(19)日本国特許庁(JP)

再 公 表 特 許(A1)

(11)国際公開番号
WO2018/159612

発行日　令和2年1月9日(2020.1.9)

(43)国際公開日　平成30年9月7日(2018.9.7)

(51)Int.Cl.	F I	テーマコード（参考）
G10L 21/007 (2013.01)	G10L 21/007	

審査請求 未請求　予備審査請求 有　（全 28 頁）

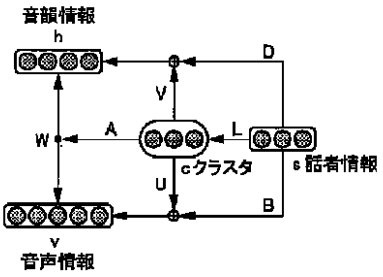
出願番号	特願2019-503021(P2019-503021)
(21)国際出願番号	PCT/JP2018/007268
(22)国際出願日	平成30年2月27日(2018.2.27)
(31)優先権主張番号	特願2017-36109(P2017-36109)
(32)優先日	平成29年2月28日(2017.2.28)
(33)優先権主張国・地域又は機関	日本国(JP)

(71)出願人　504133110
国立大学法人電気通信大学
東京都調布市調布ケ丘一丁目５番地１
(74)代理人　110000925
特許業務法人信友国際特許事務所
(72)発明者　中鹿　亘
東京都調布市調布ケ丘一丁目５番地１　国立大学法人電気通信大学内

最終頁に続く

(54)【発明の名称】声質変換装置、声質変換方法およびプログラム

(57)【要約】
パラメータ学習ユニットとパラメータ記憶ユニットと声質変換処理ユニットとを備える。パラメータ学習ユニットは、入力データを表現する可視素子と、潜在的な情報を表現した隠れ素子との間に結合重みが存在すると仮定した制限ボルツマンマシンによる確率モデルを用意する。その確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義し、それぞれの話者について、複数個の話者クラスタへの重みを推定して、パラメータを決定する。パラメータ記憶ユニットは、パラメータを記憶する。声質変換処理ユニットは、パラメータ記憶ユニットが記憶したパラメータと目標話者の話者情報とに基づいて、入力話者の音声に基づく音声情報の声質変換処理を行う。



h Phonological information
v Voice information
c Cluster
s Speaker information

【特許請求の範囲】
【請求項１】
入力話者の音声を目標話者の音声に声質変換する声質変換装置であって、
学習用の音声に基づく音声情報およびその音声情報に対応する話者情報から、声質変換のためのパラメータを決定するパラメータ学習ユニットと、
前記パラメータ学習ユニットが決定したパラメータを記憶するパラメータ記憶ユニットと、
前記パラメータ記憶ユニットが記憶したパラメータと前記目標話者の話者情報とに基づいて、前記入力話者の音声に基づく前記音声情報の声質変換処理を行う声質変換処理ユニットとを備え、
前記パラメータ学習ユニットは、音声に基づく音声情報、音声情報に対応する話者情報および音声中の音韻を表す音韻情報のそれぞれを変数とすることで、前記音声情報、前記話者情報および前記音韻情報のそれぞれの間の結合エネルギーの関係性を前記パラメータによって表す確率モデルを取得し、前記確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義するようにした
声質変換装置。
【請求項２】
さらに、前記パラメータ記憶ユニットが記憶したパラメータを前記入力話者の音声に適応して、適応後のパラメータを得る適応ユニットを備え、
前記パラメータ記憶ユニットは、前記適応ユニットで適応後のパラメータを記憶し、前記声質変換処理ユニットは、適応後のパラメータと前記目標話者の話者情報とに基づいて、前記入力話者の音声に基づく前記音声情報の声質変換処理を行う
請求項１に記載の声質変換装置。
【請求項３】
前記パラメータ学習ユニットと前記適応ユニットは共通の演算処理部で構成され、
前記学習用の音声に基づいてパラメータを決定する処理と、前記入力話者の音声に基づいて適応後のパラメータを得る処理を、前記共通の演算処理部で行うようにした
請求項２に記載の声質変換装置。
【請求項４】
前記パラメータ学習ユニットが学習する際には、複数のクラスタが最も離れるように学習し、学習した複数のクラスタの間で、話者クラスタへの重みの位置を設定する
請求項１に記載の声質変換装置。
【請求項５】
前記声質変換処理ユニットは、前記パラメータから前記目標話者の話者情報を得、得られた話者情報から前記目標話者の音声情報を得るようにした
請求項１に記載の声質変換装置。
【請求項６】
音声情報の特徴量$v=[v_1, \cdots, v_I] \in R^I$と、音韻情報の特徴量$h=[h_1, \cdots, h_J] \in \{0, 1\}^J$, $\Sigma_j h_j = 1$との間に、話者情報の特徴量$s=[s_1, \cdots, s_R] \in \{0, 1\}^R$, $\Sigma_r s_r = 1$に依存した双方な結合重み$W \in R^{I \times J}$が存在すると仮定したとき、前記話者クラスタとして、話者クラスタ$c \in R^K$を導入し、話者クラスタcを、

$$c \triangleq Ls'$$

（但し、$L \in R^{K \times R} = [\lambda_1 \cdots \lambda_R]$の各列ベクトル$\lambda_r$は、それぞれの話者クラスタへの重みを表す非負パラメータであり、$||\lambda_r||_1 = 1$, $\forall_r$の制約を課す）と表現し、音響情報の特徴量のクラスタ依存項のバイアスパラメータを$U \in R^{I \times K}$、音韻情報の特徴量のクラスタ依存項のバイアスパラメータを$V \in R^{J \times K}$、として、話者非依存項、クラスタ依存項、および話者依存項のそれぞれを、

$$\tilde{\mathbf{W}} \triangleq \mathcal{A} \circ_3^1 c\mathbf{W}$$

$$\tilde{b} \triangleq b + \mathbf{U}c + \mathbf{B}s$$

$$\tilde{d} \triangleq d + \mathbf{V}c + \mathbf{D}s$$

として示す

請求項１に記載の声質変換装置。

【請求項７】

入力話者の音声を目標話者の音声に声質変換する声質変換方法であって、

音声に基づく音声情報、音声情報に対応する話者情報および音声中の音韻を表す音韻情報のそれぞれを変数とすることで、前記音声情報、前記話者情報および前記音韻情報のそれぞれの間の結合エネルギーの関係性をパラメータによって表す確率モデルを用意し、その確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義し、それぞれの話者について、前記複数個の話者クラスタへの重みを推定して、学習用の音声についての前記パラメータを決定するパラメータ学習ステップと、

前記パラメータ学習ステップで得られたパラメータ、又は当該パラメータを前記入力話者の音声に適応した適応後のパラメータと、前記目標話者の前記話者情報とに基づいて、前記入力話者の音声に基づく前記音声情報の声質変換処理を行う声質変換処理ステップと

を含む、声質変換方法。

【請求項８】

音声に基づく音声情報、音声情報に対応する話者情報および音声中の音韻を表す音韻情報のそれぞれを変数とすることで、前記音声情報、前記話者情報および前記音韻情報のそれぞれの間の結合エネルギーの関係性をパラメータによって表す確率モデルを用意し、その確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義し、それぞれの話者について、前記複数個の話者クラスタへの重みを推定して、学習用の音声についての前記パラメータを決定して記憶するパラメータ学習ステップと、

前記パラメータ学習ステップで得られたパラメータ、又は当該パラメータを入力話者の音声に適応した適応後のパラメータと、目標話者の前記話者情報とに基づいて、前記入力話者の音声に基づく前記音声情報の声質変換処理を行う声質変換処理ステップと、

をコンピュータに実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は任意話者声質変換を可能とする声質変換装置、声質変換方法およびプログラムに関する。

【背景技術】

【０００２】

従来、入力話者音声の音韻情報を保存したまま、話者性に関する情報のみを出力話者のものへ変換させる技術である声質変換の分野では、モデルの学習時において、入力話者と出力話者の同一発話内容による音声対であるパラレルデータを使用するパラレル声質変換が主流であった。

パラレル声質変換としては、ＧＭＭ（Gaussian Mixture Model）に基づく手法、ＮＭＦ（Non-negative Matrix Factrization）に基づく手法、ＤＮＮ（Deep Neural Network）に基づく手法など、様々な統計的アプローチが提案されている（特許文献１参照）。パラレル声質変換では、パラレル制約のおかげで比較的高い精度が得られる反面、学習データとしては入力話者と出力話者の発話内容を一致させる必要があるため、利便性が損なわれてしまうという問題があった。

【０００３】

これに対して、モデルの学習時に上述のパラレルデータを使用しない非パラレル声質変

換が注目を浴びている。非パラレル声質変換は、パラレル声質変換に比べて精度面で劣るものの自由発話を用いて学習を行うことができるため利便性や実用性は高い。非特許文献１には、入力話者の音声と出力話者の音声を用いて事前に個々のパラメータを学習しておくことで、学習データに含まれる話者を入力話者または目標話者とする声質変換を可能とする技術が記載されている。

【先行技術文献】

【特許文献】

【０００４】

【特許文献１】特開２００８－５８６９６号公報

【非特許文献】

【０００５】

【非特許文献１】T. Nakashika, T. Takiguchi, and Y. Ariki: "Parallel-Data-Free, Many-To-Many Voice Conversion Using an Adaptive Restricted Boltzmann Machine," Proceedings of Machine Learning in Spoken Language Processing (MLSLP) 2015, 6 pages, 2015.

【発明の概要】

【発明が解決しようとする課題】

【０００６】

非特許文献１に記載の技術は、統計的な非パラレル声質変換アプローチとして、制限ボルツマンマシン（Restricted Boltzmann Machine：以下ＲＢＭと称する）を適用した、適応型ＲＢＭ（ＡＲＢＭ）に基づく声質変換に基づく声質変換を行う。このアプローチでは、複数の話者による音声データから自動的にそれぞれの話者固有の適応行列と、音響特徴量（メルケプストラム）から話者に依存しない潜在特徴（以下、これらを潜在的な音韻または単に音韻と呼ぶ）への射影行列を同時に推定する。これにより、入力話者の音声および入力話者の適応行列から計算した潜在的な音韻と、目標話者の適応行列を用いて音響特徴量を計算することで目標話者に近い音声を得るようにしている。

【０００７】

一度学習によって潜在的な音韻を得るための射影行列が推定されれば、新たな入力話者・目標話者に対してそれぞれの適応行列のみを推定（このステップを適応と呼ぶ）することで変換が可能となる。しかし、話者固有の適応行列は音響特徴量の二乗個のパラメータを含むため、音響特徴量の次元数や話者数が増えるほどパラメータ数が膨大となり、学習コストが掛かってしまう。そして、適応時に必要となるデータ数が多くなり、事前に学習していない話者のその場での変換が困難となってしまうといった問題が発生する。また、声質変換を利用する場面では、その場で音声を収録し、即座に変換を行いたいケースが考えられるが、従来の技術では、即座に変換することは困難であった。

【０００８】

本発明はかかる点に鑑み、各話者の発話について少ないデータ数で簡単に声質変換が可能な声質変換装置、声質変換方法およびプログラムを提供することを目的とする。

【課題を解決するための手段】

【０００９】

上記課題を解決するため、本発明の声質変換装置は、入力話者の音声を目標話者の音声に声質変換する声質変換装置であって、パラメータ学習ユニットとパラメータ記憶ユニットと声質変換処理ユニットとを備える。

パラメータ学習ユニットは、学習用の音声に基づく音声情報およびその音声情報に対応する話者情報から、声質変換のためのパラメータを決定する。

パラメータ記憶ユニットは、パラメータ学習ユニットが決定したパラメータを記憶する。

声質変換処理ユニットは、パラメータ記憶ユニットが記憶したパラメータと目標話者の話者情報とに基づいて、入力話者の音声に基づく音声情報の声質変換処理を行う。

ここで、パラメータ学習ユニットは、音声に基づく音声情報、音声情報に対応する話者

情報および音声中の音韻を表す音韻情報のそれぞれを変数とすることで、音声情報、話者情報および音韻情報のそれぞれの間の結合エネルギーの関係性をパラメータによって表す確率モデルを取得し、確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義するようにした。

【００１０】

また、本発明の声質変換方法は、入力話者の音声を目標話者の音声に声質変換する方法であって、パラメータ学習ステップと声質変換処理ステップとを含む。

パラメータ学習ステップは、音声に基づく音声情報、音声情報に対応する話者情報および音声中の音韻を表す音韻情報のそれぞれを変数とすることで、音声情報、話者情報および音韻情報のそれぞれの間の結合エネルギーの関係性をパラメータによって表す確率モデルを用意する。そして、その確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義し、それぞれの話者について、複数個の話者クラスタへの重みを推定して、学習用の音声についてのパラメータを決定する。

声質変換処理ステップは、パラメータ学習ステップで得られたパラメータ、又は当該パラメータを入力話者の音声に適応した適応後のパラメータと、目標話者の話者情報とに基づいて、入力話者の音声に基づく音声情報の声質変換処理を行う。

【００１１】

また本発明のプログラムは、上述した声質変換方法のパラメータ学習ステップと声質変換処理ステップとをコンピュータに実行させるものである。

【００１２】

本発明によれば、話者クラスタにより目標話者を設定することができるため、従来よりも非常に少ないデータ数で、入力話者音声を目標話者音声に声質変換できるようになる。

【図面の簡単な説明】

【００１３】

【図１】本発明の一実施の形態例に係る声質変換装置の構成例（例１）を示すブロック図である。

【図２】本発明の一実施の形態例に係る声質変換装置の構成例（例２）を示すブロック図である。

【図３】声質変換装置のハードウェア構成例を示すブロック図である。

【図４】従来の確率モデルを模式的に示す説明図である。

【図５】声質変換装置のパラメータ推定部が備える確率モデルを模式的に示す説明図である。

【図６】本発明の一実施の形態例に係る処理全体の流れを示すフローチャートである。

【図７】図６のステップＳ３の学習の詳細例を示すフローチャートである。

【図８】図６のステップＳ４の適応の詳細例を示すフローチャートである。

【図９】図６のステップＳ８の声質変換の詳細例を示すフローチャートである。

【図１０】本発明の一実施形態によるクラスタの重み分布の例を示す説明図である。

【図１１】声質変換装置のパラメータ推定部が備える確率モデルの別の例を示す説明図である。

【発明を実施するための形態】

【００１４】

以下、本発明の好適な一実施形態例について説明する。

【００１５】

［１．構成］

図１は、本発明の一実施形態例にかかる声質変換装置の構成例（例１）を示す図である。図１においてＰＣ等により構成される声質変換装置１は、事前に、学習用音声信号と学習用音声信号に対応する話者の情報（対応話者情報）に基づいて学習を行っておくことで、任意の話者による変換用音声信号（適応話者音声信号）を、目標話者の声質に変換し、変換済み音声信号として出力する。

学習用音声信号は、予め記録された音声データに基づく音声信号でもよく、また、マイ

クロフォン等により話者が話す音声（音波）を直接電気信号に変換したものでもよい。また、対応話者情報は、ある学習用音声信号と他の学習用音声信号とが同じ話者による音声信号か異なる話者による音声信号かを区別できるものであればよい。

【００１６】

声質変換装置１は、パラメータ学習ユニット１１と声質変換処理ユニット１２とパラメータ記憶ユニット１３とを備える。パラメータ学習ユニット１１は、学習用音声信号と対応話者情報とに基づいた学習処理により声質変換のためのパラメータを決定する。パラメータ学習ユニット１１が決定したパラメータは、パラメータ記憶ユニット１３に記憶される。パラメータ記憶ユニット１３に記憶されたパラメータは、適応処理によって、パラメータ学習ユニット１１が入力話者の適応後のパラメータに変換する。声質変換処理ユニット１２は、上述の学習処理および適応処理によりパラメータが決定された後、決定されたパラメータと目標とする話者の情報（目標話者情報）とに基づいて変換用音声信号の声質を目標話者の声質に変換し、変換済み音声信号として出力する。なお、パラメータ学習ユニット１１が学習処理と適応処理の双方を行うのは一例であり、後述する図２に示すように、パラメータ学習ユニット１１と別に適応ユニット１４を備えるようにしてもよい。

【００１７】

パラメータ学習ユニット１１は、音声信号取得部１１１と前処理部１１２と話者情報取得部１１３とパラメータ推定部１１４を備える。音声信号取得部１１１は、前処理部１１２に接続され、前処理部１１２および話者情報取得部１１３は、それぞれパラメータ推定部１１４に接続される。

【００１８】

音声信号取得部１１１は、接続された外部機器から学習用音声信号を取得するものであり、例えば、マウスやキーボード等の図示しない入力部からのユーザの操作に基づいて学習用音声信号が取得される。また、音声信号取得部１１１は、接続される不図示のマイクロフォンから、話者の発話をリアルタイムに取り込むようにしてもよい。なお、以下の説明では、パラメータ学習ユニット１１が学習用音声信号を取得してパラメータを得る処理を述べるが、パラメータ学習ユニット１１が適応話者音声信号に適応したパラメータを得る適応処理時にも、各処理部は同様の処理が行われる。適応処理の詳細については後述するが、適応処理時には、学習処理でパラメータ記憶ユニット１３に記憶されたパラメータを、適応話者音声信号に適応したパラメータとする適応化処理が行われる。

前処理部１１２は、音声信号取得部１１１で取得された学習用音声信号を単位時間ごと（以下、フレームという）に切り出し、ＭＦＣＣ（Mel-Frequency Cepstrum Coefficients：メル周波数ケプストラム係数）やメルケプストラム特徴量などのフレームごとの音声信号のスペクトル特徴量を計算した後、正規化を行うことで学習用音声情報を生成する。

【００１９】

対応話者情報取得部１１３は、音声信号取得部１１１による学習用音声信号の取得に紐付けられた対応話者情報を取得する。対応話者情報は、ある学習用音声信号の話者と他の学習用音声信号の話者とを区別できるものであればよく、例えば、図示しない入力部からのユーザの入力によって取得される。また、複数の学習用音声信号のそれぞれについて互いに話者が異なることが明らかであれば、学習用音声信号の取得に際して対応話者情報取得部１１３が自動で対応話者情報を付与してもよい。例えば、パラメータ学習ユニット１１が１０人の話し声の学習を行うと仮定すると、対応話者情報取得部１１３は、音声信号取得部１１１に入力中の学習用音声信号が１０人の内のどの話者の話し声の音声信号であるかを区別する情報（対応話者情報）を、自動的にまたはユーザからの入力により取得する。なお、ここで話し声の学習を行う人数を１０人としたのは、あくまでも一例である。パラメータ学習ユニット１１は、最低でも２人の音声が入力されれば学習が可能であるが、人数が多い方がより精度の高い学習ができることになる。

【００２０】

パラメータ推定部１１４は、音声情報推定部１１４１と話者情報推定部１１４２と音韻情報推定部１１４３とによって構成されるＲＢＭ（制限ボルツマンマシン）を適用した、

適応型ＲＢＭ（ＡＲＢＭ）の確率モデルを持ち、学習用音声信号に基づいてパラメータの推定を行う。パラメータ推定部１１４が学習処理によって推定したパラメータは、パラメータ記憶ユニット１３に記憶される。この学習処理で得たパラメータは、適応話者の音声信号がパラメータ学習ユニット１１に入力されたとき、パラメータ記憶ユニット１３からパラメータ学習ユニット１１に読み出され、そのときの適応話者の音声信号に適応したパラメータとされる。

【００２１】

パラメータ推定部１１４がパラメータを推定する際に適用される本実施形態例の確率モデルでは、各推定部１１４１，１１４２，１１４３が持つ音声情報、話者情報、および音韻情報の他に、話者の特徴から得た複数の話者クラスタの情報を持つ。すなわち、パラメータ推定部１１４は、この話者クラスタを計算する話者クラスタ計算部１１４４を有する。さらに、本実施形態例の確率モデルでは、各情報のそれぞれの間の結合エネルギーの関係性を表すパラメータを持つ。なお、以下の説明では、本実施形態例の確率モデルを、話者クラスタ適応型ＲＢＭと称する。話者クラスタ適応型ＲＢＭの詳細については後述する。

【００２２】

音声情報推定部１１４１は、音韻情報および話者情報ならびに各種パラメータを用いて音声情報を取得する。ここで、音声情報とは、それぞれの話者の音声信号の音響ベクトル（スペクトル特徴量やケプストラム特徴量など）を意味する。

【００２３】

話者情報推定部１１４２は、音声情報および音韻情報ならびに各種パラメータを用いて話者情報を推定する。ここで、話者情報とは、話者を特定するための情報であり、それぞれの話者の音声が持つ音響ベクトル情報である。すなわち、この話者情報（話者ベクトル）は、同じ話者の音声信号に対しては全て共通であり、異なる話者の音声信号に対しては互いに異なるような、音声信号の発話者を特定させるベクトルを意味している。

【００２４】

音韻情報推定部１１４３は、音声情報および話者情報ならびに各種パラメータにより音韻情報を推定する。ここで音韻情報とは、音声情報に含まれる情報の中から、学習を行う全ての話者に共通となる情報である。例えば、入力した学習用音声信号が、「こんにちは」と発話した音声の信号であるとき、この音声信号から得られる音韻情報は、その「こんにちは」と発話した言葉の情報に相当する。但し、本実施の形態例での音韻情報は、言葉に相当する情報であっても、いわゆるテキストの情報ではなく、言語の種類に限定されない音韻の情報であり、どのような言語で話者が話した場合にも共通となる、音声信号の中で潜在的に含まれる、話者情報以外の情報を表すベクトルである。

【００２５】

話者クラスタ計算部１１４４は、入力中の学習用音声信号から得た話者情報に対応したクラスタを計算する。すなわち、パラメータ推定部１１４が備える話者クラスタ適応型ＲＢＭは、話者情報を示すクラスタを複数持ち、話者クラスタ計算部１１４４は、入力中の学習用音声信号から得た話者情報に対応するクラスタを計算する。

【００２６】

また、パラメータ推定部１１４が備える話者クラスタ適応型ＲＢＭは、音声情報、話者情報、音韻情報および話者クラスタの情報を持つだけでなく、各情報のそれぞれの間の結合エネルギーの関係性をパラメータによって表すようにしている。

【００２７】

声質変換処理ユニット１２は、音声信号取得部１２１と前処理部１２２と話者情報設定部１２３と声質変換部１２４と後処理部１２５と音声信号出力部１２６とを備える。音声信号入力１２１、前処理部１２２、声質変換部１２４、後処理部１２５および音声信号出力部１２６は順次接続され、声質変換部１２４には、更にパラメータ学習ユニット１１のパラメータ推定部１１４が接続される。

【００２８】

音声信号取得部１２１は、変換用音声信号を取得し、前処理部１２２は、変換用音声信号に基づき変換用音声情報を生成する。本実施の形態例では、音声信号取得部１２１が取得する変換用音声信号は、任意の話者による変換用音声信号でよい。

音声信号取得部１２１および前処理部１２２は、上述したパラメータ学習ユニット１１の音声信号取得部１１１および前処理部１１２の構成と同じであり、別途設置することなくこれらを兼用してもよい。

【００２９】

話者情報設定部１２３は、声質変換先である目標話者を設定し目標話者情報を出力する。話者情報設定部１２３で設定される目標話者は、ここでは、パラメータ学習ユニット１１のパラメータ推定部１１４が事前に学習処理して話者情報を取得した話者の中から選ばれる。話者情報設定部１２３は、例えば、図示しないディスプレイ等に表示された複数の目標話者の選択肢（パラメータ推定部１１４が事前に学習処理した話者の一覧など）からユーザが図示しない入力部によって１つの目標話者を選択するものであってもよく、また、その際に、図示しないスピーカにより目標話者の音声を確認できるようにしてもよい。

【００３０】

声質変換部１２４は、目標話者情報に基づいて変換用音声情報に声質変換を施し、変換済み音声情報を出力する。声質変換部１２４は、音声情報設定部１２４１、話者情報設定部１２４２、音韻情報設定部１２４３、および話者クラスタ計算部１２４４を持つ。この音声情報設定部１２４１、話者情報設定部１２４２、音韻情報設定部１２４３、および話者クラスタ計算部１２４４は、上述のパラメータ推定部１１４において、話者クラスタ適応型ＲＢＭの確率モデルが持つ音声情報推定部１１４１、話者情報推定部１１４２、音韻情報推定部１１４３、および話者クラスタ計算部１１４４と同等の機能を持つ。

【００３１】

すなわち、音声情報設定部１２４１、話者情報設定部１２４２および音韻情報設定部１２４３には、それぞれ音声情報、話者情報および音韻情報が設定されるが、音韻情報設定部１２４３に設定される音韻情報は、前処理部１２２から供給される音声情報に基づいて得た情報である。一方、話者情報設定部１２４２に設定される話者情報は、パラメータ学習ユニット１１内の話者情報推定部１１４２での推定結果から取得した目標話者についての話者情報（話者ベクトル）である。音声情報設定部１２４１に設定される音声情報は、これら話者情報設定部１２４２および音韻情報設定部１２４３に設定された話者情報および音韻情報と各種パラメータとから得られる。話者クラスタ計算部１２４４は、目標話者の話者クラスタ情報を計算する。

なお、図１では声質変換部１２４を設ける構成を示したが、声質変換部１２４を別途設置することなく、パラメータ推定部１１４の各種パラメータを固定することで、パラメータ推定部１１４が声質変換の処理を実行する構成としてもよい。

【００３２】

後処理部１２５は、声質変換部１２４で得られた変換済み音声情報に逆正規化処理を施し、更に逆ＦＦＴ処理することでスペクトル情報をフレームごとの音声信号へ戻した後に結合し、変換済み音声信号を生成する。

音声信号出力部１２６は、接続される外部機器に対して変換済み音声信号を出力する。接続される外部機器としては、例えば、スピーカなどが挙げられる。

【００３３】

図２は、本発明の一実施形態例にかかる声質変換装置の別の構成例（例２）を示す図である。

図２に示す声質変換装置１は、適応話者音声信号によりパラメータの適応処理を行う適応ユニット１４を備える点が、図１に示す声質変換装置１と異なる。すなわち、図１に示す声質変換装置１では、パラメータ学習ユニット１１が、学習処理と適応処理の双方を行うようにしたのに対して、図２に示す声質変換装置１では、適応ユニット１４が適応処理を行うようにした点が異なる。

【００３４】

適応ユニット１４は、音声信号取得部１４１と前処理部１４２と適応話者情報取得部１４３とパラメータ推定部１４４を備える。音声信号取得部１４１は、適応話者音声信号を取得し、取得した音声信号を前処理部１４２に出力する。前処理部１４２は、音声信号の前処理を行って適応用音声情報を得、得られた適応用音声情報をパラメータ推定部１４４に供給する。適応話者情報取得部１４３は、適応話者についての話者情報を取得し、取得した適応話者情報をパラメータ推定部１４４に供給する。

パラメータ推定部１４４は、音声情報推定部１４４１と話者情報推定部１４４２と音韻情報推定部１４４３と話者クラスタ計算部１４４４を有し、音声情報、話者情報、音韻情報、および話者クラスタの情報を持つ。

【００３５】

適応ユニット１４で得られた適用後のパラメータは、パラメータ記憶ユニット１３に記憶した後、声質変換処理ユニット１２に供給される。あるいは、適応ユニット１４で得られた適用後のパラメータを、直接、声質変換処理ユニット１２に供給するようにしてもよい。

図２に示す声質変換装置１のその他の部分については、図１に示す声質変換装置１と同様に構成する。

【００３６】

図３は、声質変換装置１のハードウェア構成例を示す図である。ここでは、声質変換装置１をコンピュータ（ＰＣ）で構成した例を示す。

図３に示すように、声質変換装置１は、バス１０７を介して相互に接続されたＣＰＵ（中央制御ユニット：Central Processing Unit）１０１、ＲＯＭ（Read Only Memory）１０２、ＲＡＭ（Random Access Memory）１０３、ＨＤＤ（Hard Disk Drive）／ＳＳＤ（Solid State Drive）１０４、接続Ｉ／Ｆ（Interface）１０５、通信Ｉ／Ｆ１０６を備える。ＣＰＵ１０１は、ＲＡＭ１０３をワークエリアとしてＲＯＭ１０２またはＨＤＤ／ＳＳＤ１０４等に格納されたプログラムを実行することで、声質変換装置１の動作を統括的に制御する。接続Ｉ／Ｆ１０５は、声質変換装置１に接続される機器とのインターフェースである。通信Ｉ／Ｆは、ネットワークを介して他の情報処理機器と通信を行うためのインターフェースである。

【００３７】

学習用音声信号、変換用音声信号、および変換済み音声信号の入出力および設定は、接続Ｉ／Ｆ１０５または通信Ｉ／Ｆ１０６を介して行われる。パラメータ記憶ユニット１３でのパラメータの記憶は、ＲＡＭ１０３またはＨＤＤ／ＳＳＤ１０４により行われる。図１で説明した声質変換装置１の機能は、ＣＰＵ１０１において所定のプログラムが実行されることで実現される。プログラムは、記録媒体を経由して取得してもよく、ネットワークを経由して取得してもよく、ＲＯＭに組み込んで使用してもよい。また、一般的なコンピュータとプログラムの組合せでなく、ＡＳＩＣ（Application Specific Integrated Circuit）やＦＰＧＡ（Field Programmable Gate Array）などの論理回路を組むことで、声質変換装置１の構成を実現するためのハードウェア構成にしてもよい。

【００３８】

［２．話者クラスタ適応型ＲＢＭの定義］

次に、パラメータ推定部１１３および符号化部１２３が持つ確率モデルである、話者クラスタ適応型ＲＢＭについて説明する。

まず、本発明に適用される話者クラスタ適応型ＲＢＭを説明する前に、既に提案した確率モデルである、適応型ＲＢＭについて説明する。

図４は、適応型ＲＢＭのグラフ構造を模式的に示す図である。

適応型ＲＢＭの確率モデルは、音声情報ｖ、話者情報ｓおよび音韻情報ｈと、それぞれの情報の結合エネルギーの関係性を示すパラメータを持つ。ここでは、音響（メルケプストラム）情報の特徴量ｖ＝［v_1，・・・，v_I］$\in R^I$と、音韻情報の特徴量ｈ＝［h_1，・・・，h_J］$\in \{0, 1\}^J$，$\Sigma_j h_j = 1$との間に、話者特徴量ｓ＝［ｓ１，・・・，ｓＲ］$\in \{0, 1\}^R$，$\Sigma_r s_r = 1$に依存した双方向な結合重みＷ$\in R^{I \times J}$が

存在すると仮定したとき、適応型ＲＢＭの確率モデルは、次の［数１］式～［数３］式で示される条件付き確率密度関数で示される。

【００３９】

【数１】

$$p(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{s})=\frac{1}{Z}e^{-E(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{s})}$$

【００４０】

【数２】

$$E(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{s})=\frac{1}{2}\left\|\frac{\boldsymbol{v}-\tilde{\boldsymbol{b}}}{\boldsymbol{\sigma}}\right\|_2^2-\tilde{\boldsymbol{d}}^{\top}\boldsymbol{h}-(\frac{\boldsymbol{v}}{\boldsymbol{\sigma}^2})^{\top}\tilde{\mathbf{W}}\boldsymbol{h}$$

【００４１】

【数３】

$$Z=\sum_{\boldsymbol{v},\boldsymbol{h}}e^{-E(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{s})}$$

【００４２】

但し、σ∈Ｒ^{I}は音響特徴量の偏差を表すパラメータであり、ｂ∈Ｒ^{I}およびｄ∈Ｒ^{J}はそれぞれ話者特徴量ｓに依存した音響特徴量、音韻特徴量のバイアスを表す。式の中の記号の上に付けられた「~」は、該当する情報が話者に依存した情報であることを示す。なお、明細書の中では、表記上の制約のため、「~」を記号の上に付与できないので、例えばW(~)のように、記号の後に括弧で示す。「^」などの、記号の上に付与して示す他の記号についても、同様に表記する。

また、［数２］式の右辺の括線および「・^{2}」は、それぞれ要素ごとの除算、要素ごとの二乗を表す。話者依存の項W(~),ｂ(~)，ｄ(~)は、話者非依存パラメータと話者依存パラメータを用いて、下記の［数４］式～［数６］式のように定義される。

【００４３】

【数４】

$$\tilde{\mathbf{W}}\triangleq\sum_{r}\mathbf{A}_r s_r\mathbf{W}=\mathcal{A}\circ_3^1\boldsymbol{s}\mathbf{W}$$

【００４４】

【数５】

$$\tilde{\boldsymbol{b}}\triangleq\boldsymbol{b}+\sum_{r}\boldsymbol{b}_r s_r=\boldsymbol{b}+\mathbf{B}\boldsymbol{s}$$

【００４５】

【数６】

$$\tilde{\boldsymbol{d}}\triangleq\boldsymbol{d}+\sum_{r}\boldsymbol{d}_r s_r=\boldsymbol{d}+\mathbf{D}\boldsymbol{s}$$

【００４６】

ここで、Ｗ∈Ｒ$^{I\times J}$，ｂ∈Ｒ^{I}，ｄ∈Ｒ^{J}は話者非依存パラメータを表し、Ａ$_{r}$∈Ｒ$^{I\times I}$（Ａ＝｛Ａ$_{r}$｝$_{r=1}^{R}$），ｂ$_{r}$∈Ｒ^{I}（Ｂ＝［ｂ$_{1}$，・・・，ｂ$_{R}$］），ｄ$_{r}$∈Ｒ^{J}（Ｄ＝［ｄ$_{1}$，・・・，ｄ$_{R}$］）は、話者ｒに依存したパラメータを表す。また、○$_{i}^{j}$は左テンソルのモードｉ、右テンソルのモードｊに沿った内積演算を表す。

【００４７】

ここでは、音響特徴量はクリーン音声のメルケプストラムとし、発話者の違いによるパ

ラメータ変動は、話者特徴量ｓによって規定される話者依存項（［数４］式，［数５］式，［数６］式）で吸収する。したがって、音韻特徴量は話者に依存しないいずれかの要素のみがアクティブとなる観測不可能な特徴量である、音韻の情報が含まれることになる。

【００４８】

このように、適応型ＲＢＭによって音響特徴量と音韻特徴量を得ることができるが、適応型ＲＢＭでは、話者依存パラメータの数は（Ｉ^{2}Ｒ）に比例し、音響特徴量の二乗（Ｉ^{2}）が比較的大きいため、話者数が増加するほど推定するパラメータ数が膨大となり、計算に要するコストが増加してしまう。また、ある話者ｒの適応時においても、推定すべきパラメータ数が（Ｉ^{2}＋Ｉ＋Ｊ）となり、過学習を避けるために相応に多くのデータを必要とする問題があった。

【００４９】

ここで、本発明では、これらの問題を解決するために、話者クラスタ適応型ＲＢＭを適用する。

図５は、話者クラスタ適応型ＲＢＭのグラフ構造を模式的に示す図である。

話者クラスタ適応型ＲＢＭの確率モデルは、音声情報ｖ、話者情報ｓおよび音韻情報ｈと、それぞれの情報の結合エネルギーの関係性を示すパラメータの他に、話者クラスタｃ∈Ｒ^{K}を持つ。話者クラスタｃは、次の［数７］式と恒等的に表現される。

【００５０】

【数７】

$$\boldsymbol{c} \triangleq \mathbf{L}\boldsymbol{s}$$

【００５１】

但し、Ｌ∈Ｒ$^{K\times R}$＝［λ^{1}・・・λ^{R}］の各列ベクトルλ$_{r}$は、それぞれの話者クラスタへの重みを表す非負パラメータであり、||λ$_{r}$||$_{1}$＝１，∀ｒの制約を課す。

先に説明した適応型ＲＢＭ（図４）では、話者ごとに適応行列を用意したが、本発明の話者クラスタ適応型ＲＢＭではクラスタごとに適応行列を用意する。また、音響特徴量、音韻特徴量のバイアスは、話者非依存項、クラスタ依存項、話者依存項の加算で表現される。すなわち、話者依存の項Ｗ(~)，ｂ(~)，ｄ(~)は、下記の［数８］式～［数１０］式のように定義される。

【００５２】

【数８】

$$\tilde{\mathbf{W}} \triangleq \mathcal{A} \circ_3^1 \boldsymbol{c}\mathbf{W}$$

【００５３】

【数９】

$$\tilde{\boldsymbol{b}} \triangleq \boldsymbol{b} + \mathbf{U}\boldsymbol{c} + \mathbf{B}\boldsymbol{s}$$

【００５４】

【数１０】

$$\tilde{\boldsymbol{d}} \triangleq \boldsymbol{d} + \mathbf{V}\boldsymbol{c} + \mathbf{D}\boldsymbol{s}$$

【００５５】

ここで、音響情報の特徴量のクラスタ依存項のバイアスパラメータをＵ∈Ｒ$^{I\times K}$、音韻情報の特徴量のクラスタ依存項のバイアスパラメータをＶ∈Ｒ$^{J\times K}$とする。

［数８］式で示されるＡ＝｛Ａｋ｝$_{k=1}^{K}$と、先に説明した適応型ＲＢＭでの［数４］式におけるＡを比較すると、適応型ＲＢＭでは（Ｉ^{2}Ｒ）個のパラメータが含まれていたのに対して、話者クラスタ適応型ＲＢＭでは（Ｉ^{2}Ｋ）個となり、大幅にパラメータ数を削減することができる。例えば、一例としては、Ｒ＝５８、Ｉ＝３２、Ｋ＝８に設定し

た場合、先に説明した適応型ＲＢＭではパラメータ数５９３９２個になるが、話者クラスタ適応型ＲＢＭでは８１９２個になり、大幅にパラメータ数を削減できる。

【００５６】

また、先に説明した適応型ＲＢＭでは、話者一人につきＩ^2＋Ｉ＋Ｊ（＝１０７２）個のパラメータ（Ｈ＝１６の場合）であったのに対して、話者クラスタ適応型ＲＢＭでは、話者一人につきＫ＋Ｉ＋Ｊ（＝５６）個のパラメータでよい。したがって、話者クラスタ適応型ＲＢＭによると、大幅にパラメータ数を削減することができ、少ないデータで適応が可能になる。

【００５７】

話者クラスタ適応型ＲＢＭにおいても、条件付き確率ｐ（ｖ，ｈ｜ｓ）を、先に説明した［数１］式～［数３］式で定義する。このとき、条件付き確率ｐ（ｖ｜ｈ，ｓ），ｐ（ｈ｜ｖ，ｓ）は、それぞれ次の［数１１］式および［数１２］式に示すようになる。

【００５８】

【数１１】

$$p(\boldsymbol{v}|\boldsymbol{h},\boldsymbol{s}) = \mathcal{N}(\boldsymbol{v}|\tilde{\boldsymbol{b}} + \tilde{\mathbf{W}}\boldsymbol{h}, \boldsymbol{\sigma}^2)$$

【００５９】

【数１２】

$$p(\boldsymbol{h}|\boldsymbol{v},\boldsymbol{s}) = \mathcal{B}(\boldsymbol{h}|\boldsymbol{f}(\tilde{\boldsymbol{d}} + \tilde{\mathbf{W}}^{\top}\frac{\boldsymbol{v}}{\boldsymbol{\sigma}^2}))$$

【００６０】

但し、［数１１］式の右辺のＮ（・）は次元独立の多変量正規分布、［数１２］式の右辺のＢ（・）は多次元ベルヌーイ分布、ｆ（・）は要素ごとのsoftmax関数を表す。

音韻特徴量ｈは既知であり、ある話者ｒの音響特徴量の平均ベクトルμ$_r$を考えると、［数１１］式より、平均ベクトルは［数１３］式に示すようになる。

【００６１】

【数１３】

$$\begin{aligned}\boldsymbol{\mu}_r &= \boldsymbol{b} + \boldsymbol{b}_r + \mathbf{U}\boldsymbol{\lambda}_r + \mathcal{A} \circ_3^1 \boldsymbol{\lambda}_r \mathbf{W}\boldsymbol{h} \\ &= \mathbf{M}\boldsymbol{\lambda}_r' + \boldsymbol{b}_r\end{aligned}$$

【００６２】

但し、λ$_r$′＝［λ$_r$^{T} １］^{T}は、λｒの拡張ベクトルであり、Ｍ＝［μ$_1$，・・・，μ$_{K+1}$］の各列ベクトルは、［数１４］式で定義される。

【００６３】

【数１４】

$$\boldsymbol{\mu}_k = \begin{cases} \boldsymbol{u}_k + \mathbf{A}_k\mathbf{W}\boldsymbol{h} & (k = 1, \cdots, K) \\ \boldsymbol{b} & (k = K+1) \end{cases}$$

【００６４】

本発明の一実施形態例による話者クラスタ適応型ＲＢＭでは、話者依存項ｂ$_r$が存在し、話者非依存平均ベクトルμ$_k$が［数１４］式のように構造化される特徴を持つ。また、潜在的な音韻特徴量を陽に確率変数として定義している。

【００６５】

また、本発明の一実施形態例による話者クラスタ適応型ＲＢＭでは、話者非依存パラメータと話者クラスタ重みを同時に推定することができる。すなわち、Ｒ人の話者によるＮフレームの音声データ｛ｖ$_n$｜ｓ$_n$｝$_{n=1}$^{N}に対する対数尤度（［数１５］式）を最大

化するように、確率的勾配法を用いて全てのパラメータΘ＝｛Ｗ，Ｕ，Ｖ，Ａ，Ｌ，Ｂ，Ｄ，ｂ，ｄ，σ｝を同時に更新し推定することが可能である。ここでは、それぞれのパラメータの勾配は省略する。

【００６６】

【数１５】

$$\mathcal{L} = \log \prod_{n} p(\boldsymbol{v}_n|\boldsymbol{s}_n) = \sum_{n} \log \sum_{\boldsymbol{h}} p(\boldsymbol{v}_n, \boldsymbol{h}_n|\boldsymbol{s}_n)$$

【００６７】

各勾配には計算困難なモデルに対する期待値が出現するが、通常のＲＢＭの確率モデルと同様に、ＣＤ法（Contrastive Divergence法）を用いることで、効率よく近似することができる。

また、クラスタ重みの非負条件を満たすために、$\lambda_r = e^{z_r}$と置き換えて、z_rでパラメータ更新を行う。クラスタ重みはパラメータ更新後、$||\lambda_r||_1 = 1$を満たすように正則化する。

さらに、モデルの学習が行われれば、音韻特徴量およびクラスタの形成が完了したとみなし、新たな話者ｒ′について、$\Theta_{r'} = \{\lambda_{r'}, b_{r'}, d_{r'}\}$のみを更新し推定し、他のパラメータは固定する。

【００６８】

この話者クラスタ適応型ＲＢＭを声質変換に適用する際には、ある入力話者の音声の音響特徴量$v^{(i)}$および話者特徴量$s^{(i)}$、目標話者の話者特徴量$s^{(o)}$が与えられたとき、最も確率の高い音響特徴量$v^{(o)}$が目標話者の音響特徴量であるとして、［数１６］式に示すように定式化される。

【００６９】

【数１６】

$$\begin{aligned}\hat{\boldsymbol{v}}^{(o)} &\triangleq \underset{\boldsymbol{v}}{\operatorname{argmax}}\, p(\boldsymbol{v}|, \boldsymbol{v}^{(i)}, \boldsymbol{s}^{(i)}, \boldsymbol{s}^{(o)}) \\ &\simeq \underset{\boldsymbol{v}}{\operatorname{argmax}}\, p(\boldsymbol{v}|\hat{\boldsymbol{h}}, \boldsymbol{s}^{(o)}) \\ &= \boldsymbol{b} + \mathbf{B}\boldsymbol{s}^{(o)} + \mathbf{U}\mathbf{L}\boldsymbol{s}^{(o)} + \mathcal{A} \circ_3^1 \mathbf{L}\boldsymbol{s}^{(o)}\mathbf{W}\hat{\boldsymbol{h}}\end{aligned}$$

【００７０】

但し、ｈ（^）は、入力話者の音響特徴量および話者特徴量が与えられたときのｈの条件付き期待値であり、［数１７］式で表される。

【００７１】

【数１７】

$$\begin{aligned}\hat{\boldsymbol{h}} &\triangleq \mathbb{E}[\boldsymbol{h}|\boldsymbol{v}^{(i)}, \boldsymbol{s}^{(i)}] \\ &= \boldsymbol{f}(\boldsymbol{d} + \mathbf{V}\mathbf{L}\boldsymbol{s}^{(i)} + \mathbf{D}\boldsymbol{s}^{(i)} + \mathbf{W}^{\top}(\mathcal{A} \circ_3^1 \mathbf{L}\boldsymbol{s}^{(i)})^{\top}\frac{\boldsymbol{v}^{(i)}}{\boldsymbol{\sigma}^2})\end{aligned}$$

【００７２】

［３．声質変換動作］

図６は、本発明の実施形態例による声質変換処理動作を示すフローチャートである。図６に示すように、パラメータ学習処理として、声質変換装置１のパラメータ学習ユニット１１の音声信号取得部１１１と話者情報取得部１１３とは、図示しない入力部によるユーザの指示に基づいて学習用音声信号とその対応話者情報とをそれぞれ取得する（ステップＳ１）。

前処理部１１２は、音声信号取得部１１１が取得した学習用音声信号からパラメータ推定部１１４に供給する学習用音声情報を生成する（ステップＳ２）。ここでは、例えば学

習用音声信号をフレームごと（例えば、５ｍｓｅｃごと）に切り出し、切り出された学習用音声信号にＦＦＴ処理などを施すことでスペクトル特徴量（例えば、ＭＦＣＣやメルケプストラム特徴量）を算出する。そして、算出したスペクトル特徴量の正規化処理（例えば、各次元の平均と分散を用いて正規化）を行うことで学習用音声情報ｖを生成する。

生成された学習用音声情報ｖは、話者情報取得部１１３によって取得された対応話者情報ｓとともにパラメータ推定部１１４へ出力される。

【００７３】

パラメータ推定部１１４は、話者クラスタ適応型ＲＢＭの学習処理を行う（ステップＳ３）。ここでは、学習用話者情報ｓに対応した話者クラスタｃと、学習用音声情報ｖを用いて各種パラメータの推定のための学習を行う。

【００７４】

次に、ステップＳ３の詳細について、図７を参照して説明する。まず、図７に示すように、話者クラスタ適応型ＲＢＭの確率モデルにおいて、全パラメータに任意の値を入力し（ステップＳ１１）、音声情報推定部１１４１に取得した学習用音声情報ｖを入力し、話者情報推定部１１４２に取得した対応話者情報ｓを入力する（ステップＳ１２）。

そして、話者情報推定部１１４２が取得した対応話者情報ｓから、話者クラスタ計算部１１４４が話者クラスタｃを計算し、その計算した話者クラスタｃと、音声情報推定部１１４１に取得した学習用音声情報ｖを入力とする（ステップＳ１３）。

【００７５】

次に、ステップＳ１３で入力された話者クラスタｃと学習用音声情報ｖとを用いて音韻情報ｈの条件付き確率密度関数を決定し、その確率密度関数に基づいて音韻情報ｈをサンプルする（ステップＳ１４）。ここで「サンプルする」とは、条件付き確率密度関数に従うデータをランダムに１つ生成することをいい、以下、同じ意味で用いる。

【００７６】

さらに、ステップＳ１４でサンプルされた音韻情報ｈと話者クラスタｃとを用いて音声情報ｖの条件付き確率密度関数を決定し、その確率密度関数に基づいて学習用音声情報ｖをサンプルする（ステップＳ１５）。

【００７７】

次に、ステップＳ１４でサンプルされた音韻情報ｈと、ステップＳ１５でサンプルされた学習用音声情報ｖとを用いて音韻情報ｈの条件付き確率密度関数を決定し、その確率密度関数に基づいて音韻情報ｈを再サンプルする（ステップＳ１６）。

【００７８】

そして、上述の［数１５］式で示される対数尤度Ｌをそれぞれのパラメータで偏微分し、勾配法により全パラメータを更新する（ステップＳ１７）。具体的には、確率的勾配法が用いられ、サンプルされた学習用音声情報ｖ、音韻情報ｈ、および対応話者情報ｓを用いてモデルに対する期待値を近似計算することができる。

【００７９】

全パラメータを更新した後、所定の終了条件を満たしていれば（ステップＳ１８のＹＥＳ）、次のステップに進み、満たしていなければ（ステップＳ１８のＮＯ）ステップＳ１１に戻り、以降の各ステップを繰り返す（ステップＳ１８）。なお、所定の終了条件としては、例えば、これら一連のステップの繰り返し数が挙げられる。

【００８０】

再び、図６に戻り、説明を続ける。パラメータ推定部１１４は、上述の一連のステップにより推定されたパラメータを学習により決定されたパラメータとして、パラメータ記憶ユニット１３に記憶する。そして、その記憶したパラメータを、入力した適応話者音声信号に基づいて、適応後のパラメータとする適用処理を行う。この適応処理で得られた適応後のパラメータを、声質変換ユニット１２の声質変換部１２４へ引き渡す（ステップＳ４）。

【００８１】

次に、ステップＳ４での適応処理の詳細について、図８を参照して説明する。まず、図

８に示すように、話者固有パラメータとして任意の値を入力し（ステップＳ２１）、音声情報推定部１４４１に取得した適応話者音声情報ｖを入力し、話者情報推定部１４４２に取得した適応話者情報ｓを入力する（ステップＳ２２）。

そして、話者情報推定部１４４２が取得した適応話者情報ｓから、話者クラスタ計算部１４４４が話者クラスタｃを計算し、その計算した話者クラスタｃと、音声情報推定部１４４１に取得した適応話者音声情報ｖを入力とする（ステップＳ２３）。

【００８２】

次に、ステップＳ２３で入力された話者クラスタｃと適応話者音声情報ｖとを用いて音韻情報ｈの条件付き確率密度関数を決定し、その確率密度関数に基づいて音韻情報ｈをサンプルする（ステップＳ２４）。

さらに、ステップＳ２４でサンプルされた音韻情報ｈと話者クラスタｃとを用いて音声情報ｖの条件付き確率密度関数を決定し、その確率密度関数に基づいて適応話者音声情報ｖをサンプルする（ステップＳ２５）。

【００８３】

次に、ステップＳ２４でサンプルされた音韻情報ｈと、ステップＳ２５でサンプルされた適応話者音声情報ｖとを用いて音韻情報ｈの条件付き確率密度関数を決定し、その確率密度関数に基づいて音韻情報ｈを再サンプルする（ステップＳ２６）。

【００８４】

そして、上述の［数１５］式で示される対数尤度Ｌをそれぞれのパラメータで偏微分し、勾配法により適応話者に固有のパラメータを更新する（ステップＳ２７）。

【００８５】

適応話者に固有のパラメータを更新した後、所定の終了条件を満たしていれば（ステップＳ２８のＹＥＳ）、次のステップに進み、満たしていなければ（ステップＳ２８のＮＯ）ステップＳ２１に戻り、以降の各ステップを繰り返す（ステップＳ２８）。

【００８６】

再び、図６に戻り、説明を続ける。

声質変換処理として、ユーザは、図示しない入力部を操作して声質変換ユニット１２の話者情報設定部１２３において声質変換の目標となる目標話者の情報ｓ（ｏ）を設定する（ステップＳ５）。そして、音声信号取得部１２１により変換用音声信号を取得する（ステップＳ６）。

前処理部１２２は、パラメータ学習処理の場合と同じく変換用音声信号に基づいて音声情報を生成し、話者情報取得部１２３によって取得された対応話者情報ｓとともに声質変換部１２４へ出力される（ステップＳ７）。

声質変換部１２４は、話者クラスタ適応型ＲＢＭを適用して、適応話者の音声を目標話者の音声に変換する声質変換を行う（ステップＳ８）。

【００８７】

次に、ステップＳ８の詳細について、図９を参照して説明する。まず、図９に示すように、話者クラスタ適応型ＲＢＭの確率モデルにおいて、決定された全パラメータを入力し（ステップＳ３１）、音声情報設定部１２４１に音声情報ｖを入力し、話者情報設定部１２４２に入力話者情報ｓを入力し、話者クラスタ計算部１２４４が入力話者の話者クラスタｃを計算する（ステップＳ３２）。

そして、ステップＳ３２で計算された話者クラスタｃと音声情報ｖとを用いて、音韻情報ｈを推定する（ステップＳ３３）。

【００８８】

次に、声質変換部１２４は、パラメータ学習処理で学習済みの目標話者の話者情報ｓを取得し、話者クラスタ計算部１２４４が目標話者の話者クラスタｃを計算する（ステップＳ３４）。そして、ステップＳ３４で計算された目標話者の話者クラスタｃとステップＳ３３で推定した音韻情報ｈとを用いて、音声情報設定部１２４１が変換済み音声情報ｖを推定する（ステップＳ３５）。推定された変換済み音声情報ｖ（ｏ）は、後処理部１２５へ出力される。

【００８９】
再び、図６に戻り、説明を続ける。後処理部１２５は、変換済み音声情報ｖを用いて変換済み音声信号を生成する（ステップＳ９）。具体的には、正規化されている変換済み音声信号ｖに非正規化処理（ステップＳ２で説明した正規化処理に用いる関数の逆関数を施す処理）を施し、非正規化処理のなされたスペクトル特徴量を逆変換することでフレームごとの変換済み音声信号を生成し、これらフレームごとの変換済み音声信号を時刻順に結合することで変換済み音声信号を生成する。
後処理部１２５により生成された変換済み音声信号は、音声信号出力部１２６より外部へ出力される（ステップＳ１０）。変換済み音声信号を外部に接続されたスピーカで再生することにより、目標話者の音声に変換された入力音声を聞くことができる。
【００９０】
［４．評価実験例］
次に、本発明による話者クラスタ適応型ＲＢＭの効果を実証するため、声質変換実験を行った例について説明する。
確率モデルの学習には日本音響学会研究用連続音声データベース（ASJ-JIPDEC）の中からランダムにＲ＝８；１６；５８名の話者を選び、４０センテンスの音声データを用いた。学習話者の評価には、男性１名（ECL0001）を入力話者、女性１名（ECL1003）を目標話者とし、学習データとは別の１０センテンスの音声データを用いた。確率モデルの適応には、学習時に含まれない女性話者（ECL1004）、男性話者（ECL0002）をそれぞれ入力話者、目標話者とし、適応データのセンテンス数を０．２から４０まで変えて評価を行った。適応話者の評価についても適応データに含まれない１０センテンスの音声データを用いた。分析合成ツール（WORLD：URL http://ml.cs.yamanashi.ac.jp/world/index.html）によって得られたスペクトルから計算した３２次元のメルケプストラムを入力特徴量に用いた（Ｉ＝３２）。また、潜在音韻特徴量の数をＪ＝８；１６；２４、クラスタの数をＫ＝２；３；４；６；８とし、最も高い精度となるものを採用した。学習率０：０１、モーメント係数０：９、バッチサイズ１００×Ｒ、繰り返し回数１００の確率的勾配法を用いて確率モデルを学習した。
声質変換の精度を測る指標として、以下の［数１８］式で定義されるＭＤＩＲ(mel-cepstral distortion improvement ratio)の平均値を用いた。
【００９１】
【数１８】

$$MDIR[dB] = \frac{10\sqrt{2}}{\ln 10}(\|\boldsymbol{v}_o - \boldsymbol{v}_i\|_2 - \|\boldsymbol{v}_o - \hat{\boldsymbol{v}}_o\|_2)$$

【００９２】
ここで、v_o、v_i、v_o（^）は、それぞれ、入力話者とアライメントをとった目標話者音声のメルケプストラム特徴量、同アライメントをとった入力話者音声のメルケプストラム特徴量、v_iに対して声質変換を施した音声のメルケプストラム特徴量を示す。ＭＤＩＲは改善率を表し、値が大きいほど高い変換精度を示す。
まず、Ｋ＝２；Ｒ＝８およびＫ＝３；Ｒ＝１６としたとき、推定された各話者のクラスタ重みλ_rの分布を図１０Ａおよび図１０Ｂに示す。図１０Ａの例は、Ｋ＝２であり、男性のクラスタ(Cluster 1)と女性のクラスタ(Cluster 2)との２つのクラスタが自動的に形成されている。図１０Ｂの例は、Ｋ＝３であり、男性のクラスタ(Cluster 1)と女性のクラスタ(Cluster 2)の他に、さらに男女が混ざった別のクラスタ（Cluster 3）が自動的に形成されている。この図１０Ａおよび図１０Ｂにおいて、各学習者の話者クラスタの位置Ｒ１１～Ｒ１８およびＲ２１～Ｒ３０を示し、○印で示す音声は男性の音声であり、×印で示す音声は女性の音声である。
【００９３】
図１０Ａおよび図１０Ｂから分かるように、○印で示す男性の音声は、(Cluster 1)に近い位置（クラスタ重み）になり、×印で示す女性の音声は、(Cluster 2)に近い位置に

学習されており、性別の教師を与えていないにも関わらず、男性のクラスタ(Cluster 1)と女性のクラスタ(Cluster 2)が自動的に形成されていることが分かる。また、図１０Ａおよび図１０Ｂに示すように、学習データでは、二つのクラスタが最も離れるように学習されている。すなわち、互いに最も離れている話者ペアが、それぞれのクラスタ(Cluster 1及びCluster 2)と重なる位置に設定されている。そして、各クラスタが最も離れるように学習した複数のクラスタの間で、話者クラスタへの重みの位置を設定する。このように複数のクラスタが最も離れるように学習する性質は、各クラスタ（代表話者）を内分する点を自由に調節することで任意の声へ変換する際、調節の幅が広くなり好ましい。

【００９４】

次に、本発明による話者クラスタ適応型ＲＢＭによる確率モデル（ＣＡＢと示す）と、従来の非パラレル声質変換手法である適応型ＲＢＭ（ＡＲＢＭと示す）の学習話者の変換精度を比較した例を、［表１］に示す。ここでは、学習人数が８人、１６人、５８人の例を示し、値が高いほど精度が高いことを示す。

【００９５】

【表１】

# persons	8	16	58
ARBM	**3.70**	2.64	3.02
CAB	3.21	**3.06**	**3.23**

【００９６】

従来の適応型ＲＢＭ（ＡＲＢＭ）では、話者数の少ない場合では高い精度を示すが、話者数を増加させると精度が低下することが分かる。一方、話者ごとのパラメータ数を抑えた話者クラスタ適応型ＲＢＭによる確率モデル（ＣＡＢ）では、話者数を増加させても精度に変化はあまり見られない。

［表２］は、本発明による話者クラスタ適応型ＲＢＭによる確率モデルと、従来の適応型ＲＢＭ（ＡＲＢＭ）による確率モデルとの、センテンス数による変換精度を比較した例である。

【００９７】

【表２】

# sent.	0.2	0.5	1	10	40
ARBM	2.48	3.25	3.21	3.41	3.45
CAB	**3.14**	**3.54**	**3.63**	**3.60**	**3.58**

【００９８】

［表２］から明らかなように、適応に用いるセンテンス数が１以下のとき、従来モデルでは精度の低下が見られるが、話者クラスタ適応型ＲＢＭによる確率モデル（ＣＡＢ）では、０．５センテンス程度で、１０センテンス以上の場合と同等のパフォーマンスが得られる。

【００９９】

以上、本発明によれば、話者情報から話者クラスタを取得して、その話者クラスタを使って確率モデルを得るようにしたので、従来よりも非常に少ないデータ数で、入力話者音声を目標話者音声に声質変換できるようになる。

【０１００】

［５．変形例］

なお、ここまで説明した実施形態例では、目標話者の音声情報ｖと音韻情報ｎとを得る処理として、図５の話者クラスタ適応型ＲＢＭのグラフ構造に示すように、話者クラスタ

ｃが持つパラメータＡ，Ｖ，Ｕから、演算で目標話者の音声情報ｖと音韻情報ｎを得るようにした。

これに対して、図１１に示すように、話者クラスタｃが持つパラメータＡ，Ｖ，Ｕから、目標話者の話者情報ｓを得、得られた話者情報ｓを使って、話者に依存したパラメータＤ，Ａ，Ｂを得た後、これらのパラメータＤ，Ａ，Ｂから、目標話者の音声情報ｖと音韻情報ｎを得るようにしてもよい。話者に依存したパラメータＤ，Ａ，Ｂから、目標話者の音声情報ｖと音韻情報ｎを得る処理については、例えば図４の適応型ＲＢＭのグラフ構造で説明した処理が適用可能である。

この図１１に示すように、話者クラスタｃを使って目標話者の話者情報ｓを得た後、目標話者の音声情報ｖと音韻情報ｎを得るようにすることでも、図５の例と同様に、適切な目標話者の音声情報ｖと音韻情報ｎを得ることができる。この図１１に示す処理を行う場合には、目標話者の音声情報ｖと音韻情報ｎが、目標話者の話者情報ｓから得られるため、それぞれの情報の精度が向上する効果を有する。但し、データ量については、図５の例よりも増加する。

【０１０１】

また、ここまで説明した実施形態例では、学習用の音声信号による学習で、声質変換のためのパラメータを学習処理した後、適応話者音声信号の入力で、パラメータを適応話者音声信号に適応した後、適応されたパラメータを使って、目標話者の音声信号に声質変換するようにした。このようにすることで、事前に学習されていない音声信号（適応話者音声信号）を、目標話者の音声信号に声質変換することができる。これに対して、適応話者音声信号の入力を省略して、学習用の音声信号で得たパラメータを使って、学習用の音声信号を目標話者の音声信号に声質変換してもよい。

この場合には、声質変換装置１は、例えば図１に示す構成として、パラメータ学習ユニット１１での学習で得られたパラメータをパラメータ記憶ユニット１３が記憶し、声質変換処理ユニット１２は、パラメータ記憶ユニット１３が記憶したパラメータを適用して、入力音声を目標話者の音声に変換処理すればよい。

【０１０２】

また、ここまで説明した実施形態例では、学習を行う入力音声（入力話者の音声）や適応を行う入力音声として、人間の話し声の音声を処理する例について説明したが、実施形態例で説明した各情報を得る学習が可能であれば、学習用や適応を行う音声信号（入力信号）として、人間の話し声以外の様々な音とし、その音声信号を学習又は適応するようにしてもよい。例えば、サイレンの音や動物の鳴き声などのような音を学習又は適応するようにしてもよい。

【符号の説明】

【０１０３】

１・・・声質変換装置、１１・・・パラメータ学習ユニット、１２・・・声質変換処理ユニット、１３・・・パラメータ記憶ユニット、１４・・・適応ユニット、１０１・・・ＣＰＵ、１０２・・・ＲＯＭ、１０３・・・ＲＡＭ、１０４・・・ＨＤＤ／ＳＤＤ、１０５・・・接続Ｉ／Ｆ、１０６・・・通信Ｉ／Ｆ、１１１，１２１，１４１・・・音声信号取得部、１１２，１２２，１４２・・・前処理部、１１３・・・対応話者情報取得部、１１４，１４４・・・パラメータ推定部、１１４１，１４４１・・・音声情報推定部、１１４２，１４４２・・・話者情報推定部、１１４３，１４４３・・・音韻情報推定部、１１４４，１４４４・・・話者クラスタ計算部、１２３・・・話者情報設定部、１２４・・・声質変換部、１２４１・・・音声情報設定部、１２４２・・・話者情報設定部、１２４３・・・音韻情報設定部、１２４４・・・話者クラスタ計算部、１２５・・・後処理部、１２５・・・音声信号出力部

【図１】

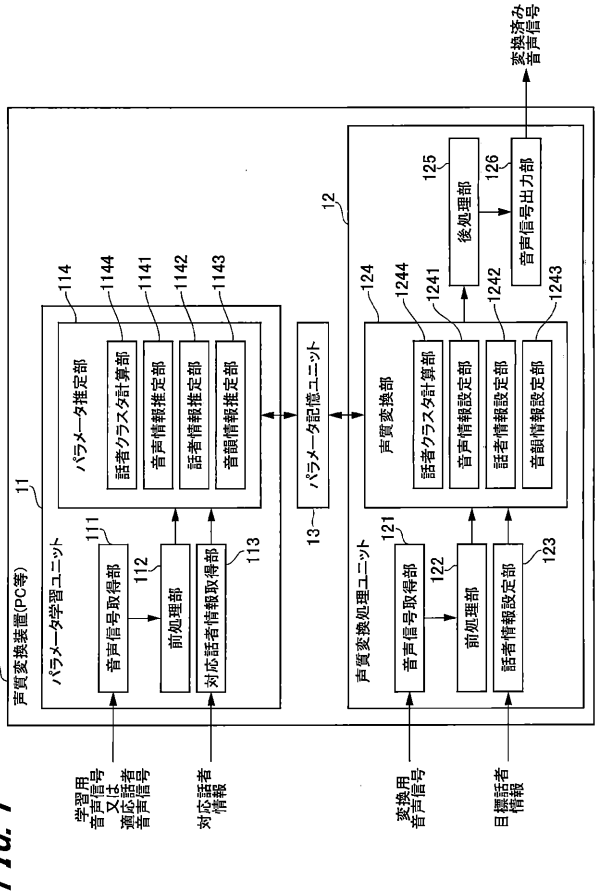


FIG. 1

【図２】

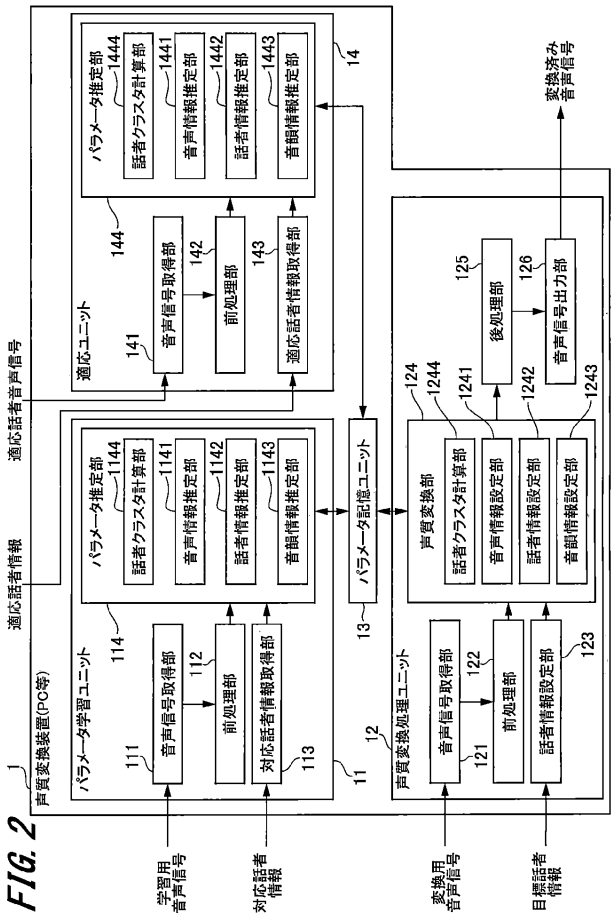
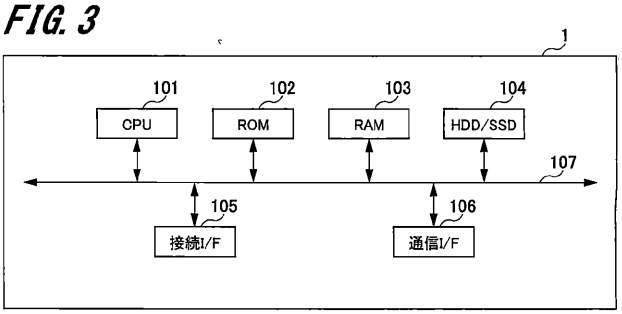


FIG. 2

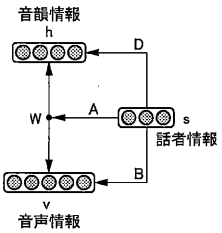
【図３】

FIG. 3



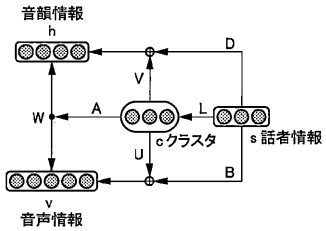
【図４】

FIG. 4



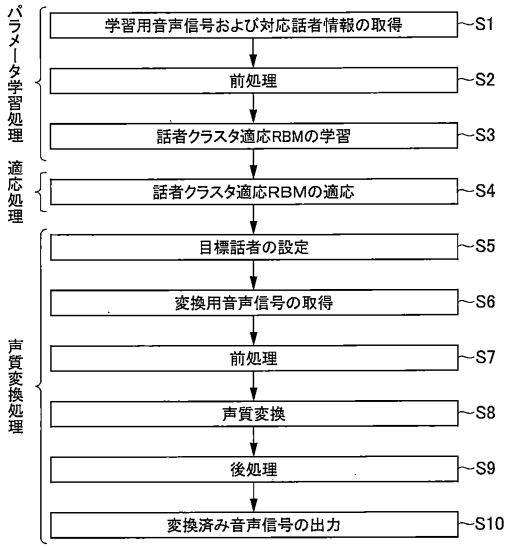
【図５】

FIG. 5



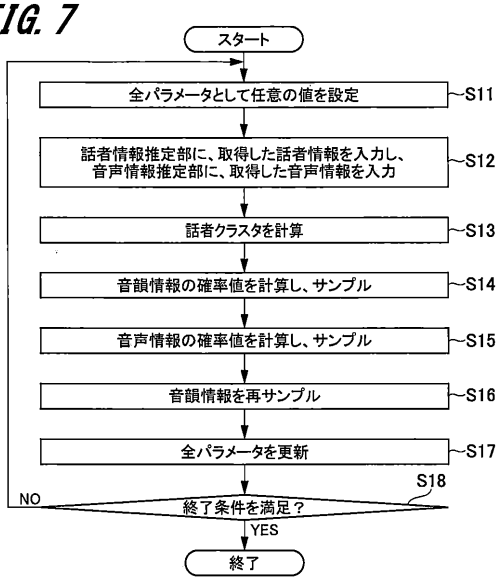
【図６】

FIG. 6



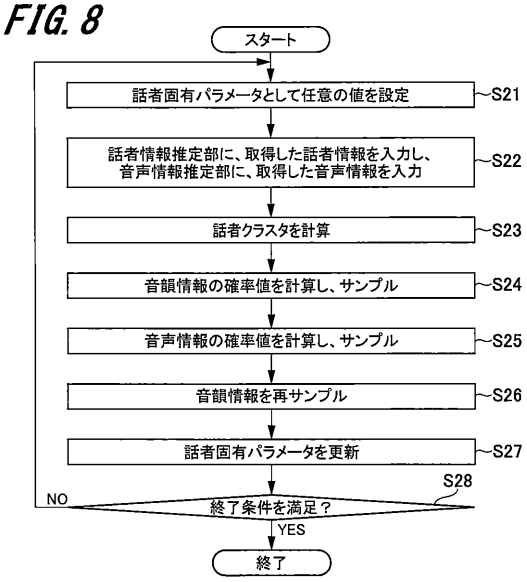
【図７】

FIG. 7



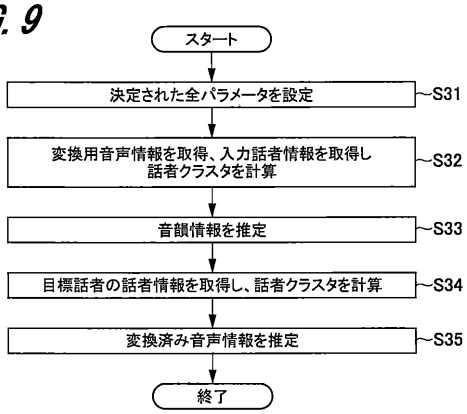
【図８】

FIG. 8



【図９】

FIG. 9



【図１０】

FIG. 10A

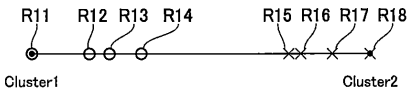
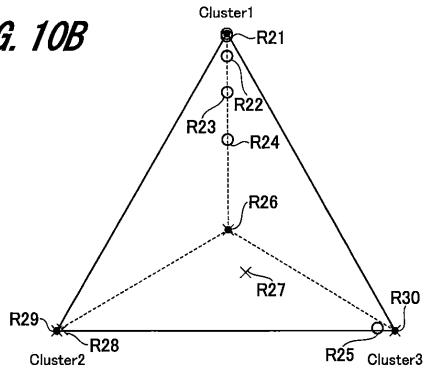
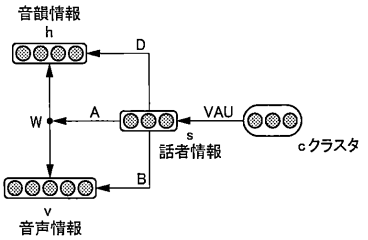


FIG. 10B



【図１１】

FIG. 11



【手続補正書】
【提出日】平成30年10月22日(2018.10.22)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
入力話者の音声を目標話者の音声に声質変換する声質変換装置であって、
学習用の音声に基づく音声情報およびその音声情報に対応する話者情報から、声質変換のためのパラメータを決定するパラメータ学習ユニットと、
前記パラメータ学習ユニットが決定したパラメータを記憶するパラメータ記憶ユニットと、
前記パラメータ記憶ユニットが記憶したパラメータと前記目標話者の話者情報とに基づいて、前記入力話者の音声に基づく前記音声情報の声質変換処理を行う声質変換処理ユニットとを備え、
前記パラメータ学習ユニットは、音声に基づく音声情報、音声情報に対応する話者情報および音声中の音韻を表す音韻情報のそれぞれを変数とすることで、前記音声情報、前記話者情報および前記音韻情報のそれぞれの間の結合エネルギーの関係性を前記パラメータによって表す確率モデルを取得し、前記確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義するようにし、
前記声質変換処理ユニットは、前記パラメータから前記目標話者の話者情報を得、得られた話者情報から前記目標話者の音声情報を得るようにした
声質変換装置。

【請求項２】
さらに、前記パラメータ記憶ユニットが記憶したパラメータを前記入力話者の音声に適応して、適応後のパラメータを得る適応ユニットを備え、
前記パラメータ記憶ユニットは、前記適応ユニットで適応後のパラメータを記憶し、前記声質変換処理ユニットは、適応後のパラメータと前記目標話者の話者情報とに基づいて、前記入力話者の音声に基づく前記音声情報の声質変換処理を行う
請求項１に記載の声質変換装置。
【請求項３】
前記パラメータ学習ユニットと前記適応ユニットは共通の演算処理部で構成され、
前記学習用の音声に基づいてパラメータを決定する処理と、前記入力話者の音声に基づいて適応後のパラメータを得る処理を、前記共通の演算処理部で行うようにした
請求項２に記載の声質変換装置。
【請求項４】
前記パラメータ学習ユニットが学習する際には、複数のクラスタが最も離れるように学習し、学習した複数のクラスタの間で、話者クラスタへの重みの位置を設定する
請求項１に記載の声質変換装置。
【請求項５】
（削除）
【請求項６】
音声情報の特徴量ｖ＝［ｖ$_{1}$，・・・，ｖ$_{I}$］∈Ｒ^{I}と、音韻情報の特徴量ｈ＝［ｈ$_{1}$，・・・，ｈ$_{J}$］∈｛０，１｝^{J}，Σ$_{j}$ｈ$_{j}$＝１との間に、話者情報の特徴量ｓ＝［ｓ$_{1}$，・・・，ｓ$_{R}$］∈｛０，１｝^{R}，Σ$_{r}$ｓ$_{r}$＝１に依存した双方な結合重みＷ∈Ｒ$^{I\times J}$が存在すると仮定したとき、前記話者クラスタとして、話者クラスタｃ∈Ｒ^{K}を導入し、話者クラスタｃを、

$$\boldsymbol{c} \triangleq \mathbf{L}\boldsymbol{s}$$

（但し、Ｌ∈Ｒ$^{K\times R}$＝［λ$_{1}$・・・λ$_{R}$］の各列ベクトルλ$_{r}$は、それぞれの話者クラスタへの重みを表す非負パラメータであり、｜｜λ$_{r}$｜｜$_{1}$＝１，∀$_{r}$の制約を課す）と表現し、音響情報の特徴量のクラスタ依存項のバイアスパラメータをＵ∈Ｒ$^{I\times K}$、音韻情報の特徴量のクラスタ依存項のバイアスパラメータをＶ∈Ｒ$^{J\times K}$、として、話者非依存項、クラスタ依存項、および話者依存項のそれぞれを、

$$\tilde{\mathbf{W}} \triangleq \mathcal{A} \circ_3^{\mathrm{T}} \mathbf{c}\mathbf{W}$$
$$\tilde{\boldsymbol{b}} \triangleq \boldsymbol{b} + \mathbf{U}\boldsymbol{c} + \mathbf{B}\boldsymbol{s}$$
$$\tilde{\boldsymbol{d}} \triangleq \boldsymbol{d} + \mathbf{V}\boldsymbol{c} + \mathbf{D}\boldsymbol{s}$$

として示す
請求項１に記載の声質変換装置。
【請求項７】
入力話者の音声を目標話者の音声に声質変換する声質変換方法であって、
音声に基づく音声情報、音声情報に対応する話者情報および音声中の音韻を表す音韻情報のそれぞれを変数とすることで、前記音声情報、前記話者情報および前記音韻情報のそれぞれの間の結合エネルギーの関係性をパラメータによって表す確率モデルを用意し、その確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義し、それぞれの話者について、前記複数個の話者クラスタへの重みを推定して、学習用の音声についての前記パラメータを決定するパラメータ学習ステップと、
前記パラメータ学習ステップで得られたパラメータ、又は当該パラメータを前記入力話者の音声に適応した適応後のパラメータと、前記目標話者の前記話者情報とに基づいて、

前記入力話者の音声に基づく前記音声情報の声質変換処理を行う声質変換処理ステップと
を含み、
前記声質変換処理ステップでの声質変換処理では、前記パラメータから前記目標話者の話者情報を得、得られた話者情報から前記目標話者の音声情報を得るようにした
声質変換方法。
【請求項８】
音声に基づく音声情報、音声情報に対応する話者情報および音声中の音韻を表す音韻情報のそれぞれを変数とすることで、前記音声情報、前記話者情報および前記音韻情報のそれぞれの間の結合エネルギーの関係性をパラメータによって表す確率モデルを用意し、その確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義し、それぞれの話者について、前記複数個の話者クラスタへの重みを推定して、学習用の音声についての前記パラメータを決定して記憶するパラメータ学習ステップと、
前記パラメータ学習ステップで得られたパラメータ、又は当該パラメータを入力話者の音声に適応した適応後のパラメータと、目標話者の前記話者情報とに基づいて、前記入力話者の音声に基づく前記音声情報の声質変換処理を行う声質変換処理ステップと、
をコンピュータに実行させるプログラムであり、
前記声質変換処理ステップでの声質変換処理は、前記パラメータから前記目標話者の話者情報を得、得られた話者情報から前記目標話者の音声情報を得るようにした
プログラム。
【手続補正２】
【補正対象書類名】明細書
【補正対象項目名】０００３
【補正方法】変更
【補正の内容】
【０００３】
声質変換が可能な声質変換装置、声質変換方法およびプログラムを提供することを目的とする。
課題を解決するための手段
［０００９］
上記課題を解決するため、本発明の声質変換装置は、入力話者の音声を目標話者の音声に声質変換する声質変換装置であって、パラメータ学習ユニットとパラメータ記憶ユニットと声質変換処理ユニットとを備える。
パラメータ学習ユニットは、学習用の音声に基づく音声情報およびその音声情報に対応する話者情報から、声質変換のためのパラメータを決定する。
パラメータ記憶ユニットは、パラメータ学習ユニットが決定したパラメータを記憶する。
声質変換処理ユニットは、パラメータ記憶ユニットが記憶したパラメータと目標話者の話者情報とに基づいて、入力話者の音声に基づく音声情報の声質変換処理を行う。
ここで、パラメータ学習ユニットは、音声に基づく音声情報、音声情報に対応する話者情報および音声中の音韻を表す音韻情報のぞれぞれを変数とすることで、音声情報、話者情報および音韻情報のそれぞれの間の結合エネルギーの関係性をパラメータによって表す確率モデルを取得し、確率モデルとして、固有の適応行列を持つ複数個の話者クラスタを定義するようにし、声質変換処理ユニットは、パラメータから目標話者の話者情報を得、得られた話者情報から目標話者の音声情報を得るようにした。
［００１０］
また、本発明の声質変換方法は、入力話者の音声を目標話者の音声に声質変換する方法であって、パラメータ学習ステップと声質変換処理ステップとを含む。
パラメータ学習ステップは、音声に基づく音声情報、音声情報に対応する話者情報および音声中の音韻を表す音韻情報のそれぞれを変数とすることで、音声情報、話者情報および音韻情報のそれぞれの間の結合エネルギーの関係性をパラメータによって表す確率モデ

ルを用意する。そして、その確率モデ
【手続補正３】
【補正対象書類名】明細書
【補正対象項目名】０００４
【補正方法】変更
【補正の内容】
【０００４】
ルとして、固有の適応行列を持つ複数個の話者クラスタを定義し、それぞれの話者について、複数個の話者クラスタへの重みを推定して、学習用の音声についてのパラメータを決定する。

声質変換処理ステップは、パラメータ学習ステップで得られたパラメータ、又は当該パラメータを入力話者の音声に適応した適応後のパラメータと、目標話者の話者情報とに基づいて、入力話者の音声に基づく音声情報の声質変換処理を行う。声質変換処理ステップでの声質変換処理では、パラメータから目標話者の話者情報を得、得られた話者情報から目標話者の音声情報を得るようにした。

［００１１］

また本発明のプログラムは、上述した声質変換方法のパラメータ学習ステップと声質変換処理ステップとをコンピュータに実行させるものである。

［００１２］

本発明によれば、話者クラスタにより目標話者を設定することができるため、従来よりも非常に少ないデータ数で、入力話者音声を目標話者音声に声質変換できるようになる。

図面の簡単な説明

［００１３］

［図１］本発明の一実施の形態例に係る声質変換装置の構成例（例１）を示すブロック図である。

［図２］本発明の一実施の形態例に係る声質変換装置の構成例（例２）を示すブロック図である。

［図３］声質変換装置のハードウェア構成例を示すブロック図である。

［図４］従来の確率モデルを模式的に示す説明図である。

［図５］声質変換装置のパラメータ推定部が備える確率モデルを模式的に示す説明図である。

［図６］本発明の一実施の形態例に係る処理全体の流れを示すフローチャートである。

［図７］図６のステップＳ３の学習の詳細例を示すフローチャートである。

［図８］図６のステップＳ４の適応の詳細例を示すフローチャートである。

［図９］図６のステップＳ８の声質変換の詳細例を示すフローチャートである。

［図１０］本発明の一実施形態によるクラスタの重み分布の例を示す説明図である。

［図１１］声質変換装置のパラメータ推定部が備える確率モデルの別の例を示す説明図である。

発明を実施するための形態

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2018/007268

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl. G10L21/007(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl. G10L21/007

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2018
Registered utility model specifications of Japan	1996-2018
Published registered utility model applications of Japan	1994-2018

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
IEEE Xplore

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	中鹿亘、滝口 哲也, 制約付き Three-Way Restricted Boltzmann Machine を用いた音響・音韻・話者情報の同時モデリング, 電子情報通信学会技術研究報告 vol. 115, no. 346, 25 November 2015, vol. 115, no. 346, pp. 7-12, (NAKASHIKA, Toru, TAKIGUCHI, Tetsuya, "Simultaneous Modelling of Acoustic, Phonetic, Speaker Features Using Improved Three-Way Restricted Boltzmann MachineIEICE technical report", IEICE technical report)	1-3, 7-8 4-6
A	JP 2016-29779 A (KDDI CORPORATION) 03 March 2016, paragraph [0005] (Family: none)	1-8
P, X	WO 2017/146073 A1 (THE UNIVERSITY OF ELECTRO-COMMUNICATIONS) 31 August 2017, claim 1 (Family: none)	1-3, 7-8

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 May 2018 (07.05.2018)	22 May 2018 (22.05.2018)
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (January 2015)

国際調査報告	国際出願番号 PCT／JP2018／007268

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G10L21/007(2013.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. G10L21/007

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922－1996年
日本国公開実用新案公報	1971－2018年
日本国実用新案登録公報	1996－2018年
日本国登録実用新案公報	1994－2018年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

IEEE Xplore

C. 関連すると認められる文献

引用文献のカテゴリー＊	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
X A	中鹿 亘、滝口 哲也, 制約付きThree－Way Restricted Boltzmann Machineを用いた音響・音韻・話者情報の同時モデリング, 電子情報通信学会技術研究報告 Vol. 115 No. 346, 2015.11.25, 第115巻、第346号, p.7-12	1-3,7-8 4-6
A	JP 2016-29779 A（KDDI株式会社）2016.03.03,［0005］（ファミリーなし）	1-8

☑ C欄の続きにも文献が列挙されている。　☐ パテントファミリーに関する別紙を参照。

＊ 引用文献のカテゴリー
「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
07. 05. 2018	22. 05. 2018

国際調査機関の名称及びあて先
日本国特許庁（ISA／JP）
郵便番号100－8915
東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員） 5Z 4188
冨澤 直樹
電話番号 03－3581－1101 内線 3591

国際調査報告

国際出願番号 ＰＣＴ／ＪＰ２０１８／００７２６８

C（続き）． 関連すると認められる文献

引用文献のカテゴリー＊	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
P, X	WO 2017/146073 A1 (国立大学法人電気通信大学) 2017.08.31, [請求項１] (ファミリーなし)	1-3,7-8

フロントページの続き

(81)指定国・地域 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT

（注）この公表は、国際事務局（ＷＩＰＯ）により国際公開された公報を基に作成したものである。なおこの公表に係る日本語特許出願（日本語実用新案登録出願）の国際公開の効果は、特許法第１８４条の１０第１項(実用新案法第４８条の１３第２項）により生ずるものであり、本掲載とは関係ありません。