

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 2 区分
 【発行日】平成 19 年 3 月 8 日 (2007.3.8)

【公開番号】特開 2001-255886 (P2001-255886A)

【公開日】平成 13 年 9 月 21 日 (2001.9.21)

【出願番号】特願 2000-64919 (P2000-64919)

【国際特許分類】

G 1 0 L 15/10 (2006.01)

G 1 0 L 11/06 (2006.01)

G 1 0 L 15/02 (2006.01)

G 1 0 L 15/20 (2006.01)

G 1 0 L 21/02 (2006.01)

【F I】

G 1 0 L 9/16 3 0 1 B

G 1 0 L 3/00 5 1 5 D

G 1 0 L 3/02 3 0 1 A

G 1 0 L 101:16

【手続補正書】

【提出日】平成 19 年 1 月 18 日 (2007.1.18)

【手続補正 1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】入力音声の特徴量を抽出する特徴量抽出ステップと、

前記入力音声の特徴量の周波数軸を、平滑化バッファに保持されている少なくとも 1 つの周波数軸変換係数から構成され、かつ、過去に求められた周波数軸変換係数列を用いて変換する周波数軸変換ステップと、

前記周波数軸変換を施した入力音声の特徴量と予め複数の話者から学習した音響モデル特徴量とをマッチングし、認識結果候補を出力するマッチングステップと、

前記認識結果候補のうち少なくとも 1 つから表現される音素系列に対して最適な周波数軸変換係数を推定する最適変換係数推定ステップと、

前記求められた最適な周波数軸変換係数と、前記平滑化バッファに保持された過去に求められた周波数軸変換係数列とを平滑化し、新たな周波数軸変換係数列を求め、当該新たな周波数軸変換係数列を平滑化バッファに保持させる変換係数平滑化ステップとを有することを特徴とする音声認識方法。

【請求項 2】前記周波数軸変換ステップは、

前記平滑化バッファに新たな周波数軸変換係数列を保持させた後は、前記新たな周波数軸変換係数列を用いて、前記入力音声の次に入力される入力音声の特徴量の周波数軸を変換することを特徴とする請求項 1 に記載の音声認識方法。

【請求項 3】前記変換係数平滑化ステップは、

前記平滑化バッファに保持された過去に求められた周波数軸変換係数の平均値と、前記最適な変換係数推定ステップで推定された最適な周波数軸変換係数との間の距離を求め、当該距離が所定以上大きい場合に、話者の交代を検知することを特徴とする請求項 1 又は 2 に記載の音声認識方法。

【請求項 4】前記変換係数平滑化ステップは、

前記平滑化バッファに保持された過去に求められた周波数軸変換係数の平均値と、前記最

適な変換係数推定ステップで推定された最適な周波数軸変換係数との間の距離を求め、当該距離が所定以上大きい場合に、前記平滑化バッファに保持された過去の周波数軸変換係数を任意の初期値を与えることによって初期化を行うことを特徴とする請求項 1 から 3 のいずれかに 1 つに記載の音声認識方法。

【請求項 5】前記変換係数平滑化ステップは、前記平滑化バッファに保持された過去に求められた周波数軸変換係数の平均値と、前記最適な変換係数推定ステップで推定された最適な周波数軸変換係数との間の距離を求め、当該距離が所定の距離を満たす場合にのみ、前記平滑化バッファに保持させることを特徴とする請求項 1 又は 2 に記載の音声認識方法。

【請求項 6】さらに、前記入力音声から無声音 / 有声音区間を弁別検出する音声区間検出ステップを有し、

前記周波数軸変換ステップにおいて、前記入力音声のうち、前記音声区間検出ステップで検出した有声音区間に対する特徴量の周波数軸を、平滑化バッファに保持されている少なくとも 1 つの周波数軸変換係数から構成される周波数軸変換係数列を用いて変換することを特徴とする請求項 1 から 5 のいずれかに 1 つに記載の音声認識方法。

【請求項 7】前記最適変換係数推定ステップは、前記認識結果候補を求めた際に使用した周波数軸変換係数を基に、複数の周波数軸変換係数候補を設定し、前記複数の周波数軸変換係数候補毎に、前記マッチングステップの認識結果候補のうち少なくとも 1 つから表現される音素系列に対して、周波数軸変換を施して得られた入力音声特徴量と予め複数の話者から学習した音響モデル特徴量との間の尤度を、マッチングにより求め、

さらに前記周波数軸変換係数候補と、前記求めた尤度との関係を二次曲線に近似し、当該二次曲線のピークにおける新たな周波数軸変換係数候補を求め、当該新たな周波数軸変換係数候補により周波数軸変換を施して得られた入力音声特徴量と予め複数の話者から学習した音響モデル特徴量との間の尤度を、マッチングにより求め、

前記周波数軸変換係数候補に対応する尤度と、前記新たな周波数軸変換係数候補に対応する尤度との中から、前記尤度が最も高い候補を、最適な周波数軸変換として選択することを特徴とする 1 から 6 のいずれか 1 つに記載の音声認識方法。

【請求項 8】入力音声の特徴量を抽出する特徴量抽出手段と、

少なくとも 1 つの周波数軸変換係数から構成され、かつ、過去に求められた周波数軸変換係数列が保持されている平滑化バッファと、

前記入力音声の特徴量の周波数軸を、平滑化バッファに保持されている周波数軸変換係数列を用いて変換する周波数軸変換手段と、

前記周波数軸変換を施した入力音声の特徴量と、予め複数の話者から学習した音響モデル特徴量とをマッチングし、認識結果候補を出力するマッチング手段と、

前記認識結果候補のうち、少なくとも 1 つから表現される音素系列に対して、少なくとも 1 つの周波数軸変換係数から構成される最適な周波数軸変換係数を推定する最適変換係数推定手段と、

前記求められた最適な周波数軸変換係数と、前記平滑化バッファに保持された過去に求められた周波数軸変換係数列とを平滑化して、新たな周波数軸変換係数列を求め、当該求めた新たな周波数軸変換係数列を平滑化バッファに保持させる変換係数平滑化手段とを備えることを特徴とする音声認識装置。

【請求項 9】前記周波数軸変換手段は、

前記平滑化バッファに新たな周波数軸変換係数列を保持させた後は、前記新たな周波数軸変換係数列を用いて、前記入力音声の次に入力される入力音声の特徴量の周波数軸を変換することを特徴とする請求項 8 に記載の音声認識装置。

【請求項 10】前記変換係数平滑化手段は、

前記平滑化バッファに保持された過去に求められた周波数軸変換係数の平均値と、前記最

適な変換係数推定ステップで推定された最適な周波数軸変換係数との間の距離を求め、当該距離が所定以上大きい場合に、話者の交代を検知することを特徴とする請求項 8 又は 9 に記載の音声認識装置。

【請求項 1 1】さらに、前記入力音声から無声音 / 有声音区間を弁別検出する音声区間検出手段を有し、

前記周波数軸変換手段において、前記入力音声のうち、前記音声区間検出手段で検出した有声音区間に対する特徴量の周波数軸を、平滑化バッファに保持されている少なくとも 1 つの周波数軸変換係数から構成される周波数軸変換係数列を用いて変換することを特徴とする請求項 8 から 1 0 のいずれかに 1 つに記載の音声認識装置。

【請求項 1 2】前記最適変換係数推定手段は、前記認識結果候補を求めた際に使用した周波数軸変換係数を基に、複数の周波数軸変換係数列候補を設定し、

前記複数の周波数軸変換係数候補毎に、前記マッチング手段の認識結果候補のうち少なくとも 1 つから表現される音素系列に対して、周波数軸変換を施して得られた入力音声特徴量と予め複数の話者から学習した音響モデル特徴量との間の尤度を、マッチングにより求め、

さらに前記周波数軸変換係数候補と、前記求めた尤度との関係を二次曲線に近似し、当該二次曲線のピークにおける新たな周波数軸変換係数候補を求め、当該新たな周波数軸変換係数候補により周波数軸変換を施して得られた入力音声特徴量と予め複数の話者から学習した音響モデル特徴量との間の尤度を、マッチングにより求め、

前記周波数軸変換係数候補と前記新たな周波数軸変換係数候補との中からの最大尤度を与える周波数軸変換係数候補を、最適な周波数軸変換として選択することを特徴とする請求項 8 から 1 1 のいずれか 1 つに記載の音声認識装置。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0 0 3 0

【補正方法】変更

【補正の内容】

【0 0 3 0】

【発明の実施の形態】

本発明の第 1 態様は、入力音声の特徴量を抽出する特徴量抽出ステップと、前記入力音声の特徴量の周波数軸を、平滑化バッファに保持されている少なくとも 1 つの周波数軸変換係数から構成され、かつ、過去に求められた周波数軸変換係数列を用いて変換する周波数軸変換ステップと、前記周波数軸変換を施した入力音声の特徴量と予め複数の話者から学習した音響モデル特徴量とをマッチングし、認識結果候補を出力するマッチングステップと、前記認識結果候補のうち少なくとも 1 つから表現される音素系列に対して最適な周波数軸変換係数を推定する最適変換係数推定ステップと、前記求められた最適な周波数軸変換係数と、前記平滑化バッファに保持された過去に求められた周波数軸変換係数列とを平滑化し、新たな周波数軸変換係数列を求め、当該新たな周波数軸変換係数列を平滑化バッファに保持させる変換係数平滑化ステップとを有するもので、発声者の音声特徴量から周波数軸上に現れる個人性を吸収することにより、認識率の向上させる作用を有する。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】0 0 3 1

【補正方法】変更

【補正の内容】

【0 0 3 1】

本発明の第 2 態様は、前記変換係数平滑化ステップは、前記平滑化バッファに保持された過去に求められた周波数軸変換係数の平均値と、前記最適な変換係数推定ステップで推定された最適な周波数軸変換係数との間の距離を求め、当該距離が所定以上大きい場合に

、話者の交代を検知することを特徴とするもので、話者交代を検出した際には交代前話者の推定周波数変換係数の影響を受けないように周波数軸変換係数を初期化するなどによって、交代後話者への最適な周波数軸変換係数を新たに求め、話者間の周波数軸変換係数の差異による、認識率の低下を防ぐ作用を有する。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0 0 3 2

【補正方法】変更

【補正の内容】

【0 0 3 2】

本発明の第 3 態様は、さらに、前記入力音声から無声音 / 有声音区間を弁別検出する音声区間検出ステップを有し、前記周波数軸変換ステップにおいて、入力音声のうち、前記音声区間検出ステップで検出した有声音区間に対する特徴量の周波数軸を、平滑化バッファに保持されている少なくとも 1 つの周波数軸変換係数から構成される周波数軸変換係数列を用いて変換することを特徴とするもので、発声内容によって変動する推定変換係数のばらつきを抑えることにより、周波数軸変換を用いた話者正規化のより高い効果を与える作用を有する。

【手続補正 5】

【補正対象書類名】明細書

【補正対象項目名】0 0 3 3

【補正方法】変更

【補正の内容】

【0 0 3 3】

本発明の第 4 態様は、前記最適変換係数推定ステップは、前記認識結果候補を求めた際に使用した周波数軸変換係数を基に、複数の周波数軸変換係数候補を設定し、前記複数の周波数軸変換係数候補毎に、前記マッチングステップの認識結果候補のうち少なくとも 1 つから表現される音素系列に対して、周波数軸変換を施して得られた入力音声特徴量と予め複数の話者から学習した音響モデル特徴量との間の尤度を、マッチングにより求め、さらに前記周波数軸変換係数候補と、前記求めた尤度との関係を二次曲線に近似し、当該二次曲線のピークにおける新たな周波数軸変換係数候補を求め、当該新たな周波数軸変換係数候補により周波数軸変換を施して得られた入力音声特徴量と予め複数の話者から学習した音響モデル特徴量との間の尤度を、マッチングにより求め、前記周波数軸変換係数候補に対応する尤度と、前記新たな周波数軸変換係数候補に対応する尤度との中から、前記尤度が最も高い候補を、最適な周波数軸変換として選択することを特徴とするもので、事前に発声を行い話者の個人性を学習する適応フェイズなどを設けず、認識時の発声そのものから学習を行う「教師なし」話者正規化を実現する作用を有する。

【手続補正 6】

【補正対象書類名】明細書

【補正対象項目名】0 0 3 4

【補正方法】変更

【補正の内容】

【0 0 3 4】

本発明の第 5 態様は、前記変換係数平滑化ステップは、前記平滑化バッファに保持された過去に求められた周波数軸変換係数の平均値と、前記最適な変換係数推定ステップで推定された最適な周波数軸変換係数との間の距離を求め、当該距離が所定の距離を満たす場合にのみ、前記平滑化バッファに保持させることを特徴とするもので、発声内容によって変動する推定変換係数のばらつきを抑えることにより、周波数軸変換を用いた話者正規化のより高い効果を与える作用を有する。

【手続補正 7】

【補正対象書類名】明細書

【補正対象項目名】 0 0 3 5

【補正方法】 変更

【補正の内容】

【 0 0 3 5 】

本発明の第 6 態様は、前記変換係数平滑化ステップは、前記平滑化バッファに保持された過去に求められた周波数軸変換係数の平均値と、前記最適な変換係数推定ステップで推定された最適な周波数軸変換係数との間の距離を求め、当該距離が所定以上大きい場合に、前記平滑化バッファに保持された過去の周波数軸変換係数を任意の初期値を与えることによって初期化を行うことを特徴とするもので、発声内容によって変動する推定変換係数のばらつきを抑えることにより、周波数軸変換を用いた話者正規化のより高い効果を与える作用を有する。

【手続補正 8】

【補正対象書類名】 明細書

【補正対象項目名】 0 0 3 6

【補正方法】 変更

【補正の内容】

【 0 0 3 6 】

本発明の第 7 態様は、入力音声の特徴量を抽出する特徴量抽出手段と、少なくとも 1 つの周波数軸変換係数から構成され、かつ、過去に求められた周波数軸変換係数列が保持されている平滑化バッファと、前記入力音声の特徴量の周波数軸を、平滑化バッファに保持されている周波数軸変換係数列を用いて変換する周波数軸変換手段と、前記周波数軸変換を施した入力音声の特徴量と、予め複数の話者から学習した音響モデル特徴量とをマッチングし、認識結果候補を出力するマッチング手段と、前記認識結果候補のうち、少なくとも 1 つから表現される音素系列に対して、少なくとも 1 つの周波数軸変換係数から構成される最適な周波数軸変換係数を推定する最適変換係数推定手段と、前記求められた最適な周波数軸変換係数と、前記平滑化バッファに保持された過去に求められた周波数軸変換係数列とを平滑化して、新たな周波数軸変換係数列を求め、当該求めた新たな周波数軸変換係数列を平滑化バッファに保持させる変換係数平滑化手段とを有することを特徴とするもので、話者の音声特徴量から周波数軸上に現れる個人性を吸収する話者正規化によって、認識率の向上させる作用を有する。

【手続補正 9】

【補正対象書類名】 明細書

【補正対象項目名】 0 0 3 7

【補正方法】 変更

【補正の内容】

【 0 0 3 7 】

本発明の第 8 態様は、前記変換係数平滑化手段は、前記平滑化バッファに保持された過去に求められた周波数軸変換係数の平均値と、前記最適な変換係数推定ステップで推定された最適な周波数軸変換係数との間の距離を求め、当該距離が所定以上大きい場合に、話者の交代を検知することを特徴とするもので、話者交代を検出した際には交代前話者の推定周波数変換係数の影響を受けないように周波数軸変換係数を初期化するなどによって、交代後話者への最適な周波数軸変換係数を新たに求め、話者間の周波数軸変換係数の差異による、認識率の低下を防ぐ作用を有する。

【手続補正 10】

【補正対象書類名】 明細書

【補正対象項目名】 0 0 3 8

【補正方法】 変更

【補正の内容】

【 0 0 3 8 】

本発明の第 9 態様は、さらに、前記入力音声から無声音 / 有声音区間を弁別検出する音

声区間検出手段を有し、前記周波数軸変換手段において、前記入力音声のうち、前記音声区間検出手段で検出した有声音区間に対する特徴量の周波数軸を、平滑化バッファに保持されている少なくとも1つの周波数軸変換係数から構成される周波数軸変換係数列を用いて変換することを特徴とするもので、発声内容によって変動する推定変換係数のばらつきを抑えることにより、周波数軸変換を用いた話者正規化のより高い効果を与える作用を有する。