

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6596376号
(P6596376)

(45) 発行日 令和1年10月23日 (2019. 10. 23)

(24) 登録日 令和1年10月4日 (2019. 10. 4)

(51) Int. Cl.	F I
G 1 O L 17/04 (2013. 01)	G 1 O L 17/04
G 1 O L 17/08 (2013. 01)	G 1 O L 17/08
G 1 O L 17/00 (2013. 01)	G 1 O L 17/00 2 O O C

請求項の数 16 (全 33 頁)

(21) 出願番号	特願2016-79449 (P2016-79449)	(73) 特許権者	000005821
(22) 出願日	平成28年4月12日 (2016. 4. 12)		パナソニック株式会社
(65) 公開番号	特開2016-206660 (P2016-206660A)		大阪府門真市大字門真1006番地
(43) 公開日	平成28年12月8日 (2016. 12. 8)	(74) 代理人	100067828
審査請求日	平成30年11月20日 (2018. 11. 20)		弁理士 小谷 悦司
(31) 優先権主張番号	特願2015-87625 (P2015-87625)	(74) 代理人	100115381
(32) 優先日	平成27年4月22日 (2015. 4. 22)		弁理士 小谷 昌崇
(33) 優先権主張国・地域又は機関	日本国 (JP)	(74) 代理人	100109438
			弁理士 大月 伸介
		(72) 発明者	辻川 美沙貴
			大阪府門真市大字門真1006番地 パナソニック株式会社内
		(72) 発明者	松井 知子
			東京都立川市緑町10-3 統計数理研究所内
			最終頁に続く

(54) 【発明の名称】 話者識別方法及び話者識別装置

(57) 【特許請求の範囲】

【請求項 1】

複数の不特定話者と、前記複数の不特定話者の音声の特徴をモデル化した複数の不特定話者モデルとを対応付けて記憶する第1のデータベースを用いて、前記第1のデータベースに記憶されていない複数の第1の話者と、前記複数の不特定話者モデルとを対応付けて記憶する第2のデータベースを作成する学習モード処理と、

前記第2のデータベースを用いて、第2の話者を識別する識別モード処理とを含み、

前記学習モード処理において、

前記複数の第1の話者のそれぞれ音声信号を取得し、

取得された前記複数の第1の話者のそれぞれの前記音声信号における特徴量と、前記第1のデータベースに記憶されている前記複数の不特定話者モデルにおける各特徴量との第1の類似度をそれぞれ算出し、

算出された前記第1の類似度が所定の値以上である複数の不特定話者モデルを特定し、

前記複数の第1の話者のそれぞれと、特定された前記複数の不特定話者モデルとを対応付けて前記第2のデータベースに記憶し、

前記識別モード処理において、

前記第2の話者の音声信号を取得し、

取得された前記第2の話者の前記音声信号における特徴量と、前記第2のデータベースに記憶されており、前記第1の話者に対応付けられている前記複数の不特定話者モデルにおける各特徴量との複数の第2の類似度を前記複数の第1の話者毎にそれぞれ算出し、

10

20

算出された前記複数の第2の類似度に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する、
話者識別方法。

【請求項2】

算出された前記複数の第2の類似度を合計した合計値を前記第1の話者毎に算出し、算出した前記合計値に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する、

請求項1記載の話者識別方法。

【請求項3】

算出された前記複数の第2の類似度のそれぞれに前記第1の類似度の順位に応じた重み付け値を乗算することにより前記複数の第2の類似度を補正し、 10

補正した前記複数の第2の類似度を合計した合計値を前記第1の話者毎に算出し、算出した前記合計値に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する、

請求項1記載の話者識別方法。

【請求項4】

前記重み付け値は、前記第1の類似度が高い順に高くする、

前記請求項3に記載の話者識別方法。

【請求項5】

算出された前記複数の第2の類似度のそれぞれに前記第1の類似度を乗算することにより前記複数の第2の類似度を補正し、 20

補正した前記複数の第2の類似度を合計した合計値を前記第1の話者毎に算出し、算出した前記合計値に基づいて、前記第2の話者が前記第2のデータベースに記憶されている前記第1の話者のいずれであるかを特定する、

請求項1記載の話者識別方法。

【請求項6】

算出された前記複数の第2の類似度のうち、所定の値以上である前記複数の第2の類似度を合計した合計値を前記第1の話者毎に算出し、算出した前記合計値に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する、 30

請求項1記載の話者識別方法。

【請求項7】

算出した前記合計値が最も高い前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人が、前記第2の話者として特定される、

請求項2～6のいずれか1項に記載の話者識別方法。

【請求項8】

前記学習モード処理において、

特定された前記複数の不特定話者モデルと、取得された前記第1の話者の前記音声信号とに基づいて、前記複数の第1の話者に対応する複数の話者モデルを新たに作成し、

作成した前記複数の話者モデルを前記複数の第1の話者に対応付けて第3のデータベースに記憶し、 40

前記識別モード処理において、

取得された前記第2の話者の音声信号における特徴量と、前記第3のデータベースに記憶されている前記第1の話者に対応付けられている前記話者モデルにおける特徴量との第3の類似度を前記第1の話者毎に算出し、

算出された前記第3の類似度に基づいて、前記第2の話者に対応する前記第3のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する、

請求項1～7のいずれか1項に記載の話者識別方法。

【請求項9】

前記第2の話者が前記第3のデータベースに記憶されている前記複数の第1の話者のい 50

ずれにも特定されない場合、取得された前記第 2 の話者の音声信号における特徴量と、前記第 2 のデータベースに記憶されており、前記第 1 の話者に対応付けられている前記複数の不特定話者モデルにおける各特徴量との複数の第 2 の類似度を前記第 1 の話者毎にそれぞれ算出し、

算出された前記複数の第 2 の類似度に基づいて、前記第 2 の話者に対応する前記第 2 のデータベースに記憶されている前記複数の第 1 の話者の中の一人を特定する、

請求項 8 記載の話者識別方法。

【請求項 10】

前記識別モード処理が行われた後、前記学習モード処理において算出された各々の前記不特定話者モデルに対応する前記第 1 の類似度と、前記識別モード処理において算出された各々の前記不特定話者モデルに対応する前記第 2 の類似度とを比較し、

前記第 1 の類似度と前記第 2 の類似度との差異が所定値以上である前記不特定話者モデルが所定数以上存在する場合、前記学習モード処理を再度行う、

請求項 1～9 のいずれか 1 項に記載の話者識別方法。

【請求項 11】

前記識別モード処理が行われた後、前記学習モード処理において算出された各々の前記不特定話者モデルに対応する前記第 1 の類似度と、前記識別モード処理において算出された各々の前記不特定話者モデルに対応する前記第 2 の類似度とを比較し、

前記第 1 の類似度と前記第 2 の類似度との差異が所定値以上である前記不特定話者モデルが所定数以上存在する場合、前記第 2 のデータベースに記憶されている前記差異が所定値以上である前記不特定話者モデルに対応する前記第 1 の類似度を前記識別モード処理において算出した前記第 2 の類似度に修正する、

請求項 1～9 のいずれか 1 項に記載の話者識別方法。

【請求項 12】

前記学習モード処理、前記識別モード処理のうちの少なくとも一方はプロセッサにより実行される、

請求項 1 に記載の話者識別方法。

【請求項 13】

複数の不特定話者と、前記複数の不特定話者の音声の特徴をモデル化した複数の不特定話者モデルとを対応付けて記憶する第 1 のデータベースを用いて、前記第 1 のデータベースに記憶されていない第 1 の話者と、前記複数の不特定話者モデルとを対応付けて記憶する第 2 のデータベースを作成する学習モード処理部と、

前記第 2 のデータベースを用いて、第 2 の話者を識別する識別モード処理部とを備え、前記学習モード処理部は、

前記複数の第 1 の話者のそれぞれの音声信号を取得する第 1 の音声取得部と、

第 1 の音声取得部によって取得された前記複数の第 1 の話者のそれぞれの前記音声信号における特徴量と、前記第 1 のデータベースに記憶されている前記複数の不特定話者モデルにおける各特徴量との第 1 の類似度をそれぞれ算出する第 1 の類似度算出部と、

前記第 1 の類似度算出部によって算出された前記第 1 の類似度が所定の値以上である複数の不特定話者モデルを特定する第 1 の特定部と、

前記複数の第 1 の話者のそれぞれと、前記第 1 の特定部によって特定された前記複数の不特定話者モデルとを対応付けて前記第 2 のデータベースに記憶する記憶処理部とを含み、

前記識別モード処理部は、

前記第 2 の話者の音声信号を取得する第 2 の音声取得部と、

前記第 2 の音声取得部によって取得された前記第 2 の話者の前記音声信号における特徴量と、前記第 2 のデータベースに記憶されており、前記第 1 の話者に対応付けられている前記複数の不特定話者モデルにおける各特徴量との複数の第 2 の類似度を前記第 1 の話者毎にそれぞれ算出する第 2 の類似度算出部と、

前記第 2 の類似度算出部によって算出された前記複数の第 2 の類似度に基づいて、前記

10

20

30

40

50

第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する第2の特定部とを含む、

話者識別装置。

【請求項14】

前記学習モード処理部、前記識別モード処理部のうちの少なくとも一方はプロセッサを含む、

請求項13に記載の話者識別装置。

【請求項15】

複数の不特定話者と、前記複数の不特定話者の音声の特徴をモデル化した複数の不特定話者モデルとを対応付けて記憶する第1のデータベースを用いて、前記第1のデータベースに記憶されていない第1の話者と、前記複数の不特定話者モデルとを対応付けて記憶する第2のデータベースを作成する学習モード処理と、

前記第2のデータベースを用いて、第2の話者を識別する識別モード処理とを含み、

前記学習モード処理において、

前記複数の第1の話者それぞれの音声信号を取得し、

取得された前記複数の第1の話者のそれぞれの前記音声信号における特徴量と、前記第1のデータベースに記憶されている前記第1の話者とは異なる前記複数の不特定話者の前記複数の不特定話者モデルにおける各特徴量との第1の類似度を算出し、

算出された前記第1の類似度が所定の値以上である複数の不特定話者モデルを特定し、

特定された前記複数の不特定話者モデルと、取得された前記複数の第1の話者のそれぞれの前記音声信号とに基づいて、前記複数の第1の話者のそれぞれに対応する話者モデルを新たに作成し、

作成した前記話者モデルを前記第1の話者に対応付けて前記第2のデータベースに記憶し、

前記識別モード処理において、

前記第2の話者の音声信号を取得し、

取得された前記第2の話者の前記音声信号における特徴量と、前記第2のデータベースに記憶されている前記第1の話者に対応付けられている話者モデルにおける特徴量との複数の第2の類似度を前記第1の話者毎に算出し、

算出された前記複数の第2の類似度に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する、

話者識別方法。

【請求項16】

前記学習モード処理、前記識別モード処理のうちの少なくとも一方はプロセッサにより実行される、

請求項15に記載の話者識別方法。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、取得した音声信号に基づいて話者を識別する話者識別方法及び話者識別装置に関するものである。

【背景技術】

【0002】

話者識別において、一般的に、予め本人の音声を収集し、学習してモデルを生成する手法がある。また、音声認識においては、話者適応として他者の音声を利用する手法がある。

【0003】

例えば、特許文献1に示す音声辞書作成方法は、新規登録ユーザの音声データを取り込んでその特徴を抽出し、予め性別、年代、声の高低等の音声の特徴要因別に作成された複数の不特定話者音声辞書の中から前記抽出された特徴に最も近い不特定話者音声辞書を選

10

20

30

40

50

択し、選択された不特定話者音声辞書を新規登録ユーザの音声データに基づいて適応化することで新規登録ユーザの個人用音声辞書を作成している。

【0004】

また、例えば、特許文献2に示す音声認識装置における話者適応化方法は、全認識対象語彙について複数話者の発声で構成された初期標準パターンセット内の各音声パターンと利用者の音声パターンとの距離情報により、各語彙ごとに標準パターンセット内から該利用者の音声パターンに近接する1個または複数の音声パターンを選択し、これらを該利用者の標準パターンセットとしている。

【先行技術文献】

【特許文献】

10

【0005】

【特許文献1】特開平11-344992号公報

【特許文献2】特開平1-161399号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

従来の話者識別では、識別対象である話者の発話が少ない場合、データ量が不十分であるため、学習により最適な話者モデルを作成することができず、話者を識別する精度が低下するという課題がある。

【0007】

20

特許文献1では、語彙を限定しない話者識別では、語彙の違い及び発話文の長さなどにより照合スコアが大きく変動し、話者識別の精度が低下するおそれがある。

【0008】

また、特許文献2では、複数のパターンに対してそれぞれ距離を算出するのみであり、最も距離が小さい単語が認識用入力音声の認識結果として判別される。そのため、未知の語彙の音声が入力された場合、音声認識の精度が低下するおそれがある。

【0009】

本開示は、上記の問題を解決するためになされたもので、識別対象である話者の話者モデルが作成されていない場合であっても、より高い精度で話者を識別することができる話者識別方法及び話者識別装置を提供することを目的とするものである。

30

【課題を解決するための手段】

【0010】

本開示の一局面に係る話者識別方法は、複数の不特定話者と、前記複数の不特定話者の音声の特徴をモデル化した複数の不特定話者モデルとを対応付けて記憶する第1のデータベースを用いて、前記第1のデータベースに記憶されていない複数の第1の話者と、前記複数の不特定話者モデルとを対応付けて記憶する第2のデータベースを作成する学習モード処理と、前記第2のデータベースを用いて、第2の話者を識別する識別モード処理とを含み、前記学習モード処理において、前記複数の第1の話者のそれぞれの音声信号を取得し、取得された前記複数の第1の話者のそれぞれの前記音声信号における特徴量と、前記第1のデータベースに記憶されている前記複数の不特定話者モデルにおける各特徴量との第1の類似度をそれぞれ算出し、算出された前記第1の類似度が所定の値以上である複数の不特定話者モデルを特定し、前記複数の第1の話者のそれぞれと、特定された前記複数の不特定話者モデルとを対応付けて前記第2のデータベースに記憶し、前記識別モード処理において、前記第2の話者の音声信号を取得し、取得された前記第2の話者の前記音声信号における特徴量と、前記第2のデータベースに記憶されており、前記第1の話者に対応付けられている前記複数の不特定話者モデルにおける各特徴量との複数の第2の類似度を前記複数の第1の話者毎にそれぞれ算出し、算出された前記複数の第2の類似度に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する。

40

【発明の効果】

50

【0011】

本開示によれば、識別対象である話者の話者モデルが作成されていない場合であっても、より高い精度で話者を識別することができる。

【図面の簡単な説明】

【0012】

【図1】図1は、本実施の形態1における話者識別装置の構成を示す図である。

【図2】図2は、本実施の形態1における話者識別装置の学習モード処理の動作について説明するためのフローチャートである。

【図3】図3は、本実施の形態1における話者識別装置の識別モード処理の動作について説明するためのフローチャートである。

【図4】図4は、本実施の形態1における話者識別装置の学習モード処理の具体的な処理の一例を説明するための模式図である。

【図5】図5は、本実施の形態1における話者識別装置の識別モード処理の具体的な処理の一例を説明するための模式図である。

【図6】図6は、本実施の形態1の第1の変形例における話者識別装置の識別モード処理の具体的な処理の一例を説明するための模式図である。

【図7】図7は、本実施の形態1の第2の変形例における話者識別装置の識別モード処理の具体的な処理の一例を説明するための模式図である。

【図8】図8は、本実施の形態1の第3の変形例における話者識別装置の識別モード処理の具体的な処理の一例を説明するための模式図である。

【図9】図9は、本実施の形態2における話者識別装置の構成を示す図である。

【図10】図10は、本実施の形態2における話者識別装置の学習モード処理の動作について説明するためのフローチャートである。

【図11】図11は、本実施の形態2における話者識別装置の識別モード処理の動作について説明するためのフローチャートである。

【図12】図12は、本実施の形態2における話者識別装置の学習モード処理の具体的な処理の一例を説明するための模式図である。

【図13】図13は、本実施の形態3における話者識別装置の構成を示す図である。

【図14】図14は、本実施の形態3における話者識別装置の学習モード処理の動作について説明するためのフローチャートである。

【図15】図15は、本実施の形態3における話者識別装置の識別モード処理の動作について説明するためのフローチャートである。

【発明を実施するための形態】

【0013】

本開示の一局面に係る話者識別方法は、複数の不特定話者と、前記複数の不特定話者の音声の特徴をモデル化した複数の不特定話者モデルとを対応付けて記憶する第1のデータベースを用いて、前記第1のデータベースに記憶されていない複数の第1の話者と、前記複数の不特定話者モデルとを対応付けて記憶する第2のデータベースを作成する学習モード処理と、前記第2のデータベースを用いて、第2の話者を識別する識別モード処理とを含み、前記学習モード処理において、前記複数の第1の話者のそれぞれの音声信号を取得し、取得された前記複数の第1の話者のそれぞれの前記音声信号における特徴量と、前記第1のデータベースに記憶されている前記複数の不特定話者モデルにおける各特徴量との第1の類似度をそれぞれ算出し、算出された前記第1の類似度が所定の値以上である複数の不特定話者モデルを特定し、前記複数の第1の話者のそれぞれと、特定された前記複数の不特定話者モデルとを対応付けて前記第2のデータベースに記憶し、前記識別モード処理において、前記第2の話者の音声信号を取得し、取得された前記第2の話者の前記音声信号における特徴量と、前記第2のデータベースに記憶されており、前記第1の話者に対応付けられている前記複数の不特定話者モデルにおける各特徴量との複数の第2の類似度を前記第1の話者毎にそれぞれ算出し、算出された前記複数の第2の類似度に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話

10

20

30

40

50

者の中の一人を特定する。

【0014】

この構成によれば、学習モード処理において、複数の第1の話者のそれぞれの音声信号が取得される。取得された複数の第1の話者のそれぞれの音声信号における特徴量と、第1のデータベースに記憶されている複数の不特定話者モデルにおける各特徴量との第1の類似度がそれぞれ算出される。算出された第1の類似度が所定の値以上である複数の不特定話者モデルが特定される。複数の第1の話者のそれぞれと、特定された複数の不特定話者モデルとが対応付けられて第2のデータベースに記憶される。識別モード処理において、第2の話者の音声信号が取得される。取得された第2の話者の音声信号における特徴量と、第2のデータベースに記憶されており、第1の話者に対応付けられている複数の不特定話者モデルにおける各特徴量との複数の第2の類似度が第1の話者毎にそれぞれ算出される。算出された複数の第2の類似度に基づいて、第2の話者に対応する第2のデータベースに記憶されている複数の第1の話者の中の一人が特定される。

10

【0015】

したがって、第1の話者の音声信号に類似する複数の不特定話者の話者モデルを用いて、第2の話者が識別されるので、識別対象である話者の話者モデルが作成されていない場合であっても、より高い精度で話者を識別することができる。

【0016】

また、上記の話者識別方法において、算出された前記複数の第2の類似度を合計した合計値を前記第1の話者毎に算出し、算出した前記合計値に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定してもよい。

20

【0017】

この構成によれば、算出された複数の第2の類似度を合計した合計値が第1の話者毎に算出され、算出された合計値に基づいて、第2の話者に対応する第2のデータベースに記憶されている複数の第1の話者の中の一人が特定される。

【0018】

したがって、複数の第2の類似度を合計した合計値に基づいて第2の話者が特定されるので、語彙の違い及び発話文の長さなどによる第2の類似度の変動を抑えることができ、より高い精度で話者を識別することができる。

30

【0019】

また、上記の話者識別方法において、算出された前記複数の第2の類似度のそれぞれに前記第1の類似度の順位に応じた重み付け値を乗算することにより前記複数の第2の類似度を補正し、補正した前記複数の第2の類似度を合計した合計値を前記第1の話者毎に算出し、算出した前記合計値に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定してもよい。

【0020】

この構成によれば、算出された複数の第2の類似度のそれぞれに第1の類似度の順位に応じた重み付け値が乗算されることにより複数の第2の類似度が補正される。補正された複数の第2の類似度を合計した合計値が第1の話者毎に算出され、算出された合計値に基づいて、第2の話者に対応する第2のデータベースに記憶されている複数の第1の話者の中の一人が特定される。

40

【0021】

したがって、複数の第2の類似度のそれぞれに第1の類似度の順位に応じた重み付け値が乗算されることにより複数の第2の類似度が補正されるので、語彙の違い及び発話文の長さなどによる第2の類似度の変動を抑えることができ、より高い精度で話者を識別することができる。

【0022】

また、上記の話者識別方法において、前記重み付け値は、前記第1の類似度が高い順に高くしてもよい。

50

【0023】

この構成によれば、重み付け値は、第1の類似度が高い順に高くなるので、語彙の違い及び発話文の長さなどによる第2の類似度の変動をより抑えることができる。

【0024】

また、上記の話者識別方法において、算出された前記複数の第2の類似度のそれぞれに前記第1の類似度を乗算することにより前記複数の第2の類似度を補正し、補正した前記複数の第2の類似度を合計した合計値を前記第1の話者毎に算出し、算出した前記合計値に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定してもよい。

【0025】

10

この構成によれば、算出された複数の第2の類似度のそれぞれに第1の類似度が乗算されることにより複数の第2の類似度が補正される。補正された複数の第2の類似度を合計した合計値が第1の話者毎に算出され、算出された合計値に基づいて、第2の話者に対応する第2のデータベースに記憶されている複数の第1の話者の中の一人が特定される。

【0026】

したがって、複数の第2の類似度のそれぞれに第1の類似度が乗算されることにより複数の第2の類似度が補正されるので、語彙の違い及び発話文の長さなどによる第2の類似度の変動を抑えることができ、より高い精度で話者を識別することができる。

【0027】

また、上記の話者識別方法において、算出された前記複数の第2の類似度のうち、所定の値以上である前記複数の第2の類似度を合計した合計値を前記複数の第1の話者毎に算出し、算出した前記合計値に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定してもよい。

20

【0028】

この構成によれば、算出された複数の第2の類似度のうち、所定の値以上である複数の第2の類似度を合計した合計値が複数の第1の話者毎に算出され、算出された合計値に基づいて、第2の話者に対応する第2のデータベースに記憶されている複数の第1の話者の中の一人が特定される。

【0029】

したがって、所定の値以上である複数の第2の類似度が合計され、所定の値より小さい第2の類似度が考慮されないので、より高い精度で話者を識別することができる。

30

【0030】

また、上記の話者識別方法において、算出した前記合計値が最も高い前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人が、前記第2の話者として特定されてもよい。

【0031】

この構成によれば、算出した合計値が最も高い第2のデータベースに記憶されている第1の話者が、第2の話者として特定されるので、容易に話者を識別することができる。

【0032】

また、上記の話者識別方法において、前記学習モード処理において、特定された前記複数の不特定話者モデルと、取得された前記複数の第1の話者のそれぞれの前記音声信号とに基づいて、前記複数の第1の話者に対応する複数の話者モデルを新たに作成し、作成した前記複数の話者モデルを前記複数の第1の話者に対応付けて第3のデータベースに記憶し、前記識別モード処理において、取得された前記第2の話者の音声信号における特徴量と、前記第3のデータベースに記憶されている前記第1の話者に対応付けられている前記話者モデルにおける特徴量との第3の類似度を前記第1の話者毎に算出し、算出された前記第3の類似度に基づいて、前記第2の話者に対応する前記第3のデータベースに記憶されている前記複数の第1の話者の中の一人を特定してもよい。

40

【0033】

この構成によれば、学習モード処理において、特定された複数の不特定話者モデルと、

50

取得された複数の第1の話者のそれぞれ音声信号とに基づいて、複数の第1の話者に対応する複数の話者モデルが新たに作成される。作成された複数の話者モデルが複数の第1の話者に対応付けて第3のデータベースに記憶される。識別モード処理において、取得された第2の話者の音声信号における特徴量と、第3のデータベースに記憶されている第1の話者に対応付けられている話者モデルにおける特徴量との第3の類似度が第1の話者毎に算出される。算出された第3の類似度に基づいて、第2の話者に対応する第3のデータベースに記憶されている複数の第1の話者の中の一人が特定される。

【0034】

したがって、取得した第1の話者の音声信号に類似する複数の他者の話者モデルを用いて第1の話者に対応する新たな話者モデルが作成されるので、新たに作成された話者モデルを用いて話者をより高い精度で識別することができる。

10

【0035】

また、上記の話者識別方法において、前記第2の話者が前記第3のデータベースに記憶されている前記第1の話者のいずれにも特定されない場合、取得された前記第2の話者の音声信号における特徴量と、前記第2のデータベースに記憶されており、前記第1の話者に対応付けられている前記複数の不特定話者モデルにおける各特徴量との複数の第2の類似度を前記第1の話者毎にそれぞれ算出し、算出された前記複数の第2の類似度に基づいて、前記第2の話者が前記第2のデータベースに記憶されている前記第1の話者のいずれであるかを特定してもよい。

【0036】

20

この構成によれば、第2の話者が第3のデータベースに記憶されている第1の話者のいずれにも特定されない場合、取得された第2の話者の音声信号における特徴量と、第2のデータベースに記憶されており、第1の話者に対応付けられている複数の不特定話者モデルにおける各特徴量との複数の第2の類似度が第1の話者毎にそれぞれ算出される。算出された複数の第2の類似度に基づいて、第2の話者が第2のデータベースに記憶されている第1の話者のいずれであるかが特定される。

【0037】

したがって、取得した第1の話者の音声信号に類似する複数の他者の話者モデルを用いて作成された新たな話者モデルを用いて第2の話者を特定することができない場合、第1の話者の音声信号に類似する複数の不特定話者の話者モデルを用いて第2の話者を識別することができる。

30

【0038】

また、上記の話者識別方法において、前記識別モード処理が行われた後、前記学習モード処理において算出された各々の前記不特定話者モデルに対応する前記第1の類似度と、前記識別モード処理において算出された各々の前記不特定話者モデルに対応する前記第2の類似度とを比較し、前記第1の類似度と前記第2の類似度との差異が所定値以上である前記不特定話者モデルが所定数以上存在する場合、前記学習モード処理を再度行ってもよい。

【0039】

この構成によれば、識別モード処理が行われた後、学習モード処理において算出された各々の不特定話者モデルに対応する第1の類似度と、識別モード処理において算出された各々の不特定話者モデルに対応する第2の類似度とが比較される。第1の類似度と第2の類似度との差異が所定値以上である不特定話者モデルが所定数以上存在する場合、学習モード処理が再度行われる。

40

【0040】

したがって、学習モード処理において算出された第1の類似度と、識別モード処理において算出された第2の類似度とが大きく異なっている場合、学習モード処理を再度行うことにより、より第1の話者に類似した複数の不特定話者モデルを特定することができ、より高い精度で話者を識別することができる。

【0041】

50

また、上記の話者識別方法において、前記識別モード処理が行われた後、前記学習モード処理において算出された各々の前記不特定話者モデルに対応する前記第1の類似度と、前記識別モード処理において算出された各々の前記不特定話者モデルに対応する前記第2の類似度とを比較し、前記第1の類似度と前記第2の類似度との差異が所定値以上である前記不特定話者モデルが所定数以上存在する場合、前記第2のデータベースに記憶されている前記差異が所定値以上である前記不特定話者モデルに対応する前記第1の類似度を前記識別モード処理において算出した前記第2の類似度に修正してもよい。

【0042】

この構成によれば、識別モード処理が行われた後、学習モード処理において算出された各々の不特定話者モデルに対応する第1の類似度と、識別モード処理において算出された各々の不特定話者モデルに対応する第2の類似度とが比較される。第1の類似度と第2の類似度との差異が所定値以上である不特定話者モデルが所定数以上存在する場合、第2のデータベースに記憶されている差異が所定値以上である不特定話者モデルに対応する第1の類似度が識別モード処理において算出された第2の類似度に修正される。

【0043】

したがって、学習モード処理において算出された第1の類似度と、識別モード処理において算出された第2の類似度とが大きく異なっている場合、第1の類似度を第2の類似度に修正することにより、より第1の話者に類似した複数の不特定話者モデルを特定することができ、より高い精度で話者を識別することができる。

【0044】

本開示の他の局面に係る話者識別装置は、複数の不特定話者と、前記複数の不特定話者の音声の特徴をモデル化した複数の不特定話者モデルとを対応付けて記憶する第1のデータベースを用いて、前記第1のデータベースに記憶されていない第1の話者と、前記複数の不特定話者モデルとを対応付けて記憶する第2のデータベースを作成する学習モード処理部と、前記第2のデータベースを用いて、第2の話者を識別する識別モード処理部とを備え、前記学習モード処理部は、前記複数の第1の話者のそれぞれ音声信号を取得する第1の音声取得部と、第1の音声取得部によって取得された前記複数の第1の話者のそれぞれの前記音声信号における特徴量と、前記第1のデータベースに記憶されている前記第1の話者とは異なる前記複数の不特定話者の前記複数の不特定話者モデルにおける各特徴量との第1の類似度をそれぞれ算出する第1の類似度算出部と、前記第1の類似度算出部によって算出された前記第1の類似度が所定の値以上である複数の不特定話者モデルを特定する第1の特定部と、前記複数の第1の話者のそれぞれと、前記第1の特定部によって特定された前記複数の不特定話者モデルとを対応付けて前記第2のデータベースに記憶する記憶処理部とを含み、前記識別モード処理部は、前記第2の話者の音声信号を取得する第2の音声取得部と、前記第2の音声取得部によって取得された前記第2の話者の前記音声信号における特徴量と、前記第2のデータベースに記憶されており、前記第1の話者に対応付けられている前記複数の不特定話者モデルにおける各特徴量との複数の第2の類似度を前記第1の話者毎にそれぞれ算出する第2の類似度算出部と、前記第2の類似度算出部によって算出された前記複数の第2の類似度に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する第2の特定部とを含む。

【0045】

この構成によれば、学習モード処理において、複数の第1の話者のそれぞれの音声信号が取得される。取得された複数の第1の話者のそれぞれの音声信号における特徴量と、第1のデータベースに記憶されている複数の不特定話者モデルにおける各特徴量との第1の類似度がそれぞれ算出される。算出された第1の類似度が所定の値以上である複数の不特定話者モデルが特定される。複数の第1の話者のそれぞれと、特定された複数の不特定話者モデルとが対応付けられて第2のデータベースに記憶される。識別モード処理において、第2の話者の音声信号が取得される。取得された第2の話者の音声信号における特徴量と、第2のデータベースに記憶されており、第1の話者に対応付けられている複数の不特

定話者モデルにおける各特微量との複数の第2の類似度が第1の話者毎にそれぞれ算出される。算出された複数の第2の類似度に基づいて、第2の話者に対応する第2のデータベースに記憶されている複数の第1の話者の中の一人が特定される。

【0046】

したがって、第1の話者の音声信号に類似する複数の不特定話者の話者モデルを用いて、第2の話者が識別されるので、識別対象である話者の話者モデルが作成されていない場合であっても、より高い精度で話者を識別することができる。

【0047】

本開示の他の局面に係る話者識別方法は、複数の不特定話者と、前記複数の不特定話者の音声の特徴をモデル化した複数の不特定話者モデルとを対応付けて記憶する第1のデータベースを用いて、前記第1のデータベースに記憶されていない第1の話者と、前記複数の不特定話者モデルとを対応付けて記憶する第2のデータベースを作成する学習モード処理と、前記第2のデータベースを用いて、第2の話者を識別する識別モード処理とを含み、前記学習モード処理において、前記複数の第1の話者のそれぞれの音声信号を取得し、取得された前記複数の第1の話者のそれぞれの前記音声信号における特微量と、前記第1のデータベースに記憶されている前記複数の不特定話者モデルにおける各特微量との第1の類似度を算出し、算出された前記第1の類似度が所定の値以上である複数の不特定話者モデルを特定し、特定された前記複数の不特定話者モデルと、取得された前記複数の第1の話者のそれぞれの前記音声信号とに基づいて、前記第1の話者に対応する話者モデルを新たに作成し、作成した前記話者モデルを前記第1の話者に対応付けて前記第2のデータベースに記憶し、前記識別モード処理において、前記第2の話者の音声信号を取得し、取得された前記第2の話者の前記音声信号における特微量と、前記第2のデータベースに記憶されている前記第1の話者に対応付けられている話者モデルにおける特微量との複数の第2の類似度を前記第1の話者毎に算出し、算出された前記複数の第2の類似度に基づいて、前記第2の話者に対応する前記第2のデータベースに記憶されている前記複数の第1の話者の中の一人を特定する。

【0048】

この構成によれば、学習モード処理において、複数の第1の話者のそれぞれの音声信号が取得される。取得された複数の第1の話者のそれぞれの音声信号における特微量と、第1のデータベースに記憶されている複数の不特定話者モデルにおける各特微量との第1の類似度が算出される。算出された第1の類似度が所定の値以上である複数の不特定話者モデルが特定される。特定された複数の不特定話者モデルと、取得された複数の第1の話者のそれぞれの音声信号とに基づいて、複数の第1の話者に対応する複数の話者モデルが新たに作成される。作成した話者モデルが第1の話者に対応付けて第2のデータベースに記憶される。識別モード処理において、第2の話者の音声信号が取得される。取得された第2の話者の音声信号における特微量と、第2のデータベースに記憶されている第1の話者に対応付けられている話者モデルにおける特微量との複数の第2の類似度が第1の話者毎に算出される。算出された複数の第2の類似度に基づいて、第2の話者に対応する第2のデータベースに記憶されている複数の第1の話者の中の一人が特定される。

【0049】

したがって、取得した第1の話者の音声信号に類似する複数の他者の話者モデルを用いて第1の話者に対応する新たな話者モデルが作成されるので、新たに作成された話者モデルを用いて話者をより高い精度で識別することができる。

【0050】

なお、これらの包括的または具体的な態様は、システム、方法、装置、集積回路、コンピュータプログラムまたはコンピュータ読み取り可能なCD-ROMなどの記録媒体で実現されてもよく、システム、方法、装置、集積回路、コンピュータプログラムまたは記録媒体の任意な組み合わせで実現されてもよい。

【0051】

以下添付図面を参照しながら、本開示の実施の形態について説明する。なお、以下の実

10

20

30

40

50

施の形態は、本開示を具体化した一例であって、本開示の技術的範囲を限定するものではない。

【0052】

(実施の形態1)

図1は、本実施の形態1における話者識別装置の構成を示す図である。話者識別装置は、例えば、テレビ、スマートホン又はカーナビゲーション装置などに内蔵される。

【0053】

図1に示す話者識別装置は、音声入力部11、第1の音声分析部12、第1の話者モデル格納部13、第1の類似度算出部14、モデル特定部15、第2の話者モデル格納部16、第2の音声分析部17、第2の類似度算出部18及び話者識別部19を備える。

10

【0054】

また、学習モード処理部1は、音声入力部11、第1の音声分析部12、第1の話者モデル格納部13、第1の類似度算出部14、モデル特定部15及び第2の話者モデル格納部16で構成される。識別モード処理部2は、音声入力部11、第2の話者モデル格納部16、第2の音声分析部17、第2の類似度算出部18及び話者識別部19で構成される。

【0055】

学習モード処理部1は、学習モード処理において、複数の不特定話者と、複数の不特定話者の音声の特徴をモデル化した複数の不特定話者モデルとを対応付けて記憶する第1の話者モデル格納部13（第1のデータベース）を用いて、第1の話者モデル格納部13に記憶されていない第1の話者と、複数の不特定話者モデルとを対応付けて記憶する第2の話者モデル格納部16（第2のデータベース）を作成する。

20

【0056】

第1の話者モデル格納部13、および第2の話者モデル格納部16は、ハードウェア構成として、それぞれ、例えば、メモリであってもよい。メモリとは、例えば、半導体メモリ、ハードディスクといったような書き込み可能な記録媒体である。

【0057】

識別モード処理部2は、識別モード処理において、第2の話者モデル格納部16を用いて、第2の話者を識別する。

【0058】

本実施の形態1における話者識別装置は、ハードウェア構成として、プロセッサとメモリとを含んでもよい。メモリには、学習モード処理部1、および識別モード処理部2のそれぞれに対応するプログラムを含んでもよい。メモリとは、例えば、半導体メモリ、ハードディスクなどの記録媒体である。プロセッサがプログラムを実行することにより、学習モード処理部1、および識別モード処理部2がそれぞれ機能する。

30

【0059】

または、本実施の形態1における話者識別装置は、学習モード処理部1、および識別モード処理部2のそれぞれの機能を実現する集積回路を含んでもよい。

【0060】

または、本実施の形態1における話者識別装置は、学習モード処理部1、および識別モード処理部2の一方に対応するプログラム、他方の機能を実現する集積回路、およびプロセッサを含んでもよい。この場合、学習モード処理部1、および識別モード処理部2の一方に対応するプログラムをプロセッサが実行することにより、学習モード処理部1、および識別モード処理部2のうちの一方が機能する。

40

【0061】

なお、本実施の形態1における話者識別装置は、学習モード処理部1によって行われる学習モード処理と、識別モード処理部2によって行われる識別モード処理とを切り替えることができる。学習モード処理と識別モード処理との切り替えは、例えば、ユーザによる入力操作により行われる。話者識別装置は、例えばユーザによる切り替えボタンの押下を受け付けることにより、学習モード処理と識別モード処理とを切り替えてもよい。また、

50

話者識別装置は、例えばユーザによる発話を音声認識することにより、学習モード処理と識別モード処理とを切り替えてもよい。

【0062】

音声入力部11は、例えばマイクロフォンで構成され、学習モード処理又は識別モード処理において、周囲の音声を収集し、収集した音声を音声信号に変換して出力する。

【0063】

第1の音声分析部12は、学習モード処理において、音声入力部11から入力された音声信号を分析し、第1の話者によって発話された音声信号を出力する。第1の音声分析部12は、第1の話者モデル格納部13に記憶されていない第1の話者の音声信号を取得する。

10

【0064】

第1の話者モデル格納部13は、複数の不特定話者と、複数の不特定話者の音声の特徴をモデル化した複数の不特定話者モデルとを対応付けて記憶する。複数の不特定話者モデルは、それぞれ、例えば、GMM (Gaussian Mixture Model) による確率モデルなどを用いて不特定話者の固有の特徴量を含むものであってもよい。

【0065】

第1の類似度算出部14は、取得された第1の話者の音声信号における特徴量と、第1の話者モデル格納部13に記憶されている複数の不特定話者モデルにおける各特徴量との第1の類似度をそれぞれ算出する。

【0066】

20

第1の類似度算出部14は、取得された第1の話者の音声信号における特徴量を算出する。ここで、特徴量は、i V e c t o rを用いて算出される。なお、i V e c t o rを用いた特徴量の算出方法については従来技術であるため、詳細な説明は省略する。また、特徴量の算出方法については、i V e c t o rに限定されず、他の算出方法を用いてもよい。また、第1の類似度算出部14は、第1の話者モデル格納部13に記憶されている複数の不特定話者モデルを読み出し、読み出した複数の不特定話者モデルにおける特徴量をそれぞれ算出する。

【0067】

なお、第1の話者モデル格納部13は、複数の不特定話者モデルに対して算出された特徴量を複数の不特定話者モデルに対応付けて予め記憶してもよい。第1の話者モデル格納部13が特徴量を記憶している場合、第1の類似度算出部14は、第1の話者モデル格納部13から特徴量を読み出す。

30

【0068】

第1の類似度算出部14は、第1の話者の音声信号における特徴量と、第1の話者モデル格納部13に記憶されている複数の不特定話者モデルにおける各特徴量との第1の類似度をそれぞれ算出する。第1の類似度算出部14は、2つの特徴量（第1の話者の音声信号における特徴量と、第1の話者モデル格納部13に記憶されている不特定話者モデルにおける特徴量）がどれだけ類似しているかを示す話者間距離を算出する。

【0069】

また、第1の類似度算出部14は、例えば、話者間距離として、第1の話者の音声信号における特徴量が、不特定話者モデルにおける各特徴量のそれぞれであると推測する尤もらしさを示す数値（尤度）を算出するのでもよい。第1の類似度算出部14は、2つの特徴量の間の近さを算出する関数を用いて、話者間距離を算出する。話者間距離は、特徴量同士が類似する度合いが高いほど、より1に近い値となり、特徴量同士が類似する度合いが低いほど、0に近い値となる。なお、類似度の算出方法については従来技術であるため、詳細な説明は省略する。また、類似度の算出方法については上記に限定されない。

40

【0070】

モデル特定部15は、第1の類似度算出部14によって算出された第1の類似度が所定の値以上である複数の不特定話者モデルを特定する。モデル特定部15は、特定した複数の不特定話者モデルを第1の話者モデル格納部13から抽出する。モデル特定部15は、

50

第1の話者を識別する識別情報と、特定した複数の不特定話者モデルとを対応付けて第2の話者モデル格納部16に記憶する。なお、モデル特定部15は、識別情報を入力音声毎に作成する。

【0071】

第2の話者モデル格納部16は、第1の話者を識別する識別情報と、モデル特定部15によって特定された複数の不特定話者モデルとを対応付けて記憶する。

【0072】

第2の音声分析部17は、識別モード処理において、音声入力部11から入力された音声信号を分析し、第2の話者によって発話された音声信号を出力する。第2の音声分析部17は、第2の話者の音声信号を取得する。第2の話者は、第1の話者と同じ話者又は第1の話者とは異なる話者である。第2の音声分析部17は、第1の音声分析部12と同じ処理を行う。

10

【0073】

第2の類似度算出部18は、第2の音声分析部17によって取得された第2の話者の音声信号における特徴量と、第2の話者モデル格納部16（第2のデータベース）に記憶されており、第1の話者に対応付けられている複数の不特定話者モデルにおける各特徴量との複数の第2の類似度を第1の話者毎にそれぞれ算出する。

【0074】

第2の類似度算出部18は、取得された第2の話者の音声信号における特徴量を算出する。ここで、特徴量は、i V e c t o rを用いて算出される。なお、i V e c t o rを用いた特徴量の算出方法については従来技術であるため、詳細な説明は省略する。また、特徴量の算出方法については、i V e c t o rに限定されず、他の算出方法を用いてもよい。また、第2の類似度算出部18は、第1の類似度算出部14による特徴量の算出方法と同じ算出方法で特徴量を算出してもよいし、第1の類似度算出部14による特徴量の算出方法とは異なる算出方法で特徴量を算出してもよい。

20

【0075】

また、第2の類似度算出部18は、第2の話者モデル格納部16に記憶されている複数の第1の話者のそれぞれに対応付けられている複数の不特定話者モデルを読み出し、読み出した複数の不特定話者モデルにおける特徴量を複数の第1の話者毎に算出する。

【0076】

30

なお、第2の話者モデル格納部16は、複数の不特定話者モデルに対して算出された特徴量を複数の不特定話者モデルに対応付けて予め記憶してもよい。第2の話者モデル格納部16が特徴量を記憶している場合、第2の類似度算出部18は、第2の話者モデル格納部16から特徴量を読み出す。

【0077】

第2の類似度算出部18は、第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者に対応付けられている複数の不特定話者モデルにおける各特徴量との複数の第2の類似度を第1の話者毎にそれぞれ算出する。第2の類似度算出部18は、2つの特徴量がどれだけ類似しているかを示す話者間距離を算出する。第2の類似度算出部18は、2つの特徴量の間の近さを算出する関数を用いて、話者間距離を算出する。話者間距離は、特徴量同士が類似している場合には1に近い値となり、特徴量同士が類似していない場合には0に近い値となる。なお、類似度の算出方法については従来技術であるため、詳細な説明は省略する。また、類似度の算出方法については上記に限定されない。

40

【0078】

また、第2の類似度算出部18は、第1の類似度算出部14による第1の類似度の算出方法と同じ算出方法で第2の類似度を算出してもよいし、第1の類似度算出部14による第1の類似度の算出方法とは異なる算出方法で第2の類似度を算出してもよい。

【0079】

話者識別部19は、第2の類似度算出部18によって算出された複数の第2の類似度に

50

基づいて、第2の話者に対応する第2の話者モデル格納部16に記憶されている複数の第1の話者の中の一人を特定する。ここで、話者識別部19は、第2の類似度算出部18によって算出された複数の第2の類似度を合計した合計値を第1の話者毎に算出し、算出した合計値に基づいて、第2の話者が第2の話者モデル格納部16に記憶されている第1の話者のいずれであるかを特定する。話者識別部19は、算出した合計値が最も高い第2の話者モデル格納部16に記憶されている第1の話者として、第2の話者を特定する。

【0080】

このように、類似度の合計値を比較することにより、語彙の違い及び文の短さに影響されることなく、話者識別の精度を高めることができる。

【0081】

10

なお、本実施の形態1において、テレビ又はスマートホンなどの端末装置が、音声入力部11を備え、話者識別装置が、音声入力部11以外の構成を備えてもよい。この場合、話者識別装置は、例えばサーバで構成され、ネットワークを介して端末装置と通信可能に接続される。サーバは、話者識別装置の機能の一部又は全部を備えてもよい。

【0082】

また、第1の話者及び第2の話者は、例えば、家庭内に配置された家電機器を操作するユーザであってもよい。

【0083】

続いて、本実施の形態1における話者識別装置の学習モード処理及び識別モード処理のそれぞれの動作について説明する。図2は、本実施の形態1における話者識別装置の学習モード処理の動作について説明するためのフローチャートである。

20

【0084】

まず、ステップS1において、第1の音声分析部12は、第1の話者モデル格納部13に記憶されていない第1の話者の音声信号を取得する。

【0085】

次に、ステップS2において、第1の類似度算出部14は、第1の音声分析部12によって取得された第1の話者の音声信号の特徴量、及び第1の話者モデル格納部13に記憶されている複数の不特定話者モデルの特徴量を算出する。

【0086】

次に、ステップS3において、第1の類似度算出部14は、算出した第1の話者の音声信号の特徴量と、算出した複数の不特定話者モデルの特徴量との第1の類似度を算出する。

30

【0087】

次に、ステップS4において、モデル特定部15は、第1の類似度算出部14によって算出された第1の類似度が所定の値以上である複数の不特定話者モデルを特定する。モデル特定部15は、特定した複数の不特定話者モデルを第1の話者モデル格納部13から抽出する。このとき、モデル特定部15は、特定した複数の不特定話者モデルを、第1の類似度が高い順に第1の話者モデル格納部13から抽出してもよい。

【0088】

次に、ステップS5において、モデル特定部15は、特定した複数の不特定話者モデルを、第1の話者を識別する識別情報に対応付けて第2の話者モデル格納部16に記憶する。

40

【0089】

図3は、本実施の形態1における話者識別装置の識別モード処理の動作について説明するためのフローチャートである。

【0090】

まず、ステップS11において、第2の音声分析部17は、第2の話者の音声信号を取得する。

【0091】

次に、ステップS12において、第2の類似度算出部18は、第2の音声分析部17に

50

よって取得された第2の話者の音声信号の特徴量、及び第2の話者モデル格納部16に記憶されており、複数の第1の話者のそれぞれに対応付けられている複数の不特定話者モデルの特徴量を算出する。

【0092】

次に、ステップS13において、第2の類似度算出部18は、第2の話者の音声信号の特徴量と、第1の話者に対応付けられている複数の不特定話者モデルの特徴量のそれぞれとの複数の第2の類似度を第1の話者毎にそれぞれ算出する。

【0093】

次に、ステップS14において、話者識別部19は、第2の類似度算出部18によって算出された複数の第2の類似度を合計した合計値を第1の話者毎に算出する。

10

【0094】

次に、ステップS15において、話者識別部19は、算出した合計値が最も高い第2の話者モデル格納部16に記憶されている第1の話者として、第2の話者を特定する。

【0095】

なお、識別モード処理が行われた後、話者識別部19は、学習モード処理において算出された各々の不特定話者モデルに対応する第1の類似度と、識別モード処理において算出された各々の不特定話者モデルに対応する第2の類似度とを比較してもよい。そして、第1の類似度と第2の類似度との差異が所定値以上である不特定話者モデルが所定数以上存在する場合、話者識別部19は、学習モード処理を再度行うように学習モード処理部1に指示し、学習モード処理部1は、学習モード処理を再度行ってもよい。

20

【0096】

また、識別モード処理が行われた後、話者識別部19は、学習モード処理において算出された各々の不特定話者モデルに対応する第1の類似度と、識別モード処理において算出された各々の不特定話者モデルに対応する第2の類似度とを比較してもよい。そして、第1の類似度と第2の類似度との差異が所定値以上である不特定話者モデルが所定数以上存在する場合、話者識別部19は、第2の話者モデル格納部16に記憶されている差異が所定値以上である不特定話者モデルに対応する第1の類似度を識別モード処理において算出した第2の類似度に修正してもよい。

【0097】

続いて、本実施の形態1における話者識別装置の具体的な処理の流れについて説明する。

30

【0098】

図4は、本実施の形態1における話者識別装置の学習モード処理の具体的な処理の一例を説明するための模式図である。

【0099】

まず、学習モード処理において、第1の話者Xが「テレビ」と発話した場合、音声入力部11は、第1の話者Xが発話した音声を集集し、収集した音声を音声信号に変換して出力する。なお、本実施の形態1では、第1の話者Xが発話する語彙は、特に限定されず、第1の話者Xは、予め決められている語彙を発話する必要はない。第1の音声分析部12は、音声入力部11から入力された音声信号を分析し、第1の話者Xによって発話された音声信号を出力する。

40

【0100】

次に、第1の類似度算出部14は、取得された第1の話者Xの音声信号と、他者の話者モデルとを比較する。すなわち、第1の類似度算出部14は、取得された第1の話者Xの音声信号における特徴量と、第1の話者モデル格納部13に記憶されている複数の不特定話者モデルA～Fにおける各特徴量との第1の類似度をそれぞれ算出する。

【0101】

次に、モデル特定部15は、第1の類似度算出部14によって算出された第1の類似度が所定の値以上である複数の不特定話者モデルを特定する。図4では、第1の類似度が0.4以上である複数の不特定話者モデルA～Cが特定されている。例えば、第1の話者X

50

の音声信号と不特定話者モデルAとの第1の類似度は、0.9であり、第1の話者Xの音声信号と不特定話者モデルBとの第1の類似度は、0.7であり、第1の話者Xの音声信号と不特定話者モデルCとの第1の類似度は、0.4である。モデル特定部15は、第1の話者Xを識別する識別情報と、特定した複数の不特定話者モデルA～Cとを対応付けたテーブルを、第2の話者モデル格納部16に記憶する。

【0102】

同様に、学習モード処理において、第1の話者Yが「音量」と発話した場合、音声入力部11は、第1の話者Yが発話した音声を集集し、収集した音声を音声信号に変換して出力する。第1の音声分析部12は、音声入力部11から入力された音声信号を分析し、第1の話者Yによって発話された音声信号を出力する。

10

【0103】

次に、第1の類似度算出部14は、取得された第1の話者Yの音声信号と、他者の話者モデルとを比較する。すなわち、第1の類似度算出部14は、取得された第1の話者Yの音声信号における特徴量と、第1の話者モデル格納部13に記憶されている複数の不特定話者モデルA～Fにおける各特徴量との第1の類似度をそれぞれ算出する。

【0104】

次に、モデル特定部15は、第1の類似度算出部14によって算出された第1の類似度が所定の値以上である複数の不特定話者モデルを特定する。図4では、第1の類似度が0.4以上である複数の不特定話者モデルD～Fが特定されている。例えば、第1の話者Yの音声信号と不特定話者モデルDとの第1の類似度は、0.8であり、第1の話者Yの音声信号と不特定話者モデルEとの第1の類似度は、0.5であり、第1の話者Yの音声信号と不特定話者モデルFとの第1の類似度は、0.4である。モデル特定部15は、第1の話者Yを識別する識別情報と、特定した複数の不特定話者モデルD～Fとを対応付けたテーブルを、第2の話者モデル格納部16に記憶する。

20

【0105】

図5は、本実施の形態1における話者識別装置の識別モード処理の具体的な処理の一例を説明するための模式図である。

【0106】

まず、識別モード処理において、第2の話者が「音量」と発話した場合、音声入力部11は、第2の話者が発話した音声を集集し、収集した音声を音声信号に変換して出力する。第2の音声分析部17は、音声入力部11から入力された音声信号を分析し、第2の話者によって発話された音声信号を出力する。

30

【0107】

次に、第2の類似度算出部18は、第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者Xに対応付けられている複数の不特定話者モデルA～Cにおける各特徴量との複数の第2の類似度を算出する。また、第2の類似度算出部18は、第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者Yに対応付けられている複数の不特定話者モデルD～Fにおける各特徴量との複数の第2の類似度を算出する。

【0108】

そして、話者識別部19は、第2の類似度算出部18によって算出された複数の第2の類似度を合計した合計値を第1の話者毎に算出する。すなわち、話者識別部19は、第2の話者の音声信号と、第1の話者Xに対応付けられている複数の不特定話者モデルA～Cのそれぞれとの複数の第2の類似度を合計した合計値を算出する。また、話者識別部19は、第2の話者の音声信号と、第1の話者Yに対応付けられている複数の不特定話者モデルD～Fのそれぞれとの複数の第2の類似度を合計した合計値を算出する。

40

【0109】

例えば、第2の話者の音声信号と不特定話者モデルAとの第2の類似度は、0.7であり、第2の話者の音声信号と不特定話者モデルBとの第2の類似度は、0.6であり、第2の話者の音声信号と不特定話者モデルCとの第2の類似度は、0.5である。そのため

50

、第2の話者と第1の話者Xとの複数の第2の類似度の合計値は、 $0.7 + 0.6 + 0.5 = 1.8$ である。また、第2の話者の音声信号と不特定話者モデルDとの第2の類似度は、 0.4 であり、第2の話者の音声信号と不特定話者モデルEとの第2の類似度は、 0.8 であり、第2の話者の音声信号と不特定話者モデルFとの第2の類似度は、 0.3 である。そのため、第2の話者と第1の話者Yとの複数の第2の類似度の合計値は、 $0.4 + 0.8 + 0.3 = 1.5$ である。

【0110】

次に、話者識別部19は、算出した合計値に基づいて、第2の話者が第2の話者モデル格納部16に記憶されている第1の話者X、Yのいずれであるかを特定する。第2の話者と第1の話者Xとの複数の第2の類似度の合計値は、 1.8 であり、第2の話者と第1の話者Yとの複数の第2の類似度の合計値は、 1.5 であり、第1の話者Xの合計値の方が第1の話者Yの合計値よりも大きい。そのため、話者識別部19は、第2の話者が第1の話者Xであると決定する。

10

【0111】

このように、第2の話者が第1の話者Xであったとしても、第1の話者Xの話者モデルが十分に学習されていない場合、当該話者モデルを用いて第2の話者を識別することができないおそれがある。そこで、本実施の形態では、第1の話者Xに類似する複数の不特定話者の複数の不特定話者モデルを用いることにより、第2の話者を正確に識別することができる。

【0112】

20

また、第1の話者の音声信号に類似する複数の不特定話者の話者モデルを用いて、第2の話者が識別されるので、識別対象である話者の話者モデルが作成されていない場合であっても、より高い精度で話者を識別することができる。

【0113】

続いて、実施の形態1の第1の変形例について説明する。

【0114】

図6は、本実施の形態1の第1の変形例における話者識別装置の識別モード処理の具体的な処理の一例を説明するための模式図である。なお、本実施の形態1の第1の変形例における話者識別装置の学習モード処理は、上記の実施の形態1と同じであるので説明を省略する。

30

【0115】

実施の形態1の第1の変形例では、複数の第2の類似度が単に合計されるのではなく、第1の類似度の大きさの順位に応じた重み付け値を乗算した複数の第2の類似度が合計される。

【0116】

実施の形態1の第1の変形例における識別モード処理において、第2の話者が「音量」と発話した場合、音声入力部11は、第2の話者が発話した音声を集集し、集集した音声を音声信号に変換して出力する。第2の音声分析部17は、音声入力部11から入力された音声信号を分析し、第2の話者によって発話された音声信号を出力する。

【0117】

40

次に、第2の類似度算出部18は、第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者Xに対応付けられている複数の不特定話者モデルA～Cにおける各特徴量との複数の第2の類似度を算出する。また、第2の類似度算出部18は、第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者Yに対応付けられている複数の不特定話者モデルD～Fにおける各特徴量との複数の第2の類似度を算出する。

【0118】

次に、話者識別部19は、第2の類似度算出部18によって算出された複数の第2の類似度のそれぞれに第1の類似度の順位に応じた重み付け値を乗算することにより複数の第2の類似度を補正する。重み付け値は、前記第1の類似度が高い順に高くする。例えば、

50

第1の話者に対応付けられている複数の不特定話者モデルの第1の類似度のうち、値が最も大きい第1の類似度には1.5の重み付け値を付与し、2番目に大きい第1の類似度には1.0の重み付け値を付与し、3番目に大きい第1の類似度には0.5の重み付け値を付与する。

【0119】

なお、学習モード処理において、モデル特定部15は、第1の話者に対して複数の不特定話者モデルを第1の類似度が大きい順に対応付けて記憶することが好ましい。これにより、識別モード処理において、第1の類似度の順位に応じた重み付け値を決定する際に、第1の類似度を算出する必要がなくなり、処理速度を向上させることができる。また、学習モード処理において、モデル特定部15は、算出された第1の類似度を、複数の不特定話者モデルに対応付けて第2の話者モデル格納部16に記憶してもよい。また、識別モード処理において、第2の類似度算出部18は、第1の話者の話者モデルの特徴量と、第1の話者に対応付けられている複数の不特定話者モデルの特徴量との第1の類似度を算出してもよい。

10

【0120】

そして、話者識別部19は、補正した複数の第2の類似度を合計した合計値を第1の話者X、Y毎に算出する。

【0121】

例えば、第2の話者の音声信号と不特定話者モデルAとの第2の類似度は、0.7であり、第2の話者の音声信号と不特定話者モデルBとの第2の類似度は、0.6であり、第2の話者の音声信号と不特定話者モデルCとの第2の類似度は、0.5である。また、不特定話者モデルAの第1の類似度は、不特定話者モデルBの第1の類似度より大きく、不特定話者モデルBの第1の類似度は、不特定話者モデルCの第1の類似度より大きい。また、重み付け値は、第1の類似度が大きい順に1.5、1.0及び0.5である。そのため、第2の話者と第1の話者Xとの補正した複数の第2の類似度の合計値は、 $0.7 \times 1.5 + 0.6 \times 1.0 + 0.5 \times 0.5 = 1.9$ である。

20

【0122】

また、例えば、第2の話者の音声信号と不特定話者モデルDとの第2の類似度は、0.4であり、第2の話者の音声信号と不特定話者モデルEとの第2の類似度は、0.8であり、第2の話者の音声信号と不特定話者モデルFとの第2の類似度は、0.3である。また、不特定話者モデルDの第1の類似度は、不特定話者モデルEの第1の類似度より大きく、不特定話者モデルEの第1の類似度は、不特定話者モデルFの第1の類似度より大きい。また、重み付け値は、第1の類似度が大きい順に1.5、1.0及び0.5である。そのため、第2の話者と第1の話者Yとの補正した複数の第2の類似度の合計値は、 $0.4 \times 1.5 + 0.8 \times 1.0 + 0.3 \times 0.5 = 1.55$ である。

30

【0123】

次に、話者識別部19は、算出した合計値に基づいて、第2の話者が第2の話者モデル格納部16に記憶されている第1の話者X、Yのいずれであるかを特定する。第2の話者と第1の話者Xとの補正した複数の第2の類似度の合計値は、1.9であり、第2の話者と第1の話者Yとの補正した複数の第2の類似度の合計値は、1.55であり、第1の話者Xの合計値の方が第1の話者Yの合計値よりも大きい。そのため、話者識別部19は、第2の話者が第1の話者Xであると決定する。

40

【0124】

このように、複数の第2の類似度のそれぞれに第1の類似度の順位に応じた重み付け値が乗算されることにより複数の第2の類似度が補正されるので、語彙の違い及び発話文の長さなどによる第2の類似度の変動を抑えることができ、より高い精度で話者を識別することができる。

【0125】

続いて、実施の形態1の第2の変形例について説明する。

【0126】

50

図7は、本実施の形態1の第2の変形例における話者識別装置の識別モード処理の具体的な処理の一例を説明するための模式図である。なお、本実施の形態1の第2の変形例における話者識別装置の学習モード処理は、上記の実施の形態1と同じであるので説明を省略する。

【0127】

実施の形態1の第2の変形例では、複数の第2の類似度が単に合計されるのではなく、第1の類似度の値に応じた重み付け値を乗算した複数の第2の類似度が合計される。

【0128】

実施の形態1の第2の変形例における識別モード処理において、第2の話者が「音量」と発話した場合、音声入力部11は、第2の話者が発話した音声を収集し、収集した音声を音声信号に変換して出力する。第2の音声分析部17は、音声入力部11から入力された音声信号を分析し、第2の話者によって発話された音声信号を出力する。

10

【0129】

次に、第2の類似度算出部18は、第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者Xに対応付けられている複数の不特定話者モデルA～Cにおける各特徴量との複数の第2の類似度を算出する。また、第2の類似度算出部18は、第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者Yに対応付けられている複数の不特定話者モデルD～Fにおける各特徴量との複数の第2の類似度を算出する。

20

【0130】

次に、話者識別部19は、第2の類似度算出部18によって算出された複数の第2の類似度のそれぞれに第1の類似度を乗算することにより複数の第2の類似度を補正する。なお、第1の類似度は、複数の不特定話者モデルに対応付けて第2の話者モデル格納部16に記憶されている。

【0131】

そして、話者識別部19は、補正した複数の第2の類似度を合計した合計値を第1の話者X、Y毎に算出する。

【0132】

例えば、第2の話者の音声信号と不特定話者モデルAとの第2の類似度は、0.7であり、第2の話者の音声信号と不特定話者モデルBとの第2の類似度は、0.6であり、第2の話者の音声信号と不特定話者モデルCとの第2の類似度は、0.5である。また、不特定話者モデルAに対応付けられている第1の類似度は、0.9であり、不特定話者モデルBに対応付けられている第1の類似度は、0.7であり、不特定話者モデルCに対応付けられている第1の類似度は、0.4である。そのため、第2の話者と第1の話者Xとの補正した複数の第2の類似度の合計値は、 $0.7 \times 0.9 + 0.6 \times 0.7 + 0.5 \times 0.4 = 1.25$ である。

30

【0133】

また、例えば、第2の話者の音声信号と不特定話者モデルDとの第2の類似度は、0.4であり、第2の話者の音声信号と不特定話者モデルEとの第2の類似度は、0.8であり、第2の話者の音声信号と不特定話者モデルFとの第2の類似度は、0.3である。また、不特定話者モデルDに対応付けられている第1の類似度は、0.8であり、不特定話者モデルEに対応付けられている第1の類似度は、0.5であり、不特定話者モデルFに対応付けられている第1の類似度は、0.4である。そのため、第2の話者と第1の話者Yとの補正した複数の第2の類似度の合計値は、 $0.4 \times 0.8 + 0.8 \times 0.5 + 0.3 \times 0.4 = 0.84$ である。

40

【0134】

次に、話者識別部19は、算出した合計値に基づいて、第2の話者が第2の話者モデル格納部16に記憶されている第1の話者X、Yのいずれであるかを特定する。第2の話者と第1の話者Xとの補正した複数の第2の類似度の合計値は、1.25であり、第2の話者と第1の話者Yとの補正した複数の第2の類似度の合計値は、0.84であり、第1の

50

話者Xの合計値の方が第1の話者Yの合計値よりも大きい。そのため、話者識別部19は、第2の話者が第1の話者Xであると決定する。

【0135】

このように、複数の第2の類似度のそれぞれに第1の類似度が乗算されることにより複数の第2の類似度が補正されるので、語彙の違い及び発話文の長さなどによる第2の類似度の変動を抑えることができ、より高い精度で話者を識別することができる。

【0136】

続いて、実施の形態1の第3の変形例について説明する。

【0137】

図8は、本実施の形態1の第3の変形例における話者識別装置の識別モード処理の具体的な処理の一例を説明するための模式図である。なお、本実施の形態1の第3の変形例における話者識別装置の学習モード処理は、上記の実施の形態1と同じであるので説明を省略する。

10

【0138】

実施の形態1の第3の変形例では、複数の第2の類似度が単に合計されるのではなく、所定の値以上である複数の第2の類似度が合計される。

【0139】

実施の形態1の第3の変形例における識別モード処理において、第2の話者が「音量」と発話した場合、音声入力部11は、第2の話者が発話した音声を集集し、集集した音声を音声信号に変換して出力する。第2の音声分析部17は、音声入力部11から入力された音声信号を分析し、第2の話者によって発話された音声信号を出力する。

20

【0140】

次に、第2の類似度算出部18は、第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者Xに対応付けられている複数の不特定話者モデルA～Cにおける各特徴量との複数の第2の類似度を算出する。また、第2の類似度算出部18は、第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者Yに対応付けられている複数の不特定話者モデルD～Fにおける各特徴量との複数の第2の類似度を算出する。

【0141】

次に、話者識別部19は、第2の類似度算出部18によって算出された複数の第2の類似度のうち、所定の値以上である複数の第2の類似度を合計した合計値を第1の話者毎に算出する。すなわち、話者識別部19は、第2の話者の音声信号と、第1の話者Xに対応付けられている複数の不特定話者モデルA～Cとの複数の第2の類似度のうち、所定の値以上である複数の第2の類似度を合計した合計値を算出する。また、話者識別部19は、第2の話者の音声信号と、第1の話者Yに対応付けられている複数の不特定話者モデルD～Fとの複数の第2の類似度のうち、所定の値以上である複数の第2の類似度を合計した合計値を算出する。

30

【0142】

例えば、話者識別部19は、0.5未満の第2の類似度を棄却し、0.5以上の複数の第2の類似度を合計する。例えば、第2の話者の音声信号と不特定話者モデルAとの第2の類似度は、0.7であり、第2の話者の音声信号と不特定話者モデルBとの第2の類似度は、0.6であり、第2の話者の音声信号と不特定話者モデルCとの第2の類似度は、0.5である。そのため、第2の話者と第1の話者Xとの複数の第2の類似度の合計値は、 $0.7 + 0.6 + 0.5 = 1.8$ である。

40

【0143】

また、例えば、第2の話者の音声信号と不特定話者モデルDとの第2の類似度は、0.4であり、第2の話者の音声信号と不特定話者モデルEとの第2の類似度は、0.8であり、第2の話者の音声信号と不特定話者モデルFとの第2の類似度は、0.3である。そのため、第2の話者と第1の話者Yとの複数の第2の類似度の合計値は、 $0.4 + 0.8 = 1.2$ である。

50

【0144】

次に、話者識別部19は、算出した合計値に基づいて、第2の話者が第2の話者モデル格納部16に記憶されている第1の話者X、Yのいずれであるかを特定する。第2の話者と第1の話者Xとの複数の第2の類似度の合計値は、1.8であり、第2の話者と第1の話者Yとの複数の第2の類似度の合計値は、1.2であり、第1の話者Xの合計値の方が第1の話者Yの合計値よりも大きい。そのため、話者識別部19は、第2の話者が第1の話者Xであると決定する。

【0145】

このように、所定の値以上である複数の第2の類似度が合計され、所定の値より小さい第2の類似度が考慮されないので、より高い精度で話者を識別することができる。

10

【0146】

(実施の形態2)

図9は、本実施の形態2における話者識別装置の構成を示す図である。話者識別装置は、例えば、テレビ、スマートフォン又はカーナビゲーション装置などに内蔵される。

【0147】

図9に示す話者識別装置は、音声入力部11、第1の音声分析部12、第1の話者モデル格納部13、第1の類似度算出部14、モデル特定部15、第2の話者モデル格納部16、第2の音声分析部17、第2の類似度算出部18、モデル作成部21、第3の話者モデル格納部22、第3の類似度算出部23及び話者識別部24を備える。

20

【0148】

また、学習モード処理部101は、音声入力部11、第1の音声分析部12、第1の話者モデル格納部13、第1の類似度算出部14、モデル特定部15、第2の話者モデル格納部16、モデル作成部21及び第3の話者モデル格納部22で構成される。識別モード処理部102は、音声入力部11、第2の話者モデル格納部16、第2の音声分析部17、第2の類似度算出部18、第3の話者モデル格納部22、第3の類似度算出部23及び話者識別部24で構成される。

【0149】

なお、実施の形態2において、実施の形態1と同じ構成については同じ符号を付し、説明を省略する。

【0150】

30

モデル作成部21は、モデル特定部15によって特定された複数の不特定話者モデルと、取得された第1の話者の音声信号とに基づいて、第1の話者に対応する話者モデルを新たに作成する。例えば、モデル作成部21は、モデル特定部15によって特定された複数の不特定話者モデルと、取得された第1の話者の音声信号とを平均化し、平均化した話者モデルを第1の話者に対応する新たな話者モデルとして作成する。なお、本実施の形態1では、平均化することにより新たな話者モデルを作成しているが、特にこれに限定されず、他の方法により新たな話者モデルを作成してもよい。モデル作成部21は、第1の話者を識別する識別情報と、作成した話者モデルとを対応付けて第3の話者モデル格納部22に記憶する。

【0151】

40

第3の話者モデル格納部22は、第1の話者を識別する識別情報と、モデル作成部21によって作成された話者モデルとを対応付けて記憶する。

【0152】

第3の類似度算出部23は、第2の音声分析部17によって取得された第2の話者の音声信号における特徴量と、第3の話者モデル格納部22（第3のデータベース）に記憶されている第1の話者に対応付けられている話者モデルにおける特徴量との第3の類似度を第1の話者毎に算出する。

【0153】

第3の類似度算出部23は、取得された第2の話者の音声信号における特徴量を算出する。ここで、特徴量は、i V e c t o rを用いて算出される。なお、i V e c t o rを用

50

いた特徴量の算出方法については従来技術であるため、詳細な説明は省略する。また、特徴量の算出方法については、i V e c t o rに限定されず、他の算出方法を用いてもよい。また、第3の類似度算出部23は、第1の類似度算出部14又は第2の類似度算出部18による特徴量の算出方法と同じ算出方法で特徴量を算出してもよいし、第1の類似度算出部14又は第2の類似度算出部18による特徴量の算出方法とは異なる算出方法で特徴量を算出してもよい。

【0154】

また、第3の類似度算出部23は、第3の話者モデル格納部22に記憶されている複数の第1の話者のそれぞれに対応付けられている話者モデルを読み出し、読み出した話者モデルにおける特徴量を複数の第1の話者毎に算出する。

10

【0155】

第3の類似度算出部23は、第2の話者の音声信号における特徴量と、第3の話者モデル格納部22に記憶されており、第1の話者に対応付けられている話者モデルにおける特徴量との第3の類似度を第1の話者毎にそれぞれ算出する。第3の類似度算出部23は、2つの特徴量がどれだけ類似しているかを示す話者間距離を算出する。第3の類似度算出部23は、2つの特徴量の間の近さを算出する関数を用いて、話者間距離を算出する。話者間距離は、特徴量同士が類似している場合には1に近い値となり、特徴量同士が類似していない場合には0に近い値となる。なお、類似度の算出方法については従来技術であるため、詳細な説明は省略する。また、類似度の算出方法については上記に限定されない。

【0156】

20

また、第3の類似度算出部23は、第1の類似度算出部14による第1の類似度の算出方法又は第2の類似度算出部18による第2の類似度の算出方法と同じ算出方法で第3の類似度を算出してもよいし、第1の類似度算出部14による第1の類似度の算出方法又は第2の類似度算出部18による第2の類似度の算出方法とは異なる算出方法で第3の類似度を算出してもよい。

【0157】

話者識別部24は、第3の類似度算出部23によって算出された第3の類似度に基づいて、第2の話者に対応する第3の話者モデル格納部22に記憶されている複数の第1の話者の中の一人を特定する。ここで、話者識別部24は、第3の類似度算出部23によって算出された第3の類似度が最も高い第3の話者モデル格納部22に記憶されている第1の話者として、第2の話者を特定する。

30

【0158】

なお、話者識別部24は、第3の類似度算出部23によって算出された第3の類似度が所定値以上であり且つ最も高い第1の話者として、第2の話者を特定してもよい。

【0159】

第2の話者が第3の話者モデル格納部22に記憶されている第1の話者のいずれにも特定されない場合、第2の類似度算出部18は、第2の音声分析部17によって取得された第2の話者の音声信号における特徴量と、第2の話者モデル格納部16に記憶されており、第1の話者に対応付けられている複数の不特定話者モデルにおける各特徴量との複数の第2の類似度を第1の話者毎にそれぞれ算出する。

40

【0160】

話者識別部24は、第2の類似度算出部18によって算出された複数の第2の類似度に基づいて、第2の話者が第2の話者モデル格納部16に記憶されている第1の話者のいずれであるかを特定する。

【0161】

続いて、本実施の形態2における話者識別装置の学習モード処理及び識別モード処理のそれぞれの動作について説明する。

【0162】

図10は、本実施の形態2における話者識別装置の学習モード処理の動作について説明するためのフローチャートである。

50

【0163】

なお、図10に示すステップS21～ステップS25の処理は、図2に示すステップS1～ステップS5の処理と同じであるので説明を省略する。

【0164】

ステップS26において、モデル作成部21は、モデル特定部15によって特定された複数の不特定話者モデルと、取得された第1の話者の音声信号とに基づいて、第1の話者に対応する話者モデルを新たに作成する。

【0165】

次に、ステップS27において、モデル作成部21は、第1の話者を識別する識別情報と、作成した話者モデルとを対応付けて第3の話者モデル格納部22に記憶する。

10

【0166】

図11は、本実施の形態2における話者識別装置の識別モード処理の動作について説明するためのフローチャートである。

【0167】

まず、ステップS31において、第2の音声分析部17は、第2の話者の音声信号を取得する。

【0168】

次に、ステップS32において、第3の類似度算出部23は、第2の音声分析部17によって取得された第2の話者の音声信号の特徴量、及び第3の話者モデル格納部22に記憶されている第1の話者に対応付けられている話者モデルの特徴量を算出する。

20

【0169】

次に、ステップS33において、第3の類似度算出部23は、第2の話者の音声信号の特徴量と、第1の話者に対応付けられている話者モデルの特徴量との第3の類似度を第1の話者毎にそれぞれ算出する。

【0170】

次に、ステップS34において、話者識別部24は、算出された複数の第3の類似度の中に、所定値以上の第3の類似度があるか否かを判断する。なお、所定値は、2つの話者モデルが類似していると判断することが可能な値であり、例えば、0.5である。

【0171】

ここで、所定値以上の第3の類似度があると判断された場合（ステップS34でYES）、ステップS35において、話者識別部24は、所定値以上である第3の類似度のうち、値が最も高い第3の類似度に対応する第1の話者として、第2の話者を特定する。

30

【0172】

一方、所定値以上の第3の類似度がないと判断された場合（ステップS34でNO）、ステップS36において、第2の類似度算出部18は、第2の音声分析部17によって取得された第2の話者の音声信号の特徴量、及び第2の話者モデル格納部16に記憶されており、複数の第1の話者のそれぞれに対応付けられている複数の不特定話者モデルの特徴量を算出する。なお、ステップS32で第2の話者の音声信号の特徴量が既に算出されているため、第2の類似度算出部18は、第2の話者の音声信号の特徴量を算出しなくてもよい。

40

【0173】

なお、図11に示すステップS37～ステップS39の処理は、図3に示すステップS13～ステップS15の処理と同じであるので説明を省略する。

【0174】

続いて、本実施の形態2における話者識別装置の具体的な処理の流れについて説明する。

【0175】

図12は、本実施の形態2における話者識別装置の学習モード処理の具体的な処理の一例を説明するための模式図である。

【0176】

50

まず、学習モード処理において、第1の話者Xが「テレビ」と発話した場合、音声入力部11は、第1の話者Xが発話した音声を集集し、集集した音声を音声信号に変換して出力する。なお、本実施の形態1では、第1の話者Xが発話する語彙は、特に限定されず、第1の話者Xは、予め決められている語彙を発話する必要はない。第1の音声分析部12は、音声入力部11から入力された音声信号を分析し、第1の話者Xによって発話された音声信号を出力する。

【0177】

次に、第1の類似度算出部14は、取得された第1の話者Xの音声信号と、他者の話者モデルとを比較する。すなわち、第1の類似度算出部14は、取得された第1の話者Xの音声信号における特徴量と、第1の話者モデル格納部13に記憶されている複数の不特定話者モデルA～Fにおける各特徴量との第1の類似度をそれぞれ算出する。

10

【0178】

次に、モデル特定部15は、第1の類似度算出部14によって算出された第1の類似度が所定の値以上である複数の不特定話者モデルを特定する。図12では、第1の類似度が0.4以上である複数の不特定話者モデルA～Cが特定されている。例えば、第1の話者Xの音声信号と不特定話者モデルAとの第1の類似度は、0.9であり、第1の話者Xの音声信号と不特定話者モデルBとの第1の類似度は、0.7であり、第1の話者Xの音声信号と不特定話者モデルCとの第1の類似度は、0.4である。モデル特定部15は、第1の話者Xを識別する識別情報と、特定した複数の不特定話者モデルA～Cとを対応付けたテーブルを、第2の話者モデル格納部16に記憶する。

20

【0179】

次に、モデル作成部21は、モデル特定部15によって特定された複数の不特定話者モデルA～Cと、取得された第1の話者Xの音声信号とに基づいて、第1の話者Xに対応する話者モデルX'を新たに作成する。モデル作成部21は、第1の話者Xを識別する識別情報と、第1の話者Xに対応する新たに作成した話者モデルX'とを対応付けて第3の話者モデル格納部22に記憶する。

【0180】

同様に、学習モード処理において、第1の話者Yが「音量」と発話した場合、音声入力部11は、第1の話者Yが発話した音声を集集し、集集した音声を音声信号に変換して出力する。第1の音声分析部12は、音声入力部11から入力された音声信号を分析し、第1の話者Yによって発話された音声信号を出力する。

30

【0181】

次に、第1の類似度算出部14は、取得された第1の話者Yの音声信号と、他者の話者モデルとを比較する。すなわち、第1の類似度算出部14は、取得された第1の話者Yの音声信号における特徴量と、第1の話者モデル格納部13に記憶されている複数の不特定話者モデルA～Fにおける各特徴量との第1の類似度をそれぞれ算出する。

【0182】

次に、モデル特定部15は、第1の類似度算出部14によって算出された第1の類似度が所定の値以上である複数の不特定話者モデルを特定する。図12では、第1の類似度が0.4以上である複数の不特定話者モデルD～Fが特定されている。例えば、第1の話者Yの音声信号と不特定話者モデルDとの第1の類似度は、0.8であり、第1の話者Yの音声信号と不特定話者モデルEとの第1の類似度は、0.5であり、第1の話者Yの音声信号と不特定話者モデルFとの第1の類似度は、0.4である。モデル特定部15は、第1の話者Yを識別する識別情報と、特定した複数の不特定話者モデルD～Fとを対応付けたテーブルを、第2の話者モデル格納部16に記憶する。

40

【0183】

次に、モデル作成部21は、モデル特定部15によって特定された複数の不特定話者モデルD～Fと、取得された第1の話者Yの音声信号とに基づいて、第1の話者Yに対応する話者モデルY'を新たに作成する。モデル作成部21は、第1の話者Yを識別する識別情報と、第1の話者Yに対応する新たに作成した話者モデルY'とを対応付けて第3の話

50

者モデル格納部 2 2 に記憶する。

【0184】

このように、本実施の形態 2 では、取得した第 1 の話者の音声信号に類似する複数の他者の話者モデルを用いて第 1 の話者に対応する新たな話者モデルが作成されるので、新たに作成された話者モデルを用いて話者をより高い精度で識別することができる。

【0185】

(実施の形態 3)

図 1 3 は、本実施の形態 3 における話者識別装置の構成を示す図である。話者識別装置は、例えば、テレビ、スマートホン又はカーナビゲーション装置などに内蔵される。

【0186】

図 1 3 に示す話者識別装置は、音声入力部 1 1、第 1 の音声分析部 1 2、第 1 の話者モデル格納部 1 3、第 1 の類似度算出部 1 4、モデル特定部 1 5、第 2 の音声分析部 1 7、モデル作成部 2 1、第 3 の話者モデル格納部 2 2、第 3 の類似度算出部 2 3 及び話者識別部 2 4 を備える。

【0187】

また、学習モード処理部 2 0 1 は、音声入力部 1 1、第 1 の音声分析部 1 2、第 1 の話者モデル格納部 1 3、第 1 の類似度算出部 1 4、モデル特定部 1 5、モデル作成部 2 1 及び第 3 の話者モデル格納部 2 2 で構成される。識別モード処理部 2 0 2 は、音声入力部 1 1、第 2 の音声分析部 1 7、第 3 の話者モデル格納部 2 2、第 3 の類似度算出部 2 3 及び話者識別部 2 4 で構成される。

【0188】

なお、実施の形態 3 において、実施の形態 1 及び実施の形態 2 と同じ構成については同じ符号を付し、説明を省略する。

【0189】

実施の形態 3 が実施の形態 2 と異なる点は、実施の形態 3 の話者識別装置が、実施の形態 2 における第 2 の話者モデル格納部 1 6 及び第 2 の類似度算出部 1 8 を備えていない点である。実施の形態 3 において、モデル特定部 1 5 は、特定した複数の不特定話者モデルを、第 1 の話者を識別する識別情報に対応付けて第 2 の話者モデル格納部 1 6 に記憶しない。そのため、実施の形態 3 の識別モード処理では、特定した複数の不特定話者モデルと第 2 の話者の音声信号とに基づいて新たに作成された話者モデルのみを用いて、話者が識別される。

【0190】

モデル作成部 2 1 は、モデル特定部 1 5 によって特定された複数の不特定話者モデルと、取得された第 1 の話者の音声信号とに基づいて、第 1 の話者に対応する話者モデルを新たに作成する。モデル作成部 2 1 は、第 1 の話者を識別する識別情報と、作成した話者モデルとを対応付けて第 3 の話者モデル格納部 2 2 に記憶する。

【0191】

第 3 の話者モデル格納部 2 2 は、ハードウェア構成として、例えば、メモリであってもよい。メモリとは、例えば、半導体メモリ、ハードディスクといったような書き込み可能な記録媒体である。

【0192】

第 3 の類似度算出部 2 3 は、第 2 の音声分析部 1 7 によって取得された第 2 の話者の音声信号における特徴量と、第 3 の話者モデル格納部 2 2 (第 3 のデータベース) に記憶されている第 1 の話者に対応付けられている話者モデルにおける特徴量との第 3 の類似度を第 1 の話者毎に算出する。

【0193】

話者識別部 2 4 は、第 3 の類似度算出部 2 3 によって算出された第 3 の類似度に基づいて、第 2 の話者に対応する第 3 の話者モデル格納部 2 2 に記憶されている複数の第 1 の話者の中の一人を特定する。

【0194】

続いて、本実施の形態 3 における話者識別装置の学習モード処理及び識別モード処理のそれぞれの動作について説明する。

【0195】

図 14 は、本実施の形態 3 における話者識別装置の学習モード処理の動作について説明するためのフローチャートである。

【0196】

なお、図 14 に示すステップ S 4 1 ～ステップ S 4 4 の処理は、図 2 に示すステップ S 1 ～ステップ S 4 の処理と同じであるので説明を省略する。

【0197】

ステップ S 4 5 において、モデル作成部 2 1 は、モデル特定部 1 5 によって特定された複数の不特定話者モデルと、取得された第 1 の話者の音声信号とに基づいて、第 1 の話者に対応する話者モデルを新たに作成する。

10

【0198】

次に、ステップ S 4 6 において、モデル作成部 2 1 は、第 1 の話者を識別する識別情報と、作成した話者モデルとを対応付けて第 3 の話者モデル格納部 2 2 に記憶する。

【0199】

図 15 は、本実施の形態 3 における話者識別装置の識別モード処理の動作について説明するためのフローチャートである。

【0200】

なお、図 15 に示すステップ S 5 1 ～ステップ S 5 5 の処理は、図 11 に示すステップ S 3 1 ～ステップ S 3 5 の処理と同じであるので説明を省略する。

20

【0201】

ステップ S 5 4 において所定値以上の第 3 の類似度がないと判断された場合（ステップ S 5 4 で NO）、ステップ S 5 6 において、話者識別部 2 4 は、第 2 の話者を特定することができないことを通知する。話者識別装置が表示部を備えている場合、話者識別部 2 4 は、第 2 の話者を特定することができないことを示す画面を表示部に表示する。また、話者識別装置が音声出力部を備えている場合、話者識別部 2 4 は、第 2 の話者を特定することができないことを示す音声を音声出力部から出力する。なお、所定値以上の第 3 の類似度がないと判断された場合、第 2 の話者を特定することができないことを通知することなく、処理を終了してもよい。

30

【0202】

このように、本実施の形態 3 では、取得した第 1 の話者の音声信号に類似する複数の他者の話者モデルを用いて第 1 の話者に対応する新たな話者モデルが作成されるので、新たに作成された話者モデルを用いて話者をより高い精度で識別することができる。

【0203】

なお、上記各実施の形態において、各構成要素は、専用のハードウェアで構成されるか、各構成要素に適したソフトウェアプログラムを実行することによって実現されてもよい。各構成要素は、CPU またはプロセッサなどのプログラム実行部が、ハードディスクまたは半導体メモリなどの記録媒体に記録されたソフトウェアプログラムを読み出して実行することによって各構成要素の動作または機能が実現されてもよい。

40

【0204】

各構成要素の動作または機能を実現するときに利用するデータは、半導体メモリ、または、ハードディスクなどの書き込み可能な記録媒体に記録されてもよい。

【産業上の利用可能性】

【0205】

本開示に係る話者識別方法及び話者識別装置は、識別対象である話者の話者モデルが作成されていない場合であっても、より高い精度で話者を識別することができ、取得した音声信号に基づいて話者を識別する話者識別方法及び話者識別装置として有用である。

【符号の説明】

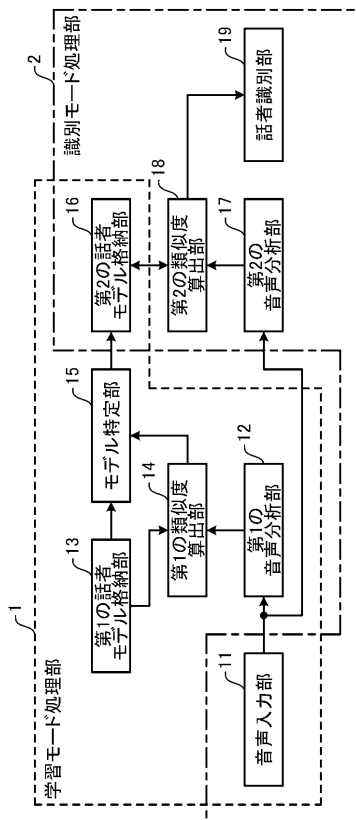
【0206】

50

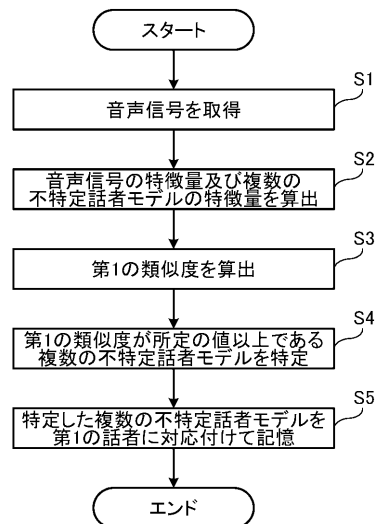
- 1, 101, 201 学習モード処理部
 2, 102, 202 識別モード処理部
 11 音声入力部
 12 第1の音声分析部
 13 第1の話者モデル格納部
 14 第1の類似度算出部
 15 モデル特定部
 16 第2の話者モデル格納部
 17 第2の音声分析部
 18 第2の類似度算出部
 19, 24 話者識別部
 21 モデル作成部
 22 第3の話者モデル格納部
 23 第3の類似度算出部

10

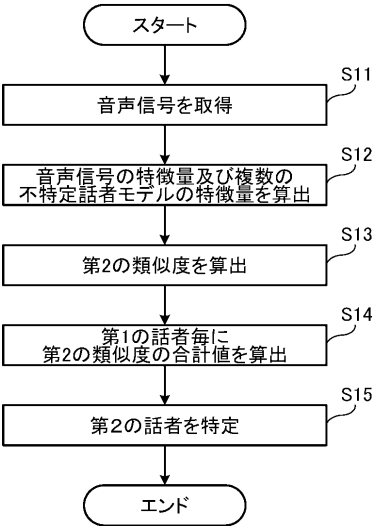
【図1】



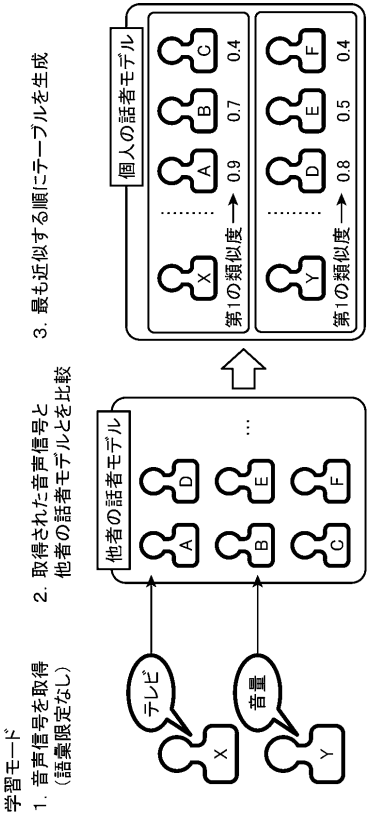
【図2】



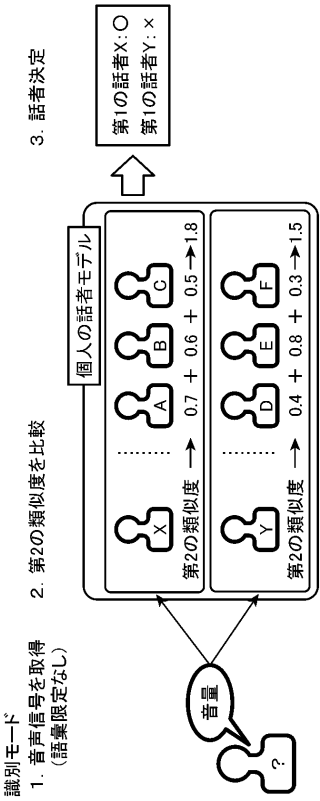
【図 3】



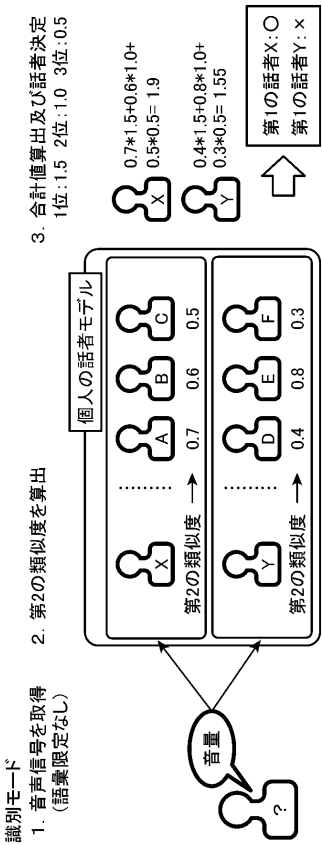
【図 4】



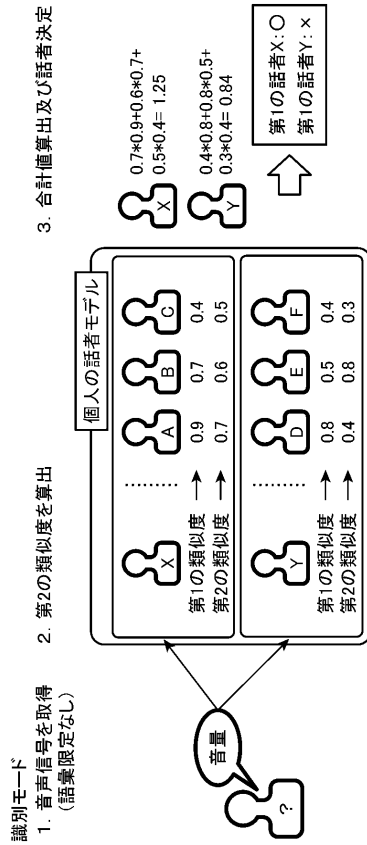
【図 5】



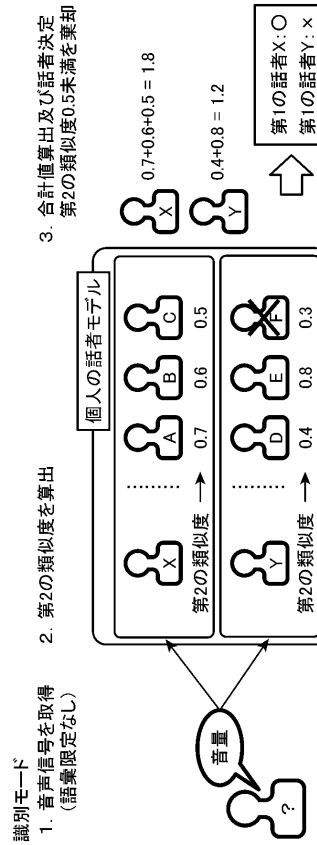
【図 6】



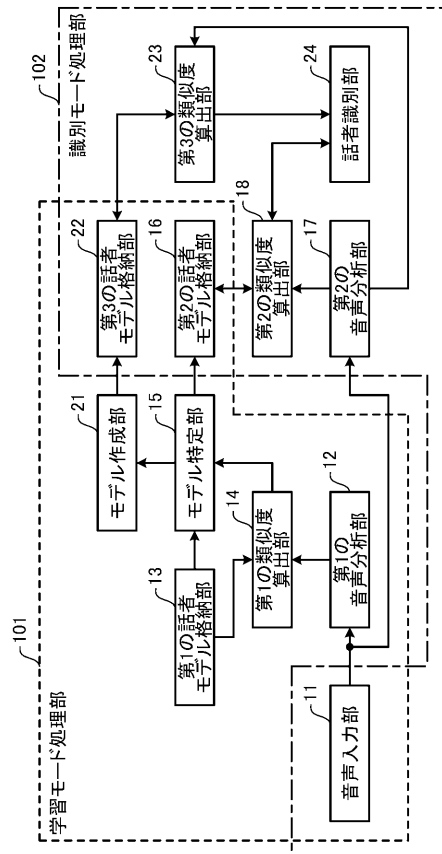
【図 7】



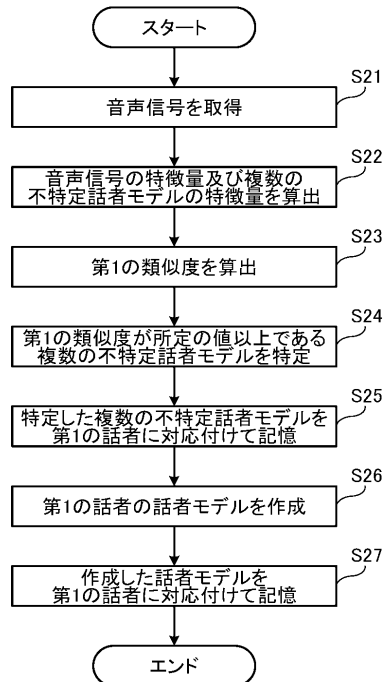
【图 8】



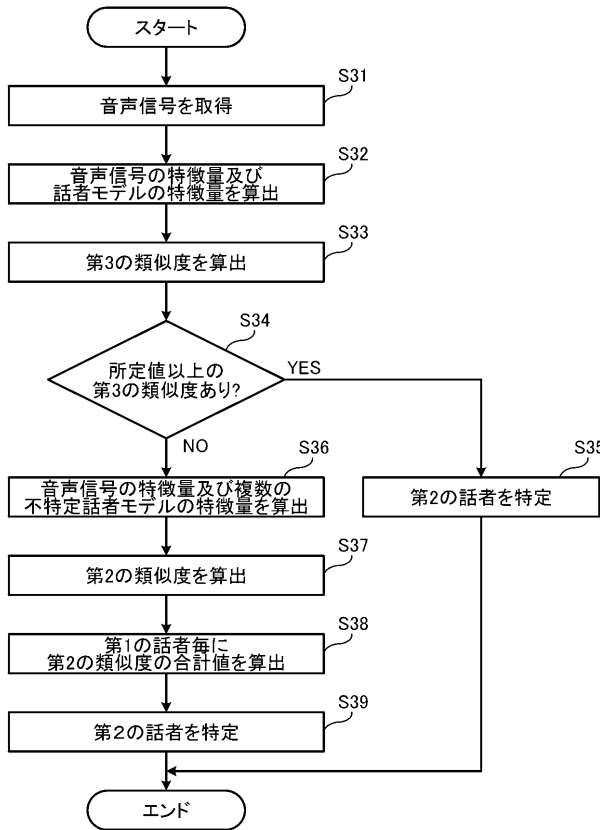
【図 9】



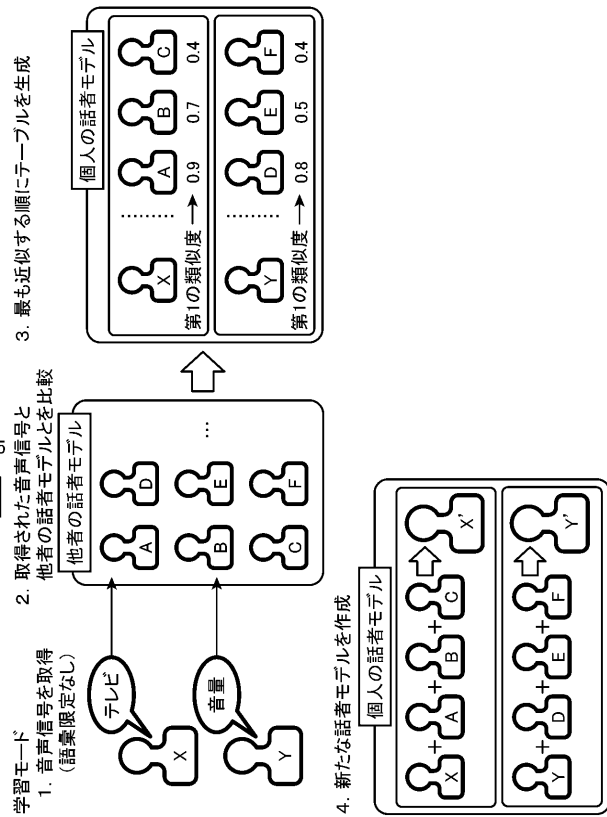
【図 10】



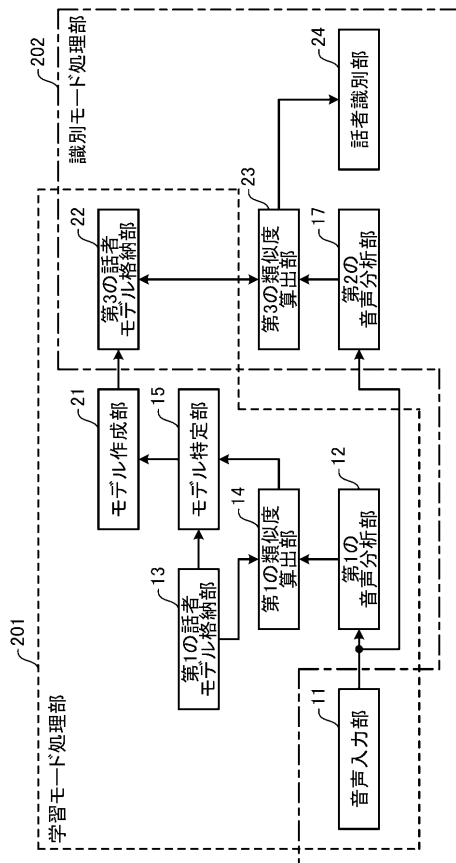
【図 1 1】



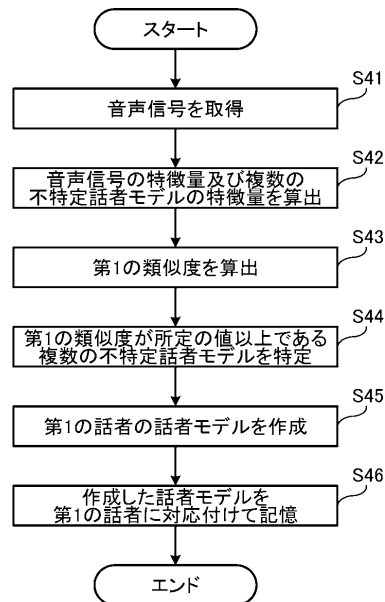
【図 1 2】



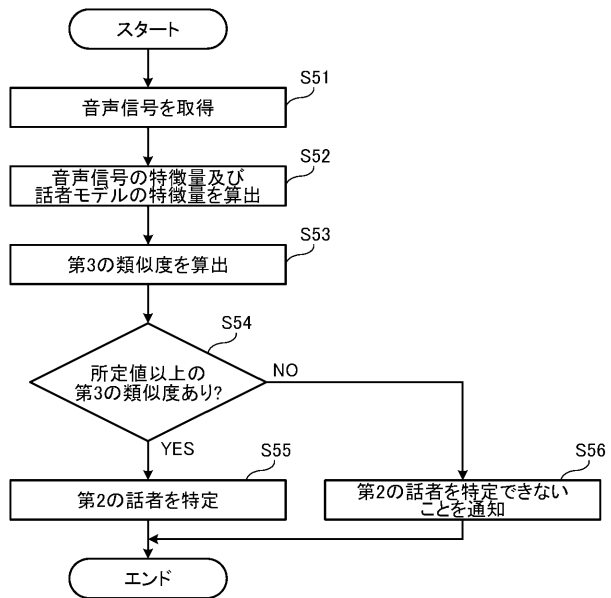
【図 1 3】



【图 14】



【図 15】



フロントページの続き

審査官 千本 潤介

(56)参考文献 特開平11-344992 (JP, A)

特表2014-502374 (JP, A)

赤津 達也,加藤 正治,小坂 哲夫,好田 正紀,音素クラスHMMを使用した話者ベクトルに基づく話者識別法の検討,情報処理学会研究報告,日本,社団法人情報処理学会,2007年12月20日,第2007巻第129号,pp.229-234

(58)調査した分野(Int.Cl., DB名)

G10L 17/00-17/26