(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5194197号
(P5194197)
(45)発行日　平成25年5月8日(2013.5.8)　(24)登録日　平成25年2月8日(2013.2.8)

(51)Int.Cl.　　F I
G10L 21/007 (2013.01)　G10L 21/04 120D
G10L 25/15 (2013.01)　G10L 11/00 101D
G10L 25/75 (2013.01)　G10L 11/00 501

請求項の数 17　(全 39 頁)

(21)出願番号　特願2012-551826 (P2012-551826)
(86)(22)出願日　平成24年7月12日(2012.7.12)
(86)国際出願番号　PCT/JP2012/004517
(87)国際公開番号　WO2013/008471
(87)国際公開日　平成25年1月17日(2013.1.17)
審査請求日　平成24年11月14日(2012.11.14)
(31)優先権主張番号　特願2011-156042 (P2011-156042)
(32)優先日　平成23年7月14日(2011.7.14)
(33)優先権主張国　日本国(JP)

早期審査対象出願

(73)特許権者　000005821
パナソニック株式会社
大阪府門真市大字門真１００６番地
(74)代理人　100109210
弁理士　新居　広守
(72)発明者　釜井　孝浩
日本国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　廣瀬　良文
日本国大阪府門真市大字門真１００６番地
パナソニック株式会社内

審査官　山下　剛史

最終頁に続く

(54)【発明の名称】声質変換システム、声質変換装置及びその方法、声道情報生成装置及びその方法

(57)【特許請求の範囲】
【請求項１】
声道の形状を示す声道形状情報を用いて入力音声の声質を変換する声質変換システムであって、
互いに種類が異なる複数の母音の音声を受け付ける母音受付部と、
前記母音受付部によって受け付けられた複数の母音の音声を分析することにより、前記母音の種類毎に、第１声道形状情報を生成する分析部と、
前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する混合部と、
入力音声の声道形状情報及び音源情報を取得し、前記入力音声に含まれる母音の声道形状情報と、前記入力音声に含まれる母音と同じ種類の母音の前記第２声道形状情報とを混合することにより、前記入力音声の声道形状情報を変換し、変換後の前記入力音声の声道形状情報と前記入力音声の音源情報とを用いて合成音を生成することにより、前記入力音声の声質を変換する合成部とを備える
声質変換システム。
【請求項２】
前記混合部は、
前記母音の種類毎に生成された複数の第１声道形状情報を平均することにより、１つの平均声道形状情報を算出する平均声道情報算出部と、

前記母音受付部によって受け付けられた母音の種類毎に、当該母音の第１声道形状情報と前記平均声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する混合声道情報生成部とを備える
請求項１に記載の声質変換システム。
【請求項３】
平均声道情報算出部は、前記複数の第１声道形状情報を重み付き算術平均することにより、前記平均声道形状情報を算出する
請求項２に記載の声質変換システム。
【請求項４】
前記混合部は、前記入力音声に含まれる母音の局所的発話速度が大きいほど、前記入力音声に含まれる母音と同じ種類の母音の前記第２声道形状情報が前記母音の種類毎に生成された複数の第１声道形状情報の平均に近付くように、前記第２声道形状情報を生成する
請求項１～３のいずれか１項に記載の声質変換システム。
【請求項５】
前記混合部は、母音の種類に応じて設定された混合比率を用いて、前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合する
請求項１～４のいずれか１項に記載の声質変換システム。
【請求項６】
前記混合部は、ユーザーによって設定された混合比率を用いて、前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合する
請求項１～５のいずれか１項に記載の声質変換システム。
【請求項７】
前記混合部は、前記入力音声の言語種類に応じて設定された混合比率を用いて、前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合する
請求項１～６のいずれか１項に記載の声質変換システム。
【請求項８】
前記声質変換システムは、さらに、
前記入力音声の声道形状情報及び音源情報が記憶されている入力音声記憶部を備え、
前記合成部は、前記入力音声記憶部から、前記入力音声の声道形状情報及び音源情報を取得する
請求項１～７のいずれか１項に記載の声質変換システム。
【請求項９】
入力音声の声質を変換する際に用いられる、声道の形状を示す声道形状情報を生成する声道情報生成装置であって、
互いに種類が異なる複数の母音の音声を分析することにより、前記母音の種類毎に、第１声道形状情報を生成する分析部と、
前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する混合部とを備える
声道情報生成装置。
【請求項１０】
さらに、
前記母音の種類毎に、前記第２声道形状情報を用いて合成音を生成する合成部と、
前記合成音を音声として出力する出力部とを備える
請求項９に記載の声道情報生成装置。
【請求項１１】
声道の形状を示す声道形状情報を用いて入力音声の声質を変換する声質変換装置であっ

て、
母音の種類毎に、当該母音の第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合することにより生成された第２声道形状情報を記憶している母音声道情報記憶部と、
入力音声に含まれる母音の声道形状情報と、前記入力音声に含まれる母音と同じ種類の母音の前記第２声道形状情報とを混合することにより、前記入力音声の声道形状情報を変換し、変換後の前記入力音声の声道形状情報と前記入力音声の音源情報とを用いて合成音を生成することにより、前記入力音声の声質を変換する合成部とを備える
声質変換装置。
【請求項１２】
声道の形状を示す声道形状情報を用いて入力音声の声質を変換する声質変換方法であって、
互いに種類が異なる複数の母音の音声を受け付ける母音受付ステップと、
前記母音受付ステップにおいて受け付けられた複数の母音の音声を分析することにより、前記母音の種類毎に第１声道形状情報を生成する分析ステップと、
前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する混合ステップと、
入力音声に含まれる母音の声道形状情報と、前記入力音声に含まれる母音と同じ種類の母音の前記第２声道形状情報とを混合することにより、前記入力音声の声道形状情報を変換する変換ステップと、
変換後の前記入力音声の声道形状情報と前記入力音声の音源情報とを用いて合成音を生成することにより、前記入力音声の声質を変換する合成ステップとを含む
声質変換方法。
【請求項１３】
入力音声の声質を変換する際に用いられる、声道の形状を示す声道形状情報を生成する声道情報生成方法であって、
互いに種類が異なる複数の母音の音声を分析することにより、前記母音の種類毎に第１声道形状情報を生成する分析ステップと、
前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する混合ステップとを含む
声道情報生成方法。
【請求項１４】
声道の形状を示す声道形状情報を用いて入力音声の声質を変換する声質変換方法であって、
入力音声に含まれる母音の声道形状情報と、前記入力音声に含まれる母音と同じ種類の母音の第１声道形状情報及び前記入力音声に含まれる母音と異なる種類の母音の第１声道形状情報を混合することにより生成された、前記入力音声に含まれる母音と同じ種類の母音の第２声道形状情報とを混合することにより、前記入力音声の声道形状情報を変換する変換ステップと、
変換後の前記入力音声の声道形状情報と前記入力音声の音源情報とを用いて合成音を生成することにより、前記入力音声の声質を変換する合成ステップとを含む
声質変換方法。
【請求項１５】
請求項１２に記載の声質変換方法をコンピュータに実行させるためのプログラム。
【請求項１６】
請求項１３に記載の声道情報生成方法をコンピュータに実行させるためのプログラム。
【請求項１７】
請求項１４に記載の声質変換方法をコンピュータに実行させるためのプログラム。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、声質変換技術に関する。
【背景技術】
【０００２】
従来の声質変換技術としては、互いに異なる２つの話し方（例えば感情）で発声された同一内容の音声の対を大量に用意し、それらから２つの話し方の間の変換規則を学習する技術がある（例えば、特許文献１参照）。特許文献１に記載の声質変換技術では、学習モデルに基づいて無感情音声から感情音声への変換を行うことができる。
【０００３】
特許文献２記載の声質変換技術では孤立発声された少量の母音から特徴量を抽出することによって目的の音声への変換を実現している。
【先行技術文献】
【特許文献】
【０００４】
【特許文献１】特開平７－７２９００号公報
【特許文献２】国際公開第２００８／１４２８３６号
【発明の概要】
【発明が解決しようとする課題】
【０００５】
しかしながら、上記の声質変換技術では、入力音声を滑らかで自然な音声に変換することができない場合がある。
【０００６】
そこで、本発明は、入力音声を滑らかで自然な音声に変換することができる声質変換システムを提供する。
【課題を解決するための手段】
【０００７】
本発明の一態様に係る声質変換システムは、声道の形状を示す声道形状情報を用いて入力音声の声質を変換する声質変換システムであって、互いに種類が異なる複数の母音の音声を受け付ける母音受付部と、前記母音受付部によって受け付けられた複数の母音の音声を分析することにより、前記母音の種類毎に、第１声道形状情報を生成する分析部と、前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する混合部と、入力音声の声道形状情報及び音源情報を取得し、前記入力音声に含まれる母音の声道形状情報と、前記入力音声に含まれる母音と同じ種類の母音の前記第２声道形状情報とを混合することにより、前記入力音声の声道形状情報を変換し、変換後の前記入力音声の声道形状情報と前記入力音声の音源情報とを用いて合成音を生成することにより、前記入力音声の声質を変換する合成部とを備える。
【０００８】
なお、これらの全般的または具体的な態様は、システム、方法、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（Ｃｏｍｐａｃｔ　Ｄｉｓｃ　Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）などの記録媒体で実現されてもよく、システム、方法、集積回路、コンピュータプログラムおよび記録媒体の任意な組み合わせで実現されても良い。
【発明の効果】
【０００９】
本発明の一態様に係る声質変換システムによれば、入力音声を滑らかで自然な音声に変換することができる。
【図面の簡単な説明】

【００１０】
【図１】図１は、母音のスペクトル包絡の一例を示す模式図である。
【図２Ａ】図２Ａは、孤立母音の第１及び第２フォルマント周波数の分布を示す図である。
【図２Ｂ】図２Ｂは、文中母音の第１及び第２フォルマント周波数の分布を示す図である。
【図３】図３は、人間の声道についての音響管モデルを示す図である。
【図４Ａ】図４Ａは、孤立母音と平均声道形状情報との関係を示す図である。
【図４Ｂ】図４Ｂは、文中母音と平均声道形状情報との関係を示す図である。
【図５Ａ】図５Ａは、孤立母音の第１及び第２フォルマント周波数の平均を示す図である。
【図５Ｂ】図５Ｂは、文中母音の第１及び第２フォルマント周波数の平均を示す図である。
【図６】図６は、文中母音のＦ１－Ｆ２平均、孤立母音のＦ１－Ｆ２平均、及び平均声道形状情報の各々と、複数の文中母音の第１及び第２フォルマント周波数との二乗平均平方根誤差を示す図である。
【図７】図７は、Ｆ１－Ｆ２平面における各孤立母音の位置を平均声道形状情報の位置に向かって移動させたときの効果を説明するための図である。
【図８】図８は、実施の形態１における声質変換システムの構成図である。
【図９】図９は、実施の形態１における分析部の詳細な構成の一例を示す図である。
【図１０】図１０は、実施の形態１における合成部の詳細な構成の一例を示す図である。
【図１１Ａ】図１１Ａは、実施の形態１における声質変換システムの処理動作を示すフローチャートである。
【図１１Ｂ】図１１Ｂは、実施の形態１における声質変換システムの処理動作を示すフローチャートである。
【図１２】図１２は、実施の形態１における声質変換システムの処理動作を示すフローチャートである。
【図１３Ａ】図１３Ａは、日本語の入力音声の声質を変換したときの実験結果を示す図である。
【図１３Ｂ】図１３Ｂは、英語の入力音声の声質を変換したときの実験結果を示す図である。
【図１４】図１４は、Ｆ１－Ｆ２平面に英語の１３母音を配置した図である。
【図１５】図１５は、実施の形態１における母音受付部の一例を示す図である。
【図１６】図１６は、全ての孤立母音の第１及び第２フォルマント周波数を比率ｑで移動させた場合にＦ１－Ｆ２平面上で形成される多角形を示す図である。
【図１７】図１７は、声道長変換比率ｒで声道断面積関数を伸縮する変換方法について説明するための図である。
【図１８】図１８は、声道長変換比率ｒで声道断面積関数を伸縮する変換方法について説明するための図である。
【図１９】図１９は、声道長変換比率ｒで声道断面積関数を伸縮する変換方法について説明するための図である。
【図２０】図２０は、実施の形態２における声質変換システムの構成図である。
【図２１】図２１は、実施の形態２における声道情報生成装置が出力する各母音の音声を説明するための図である。
【図２２】図２２は、実施の形態３における声質変換システムの構成図である。
【図２３】図２３は、他の実施の形態に係る声質変換システムの構成図である。
【図２４】図２４は、特許文献１における声質変換装置の構成図である。
【図２５】図２５は、特許文献２における声質変換装置の構成図である。
【発明を実施するための形態】
【００１１】

（本発明の基礎となった知見）
機器やインタフェースにおいて音声出力機能は、操作方法や機器の状態をユーザーに知らせるなどの重要な役割を担っている。また、情報機器においては、音声出力機能は、ネットワークを介して取得したテキスト情報などを読み上げる機能としても用いられる。
【００１２】
さらに最近では、機器が擬人化されるとともに特徴的な声を出力することが求められる場合も増えている。例えば、人は、人型ロボットに人格を感じるため、人型ロボットが単調な合成音声で話したときには違和感を覚えることが多い。
【００１３】
また、有名人やアニメのキャラクターの声で好きな言葉を喋らせる事ができるサービスが登場している。このようなサービスを提供するためのアプリケーションでは、話す内容以上に声の特徴がニーズの中心となっている。
【００１４】
このように、音声出力機能への要求は、かつての明瞭性あるいは正確性から、声の種類が選べること、あるいは好みの声に変化させられることへと広がっている。
【００１５】
さて、このような音声出力機能を実現する手段としては、人が話した声を録音して再生する録音再生方式と、テキストや発音記号から音声波形を生成する音声合成方式とがある。録音再生方式は、音が良いのが長所であるが、記憶容量が大きくなることと状況に応じて発話させる内容が変えられないこととが短所である。
【００１６】
一方、音声合成方式は、テキストで発話内容を変えることができるので記憶容量の増大は避けられるが、音質やイントネーションの自然さという点において録音再生方式には及ばない。したがって、メッセージの種類が少ない場合は録音再生方式が選ばれ、多い場合は音声合成方式が選ばれることが多い。
【００１７】
ところが、いずれの方式を用いても、声の種類は予め用意した種類に限られる。すなわち、男性と女性など２種類の声を使いたい場合は、両方の声を録音しておくか両方の声の音声合成部を用意する必要があり、機器のコストや開発のコストが増大する。まして、好みの声に調整したり変えたりすることは不可能である。
【００１８】
そこで、声の特徴を別の話者の声の特徴に近似させる声質変換技術の要求が高まっている。
【００１９】
上述したように、従来の声質変換技術としては、互いに異なる２つの話し方（例えば感情）で発声された同一内容の音声の対を大量に用意し、それらから２つの話し方の間の変換規則を学習する技術がある（例えば、特許文献１参照）。
【００２０】
図２４は、特許文献１に記載の声質変換装置の構成図である。
【００２１】
この図に示す声質変換装置は、音響的分析部２００２と、スペクトルのＤＰ（Ｄｙｎａｍｉｃ　Ｐｒｏｇｒａｍｍｉｎｇ）マッチング部２００４と、各音素の時間長伸縮部２００６と、ニューラルネットワーク部２００８とを備える。
【００２２】
ニューラルネットワーク部２００８は、無感情な音声の音響的特徴パラメータを、感情を伴った音声の音響的特徴パラメータに変換するための学習を行う。その後、学習済みの当該ニューラルネットワーク部２００８を用いて無感情な音声に感情が付与される。
【００２３】
スペクトルのＤＰマッチング部２００４は、音響的分析部２００２で抽出された特徴パラメータのうち、スペクトルの特徴パラメータについて、無感情の音声と感情を伴った音

声との間の類似度を時々刻々調べる。そして、スペクトルのＤＰマッチング部２００４は、同一の音素毎の時間的な対応をとることによって、無感情音声に対する感情音声の音素毎の時間的な伸縮率を求める。

【００２４】

各音素の時間長伸縮部２００６は、スペクトルのＤＰマッチング部２００４で得られた音素毎の時間的な伸縮率に応じて、感情音声の特徴パラメータの時系列を時間的に正規化して無感情音声の特徴パラメータの時系列に合うようにする。

【００２５】

ニューラルネットワーク部２００８は、学習時においては、時々刻々と入力層に与えられる無感情音声の音響的特徴パラメータと出力層に与えられる感情音声の音響的特徴パラメータとの違いを学習する。

【００２６】

また、ニューラルネットワーク部２００８は、感情の付与時においては、学習時に決定されたネットワーク内部の重み係数を用いて、時々刻々と入力層に与えられる無感情音声の音響的特徴パラメータから感情音声の音響的特徴パラメータを推定する計算を行なう。以上により、声質変換装置は、学習モデルに基づいて無感情音声から感情音声への変換を行う。

【００２７】

しかしながら、特許文献１の技術では、予め決められた学習用文章と同一の内容の文章の音声を、目標とする感情を伴った発声で収録する必要がある。したがって、話者変換に用いる場合は、目標とする話者（目標話者）に予め決められた学習用文章を全て発話してもらう必要がある。したがって、目標話者に対する負担が大きくなることという課題がある。

【００２８】

そこで、目標話者の発声負担が少なくなる技術として、少量の音声から目標話者の特徴量を抽出して用いる技術が提案されている（例えば、特許文献２参照）。

【００２９】

図２５は、特許文献２に記載の声質変換装置の構成図である。

【００３０】

この図に示す声質変換装置は、入力音声の母音の声道情報を入力された変換比率で目標話者の母音の声道情報に変換することにより、入力音声の声質を変換する。ここで、声質変換装置は、目標母音声道情報保持部２１０１と、変換比率入力部２１０２と、母音変換部２１０３と、子音声道情報保持部２１０４と、子音選択部２１０５と、子音変形部２１０６と、合成部２１０７とを備える。

【００３１】

目標母音声道情報保持部２１０１には、目標話者が発声した代表的な母音から抽出された目標母音声道情報が保持されている。母音変換部２１０３は、入力音声の母音区間の声道情報を、目標母音声道情報を用いて変換する。

【００３２】

この時、母音変換部２１０３は、変換比率入力部２１０２から与えられた変換比率に基づいて、入力音声の母音区間の声道情報と目標母音声道情報とを混合する。子音選択部２１０５は、前後の母音との接続性を考慮して子音声道情報保持部２１０４から子音の声道情報を選択する。そして、子音変形部２１０６は、選択された子音の声道情報を、前後の母音になめらかに繋がるように変形する。合成部２１０７は、入力音声の音源情報と、母音変換部２１０３、子音選択部２１０５及び子音変形部２１０６により変形された声道情報とを用いて、合成音を生成する。

【００３３】

しかしながら、特許文献２の技術では目標音声の声道情報として孤立発声された母音の声道情報を用いているので、変換された音声は滑らかさに欠け、ぎこちない印象となる。これは、別々に発声した母音の特徴と、文として連続して発声された音声中の母音の特徴

との間に違いがあることに起因する。したがって、日常会話などの音声を対象に声質変換を行うと自然性の低下が著しくなる。

【００３４】

以上説明したように、従来の声質変換技術では、少量の目標音声のサンプルを用いて入力音声の声質を変換する場合に、滑らかで自然な音声に変換することができなかった。すなわち、特許文献１の技術では、多量の同一内容の発声音声対から変換規則を学習する必要があるために、目標話者による大量の発声が必要になるという課題があった。一方、特許文献２の技術では、目標話者による母音の音声の入力のみで声質変換が可能であるという利点を有するが、利用できる音声特徴量が孤立発声された母音のものであるために生成される音声の自然性が低いという課題があった。

【００３５】

このような課題を鑑みて、本願発明者らが見出した知見を以下に説明する。

【００３６】

孤立して発声された音声（ｄｉｓｃｒｅｔｅ　ｕｔｔｅｒａｎｃｅ　ｓｐｅｅｃｈ）に含まれる母音は、文章として発声された音声に含まれる母音と異なる特徴を有する。例えば、「あ（ａ）」のみ発声したときの母音である「ａ」は、日本語の「こんにちは／ｋｏ　Ｎ　ｎｉ　ｃｈｉ　ｗａ／」に含まれる文末の「ａ」とは、異なる特徴を有する。また、「え（ｅ）」のみ発声した時の母音である「ｅ」は、英語の「Ｈｅｌｌｏ」に含まれる「ｅ」とは、異なる特徴を有する。

【００３７】

以下、孤立して発声することを「孤立発声」とも表記し、文章として連続して発声することを「連続発声」又は「文発声」とも表記する。また、孤立発声された母音を「孤立母音」とも表記し、文章として連続発声された母音を「文中母音」とも表記する。本願発明者らは、鋭意研究を行った結果、孤立発声の母音と文発声の母音の違いに関する新たな知見を見出した。以下、説明する。

【００３８】

図１は、母音のスペクトル包絡の一例を示す模式図である。図１において、縦軸はパワーを示し、横軸は周波数を示す。図１に示すように、母音のスペクトルは複数のピークを有する。この複数のピークは、声道の共振に対応する。最も小さい周波数のピークは、第１フォルマントと呼ばれる。２番目に小さい周波数のピークは、第２フォルマントと呼ばれる。それぞれのピークの位置に対応する周波数（中心周波数）を、それぞれ第１フォルマント周波数、第２フォルマント周波数と呼ぶ。母音の種類は、主に、第１フォルマント周波数と第２フォルマント周波数との関係で決まる。

【００３９】

図２Ａは、孤立母音の第１及び第２フォルマント周波数の分布を示す。図２Ｂは、文中母音の第１及び第２フォルマント周波数の分布を示す。図２Ａ及び図２Ｂにおいて、横軸は第１フォルマント周波数を示し、縦軸は第２フォルマント周波数を示す。図２Ａ及び図２Ｂに示す第１及び第２フォルマント周波数で定義された２次元平面をＦ１－Ｆ２平面と呼ぶ。

【００４０】

具体的には、図２Ａは、ある話者が日本語の５母音を孤立発声したときの母音の第１及び第２フォルマント周波数を示す。また、図２Ｂは、同じ話者が日本語の文章を連続発声したときの母音の第１及び第２フォルマント周波数を示す。図２Ａ及び図２Ｂにおいて、／ａ／／ｉ／／ｕ／／ｅ／／ｏ／の５つの母音は、異なる記号で示されている。

【００４１】

図２Ａに示すように、５つの孤立母音を結ぶ点線の形状は、五角形となる。また、／ａ／／ｉ／／ｕ／／ｅ／／ｏ／の５つの孤立母音は、Ｆ１－Ｆ２平面において互いに離れて配置される。これは、／ａ／／ｉ／／ｕ／／ｅ／／ｏ／の５つの孤立母音は、それぞれ異なる特徴を有することを意味する。例えば／ａ／と／ｉ／の孤立母音は、／ａ／と／ｏ／の孤立母音よりも大きく離れていることが分かる。

【００４２】
しかし、図２Ｂに示すように、５つの文中母音は、Ｆ１－Ｆ２平面において互いの位置が近付いている。つまり、図２Ｂに示す文中母音の位置は、図２Ａに示す孤立母音の位置よりも五角形の中心又は重心に近付いている。
【００４３】
文中母音では、その母音の前後の音素又は子音との調音が行なわれる。そのために、それぞれの文中母音に発声の怠け（ｒｅｄｕｃｔｉｏｎ　ｏｆ　ａｒｔｉｃｕｌａｔｉｏｎ）が生じる。このため、文章として連続発声されたときの個々の母音は曖昧な発音になる。ただし、文章全体を通して音声は、なめらかで自然に聞こえる。
【００４４】
逆に、孤立母音と同じように、１つ１つの文中母音がはっきりと発音された場合、調音運動が不自然になる。その結果、文章全体を通して音声は、滑らかではなく、ぎこちなく聞こえる。しがって、連続音声を合成する際には、発声の怠けを模擬する音声を用いることが重要である。
【００４５】
発声の怠けを実現するためには、文発声の音声から母音特徴量を抽出しても良い。しかし、そのためには多くの文発声の音声を用意する必要があるため、実用上使いやすさが大きく損なわれる。さらに、文中母音は、前後の音韻の影響を強く受ける。前後の音韻（音韻環境）が近い母音を用いなければ、自然性が損なわれた音声となる。このため、膨大な量の文発声の音声が必要となる。例えば、数十文程度の文発声の音声では、必要十分な量とはならない。
【００４６】
本願発明者らは、（１）少量の音声を用意すれば良いという利便性を活かすために、孤立母音の特徴量を取得し、（２）発声の怠けを模擬するために、Ｆ１－Ｆ２平面において孤立母音によって形成される五角形を縮小する方向に孤立母音の特徴量を移動させるという知見を見出した。この知見に基づく、具体的な方法を説明する。
【００４７】
１つ目の方法は、Ｆ１－Ｆ２平面において各母音を五角形の重心に向けて移動させる方法である。ここで、Ｆ１－Ｆ２平面上での第ｉ母音の位置ベクトルｂを式（１）のように定義する。
【００４８】
【数１】

$$\mathbf{b}_i = \begin{bmatrix} f1_i & f2_i \end{bmatrix} \qquad (1)$$

【００４９】
ここで、$f1_i$は、第ｉ母音の第１フォルマント周波数を示し、$f2_i$は、第ｉ母音の第２フォルマント周波数を示す。ｉは母音の種類を表すインデックスである。５母音の場合は、$1 \leqq i \leqq 5$となる。
【００５０】
重心ｇは、下記の式（２）で表される。
【００５１】
【数２】

$$\mathbf{g} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{b}_i \qquad (2)$$

【００５２】
ここで、Ｎは母音の種類の数である。すなわち、重心ｇは、母音の位置ベクトルの算術平均である。続いて、第ｉ母音の位置ベクトルを下記の式（３）のように変換する。

【００５３】
【数３】

$$\hat{\mathbf{b}}_i = a\mathbf{g} + (1-a)\mathbf{b} \qquad (3)$$

【００５４】
ここで、ａは、０から１の間の値であり、母音の位置ベクトルｂを重心ｇに近付ける度合いを表す曖昧化度合い係数である。曖昧化度合い係数ａが１に近いほど、全ての母音は重心ｇに近付く。その結果、母音の位置ベクトルｂの違いも小さくなる。言い換えれば、図２Ａに示すＦ１－Ｆ２平面上において、各母音の音響的特徴が曖昧になる。
【００５５】
上記の考え方により、母音の曖昧化ができる。しかし、フォルマント周波数を直接変更することは、問題がある。図２Ａには、第１フォルマント周波数と第２フォルマント周波数のみが示されている。しかし、孤立母音と文中母音とでは、第１及び第２フォルマント周波数だけではなく、他の物理量も異なっている。他の物理量は、例えば、第２フォルマント周波数よりも高次のフォルマント周波数又は各フォルマントのバンド幅などである。したがって、例えば、母音の第２フォルマント周波数のみをより高い周波数に変化させたとき、第２フォルマント周波数が第３フォルマント周波数に接近しすぎることが考えられる。
【００５６】
その結果、スペクトル包絡において異常に鋭いピークが現れ、合成フィルタが発振する、又は合成音の振幅が異常に大きくなる可能性がある。このような場合は、正常な音声を合成することができない。
【００５７】
音声の声質を変換する場合、音声の特徴を表す複数のパラメータがバランスを保った状態で変化しなければ、変換後の音声が妥当ではない音となってしまう。したがって、第１フォルマント周波数及び第２フォルマント周波数の２つのパラメータだけを変化させた場合、複数のパラメータのバランスが崩れ、著しく音質が劣化する。
【００５８】
この課題を解決するために、本願発明者らは、フォルマント周波数を直接変更するのではなく、声道形状を変形させることで母音を曖昧化する方法を見出した。
【００５９】
（声道断面積関数）
声道形状を示す情報（以下、「声道形状情報」という）としては、例えば、声道断面積関数がある。図３は、人間の声道についての音響管モデルを示す。人間の声道とは、声帯から口唇までの空間である。
【００６０】
図３の（ａ）において、縦軸は断面積の大きさを示し、横軸は音響管のセクション番号を示す。ここで、音響管のセクション番号とは、声道の中の位置を示す。横軸の左端は、口唇（Ｌｉｐ）の位置に対応し、横軸の右端は、声門（ｇｌｏｔｔｉｓ）の位置に対応する。
【００６１】
図３の（ａ）に示す音響管モデルは、複数の円形の音響管が縦続接続されている。声道の断面積を、各セクションの音響管の断面積として、声道形状を模擬している。ここで、声道の長さ方向の位置と、その位置に対応する断面積の大きさとの関係を声道断面積関数と呼ぶ。
【００６２】
声道の断面積は、ＬＰＣ分析に基づくＰＡＲＣＯＲ係数と一意に対応することが知られている。下記の式（４）により、ＰＡＲＣＯＲ係数を、声道の断面積に変換できる。以下、ＰＡＲＣＯＲ係数k_iを、声道形状情報の一例として説明する。ただし、声道形状情報

は、ＰＡＲＣＯＲ係数に限定されるものではなく、ＰＡＲＣＯＲ係数に等価なＬＳＰ（Ｌｉｎｅ　Ｓｐｅｃｔｒｕｍ　Ｐａｉｒｓ）やＬＰＣなどであっても良い。また、上述の音響管モデルにおける音響管の間の反射係数とＰＡＲＣＯＲ係数とは、符号が反転していることが違うだけである。このため、声道形状情報として反射係数が用いられても良い。

【００６３】

【数４】

$$\frac{A_i}{A_{i+1}} = \frac{1-k_i}{1+k_i} \qquad (4)$$

【００６４】

ここで、A_iは、図３の（ｂ）に示す第ｉ区間の音響管の断面積であり、k_iは、第ｉ番目と第ｉ＋１番目との境界のＰＡＲＣＯＲ係数（反射係数）である。

【００６５】

ＰＡＲＣＯＲ係数は、ＬＰＣ分析により分析された線形予測係数α_iを用いて算出することができる。具体的には、ＰＡＲＣＯＲ係数は、Ｌｅｖｉｎｓｏｎ－Ｄｕｒｂｉｎ－Ｉｔａｋｕｒａアルゴリズムを用いることにより算出される。なお、ＰＡＲＣＯＲ係数は次の特徴を有する。

・線形予測係数は分析次数ｐに依存するが、ＰＡＲＣＯＲ係数は分析の次数に依存しない。

・低次係数の値の変動はスペクトルへの影響が大きく、高次になるにつれて値の変動がスペクトルに与える影響が小さくなる。

・高次係数の値の変動のスペクトルへの影響は全周波数帯域に渡って平坦なものである。

【００６６】

なお、声道形状情報は、必ずしも声道の断面積を示す情報である必要はなく、声道の各セクションの容積を示す情報であっても良い。

【００６７】

（声道形状の変形）

次に、声道形状の変形について説明する。上述のように、声道の形状は、式（４）に示すＰＡＲＣＯＲ係数から求められる。ここでは、声道形状を変形するために、複数の声道形状情報を混合する。具体的には、複数の声道断面積関数の加重平均を求める代わりに、複数のＰＡＲＣＯＲ係数ベクトルの加重平均を求める。第ｉ母音のＰＡＲＣＯＲ係数ベクトルは、式（５）で表される。

【００６８】

【数５】

$$\mathbf{k}_i = \begin{pmatrix} k_1^i & k_2^i & \cdots & k_M^i \end{pmatrix} \qquad (5)$$

【００６９】

複数の母音のＰＡＲＣＯＲ係数ベクトルの加重平均は式（６）で表される。

【００７０】

【数６】

$$\bar{\mathbf{k}} = \sum_i w_i \mathbf{k}_i \qquad \sum_i w_i = 1 \qquad (6)$$

【００７１】

ここでw_iは、重み係数である。混合したい母音の声道形状情報が２つの場合、重み係

数は、２つの声道形状情報の混合比に対応する。

【００７２】

（声道形状情報の曖昧化）

次に、母音を曖昧化するために複数の母音の声道形状情報を混合する手順を説明する。

【００７３】

まず、Ｎ個の種類の母音の平均声道形状情報を式（７）で求める。つまり、各母音の声道形状情報が示す値（ここではＰＡＲＣＯＲ係数）の算術平均を算出することにより、平均声道形状情報を生成する。

【００７４】

【数７】

$$\bar{\mathbf{k}} = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{k}_i \qquad (7)$$

【００７５】

次に、第ｉ母音の曖昧化度合い係数ａを用いて、第ｉ母音の声道形状情報を曖昧化後の声道形状情報に変換する。すなわち、各母音の声道形状情報が示す値を平均声道形状情報が示す値に近付けることにより、曖昧化後の各母音の声道形状情報を生成する。つまり、第ｉ母音の声道形状情報と他の母音の声道形状情報とを混合して、曖昧化後の声道形状情報を生成する。

【００７６】

【数８】

$$\hat{\mathbf{k}}_i = a\bar{\mathbf{k}} + (1-a)\mathbf{k}_i \qquad (8)$$

$\mathbf{k}_i$：曖昧化前の母音の声道形状情報、$\hat{\mathbf{k}}_i$：曖昧化後の母音の声道形状情報

【００７７】

このようにして生成された曖昧化後の母音の声道形状情報を用いて音声の合成を行うことで、音質を劣化させず、発声の怠けを再現することができる。

【００７８】

以下に、実際に実験を行った結果について説明する。

【００７９】

図４Ａは、孤立母音と平均声道形状情報との関係を示す。また、図４Ｂは、文中母音と平均声道形状情報との関係を示す。図４Ａ及び図４Ｂにおいて、平均声道形状情報は、式（７）に従って、図２Ａに示す孤立母音の情報を用いて求めたなお、図４Ａ及び図４Ｂに示す星印は、平均声道形状情報を用いて合成された母音の第１及び第２フォルマント周波数を示す。

【００８０】

図４Ａにおいて、平均声道形状情報は、５つの母音によって形成される五角形の重心近傍に位置する。図４Ｂにおいて、平均声道形状情報は、文中母音が分布する領域の中心近傍に位置する。

【００８１】

図５Ａは、孤立母音（図２Ａに示す１５個の母音）の第１及び第２フォルマント周波数の平均を示す。また、図５Ｂは、文中母音（図２Ｂに示す９５個の母音）の第１及び第２フォルマント周波数の平均を示す。なお、以下において、第１及び第２フォルマント周波数の平均をＦ１－Ｆ２平均とも呼ぶ。

【００８２】

図５Ａ及び図５Ｂにおいて、第１フォルマント周波数及び第２フォルマント周波数の平均は、破線で示されている。また、図５Ａ及び図５Ｂには、図４Ａ及び図４Ｂに示した平均声道形状情報も星印で示されている。

【００８３】
式（７）を用いて求めた図４Ａに示す平均声道形状情報の位置は、図５Ａに示す孤立母音のＦ１－Ｆ２平均の位置よりも、図５Ｂに示す文中母音のＦ１－Ｆ２の平均の位置に近い。したがって、式（７）及び式（８）を用いて求めた平均声道形状情報は、孤立母音のＦ１－Ｆ２の平均よりも、実際の発声の怠けに近似している。以下に、具体的な座標値を用いて説明する。
【００８４】
図６は、文中母音のＦ１－Ｆ２平均、孤立母音のＦ１－Ｆ２平均、及び平均声道形状情報の各々と、複数の文中母音の第１及び第２フォルマント周波数との二乗平均平方根誤差（ＲＭＳＥ：ｒｏｏｔ　ｍｅａｎ　ｓｑｕａｒｅ　ｅｒｒｏｒ）を示す。
【００８５】
図６に示すように、平均声道形状情報のＲＭＳＥは、孤立母音のＦ１－Ｆ２平均のＲＭＳＥよりも、文中母音のＦ１－Ｆ２平均のＲＭＳＥに近い。ただし、ＲＭＳＥが近いことだけが、音声の自然さに貢献するとは言えないが、発声の怠けの近似度合いを表す指標として見ることはできる。
【００８６】
次に、図７は、式（８）を用いて、Ｆ１－Ｆ２平面における各孤立母音の位置を平均声道形状情報の位置に向かって移動させたときの効果を説明するための図である。図７において、大きい白丸はａ＝０の場合の各母音の位置、小さい白丸はａ＝１の場合の各母音の位置すなわち平均声道形状における位置を表し、黒い点はａを０．１刻みで大きくしていった場合の各母音の位置を表している。全ての母音が孤立母音の位置から平均声道形状における母音の位置に向かって連続的に移動している。このように、声道形状情報を混合して声道形状を変形することにより、第１及び第２フォルマント周波数は平均化及び曖昧化が可能であることが分かった。
【００８７】
そこで、本発明の一態様に係る声質変換システムは、声道の形状を示す声道形状情報を用いて入力音声の声質を変換する声質変換システムであって、互いに種類が異なる複数の母音の音声を受け付ける母音受付部と、前記母音受付部によって受け付けられた複数の母音の音声を分析することにより、前記母音の種類毎に、第１声道形状情報を生成する分析部と、前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する混合部と、入力音声の声道形状情報及び音源情報を取得し、前記入力音声に含まれる母音の声道形状情報と、前記入力音声に含まれる母音と同じ種類の母音の前記第２声道形状情報とを混合することにより、前記入力音声の声道形状情報を変換し、変換後の前記入力音声の声道形状情報と前記入力音声の音源情報とを用いて合成音を生成することにより、前記入力音声の声質を変換する合成部とを備える。
【００８８】
この構成によれば、母音の種類毎に、複数の第１声道形状情報を混合して第２声道形状情報を生成することができる。つまり、少量の音声のサンプルから母音の種類毎に第２声道形状情報を生成することができる。このように母音の種類毎に生成された第２声道形状情報は、曖昧化された母音の声道形状情報に相当する。したがって、第２声道形状情報を用いて入力音声の声質を変換することにより、入力音声を滑らかで自然な音声に変換することが可能となる。
【００８９】
また例えば、前記混合部は、前記母音の種類毎に生成された複数の第１声道形状情報を平均することにより、１つの平均声道形状情報を算出する平均声道情報算出部と、前記母音受付部によって受け付けられた母音の種類毎に、当該母音の第１声道形状情報と前記平均声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する混合声道情報生成部とを備えても良い。
【００９０】

この構成によれば、第２声道形状情報を平均声道形状情報に容易に近付けることが可能となる。

【００９１】

また例えば、平均声道情報算出部は、前記複数の第１声道形状情報を重み付き算術平均することにより、前記平均声道形状情報を算出しても良い。

【００９２】

この構成によれば、複数の第１声道形状情報の重み付き算術平均を平均声道形状情報として算出することができる。したがって、例えば、目標話者の発声の怠けの特徴に応じて第１声道形状情報に重み付けすることにより、入力音声をより滑らかで自然な目標話者の音声に変換することも可能となる。

【００９３】

また例えば、前記混合部は、前記入力音声に含まれる母音の局所的発話速度が大きいほど、前記入力音声に含まれる母音と同じ種類の母音の前記第２声道形状情報が前記母音の種類毎に生成された複数の第１声道形状情報の平均に近付くように、前記第２声道形状情報を生成しても良い。

【００９４】

この構成によれば、入力音声に含まれる母音の局所的発話速度に応じて複数の第１声道形状情報の混合比率を設定することができる。文中母音の曖昧化度合いは、局所的発話速度に依存する。したがって、入力音声をより滑らかで自然な音声に変換することが可能となる。

【００９５】

また例えば、前記混合部は、母音の種類に応じて設定された混合比率を用いて、前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合しても良い。

【００９６】

この構成によれば、母音の種類に応じて、複数の第１声道形状情報の混合比率を設定することができる。文中母音の曖昧化度合いは、母音の種類に依存する。したがって、入力音声をより滑らかで自然な音声に変換することが可能となる。

【００９７】

また例えば、前記混合部は、ユーザーによって設定された混合比率を用いて、前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合しても良い。

【００９８】

この構成によれば、複数の母音の曖昧化度合いを、ユーザーの好みにあわせて設定することができる。

【００９９】

また例えば、前記混合部は、前記入力音声の言語種類に応じて設定された混合比率を用いて、前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合しても良い。

【０１００】

この構成によれば、入力音声の言語種類に応じて、複数の第１声道形状情報の混合比率を設定することができる。文中母音の曖昧化度合いは、入力音声の言語種類に依存する。したがって、各言語にふさわしい曖昧化度合いを設定することができる。

【０１０１】

また例えば、前記声質変換システムは、さらに、前記入力音声の声道形状情報及び音源情報が記憶されている入力音声記憶部を備え、前記合成部は、前記入力音声記憶部から、前記入力音声の声道形状情報及び音源情報を取得しても良い。

【０１０２】

本発明の一態様に係る声道情報生成装置は、入力音声の声質を変換する際に用いられる、声道の形状を示す声道形状情報を生成する声道情報生成装置であって、互いに種類が異

なる複数の母音の音声を分析することにより、前記母音の種類毎に、第１声道形状情報を生成する分析部と、前記母音の種類毎に、当該母音の前記第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する混合部とを備える。

【０１０３】

この構成によれば、母音の種類毎に、複数の第１声道形状情報を混合して第２声道形状情報を生成することができる。つまり、少量の音声のサンプルから母音の種類毎に第２声道形状情報を生成することができる。このように母音の種類毎に生成された第２声道形状情報は、曖昧化された母音の声道形状情報に相当する。したがって、第２声道形状情報が声質変換装置に出力されれば、声質変換装置は、第２声道形状情報を用いて入力音声を滑らかで自然な音声に変換することができる。

【０１０４】

また例えば、さらに、前記母音の種類毎に、前記第２声道形状情報を用いて合成音を生成する合成部と、前記合成音を音声として出力する出力部とを備えても良い。

【０１０５】

この構成によれば、母音の種類毎に第２声道形状情報を用いて生成された合成音を音声として出力することができる。したがって、従来の声質変換装置を用いて、入力音声を滑らかで自然な音声に変換することができる。

【０１０６】

本発明の一態様に係る声質変換装置は、声道の形状を示す声道形状情報を用いて入力音声の声質を変換する声質変換装置であって、母音の種類毎に、当該母音の第１声道形状情報と、当該母音と異なる種類の母音の前記第１声道形状情報とを混合することにより生成された第２声道形状情報を記憶している母音声道情報記憶部と、入力音声に含まれる母音の声道形状情報と、前記入力音声に含まれる母音と同じ種類の母音の前記第２声道形状情報とを混合することにより、前記入力音声の声道形状情報を変換し、変換後の前記入力音声の声道形状情報と前記入力音声の音源情報とを用いて合成音を生成することにより、前記入力音声の声質を変換する合成部とを備える。

【０１０７】

この構成によれば、上記声質変換システムと同様の効果を奏することができる。

【０１０８】

なお、これらの全般的または具体的な態様は、方法、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、方法、集積回路、コンピュータプログラムおよび記録媒体の任意な組み合わせで実現されても良い。

【０１０９】

以下本発明の実施の形態について、図面を参照しながら説明する。

【０１１０】

なお、以下で説明する実施の形態は、いずれも本発明の一具体例を示す。以下の実施の形態で示される数値、形状、材料、構成要素、構成要素の配置位置及び接続形態、ステップ、ステップの順序などは、一例であり、本発明を限定する主旨ではない。また、以下の実施の形態における構成要素のうち、最上位概念を示す独立請求項に記載されていない構成要素については、任意の構成要素として説明される。

【０１１１】

（実施の形態１）

図８は、実施の形態１における声質変換システム１００の構成図である。

【０１１２】

声質変換システム１００は、声道の形状を示す声道形状情報を用いて入力音声の声質を変換する。図８に示すように、声質変換システム１００は、入力音声記憶部１０１と、母音受付部１０２と、分析部１０３と、第１母音声道情報記憶部１０４と、混合部１０５と、第２母音声道情報記憶部１０７、合成部１０８と、出力部１０９と、混合比率入力部１

１０と、変換比率入力部１１１とを備える。それぞれの構成要素は、有線又は無線で接続されており、互いに情報を送受信する。以下、各構成要素について、説明する。

【０１１３】

（入力音声記憶部１０１）

入力音声記憶部１０１は、入力音声情報と、入力音声情報と対応付けられた付属情報とを記憶している。入力音声情報とは、変換対象となる入力音声に関する情報である。具体的には、入力音声情報は、複数の音素で構成される音声の情報である。例えば、ある歌手が歌った音声等を予め録音しておくことにより、入力音声情報が準備される。より具体的には、入力音声記憶部１０１は、入力音声情報を声道情報と音源情報とに分離した形式で記憶している。

【０１１４】

付属情報は、入力音声において音素の境界を示す時間の情報と、音素の種類の情報とを含む。

【０１１５】

（母音受付部１０２）

母音受付部１０２は、母音の音声を受け付ける。本実施の形態では、母音受付部１０２は、入力音声と同じ言語の母音の音声であって、互いに種類が異なる複数の母音の音声を受け付ける。互いに種類が異なる複数の母音の音声とは、複数の異なる種類の母音を含んでいれば良く、同じ種類の複数の母音を含んでも良い。

【０１１６】

母音受付部１０２は、分析部１０３に、母音の音声に対応する電気信号である母音の音響信号を送信する。

【０１１７】

母音受付部１０２は、例えば、話者が発した音声を受け付ける場合は、マイクロホンを有する。母音受付部１０２は、例えば、予め電気信号に変換されている音響信号を受け付ける場合、オーディオ回路及びアナログデジタル変換器を有する。母音受付部１０２は、例えば、予め音響信号がデジタルデータに変換された音響データを受け付ける場合、データ読出器を有する。

【０１１８】

なお、母音受付部１０２は、表示部を備えても良い。表示部は、目標話者に発声させたい単母音又は文章と、発声タイミングとを表示する。

【０１１９】

また、母音受付部１０２が受け付ける音声は、孤立発声された母音であっても良い。例えば、母音受付部１０２は、代表的な母音の音響信号を受け付けても良い。代表的な母音は、言語により異なる。例えば、日本語の代表的な母音とは、／ａ／／ｉ／／ｕ／／ｅ／／ｏ／の５種類の母音である。英語の代表的な母音は、以下に国際音声記号（ＩＰＡ：Ｉｎｔｅｒｎａｔｉｏｎａｌ　Ｐｈｏｎｅｔｉｃ　Ａｌｐｈａｂｅｔ）で示す１３種類の母音である。

【０１２０】

【数９】

$$[i][u][ɪ][ʊ][e][o][ə][ɛ][ʌ][ɔ][æ][ɑ][ɒ]$$

【０１２１】

母音受付部１０２は、例えば日本語の母音の音声を受け付ける場合は、／ａ／／ｉ／／ｕ／／ｅ／／ｏ／の５種類の母音を目標話者に孤立発声（すなわち各母音の間を開けて発声）させることで、母音の音声を受け付ける。このように話者に母音を孤立発声してもらうことにより、分析部１０３は、パワー情報を用いて母音区間を切り出すことが可能となる。

【０１２２】

ただし、母音受付部１０２は、必ずしも孤立発声された母音の音声を受け付ける必要はない。母音受付部１０２は、文章として連続発声された母音を受け付けても良い。例えば話者が緊張していて意識的にはっきりとした発声が行われた場合は、文章として連続発声された母音も、孤立発声された母音に近い音声になることがある。母音受付部１０２が文発声の母音を受け付ける場合は、例えば５母音を含む文章（例えば「本日は晴天なり」など）を話者に発声させれば良い。この場合、分析部１０３は、ＨＭＭ（Ｈｉｄｄｅｎ－Ｍａｒｋｏｖ－Ｍｏｄｅｌ）などを用いた音素自動セグメンテーション技術によって母音区間を切り出すことができる。

【０１２３】

（分析部１０３）

分析部１０３は、母音受付部１０２から母音の音響信号を受け付ける。分析部１０３は、母音受付部１０２で受け付けられた母音の音響信号に対して、付属情報を付与する。さらに、分析部１０３は、例えばＬＰＣ（Ｌｉｎｅａｒ　Ｐｒｅｄｉｃｔｉｖｅ　Ｃｏｄｉｎｇ）分析やＡＲＸ（Ａｕｔｏ－ｒｅｇｒｅｓｓｉｖｅ　Ｅｘｏｇｅｎｏｕｓ）分析などの分析方法を用いて各母音の音響信号を分析することにより、各母音の音響信号を声道情報と音源情報とに分離する。

【０１２４】

声道情報には、母音が発声された時の声道の形状を示す声道形状情報が含まれる。分析部１０３によって分離された声道情報に含まれる声道形状情報を第１声道形状情報と呼ぶ。つまり、分析部１０３は、母音受付部１０２によって受け付けられた複数の母音の音声を分析することにより、母音の種類毎に、第１声道形状情報を生成する。

【０１２５】

第１声道形状情報の例としては、上述のＬＰＣの他に、ＰＡＲＣＯＲ係数、ＰＡＲＣＯＲ係数と等価なＬＳＰ（Ｌｉｎｅ　Ｓｐｅｃｔｒｕｍ　Ｐａｉｒｓ）などがある。また、音響管モデルにおける音響管の間の反射係数とＰＡＲＣＯＲ係数との関係は、符号が反転していることのみである。このため、反射係数そのものを第１声道形状情報として用いても良い。

【０１２６】

付属情報は、各母音の種類（／ａ／／ｉ／など）と、母音区間中心の時刻とを含む。分析部１０３は、第１母音声道情報記憶部１０４に、母音の種類毎に、少なくとも母音の第１声道形状情報を格納する。

【０１２７】

次に、母音の第１声道形状情報の生成方法の一例を説明する。

【０１２８】

図９は、実施の形態１における分析部１０３の詳細な構成の一例を示す。分析部１０３は、母音安定区間抽出部１０３１と母音声道情報作成部１０３２とを備える。

【０１２９】

母音安定区間抽出部１０３１は、入力された母音を含む音声から孤立母音の区間（母音区間）を抽出することにより母音区間中心の時刻を算出する。母音区間の抽出方法は特に限定される必要はない。例えば、母音安定区間抽出部１０３１は、パワーが一定以上の区間を安定区間とし、当該安定区間を母音区間として抽出するようにしても良い。

【０１３０】

母音声道情報作成部１０３２は、母音安定区間抽出部１０３１により抽出された孤立母音の母音区間中心に対して、母音の声道形状情報を作成する。例えば、母音声道情報作成部１０３２は、上述のＰＡＲＣＯＲ係数を第１声道形状情報として算出する。母音声道情報作成部１０３２は、第１母音声道情報記憶部１０４に、母音の第１声道形状情報を格納する。

【０１３１】

（第１母音声道情報記憶部１０４）

第１母音声道情報記憶部１０４は、母音の種類毎に、少なくとも母音の第１声道形状情

報を記憶している。つまり、第１母音声道情報記憶部１０４は、分析部１０３によって母音の種類毎に生成された複数の第１声道形状情報を記憶している。

【０１３２】

（混合部１０５）

混合部１０５は、母音の種類毎に、当該母音の第１声道形状情報と、当該母音と異なる種類の母音の第１声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する。具体的には、混合部１０５は、母音の種類毎に、当該母音の第２声道形状情報が当該母音の第１声道形状情報よりも平均声道形状情報に近付くように、当該母音の第２声道形状情報を生成する。このように生成される第２声道形状情報は、曖昧化された声道形状情報に相当する。

【０１３３】

なお、平均声道形状情報とは、母音の種類毎に生成された複数の第１声道形状情報の平均である。また、複数の声道形状情報を混合するとは、複数の声道形状情報の各々が示す値又はベクトルを重み付け加算することを意味する。

【０１３４】

ここで、混合部１０５の詳細な構成の一例を説明する。混合部１０５は、例えば、平均声道情報算出部１０５１と混合声道情報生成部１０５２とを備える。

【０１３５】

（平均声道情報算出部１０５１）

平均声道情報算出部１０５１は、第１母音声道情報記憶部１０４に記憶された複数の第１声道形状情報を取得する。平均声道情報算出部１０５１は、取得した複数の第１声道形状情報を平均することにより、１つの平均声道形状情報を算出する。具体的な処理については、後述する。平均声道情報算出部１０５１は、混合声道情報生成部１０５２に平均声道形状情報を送信する。

【０１３６】

（混合声道情報生成部１０５２）

混合声道情報生成部１０５２は、平均声道情報算出部１０５１から平均声道形状情報を受信する。また、混合声道情報生成部１０５２は、第１母音声道情報記憶部１０４に記憶された複数の第１声道形状情報を取得する。

【０１３７】

そして、混合声道情報生成部１０５２は、母音受付部１０２によって受け付けられた母音の種類毎に、当該母音の第１声道形状情報と平均声道形状情報とを混合することにより、当該母音の第２声道形状情報を生成する。具体的には、混合声道情報生成部１０５２は、母音の種類毎に、第１声道形状情報を平均声道形状情報に近付ける処理を行うことにより、第２声道形状情報を生成する。

【０１３８】

第１声道形状情報と平均声道形状情報との混合比率は、母音の曖昧化度合いに応じて設定されれば良い。本実施の形態では、混合比率は、式（８）における曖昧化度合い係数ａに相当する。つまり、混合比率は、値が大きいほど曖昧化度合いが高くなる。混合声道情報生成部１０５２は、混合比率入力部１１０から入力された混合比率を用いて、第１声道形状情報と平均声道形状情報とを混合する。

【０１３９】

なお、混合声道情報生成部１０５２は、予め記憶されている混合比率を用いて、第１声道形状情報と平均声道形状情報とを混合しても良い。この場合、声質変換システム１００は、必ずしも混合比率入力部１１０を備える必要はない。

【０１４０】

ある種類の母音の第２声道形状情報を平均声道形状情報に近付けた場合、その種類の母音の第２声道形状情報は、他の種類の母音の第２声道形状情報に近付く。すなわち、第２声道形状情報が平均声道形状情報により近付くように混合比率が設定されれば、混合声道情報生成部１０５２は、より曖昧化された第２声道形状情報を生成することができる。こ

のようなより曖昧化された第２声道形状情報を用いて生成された合成音は、滑舌が悪い音声となる。例えば、幼児の声に入力音声の声質を変換するときには、このように第２声道形状情報が平均声道形状情報に近付くように混合比率が設定されることが有効である。

【０１４１】

また、第２声道形状情報を平均声道形状情報にあまり近付けない場合、第２声道形状情報は、孤立母音の声道形状情報に近くなる。例えば、口を大きく開けてはっきり調音する傾向にある歌声に入力音声の声質を変換するときには、このように第２声道形状情報が平均声道形状情報にあまり近付かないように混合比率が設定されることが適している。

【０１４２】

混合声道情報生成部１０５２は、第２母音声道情報記憶部１０７に、母音の種類毎の第２声道形状情報を格納する。

【０１４３】

（第２母音声道情報記憶部１０７）

第２母音声道情報記憶部１０７は、母音の種類別に、第２声道形状情報を記憶している。つまり、第２母音声道情報記憶部１０７は、混合部１０５によって母音の種類毎に生成された複数の第２声道形状情報を記憶している。

【０１４４】

（合成部１０８）

合成部１０８は、入力音声記憶部１０１に記憶されている入力音声情報を取得する。また、合成部１０８は、第２母音声道情報記憶部１０７に記憶されている母音の種類毎の第２声道形状情報を取得する。

【０１４５】

そして、合成部１０８は、入力音声情報に含まれる母音の声道形状情報と、入力音声情報に含まれる母音と同じ種類の母音の第２声道形状情報とを混合することにより、入力音声の声道形状情報を変換する。その後、合成部１０８は、入力音声の変換後の声道形状情報と、入力音声記憶部１０１に記憶されている入力音声の音源情報とを用いて合成音を生成することにより、入力音声の声質を変換する。

【０１４６】

具体的には、合成部１０８は、変換比率入力部１１１から入力された変換比率を混合比率として用いて、入力音声情報に含まれる母音の声道形状情報と、当該母音と同じ種類の母音の第２声道形状情報とを混合する。この変換比率は、入力音声を変化させる度合いに応じて設定されれば良い。

【０１４７】

なお、合成部１０８は、予め記憶されている変換比率を用いて、入力音声情報に含まれる母音の声道形状情報と当該母音と同じ種類の母音の第２声道形状情報とを混合しても良い。この場合、声質変換システム１００は、必ずしも変換比率入力部１１１を備える必要はない。

【０１４８】

合成部１０８は、このように生成された合成音の信号を出力部１０９に送信する。

【０１４９】

ここで、合成部１０８の詳細な構成の一例を説明する。なお、以下に説明する合成部１０８の詳細な構成は、特許文献２と同様の構成である。

【０１５０】

図１０は、実施の形態１における合成部１０８の詳細な構成の一例を示す。合成部１０８は、母音変換部１０８１と、子音選択部１０８２と、子音声道情報記憶部１０８３と、子音変形部１０８４と、音声合成部１０８５とを備える。

【０１５１】

母音変換部１０８１は、入力音声記憶部１０１から、音素境界付き声道情報と音源情報とを取得する。

【０１５２】

音素境界付き声道情報とは、入力音声の声道情報に、入力音声に対応する音素情報と各音素の時間長の情報とが付された情報である。母音変換部１０８１は、母音区間ごとに該当する母音の第２声道形状情報を第２母音声道情報記憶部１０７から読み出す。そして、母音変換部１０８１は、母音区間の声道形状情報と読み出した第２声道形状情報とを混合することにより、入力音声の母音部の声質変換を行なう。この時の変換度合いは、変換比率入力部１１１から入力された変換比率に基づく。

【０１５３】

子音選択部１０８２は、前後の母音との接続性を考慮して子音声道情報記憶部１０８３から子音の声道情報を選択する。そして、子音変形部１０８４は、選択された子音の声道情報を、前後の母音になめらかに繋がるように変形する。音声合成部１０８５は、入力音声の音源情報と、母音変換部１０８１、子音選択部１０８２及び子音変形部１０８４により変形された声道情報とを用いて、合成音を生成する。

【０１５４】

このように、特許文献２における目標母音声道情報を第２声道形状情報に置き換えて声質変換が実行される。

【０１５５】

（出力部１０９）

出力部１０９は、合成部１０８から合成音信号を受信する。出力部１０９は、合成音信号を合成音として出力する。出力部１０９は、例えば、スピーカで構成される。

【０１５６】

（混合比率入力部１１０）

混合比率入力部１１０は、混合声道情報生成部１０５２で用いる混合比率を受け付ける。混合比率入力部１１０は、混合声道情報生成部１０５２に、受け付けた混合比率を送信する。

【０１５７】

（変換比率入力部１１１）

変換比率入力部１１１は、合成部１０８で用いる変換比率を受け付ける。変換比率入力部１１１は、合成部１０８に、受け付けた変換比率を送信する。

【０１５８】

次に、以上のように構成された声質変換システム１００における各種動作について説明する。

【０１５９】

図１１Ａ、図１１Ｂ及び図１２は、実施の形態１における声質変換システム１００の処理動作を示すフローチャートである。

【０１６０】

具体的には、図１１Ａは、声質変換システム１００において母音の音声を受け付けてから第２声道形状情報を生成するまでの処理の流れを示す。また、図１１Ｂは、図１１Ａに示す第２声道形状情報生成処理（Ｓ６００）の詳細を示す。また、図１２は、実施の形態１において入力音声の声質を変換する処理の流れを示す。

【０１６１】

（ステップＳ１００）

母音受付部１０２は、目標話者が発声した母音が含まれる音声を受け付ける。母音が含まれる音声とは、例えば、日本語の場合、日本語の５母音を「アー、イー、ウー、エー、オー」と発声したときの音声である。各母音の間隔は、５００ｍｓ程度であれば良い。

【０１６２】

（ステップＳ２００）

分析部１０３は、母音受付部１０２が受け付けた音声に含まれる１つの母音の声道形状情報を第１声道形状情報として生成する。

【０１６３】

（ステップＳ３００）

分析部１０３は、生成された第１声道形状情報を、第１母音声道情報記憶部１０４に格納する。

【０１６４】

（ステップＳ４００）

分析部１０３は、母音受付部１０２が受け付けた音声に含まれる全ての種類の母音について、第１声道形状情報が生成されたか否かを判定する。例えば、分析部１０３は、母音受付部１０２が受け付けた音声に含まれる母音の種類情報を取得する。さらに、分析部１０３は、取得した母音の種類情報を参照して、音声に含まれる全ての種類の母音の第１声道形状情報が第１母音声道情報記憶部１０４に記憶されているか否かを判定する。ここで、全ての種類の母音の第１声道形状情報が第１母音声道情報記憶部１０４に記憶されている場合に、分析部１０３は、完了と判断する。一方、いずれかの種類の母音の第１声道形状情報が記憶されていない場合には、分析部１０３は、ステップＳ２００の処理を行う。

【０１６５】

（ステップＳ５００）

平均声道情報算出部１０５１は、第１母音声道情報記憶部１０４に記憶されている全ての種類の母音の第１声道形状情報を用いて、１つの平均声道形状情報を算出する。

【０１６６】

（ステップＳ６００）

混合声道情報生成部１０５２は、ステップＳ１００で受け付けられた音声に含まれる母音の種類毎に、平均声道形状情報と、第１母音声道情報記憶部１０４に記憶されている第１声道形状情報とを用いて、第２声道形状情報を生成する。

【０１６７】

ここで、図１１Ｂを用いて、ステップＳ６００の詳細を説明する。

【０１６８】

（ステップＳ６０１）

混合声道情報生成部１０５２は、第１母音声道情報記憶部１０４に記憶されている１つの母音の第１声道形状情報に平均声道形状情報を混合することによって、当該母音の第２声道形状情報を生成する。

【０１６９】

（ステップＳ６０２）

混合声道情報生成部１０５２は、第２母音声道情報記憶部１０７に、ステップＳ６０１で生成された第２声道形状情報を格納する。

【０１７０】

（ステップＳ６０３）

混合声道情報生成部１０５２は、ステップＳ１００で受け付けられた音声に含まれる全ての種類の母音について、ステップＳ６０２の処理が行われたか否かを判定する。例えば、混合声道情報生成部１０５２は、母音受付部１０２が受け付けた音声に含まれる母音の種類情報を取得する。そして、混合声道情報生成部１０５２は、取得した母音の種類情報を参照して、音声に含まれる全ての種類の母音の第２声道形状情報が第２母音声道情報記憶部１０７に記憶されているか否かを判定する。

【０１７１】

ここで、全ての種類の母音の第２声道形状情報が第２母音声道情報記憶部１０７に記憶されている場合に、混合声道情報生成部１０５２は、完了と判断する。一方、いずれかの種類の母音の第２声道形状情報が第２母音声道情報記憶部１０７に記憶されていない場合には、混合声道情報生成部１０５２は、ステップＳ６０１の処理を行う。

【０１７２】

次に、このように母音の種類毎に生成された第２声道形状情報を用いて入力音声の声質を変換する処理について図１２を用いて説明する。

【０１７３】

（ステップＳ８００）

合成部１０８は、第２母音声道情報記憶部１０７に記憶されている第２声道形状情報を用いて、入力音声記憶部１０１に記憶されている入力音声の声道形状情報を変換する。具体的には、合成部１０８は、入力音声に含まれる母音の声道形状情報を、入力音声に含まれる母音と同じ種類の母音の第２声道形状情報と混合することにより、入力音声の声道形状情報を変換する。

【０１７４】

（ステップＳ９００）

合成部１０８は、ステップＳ８００で変換された入力音声の声道形状情報と、入力音声記憶部１０１に記憶されている入力音声の音源情報とを用いて、合成音を生成する。これにより、入力音声の声質が変換された合成音が生成される。つまり、声質変換システム１００は、入力音声の特徴を変化させることができる。

【０１７５】

（実験結果）

次に、実際に入力音声の声質を変換する実験を行って効果を確認した結果について説明する。図１３Ａは、日本語の入力音声の声質を変換したときの実験結果を示す。ここでは、入力音声は、ある女性話者によって文発声された音声である。また、目標話者は、入力音声を発声した女性話者とは別の女性話者である。図１３Ａには、その目標話者が孤立発声した母音に基づいて入力音声の声質が変換された結果が示されている。

【０１７６】

図１３Ａの（ａ）は、従来技術で声質変換されたスペクトログラムを示す。図１３Ａの（ｂ）は、本実施の形態における声質変換システム１００により声質変換されたスペクトログラムを示す。本実験では、式（８）における曖昧化度合い係数ａ（混合比率）として、「０．３」を用いた。

【０１７７】

また、発話内容は、日本語の「ねえご隠居さん、昔から鶴は千年、亀は万年なんてことを言いますね」（／ｎｅ ｅ ｇｏ ｉ Ｎ ｋｙｏ ｓａ Ｎ, ｍｕ ｋａ ｓｈｉ ｋａ ｒａ, ｔｓｕ ｒｕ ｗａ ｓｅ Ｎ ｎｅ Ｎ, ｋａ ｍｅ ｗａ ｍａ Ｎ ｎｅ Ｎ ｎａ Ｎ ｔｅ ｋｏ ｔｏ ｏ ｉ ｉ ｍａ ｓｕ ｎｅ／、“Ｈｉ ｄａｄｄｙ. Ｔｈｅｙ ｓａｙ ｃｒａｎｅ ｌｉｖｅｓ ｌｏｎｇｅｒ ｔｈａｎ ａ ｔｈｏｕｓａｎｄ ｙｅａｒｓ, ａｎｄ ｔｏｒｔｏｉｓｅ ｌｉｖｅｓ ｌｏｎｇｅｒ ｔｈａｎ ｔｅｎ ｔｈｏｕｓａｎｄ ｙｅａｒｓ, ｄｏｎ’ｔ ｔｈｅｙ？”）である。

【０１７８】

図１３Ａの（ｂ）では、（ａ）と比べて、全体に時間方向のフォルマント軌跡が滑らかになっており、連続発声として自然性が改善している。特に、図１３Ａにおいて白線で囲んだ部分は、（ａ）と（ｂ）との間で顕著な違いが見られる。

【０１７９】

図１３Ｂは、英語の入力音声の声質を変換したときの実験結果を示す。具体的には、図１３Ｂの（ａ）は、従来技術で声質変換されたスペクトログラムを示す。図１３Ｂの（ｂ）は、本実施の形態における声質変換システム１００によって声質変換されたスペクトログラムを示す。

【０１８０】

図１３Ｂにおいて、入力音声の話者と目標話者とは、図１３Ａと同様である。また、曖昧化度合い係数ａも図１３Ａと同様である。

【０１８１】

発話内容は、英語の“Ｗｏｒｋ ｈａｒｄ ｔｏｄａｙ.”である。なお、英語の発話内容は、カタカナの「ワークハードトゥデイ」と言う文字列に置き換えられ、日本語の音素で合成音が生成されている。

【０１８２】

声質変換後の音声の韻律（すなわちイントネーションパターン）は、入力音声の韻律と

同じになるため、日本語の音素を用いて声質変換しても、声質変換後の音声はある程度英語らしく聞こえる。しかし、英語の母音は日本語に比べて数が多いため、日本語の代表的な母音だけでは、英語の母音を表現できないという問題がある。

【０１８３】

そこで、本実施の形態に示す技術で母音を曖昧化することによって、日本語らしさを低減し、結果として英語音声として自然さを増すことが可能となる。特に、以下にＩＰＡで示す、曖昧母音であるｓｃｈｗａは、日本語の５母音とは全く異なり、Ｆ１－Ｆ２平面において日本語の５母音によって形成される五角形の重心付近に位置するために、本実施の形態による曖昧化の効果が大きい。

【０１８４】

【数１０】

$$[\text{ə}]$$

【０１８５】

図１３Ｂにおいて、特に白線で囲んだ部分は、（ａ）と（ｂ）との間で顕著な違いが見られる。時刻１．２秒においては、第１及び第２フォルマント周波数だけではなく、第３フォルマント周波数にも違いが現れていることが分かる。実際に合成音を聞いた印象では、（ａ）はカタカナをそのまま話しているような感じであり、（ｂ）は英語として受け入れやすい感じであった。また、（ａ）は英語を話すときに力を入れて調音している感じであり、（ｂ）はリラックスして話している感じであった。

【０１８６】

ところで、発声の怠けは発話速度によって変化する。ゆっくり発話するときには、各母音は、孤立母音と同様に正確に調音される。この特徴は、歌を歌う場合などに顕著に現れる。入力音声が歌声の場合は、声質変換システム１００は、孤立母音をそのまま用いて声質変換を行っても、違和感のない合成音を生成することが可能である。

【０１８７】

一方で、会話調の話し方で早く発話するときには、顎や舌などの調音器官の動きが発話速度に追いつかないために、発声の怠けが大きくなる。そこで、曖昧化度合い（混合比率）は、当該音韻周辺の局所的発話速度に応じて設定されても良い。つまり、混合部１０５は、入力音声に含まれる母音の局所的発話速度が大きいほど、入力音声に含まれる母音と同じ種類の母音の第２声道形状情報が平均声道形状情報に近付くように、第２声道形状情報を生成しても良い。これにより、入力音声をより滑らかで自然な音声に変換することが可能となる。

【０１８８】

具体的には、式（８）の曖昧化度合い係数ａ（混合比率）は、例えば次の式（９）ように局所的発話速度ｒ（単位は１秒あたりの音素数など）の関数として設定されれば良い。

【０１８９】

【数１１】

$$a = a_0 + h(r - r_0) \qquad (9)$$

【０１９０】

ここで、a_0は基準の曖昧化度合いを表す値であり、r_0は基準の発話速度（単位はｒと同じ）である。また、ｈは所定の値であり、ｒによってａを変化させる感度である。

【０１９１】

なお、文中母音は、Ｆ１－Ｆ２平面において、孤立母音よりも多角形の内側に移動するが、その度合いは母音によって異なる。例えば図４Ａ及び図４Ｂにおいて、／ｏ／は比較的変化が少ないが、／ａ／は少数の外れ値を除いて大きく内側に移動している。また、／ｉ／も多くが特定の方向に移動しているが、／ｕ／は移動する方向もまちまちである。

【０１９２】

そこで、母音によって曖昧化度合い（混合比率）を変えることも有効と考えられる。つまり、混合部１０５は、母音の種類に応じて設定された混合比率を用いて、母音の種類毎に、当該母音の第１声道形状情報と、当該母音と異なる種類の母音の第１声道形状情報とを混合しても良い。この場合、／ｏ／の曖昧化度合いを小さく、／ａ／の曖昧化度合いを大きくしても良い。また／ｉ／は曖昧化度合いを大きく、／ｕ／はどちらの方向に移動させれば良いか分からないために曖昧化度合いを小さくしても良い。これらの傾向は個人によって異なる可能性があるので、目標話者が誰であるかによって曖昧化度合いを変えても良い。

【０１９３】

もちろん、曖昧化度合いは、ユーザーの好みによって変えられても構わない。この場合、ユーザーは、混合比率入力部１１０を介して、母音の種類毎に、好みの曖昧化度合いを示す混合比率を入力すれば良い。つまり、混合部１０５は、ユーザーによって設定された混合比率を用いて、母音の種類毎に、当該母音の第１声道形状情報と、当該母音と異なる種類の母音の第１声道形状情報とを混合しても良い。

【０１９４】

また、平均声道情報算出部１０５１は、式（７）に示すように、複数の第１声道形状情報の算術平均（相加平均）を算出することにより、平均声道形状情報を算出したが、必ずしも式（７）のように平均声道形状情報を算出する必要はない。例えば、平均声道情報算出部１０５１は、式（６）の重み係数w_iを不均一にして、平均声道形状情報を算出しても良い。

【０１９５】

つまり、平均声道形状情報は、互いに種類が異なる複数の母音の第１声道形状情報の重み付き算術平均であっても構わない。例えば、個人ごとに発声の怠けの特徴を調べて、その個人の発声の怠けを近似するように重み係数の調整を行なうことは効果的である。例えば、目標話者の発声の怠けの特徴に応じて第１声道形状情報に重み付けすることにより、入力音声をより滑らかで自然な目標話者の音声に変換することも可能となる。

【０１９６】

また、平均声道情報算出部１０５１は、式（７）のような相加平均ではなく、相乗平均や調和平均を平均声道形状情報として算出しても構わない。具体的には、式（１０）のようにＰＡＲＣＯＲ係数の平均ベクトルを表すと、平均声道情報算出部１０５１は、式（１１）のように、複数の母音の第１声道形状情報の相乗平均を平均声道形状情報として算出しても良い。また、平均声道情報算出部１０５１は、式（１２）のように、複数の母音の第１声道形状情報の調和平均を平均声道形状情報として算出しても良い。

【０１９７】

【数１２】

$$\bar{\mathbf{k}} = \left(\bar{k}_1 \quad \bar{k}_2 \quad \cdots \quad \bar{k}_M \right) \qquad (10)$$

【０１９８】

【数１３】

$$\bar{k}_m = \sqrt[N]{\prod_{i=1}^{N} k_m^i} = \sqrt[N]{k_m^1 k_m^2 \cdots k_m^N} \qquad (11)$$

【０１９９】

【数１４】

$$\bar{k}_m = \frac{N}{\sum_{i=1}^{N} \frac{1}{k_m^i}} = \frac{N}{\frac{1}{k_m^1} + \frac{1}{k_m^2} + \cdots + \frac{1}{k_m^N}} \qquad (12)$$

【０２００】

要するに、複数の母音の第１声道形状情報の平均は、各母音の第１声道形状情報と混合されたときに、Ｆ１－Ｆ２平面における母音の分布範囲が縮小されるように算出されれば良い。

【０２０１】

例えば日本語の５母音／ａ／、／ｉ／、／ｕ／、／ｅ／、／ｏ／の場合、式（７）や式（１１）、式（１２）のような平均声道形状を求めることは必ずしも必要ではない。例えば、ある母音と別の母音を混合することによってその母音を５角形の重心に近づける操作が行なわれても良い。例えば母音／ａ／のあいまい化を行う場合、／ａ／とは別の種類の母音を少なくとも２つ選び、選ばれた２つの母音を用いて所定の重みで混合を行っても良い。Ｆ１－Ｆ２平面上で５母音が形成する五角形が凸五角形（全ての内角の大きさが二直角より小さい五角形）であれば、／ａ／と他の任意の２つの母音を混合して作られた母音は必ずこの５角形の内側に位置する。多くの場合、日本語の５母音が形成する五角形は凸五角形であり、この方法によって母音を曖昧化できる。

【０２０２】

また、上述したように英語には日本語よりも母音の数が多いため、Ｆ１－Ｆ２平面において母音間の距離が小さい傾向にある。この傾向は言語によって異なるので、曖昧化度合い係数は、言語に応じて設定されることが望ましい。つまり、混合部１０５は、入力音声の言語種類に応じて定められた混合比率を用いて、母音の種類毎に、当該母音の第１声道形状情報と、当該母音と異なる種類の母音の第１声道形状情報とを混合しても良い。これにより、各言語にふさわしい曖昧化度合いを設定することができ、入力音声をより滑らかで自然な音声に変換することが可能となる。

【０２０３】

英語の母音種類は日本語よりも多いため、Ｆ１－Ｆ２平面での多角形は日本語の多角形よりも複雑である。図１４は、Ｆ１－Ｆ２平面に英語の１３母音を配置した図である。なお、図１４は、「Ghonim, A., Smith, J. and Wolfe, J. (2007) “The sounds of world English”, http://www.phys.unsw.edu.au/swe」から引用した。英語では母音のみを発声することは難しいので、［ｈ］と［ｄ］で挟まれた仮想的な単語で母音が表されている。１３母音を全て加算平均して求めた平均声道形状と各母音を混合した場合、各母音が重心に近づく方向に移動するため曖昧化される。

【０２０４】

しかし、日本語の場合に述べたように、全ての母音を用いて平均声道形状を求めることは必ずしも必要ではない。図１４の配置を用いると、“ｈｅｅｄ”、“ｈａｉｒｅｄ”、“ｈａｄ”、“ｈａｒｄ”、“ｈｏｄ”、“ｈｏｗｄ”、“ｗｈｏｄ”を用いて凸多角形を構成することができる。この多角形の辺に近い母音は日本語と同様に、当該母音をそれとは別の少なくとも２母音を選び混合することで曖昧化が可能である。一方、多角形の内部に位置する母音（図では“ｈｅａｒｄ”）については、それらがもともと曖昧な音であるためにそのまま利用する。

【０２０５】

このように、本実施の形態における声質変換システム１００によれば、少量の母音を入力するだけで滑らかな文発声の音声を生成することができる。さらに、日本語母音を用いて英語の音声を生成することができるなど、飛躍的に柔軟な声質変換が可能になる。

【０２０６】

つまり、本実施の形態における声質変換システム１００によれば、母音の種類毎に、複数の第１声道形状情報を混合して第２声道形状情報を生成することができる。つまり、少量の音声のサンプルから母音の種類毎に第２声道形状情報を生成することができる。このように母音の種類毎に生成された第２声道形状情報は、曖昧化された母音の声道形状情報に相当する。したがって、第２声道形状情報を用いて入力音声の声質を変換することにより、入力音声を滑らかで自然な音声に変換することが可能となる。

【０２０７】

なお、母音受付部１０２は、前述したとおり典型的にはマイクロホンを有するが、さらに、ユーザーに発声内容とタイミングとを指示するための表示装置（ｐｒｏｍｐｔｅｒ）を有することが望ましい。具体例としては、図１５に示すように、母音受付部１０２は、マイクロホン１０２１と、マイクロホン１０２１の近傍に配置された液晶ディスプレイなどの表示部１０２２とから構成されても良い。この場合、表示部１０２２は、目標話者に発声させる内容１０２３（この場合は母音）とタイミング１０２４とを表示すれば良い。

【０２０８】

なお、本実施の形態では、混合部１０５は、平均声道形状情報を算出していたが、必ずしも平均声道形状情報を算出する必要はない。例えば、混合部１０５は、母音の種類毎に、当該母音の第１声道形状情報と、当該母音とは異なる種類の母音の声道形状情報とを所定の混合比率で混合することにより、当該母音の第２声道形状情報を生成すれば良い。このとき、所定の混合比率は、第２声道形状情報が第１声道形状情報よりも平均声道形状情報に近付くように設定されれば良い。

【０２０９】

つまり、混合部１０５は、Ｆ１－Ｆ２平面上で母音間の距離が近付くように第２声道形状情報が生成されれば、どのように複数の第１声道形状情報が混合されても構わない。例えば、混合部１０５は、入力音声においてある母音から別の母音に遷移する時に声道形状情報が急峻に変わらないように第２声道形状情報を生成しても良い。つまり、混合部１０５は、入力音声に含まれる母音の並びに適応して混合比率を変化させながら、入力音声に含まれる母音と同じ種類の母音の第１声道形状情報と、入力音声に含まれる母音と異なる種類の母音の第１声道形状情報とを混合しも良い。その結果、第２声道形状情報から得られる母音のＦ１－Ｆ２平面における位置は、同じ種類の母音であっても、多角形領域内で動くことになる。これは、ＰＡＲＣＯＲ係数の時系列を移動平均法などにより平滑化することで実現可能である。

【０２１０】

（実施の形態１の変形例）

次に、実施の形態１の変形例について説明する。

【０２１１】

実施の形態１では、母音受付部１０２は、当該言語における代表的な全ての種類の母音（日本語では５母音）を受け付けていたが、本変形例では、母音受付部１０２は、必ずしも全ての種類の母音を受け付ける必要はない。本変形例では、実施の形態１よりも少ない種類の母音で声質変換を実現する。以下、その方法について説明する。

【０２１２】

母音の種類は第１フォルマント周波数と第２フォルマント周波数とで特徴付けられるが、それらの値は個人によって異なっている。それでも、同一の母音と知覚される理由を説明するモデルとして、第１フォルマント周波数と第２フォルマント周波数との比によって母音が特徴付けられるとみなしたモデルがある。ここで、第ｉ母音の第１フォルマント周波数$ｆ１_{i}$及び第２フォルマント周波数$ｆ２_{i}$からなるベクトル$ｖ_{i}$を式（１３）で表すとし、第１フォルマント周波数と第２フォルマント周波数との比を保ったままベクトルｖｉを移動したベクトル$ｖ_{i}$’を式（１４）で表すとする。

【０２１３】

【数１５】

$$\mathbf{v}_i = [f1_i \quad f2_i] \qquad (13)$$

【０２１４】

【数１６】

$$\mathbf{v}_i' = q\mathbf{v}_i = q[f1_i \quad f2_i] = [qf1_i \quad qf2_i] \qquad (14)$$

【０２１５】

ｑはベクトルｖ$_i$とベクトルｖ$_i$’との比率である。上述のモデルに基づけば、比率ｑの値を変化させてもベクトルｖ$_i$とベクトルｖ$_i$’とは同じ母音として知覚される。

【０２１６】

このように、全ての孤立母音の第１及び第２フォルマント周波数を比率ｑで移動した場合、Ｆ１－Ｆ２平面上で母音の第１及び第２フォルマント周波数によって形成される多角形は、図１６に示すように互いに相似となる。図１６では、元の多角形Ａと、ｑ＞１の時の多角形Ｂと、ｑ＜１の時の多角形Ｃ及びＤとが表されている。

【０２１７】

このように第１フォルマント周波数ｆ１$_i$と第２フォルマント周波数ｆ２$_i$との比を保ったまま声道形状を変形する方法としては、声道の長さを変更するという方法がある。声道長を１／ｑ倍にすれば、全てのフォルマントの周波数がｑ倍になる。そこで、まず声道長変換比率ｒ＝１／ｑを求め、次に声道長変換比率ｒで声道断面積関数を伸縮するような変換を行なう。

【０２１８】

まず、声道長変換比率ｑを求める方法について説明する。

【０２１９】

ＰＡＲＣＯＲ係数は、分析次数が十分高ければ高次の係数になるに従って絶対値が小さくなる傾向にある。特に、声帯の位置に相当するセクション番号以上の次数では小さな値が続く。そこで、高次の係数から順に低い次数へと値を検査し、絶対値がある閾値を超えたところを声帯位置とみなし、その次数ｋを記憶しておく。この方法により、あらかじめ用意された母音から取り出したｋをｋａ、入力された母音から取り出したｋをｋｂとすれば、声道長変換比率ｒは、式（１５）のように計算することができる。

【０２２０】

【数１７】

$$r = \frac{kb}{ka} \qquad (15)$$

【０２２１】

次に、声道長変換比率ｒで声道断面積関数を伸縮する変換方法について説明する。

【０２２２】

図１７は、ある母音の声道断面積関数を示す。横軸は、口唇から声帯へ向かっての距離をセクション番号で表す。縦軸は、声道断面積を表す。破線は、声道断面積をスプライン関数などにより内挿して連続値にしたものである。

【０２２３】

連続値になった声道断面積関数を新たなセクション間隔１／ｒでサンプリングし（図１８）、サンプリングされた値を元のセクション間隔で配置しなおす（図１９）。図１９の例では、声道末端部分（声帯側）に余剰セクションが生まれるが（図１９の網掛け部分）、余剰セクションの部分は一定の断面積にしておく。これは、声道長を超えるセクションではＰＡＲＣＯＲ係数の絶対値が非常に小さい値になるからである。つまり、ＰＡＲＣＯ

Ｒ係数の符号を反転したものはセクション間の反射係数であり、反射係数が０であると言うことはセクション間の断面積に差がないことを意味するからである。

【０２２４】

上記の例では、声道長を短くする場合（ｒ＜１）の変換方法を示した。一方、声道長を長くする場合（ｒ＞１）は、声道末端部分（声帯側）には収まりきらないセクションが生まれるが、これらのセクションの値は捨てる。捨てるＰＡＲＣＯＲ係数の絶対値が小さくなるように、元々の分析次数を高めにとっておくと良い。例えばサンプリング周波数１０ｋＨｚの音声に対して通常のＰＡＲＣＯＲ分析では次数を１０前後にするが、２０などの高い値にしておけば良い。

【０２２５】

このような方法で、入力された単一の母音と、あらかじめ用意された母音から、全ての母音の声道形状情報を推定することが可能である。つまり、母音受付部１０２は、全ての種類の母音を受け付ける必要がなくなる。

【０２２６】

（実施の形態２）

次に、実施の形態２について説明する。

【０２２７】

本実施の形態では、声質変換システムが２つの装置によって構成される点が、実施の形態１における声質変換システムと異なる。以下において、実施の形態１と異なる点を中心に説明する。

【０２２８】

図２０は、実施の形態２における声質変換システム２００の構成図である。図２０において、図８と同じ機能を有する構成要素については同じ符号を用い、適宜説明を省略する。

【０２２９】

図２０に示すように、声質変換システム２００は、声道情報生成装置２０１と声質変換装置２０２とを備える。

【０２３０】

声道情報生成装置２０１は、入力音声の声質を変換する際に用いられる、声道の形状を示す第２声道形状情報を生成する。声道情報生成装置２０１は、母音受付部１０２と、分析部１０３と、第１母音声道情報記憶部１０４と、混合部１０５と、混合比率入力部１１０と、第２母音声道情報記憶部１０７と、合成部１０８ａと、出力部１０９とを備える。

【０２３１】

合成部１０８ａは、母音の種類毎に、第２母音声道情報記憶部１０７に記憶されている第２声道形状情報を用いて合成音を生成する。そして、合成部１０８ａは、生成した合成音の信号を出力部１０９に送信する。声道情報生成装置２０１の出力部１０９は、母音の種類毎に生成された合成音の信号を音声として出力する。

【０２３２】

図２１は、実施の形態２における声道情報生成装置２０１が出力する母音の音声を説明するための図である。図２１では、声道情報生成装置２０１の母音受付部１０２によって受け付けられる複数の母音の音声によりＦ１－Ｆ２平面に形成される五角形を実線で表わす。また、声道情報生成装置２０１の出力部１０９によって母音の種類毎に出力される音声によりＦ１－Ｆ２平面に形成される五角形を破線で表わす。

【０２３３】

図２１から明らかなように、声道情報生成装置２０１の出力部１０９は、曖昧化された母音の音声を出力する。

【０２３４】

声質変換装置２０２は、声道形状情報を用いて入力音声の声質を変換する。声質変換装置２０２は、母音受付部１０２と、分析部１０３と、第１母音声道情報記憶部１０４と、入力音声記憶部１０１と、合成部１０８ｂと、変換比率入力部１１１と、出力部１０９と

を備える。この声質変換装置２０２は、図２５に示す特許文献２の声質変換装置と同様の構成である。

【０２３５】

合成部１０８ｂは、第１母音声道情報記憶部１０４に記憶されている第１声道形状情報を用いて、入力音声の声質を変換する。ただし、本実施の形態では、声質変換装置２０２の母音受付部１０２は、声道情報生成装置２０１によって曖昧化された母音の音声を受け付けている。つまり、声質変換装置２０２の第１母音声道情報記憶部１０４に記憶されている第１声道形状情報は、実施の形態１における第２声道形状情報に相当する。したがって、声質変換装置２０２の出力部１０９は、実施の形態１と同様の音声を出力する。

【０２３６】

以上のように、本実施の形態における声質変換システム２００によれば、声道情報生成装置２０１と声質変換装置２０２との２つの装置によって構成することができる。そして、声質変換装置２０２は、従来の声質変換装置と同様の構成にすることができる。つまり、本実施の形態における声質変換システム２００によれば、実施の形態１と同様の効果を、従来の声質変換装置を用いて実現することが可能となる。

【０２３７】

（実施の形態３）

次に、実施の形態３について説明する。

【０２３８】

本実施の形態では、声質変換システムが２つの装置によって構成される点が、実施の形態１における声質変換システムと異なる。以下において、実施の形態１と異なる点を中心に説明する。

【０２３９】

図２２は、実施の形態３における声質変換システム３００の構成図である。図２２において、図８と同じ機能を有する構成要素については同じ符号を用い、適宜説明を省略する。

【０２４０】

図２２に示すように、声質変換システム３００は、声道情報生成装置３０１と声質変換装置３０２とを備える。

【０２４１】

声道情報生成装置３０１は、第１母音声道情報記憶部１０４と、混合部１０５と、混合比率入力部１１０とを備える。声質変換装置３０２は、入力音声記憶部１０１と、母音受付部１０２と、分析部１０３と、合成部１０８と、出力部１０９と、変換比率入力部１１１と、母音声道情報記憶部３０３と、母音声道情報入出力切替部３０４とを備える。

【０２４２】

母音声道情報入出力切替部３０４は、第１のモード又は第２のモードで動作する。具体的には、母音声道情報入出力切替部３０４は、第１のモードでは、母音声道情報記憶部３０３に記憶されている第１声道形状情報を第１母音声道情報記憶部１０４に出力する。一方、母音声道情報入出力切替部３０４は、第２のモードでは、混合部１０５から出力された第２声道形状情報を、母音声道情報記憶部３０３に格納する。

【０２４３】

母音声道情報記憶部３０３には、第１声道形状情報及び第２声道形状情報が格納される。つまり、母音声道情報記憶部３０３は、実施の形態１における第１母音声道情報記憶部１０４及び第２母音声道情報記憶部１０７に相当する。

【０２４４】

以上、本実施の形態における声質変換システムによれば、母音を曖昧化する機能を有する声道情報生成装置３０１を独立した装置として構成することができる。そして、声道情報生成装置３０１は、マイクロホンなどが不要であるので、コンピュータソフトウェアとして実現することができる。したがって、声道情報生成装置３０１は、声質変換装置３０２の性能を高めるために後付けするソフトウェア（いわゆるプラグイン）として提供する

ことができる。

【０２４５】

また、声道情報生成装置３０１は、サーバーアプリケーションとして実現することもできる。この場合、声道情報生成装置３０１は、ネットワークを介して声質変換装置３０２と接続されれば良い。

【０２４６】

以上、本発明の一態様に係る声質変換システム、声質変換装置、及び声道情報生成装置について、実施の形態に基づいて説明したが、本発明は、これらの実施の形態に限定されるものではない。本発明の趣旨を逸脱しない限り、当業者が思いつく各種変形を本実施の形態に施したもの、あるいは異なる実施の形態における構成要素を組み合わせて構築される形態も、本発明の範囲内に含まれる。

【０２４７】

例えば、上記実施の形態１～３において、声質変換システムは、複数の構成要素を備えていたが、必ずしもそれらの構成要素のすべてを備える必要はない。例えば、声質変換システムは、図２３に示すように構成されても良い。

【０２４８】

図２３は、他の実施の形態に係る声質変換システム４００の構成図である。なお、図２３において、図８と同様の構成要素については、同一の符号を付し、適宜説明を省略する。

【０２４９】

図２３に示す声質変換システム４００は、声道情報生成装置４０１と声質変換装置４０２とを備える。なお、図２３において、図８と同様の構成要素については、同一の符号を付し、説明を省略する。

【０２５０】

図２３に示す声質変換システム４００は、分析部１０３及び混合部１０５を有する声道情報生成装置４０１と、第２母音声道情報記憶部１０７及び合成部１０８を有する声質変換装置４０２とを備える。なお、声質変換システム４００は、必ずしも第２母音声道情報記憶部１０７を備える必要はない。

【０２５１】

声質変換システム４００は、このように構成されても、曖昧化された声道形状情報である第２声道形状情報を用いて入力音声の声質を変換することができるので、実施の形態１における声質変換システム１００と同様の効果を奏することができる。

【０２５２】

また、上記各実施の形態における声質変換システム、声質変換装置、又は声道情報生成装置が備える構成要素の一部又は全部は、１個のシステムＬＳＩ（Ｌａｒｇｅ　Ｓｃａｌｅ　Ｉｎｔｅｇｒａｔｉｏｎ：大規模集積回路）から構成されているとしても良い。

【０２５３】

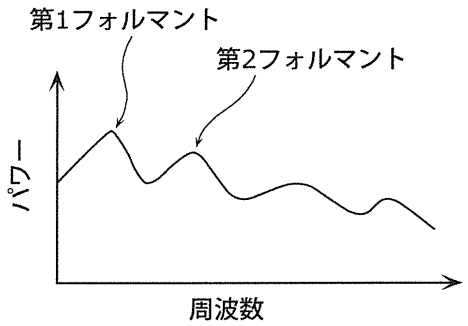
システムＬＳＩは、複数の構成要素を１個のチップ上に集積して製造された超多機能ＬＳＩであり、具体的には、マイクロプロセッサ、ＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）、ＲＡＭ（Ｒａｎｄａｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）などを含んで構成されるコンピュータシステムである。前記ＲＯＭには、コンピュータプログラムが記憶されている。前記マイクロプロセッサが、前記コンピュータプログラムに従って動作することにより、システムＬＳＩは、その機能を達成する。

【０２５４】

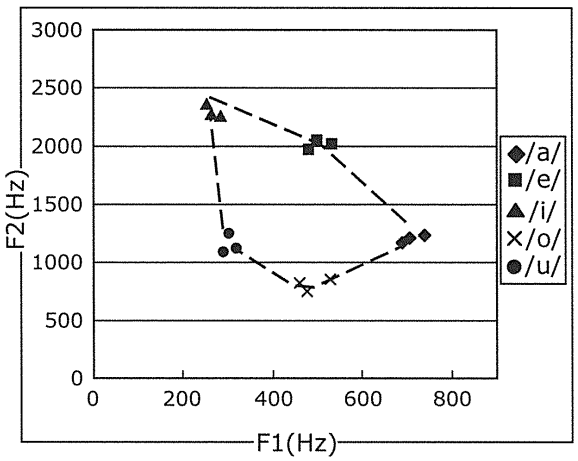
なお、ここでは、システムＬＳＩとしたが、集積度の違いにより、ＩＣ、ＬＳＩ、スーパーＬＳＩ、ウルトラＬＳＩと呼称されることもある。また、集積回路化の手法はＬＳＩに限るものではなく、専用回路又は汎用プロセッサで実現しても良い。ＬＳＩ製造後に、プログラムすることが可能なＦＰＧＡ（Ｆｉｅｌｄ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｇａｔｅ　Ａｒｒａｙ）、あるいはＬＳＩ内部の回路セルの接続や設定を再構成可能なリコンフィギュラブル・プロセッサを利用しても良い。

【０２５５】
さらには、半導体技術の進歩又は派生する別技術によりＬＳＩに置き換わる集積回路化の技術が登場すれば、当然、その技術を用いて機能ブロックの集積化を行っても良い。バイオ技術の適用等が可能性としてありえる。
【０２５６】
また、本発明の一態様は、このような特徴的な構成要素を備える声質変換システム、声質変換装置、又は声道情報生成装置だけでなく、声質変換システム、声質変換装置、又は声道情報生成装置に含まれる特徴的な処理部をステップとする声質変換方法又は声道情報生成方法であっても良い。また、本発明の一態様は、声質変換方法又は声道情報生成方法に含まれる特徴的な各ステップをコンピュータに実行させるコンピュータプログラムであっても良い。そして、そのようなコンピュータプログラムを、ＣＤ－ＲＯＭ等のコンピュータ読取可能な非一時的な記録媒体あるいはインターネット等の通信ネットワークを介して流通させても良い。
【産業上の利用可能性】
【０２５７】
本発明の一態様に係る声質変換システムは、音声加工ツール、ゲーム、家電製品等の音声ガイド、ロボットの音声出力等として有用である。また、ある人の声を別の人の声に変換する用途ではなくとも、テキスト音声合成の出力を滑らかで聞きやすい印象にするための用途にも応用できる。
【符号の説明】
【０２５８】
１００、２００、３００、４００　声質変換システム
１０１　入力音声記憶部
１０２　母音受付部
１０３　分析部
１０４　第１母音声道情報記憶部
１０５　混合部
１０７　第２母音声道情報記憶部
１０８、１０８ａ、１０８ｂ　合成部
１０９　出力部
１１０　混合比率入力部
１１１　変換比率入力部
２０１、３０１、４０１　声道情報生成装置
２０２、３０２、４０２　声質変換装置
３０３　母音声道情報記憶部
３０４　母音声道情報入出力切替部
１０２１　マイクロホン
１０２２　表示部
１０３１　母音安定区間抽出部
１０３２　母音声道情報作成部
１０５１　平均声道情報算出部
１０５２　混合声道情報生成部
１０８１　母音変換部
１０８２　子音選択部
１０８３　声道情報記憶部
１０８４　子音変形部
１０８５　音声合成部

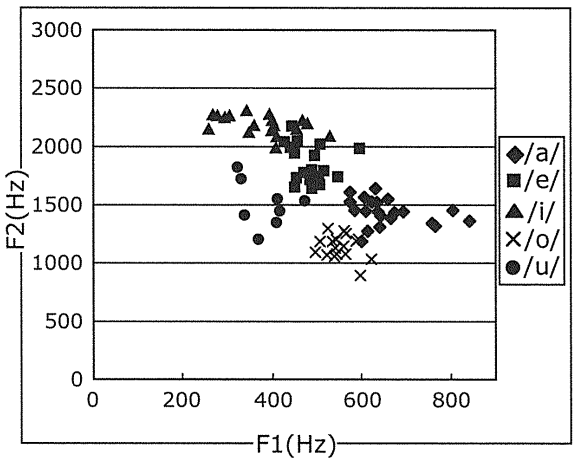
【図１】


第1フォルマント
第2フォルマント
パワー
周波数

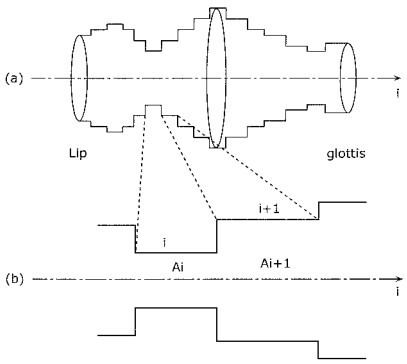
【図２Ａ】


3000
2500
2000
1500
1000
500
0
F2(Hz)
0
200
400
600
800
F1(Hz)
◆/a/
■/e/
▲/i/
×/o/
●/u/

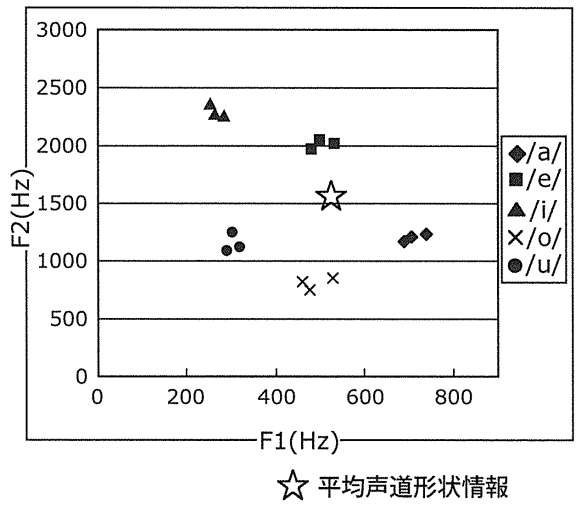
【図２Ｂ】


3000
2500
2000
1500
1000
500
0
F2(Hz)
0
200
400
600
800
F1(Hz)
◆/a/
■/e/
▲/i/
×/o/
●/u/

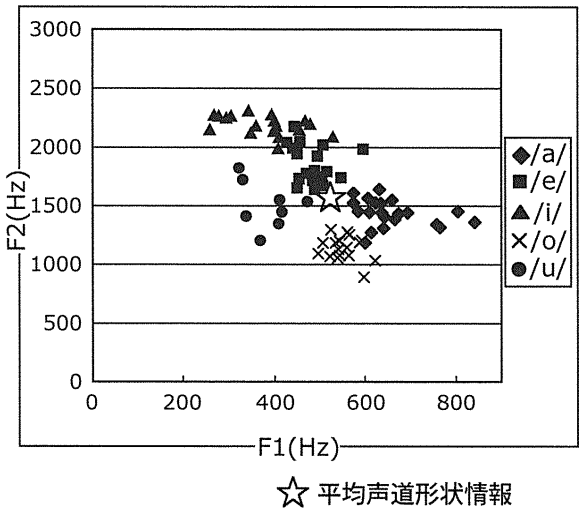
【図３】


(a)
i
Lip
glottis
i+1
i
Ai
Ai+1
(b)
i

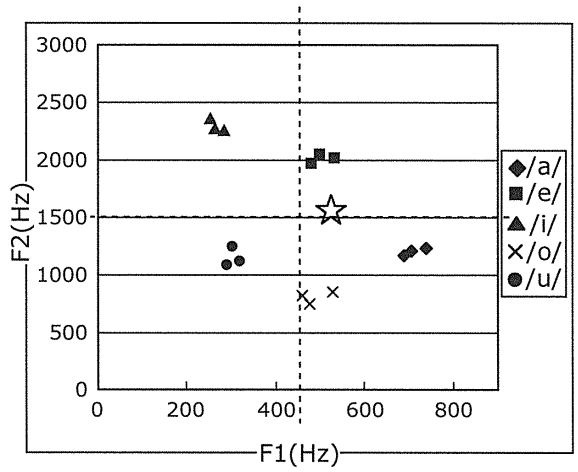
【図４Ａ】



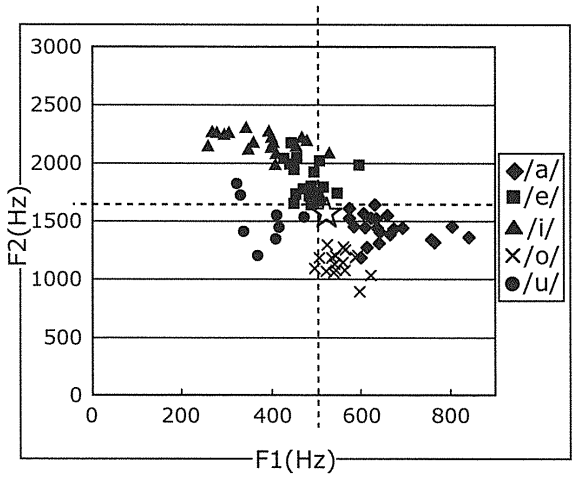
【図４Ｂ】



【図５Ａ】



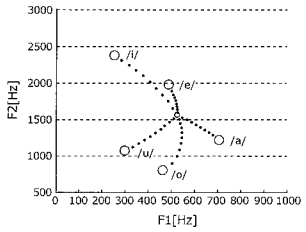
【図５Ｂ】



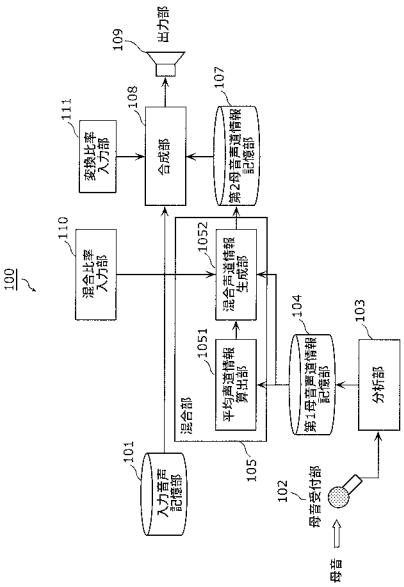
【図６】

	平方根二乗誤差(RMSE)
文中母音のF1-F2平均	399.5
孤立母音のF1-F2平均	422.5
孤立母音の平均声道形状情報	404.0

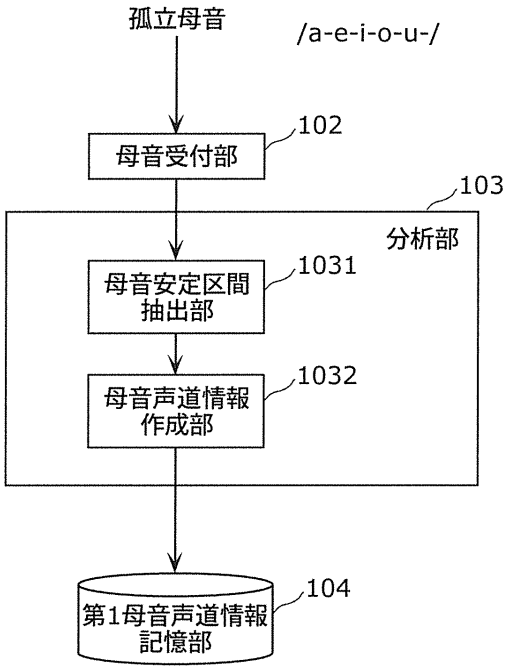
【図７】



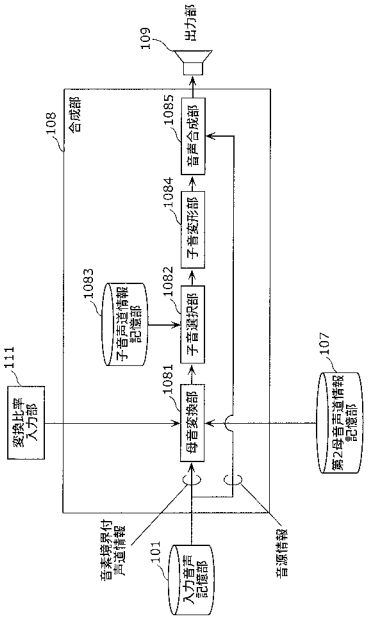
【図８】



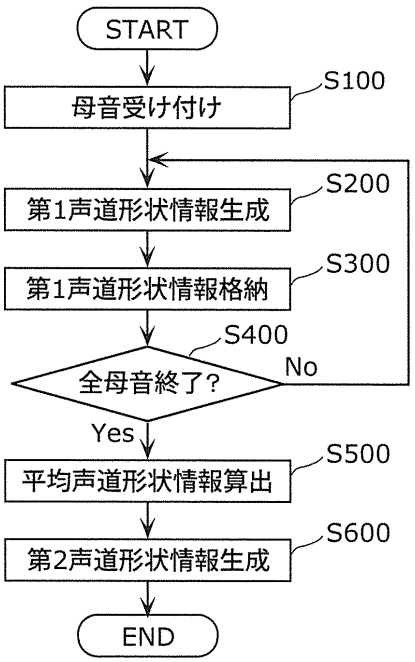
【図９】



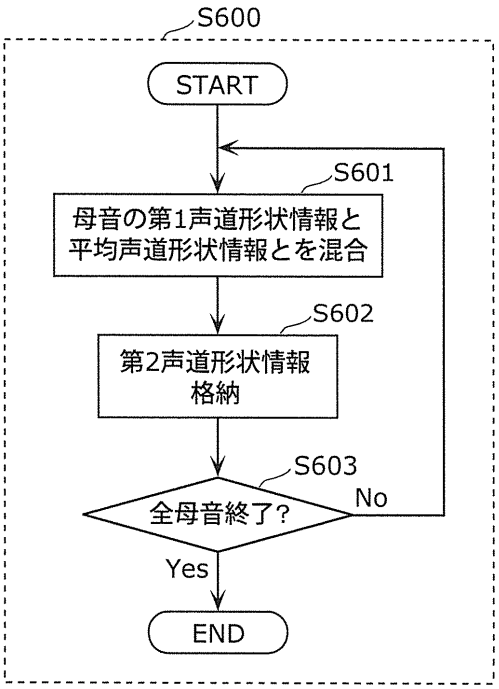
【図１０】



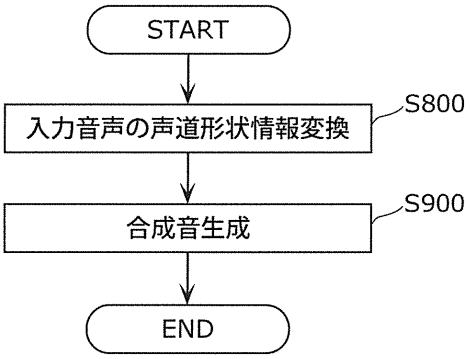
【図１１Ａ】



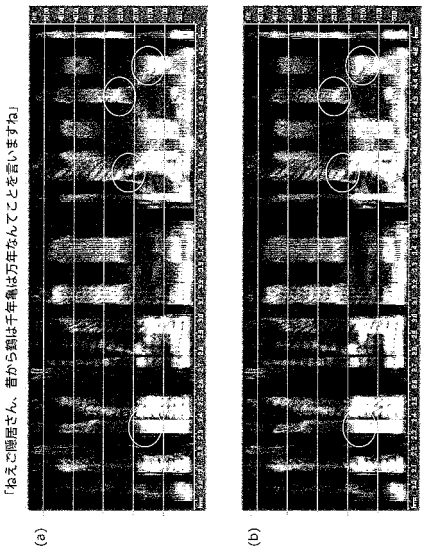
【図１１Ｂ】



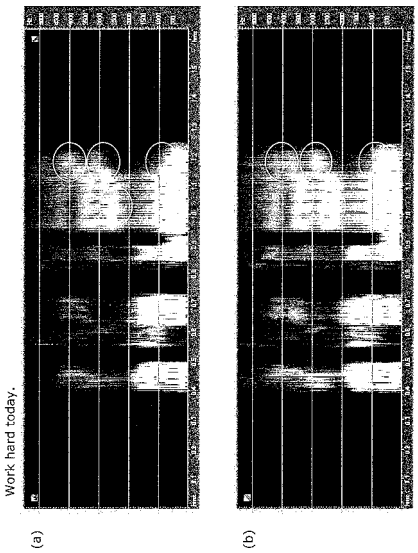
【図１２】



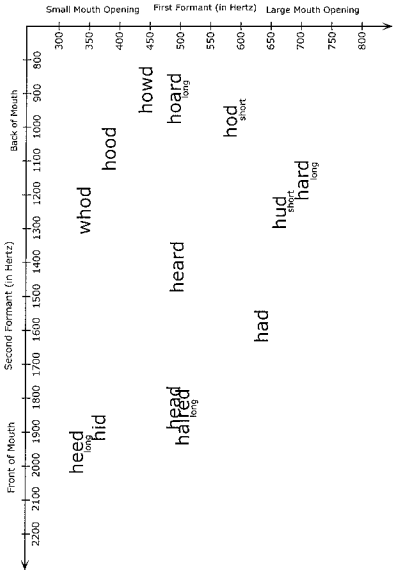
【図１３Ａ】



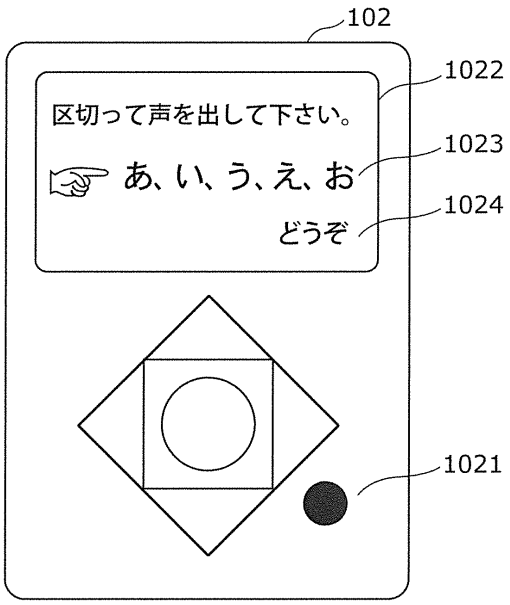
【図１３Ｂ】



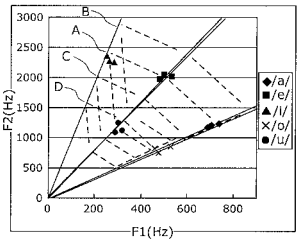
【図１４】



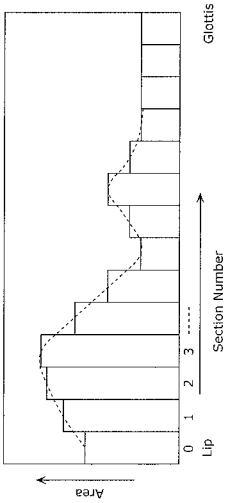
【図１５】



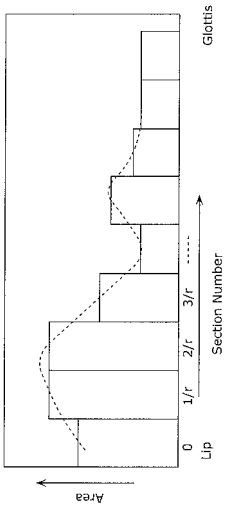
【図１６】



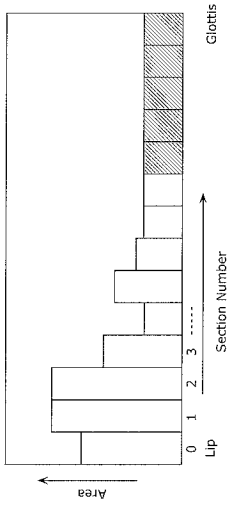
【図１７】



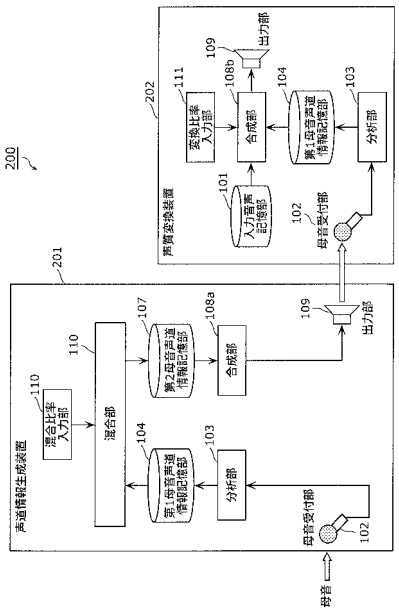
【図１８】



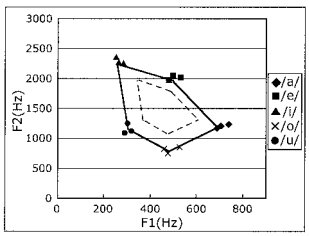
【図１９】



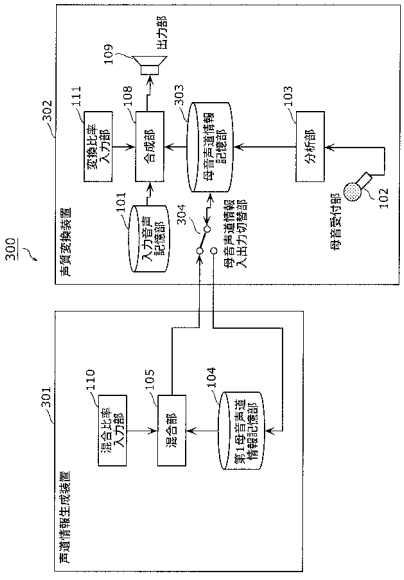
【図２０】



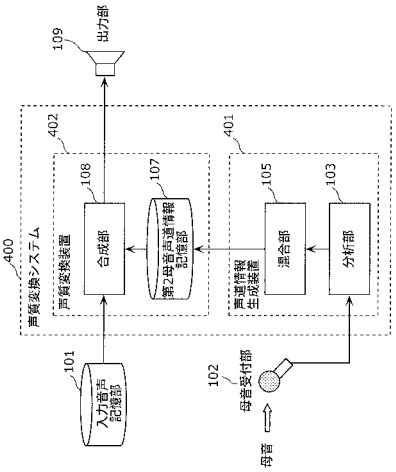
【図２１】



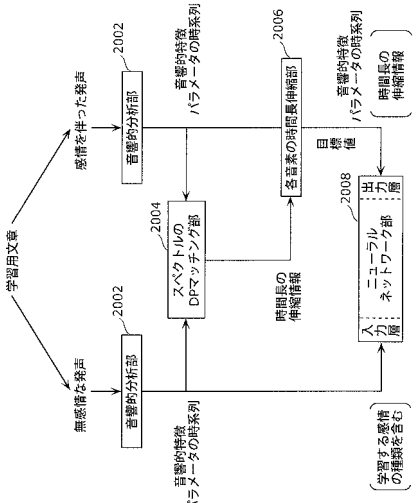
【図２２】



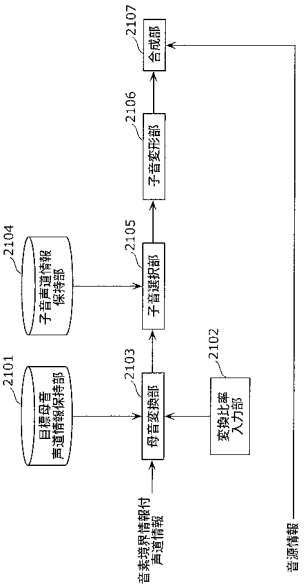
【図２３】


400
声質変換システム
402
声質変換装置
108
合成部
107
第２母音声道情報記憶部
401
声道情報生成装置
105
混合部
103
分析部
109
出力部
101
入力音声記憶部
102
母音受付部
母音

【図２４】


学習用文章
感情を伴った発声
無感情な発声
2002
音響的分析部
音響的特徴パラメータの時系列
2004
スペクトルのDPマッチング部
2006
各音素の時間長伸縮部
目標値
音響的特徴パラメータの時系列
時間長の伸縮情報
2008
出力層
ニューラルネットワーク部
入力層
学習する感情の種類を含む

【図２５】


2101
目標母音声道情報保持部
2104
子音声道情報保持部
2102
変換比率入力部
2103
母音変換部
2105
子音選択部
2106
子音変形部
2107
合成部
音素境界情報付声道情報
音源情報

フロントページの続き

(56)参考文献 国際公開第２００８／１４８５４７（WO，Ａ１）
特開２００７－５０１４３（ＪＰ，Ａ）
特開２００１－２８２３００（ＪＰ，Ａ）
特開２００６－３３０３４３（ＪＰ，Ａ）
国際公開第２００８／１４２８３６（WO，Ａ１）
国際公開第２０１０／０３５４３８（WO，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
G10L 21/00-25/93