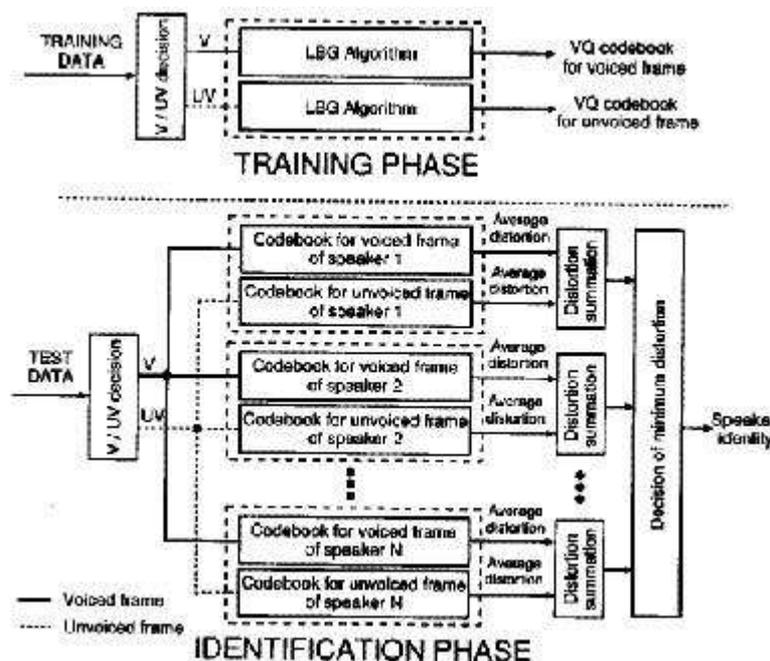
【技術内容】

入力音声をベクトル量子化(Vector Quantization : VQ)したときの、量子化歪の量を用いて照合する方法である。まず登録話者ごとに充分長い文を発声させ、その短時間スペクトルの集合をクラスタ化し、各クラスタの重心(セントロイド) を符号帳(コードブック)の要素とする。入力音声は各登録話者の符号帳でベクトル量子化し、入力音声全体の平均量子化歪を求め、話者識別は、最も平均量子化歪の小さい符号帳の話者を選択する。話者認証は、平均量子化歪を閾値と比較して判定する。

図は、LBG アルゴリズムを用いて符号帳を作成し、話者を識別するシステムの例である。入力音声はまず有声音 / 無声音区間に分け、それぞれを LBG アルゴリズムにより話者ごとの符号帳を作成する。識別では同様に入力音声を有声音 / 無声音区間に分け、それぞれを符号帳によりベクトル量子化し、総合的な量子化歪の最も小さい登録話者とで最終的な判定を行う。

【図】

ベクトル量子化歪によるテキスト独立型話者認識システム



出典 : 「線形予測残差スペクトルの調波構造に含まれる個人性情報を用いた話者認識」, 「電子情報通信学会論文誌 A Vol.J80-A No.9 pp.1360-1367」, 1997 年 9 月発行、早川昭二 武田一哉 板倉文忠著、電子情報通信学会発行、1360 頁 図 6 ベクトル量子化(VQ)ひずみを利用した話者識別システム

【出典 / 参考資料】

・「音声情報処理」, 「電子情報通信工学シリーズ」, 1998 年 6 月 30 日発行、古井貞熙著、森北出版株式会社発行、142 頁

・「線形予測残差スペクトルの調波構造に含まれる個人性情報を用いた話者認識」, 「電子情報通信学会論文誌 A Vol.J80-A No.9 pp.1360-1367」, 1997 年 9 月発行、早川昭二 武田一哉 板倉文忠著、電子情報通信学会発行

【技術分類】 6 - 2 - 3 - 2 音声 / 照合・判定技術 / 話者照合方式 / テキスト独立方式

【 F I 】 G10L3/00

【技術名称】 6 - 2 - 3 - 2 - 2 HMM 方式と GMM 方式

【技術内容】

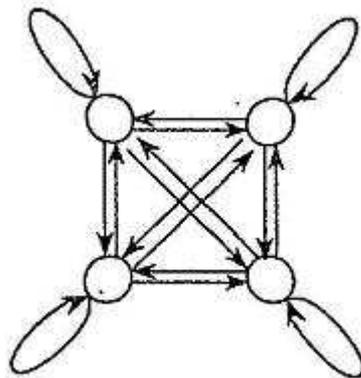
HMM を用いて話者認識する方法について先ず説明する。テキスト独立方式では任意の発声内容を表現する必要があるため、図のように、任意の状態からスタートして任意の状態に遷移し、任意の状態を終端とする、エルゴード型 HMM で表現されることが多かった。このモデルから、入力音声のスペクトル系列が出現する尤度を計算し、話者を認識する。HMM の学習は、話者ごとに、学習用に用意された文章や多数の単語を発声した音声を用いる。

尚、状態間の遷移は発声内容により変化するため、テキスト独立方式の場合、遷移確率の評価は無意味であり、出力確率のみが意味を持つ。この場合、全状態でのガウス分布の総数さえ同じなら、状態数がいくつであっても同じ結果となる。従って、現在ではエルゴード型 HMM が使われることはほとんどなく、混合ガウス分布で記述した 1 つの状態のみを用いる HMM、すなわち混合ガウス分布モデル(Gaussian Mixtures Model : GMM)がよく用いられる。

テキスト独立型話者認識では、入力音声と標準パターンやモデルとの時間整合を必要としないことが、DP マッチングなどにより時間整合を行なうテキスト依存型との大きな違いである。それ故テキスト独立型のほとんどは、時間整合を行なわない簡易型のテキスト依存型話者認識としても使用できる。

【図】

エルゴード型 HMM の例



出典:「HMM による話者認識」,「電子情報通信学会技術研究報告 音声 SP95-111 pp.17-24」, 1996 年 1 月発行、松井知子著、電子情報通信学会発行、19 頁 図 3 エルゴード型の HMM

【出典 / 参考資料】

・「音声情報処理」,「電子情報通信工学シリーズ」, 1998 年 6 月 30 日発行、古井貞熙著、森北出版株式会社発行、143 頁

・「HMM による話者認識」,「電子情報通信学会技術研究報告 音声 SP95-111 pp.17-24」, 1996 年 1 月発行、松井知子著、電子情報通信学会発行

【技術分類】 6 - 2 - 3 音声 / 照合・判定技術 / 話者照合方式

【 F I 】 G10L3/00

【技術名称】 6 - 2 - 3 - 3 テキスト指定方式

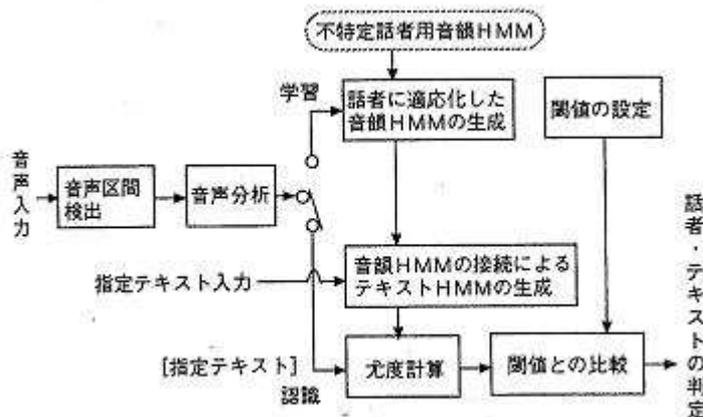
【技術内容】

発声内容をあらかじめ決めておかず、その時々に応じてシステムの側から指定し、照合する方法である。録音機材などによる成りすましを防ぐために考案された。テキスト指定型話者認識の基本的構成を図1に示す。どのような文章にも対応できるように、あらかじめ登録話者ごとに、各音素のスペクトルを表すモデル(音韻モデル)を作成しておく必要がある。しかし個々の登録者の学習用音声には限りがあるため、多数話者の音声を基とした不特定話者音韻モデルを種として用いる。この音韻モデルはHMMで表現されており、音素性の情報を十分に持っているが、声の個人性に関する情報は持っていない。これを、図2のように各話者の学習用音声を用いて、その話者に合うように自動的に適応化する。

照合時には、指定したテキストに従って話者の音韻モデルを接続し、文音声モデルを作成する。入力音声をその文音声モデルと比較し、類似性が大きいとき受理する。

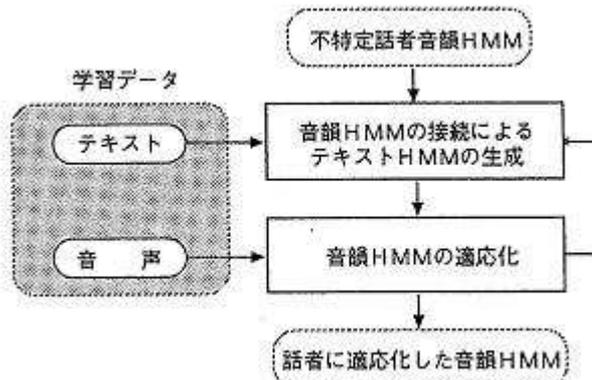
【図】

図1 テキスト指定型話者認識の原理



出典:「HMMによる話者認識」,「電子情報通信学会技術研究報告 音声 SP95-111 pp.17-24」,1996年1月発行、松井知子著、電子情報通信学会発行、20頁 図4 テキスト指定型話者照合システムの構成

図2 音韻モデルの作成法



出典:「HMMによる話者認識」,「電子情報通信学会技術研究報告 音声 SP95-111 pp.17-24」,1996

年 1 月発行、松井知子著、電子情報通信学会発行、20 頁 図 5 話者に適応した音韻モデルの作成法

【出典 / 参考資料】

・「音声情報処理」, 「電子情報通信工学シリーズ」, 1998 年 6 月 30 日発行、古井貞熙著、森北出版株式会社発行、144 頁

・「HMM による話者認識」, 「電子情報通信学会技術研究報告 音声 SP95-111 pp.17-24」, 1996 年 1 月発行、松井知子著、電子情報通信学会発行