

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2005-234337
(P2005-234337A)

(43) 公開日 平成17年9月2日(2005.9.2)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
G 1 0 L 13/06	G 1 0 L 5/04	5 D 0 4 5
G 1 0 L 13/08	G 1 0 L 3/00	

審査請求 未請求 請求項の数 6 O L (全 13 頁)

(21) 出願番号	特願2004-44852 (P2004-44852)	(71) 出願人	000004075 ヤマハ株式会社 静岡県浜松市中沢町10番1号
(22) 出願日	平成16年2月20日(2004.2.20)	(74) 代理人	100098084 弁理士 川▲崎▼ 研二
		(72) 発明者	川▲原▼ 毅彦 静岡県浜松市中沢町10番1号 ヤマハ株式会社内
		(72) 発明者	劔持 秀紀 静岡県浜松市中沢町10番1号 ヤマハ株式会社内
		Fターム(参考)	5D045 AA07 AA08 AA09

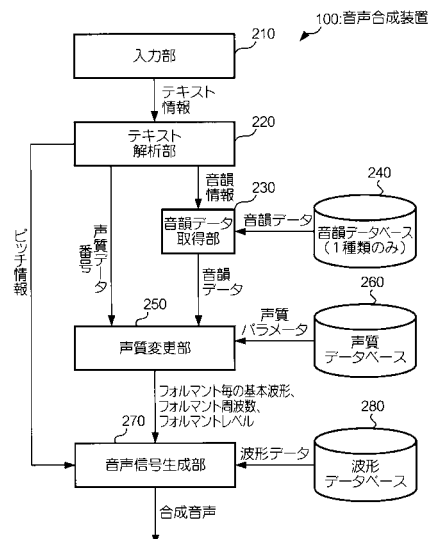
(54) 【発明の名称】 音声合成装置、音声合成方法、及び音声合成プログラム

(57) 【要約】

【課題】 ハードウェア資源に大きな制約が課されている環境下においても、様々な種類の声質の合成音声を生産することができる音声合成装置等を提供する。

【解決手段】 1種類の音韻データを備えた音声合成装置100に声質変更部250及び声質データベース260を設ける。声質変更部250は、テキスト解析部220から供給される声質データ番号を検索キーとして声質データベース260を検索し、声質パラメータを取得する。声質変更部250は、取得した声質パラメータに基づいて、音韻データ取得部230によって取得される音韻データに示される各音韻の声質を変更する。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

入力されるテキスト情報から、合成音声の音韻を指定する音韻指定情報及び該合成音声の声質を指定する声質指定情報を取得する取得手段と、

特定話者の各音韻をあらゆる一連の音韻データを記憶する第 1 記憶手段と、

前記各音韻の声質を変更するために必要な情報であって、音韻データの加工内容をあらゆる音韻データ加工情報を複数種類記憶する第 2 記憶手段と、

前記第 1 記憶手段から、前記音韻指定情報に示される音韻に対応する音韻データを抽出する第 1 抽出手段と、

前記第 2 記憶手段から、前記声質指定情報に示される声質に対応する音韻データ加工情報を抽出する第 2 抽出手段と、

前記抽出された音韻データを、前記抽出された音韻データ加工情報に基づいて加工し、前記合成音声を生成する生成手段と

を具備することを特徴とする音声合成装置。

【請求項 2】

前記各音韻データは、それぞれ対応する音韻のフォルマントを複数備え、

前記音韻データ加工情報には、前記フォルマントの変更内容をあらゆるフォルマント変更情報が含まれ、

前記生成手段は、前記音韻データに示される音韻の各フォルマントを、前記フォルマント変更情報に基づいて変更し、変更後の各フォルマントを加算した後の信号波形に基づいて前記合成音声を生成することを特徴とする請求項 1 に記載の音声合成装置。

【請求項 3】

前記フォルマントは、フォルマント周波数とフォルマントレベルとの対によって構成され、

前記フォルマント変更情報には、前記フォルマント周波数の変更内容をあらゆるフォルマント周波数変更情報と前記フォルマントレベルの変更内容をあらゆるフォルマントレベル変更情報が含まれ、

前記生成手段は、前記音韻データに示される音韻の各フォルマント周波数と各フォルマントレベルのそれぞれを、前記フォルマント周波数変更情報及び前記フォルマントレベル変更情報に基づいて変更することにより、前記変更後の各フォルマントを得ることを特徴とする請求項 2 に記載の音声合成装置。

【請求項 4】

前記取得手段は、前記テキスト情報から前記音韻指定情報及び前記声質指定情報を取得するほか、前記合成音声のピッチを指定するピッチ指定情報を取得し、

前記生成手段は、前記変更後の各フォルマントを加算した後の信号波形に対して前記ピッチ指定情報に示されるピッチを与えることにより、前記合成音声を得ることを特徴とする請求項 2 または 3 に記載の音声合成装置。

【請求項 5】

入力されるテキスト情報から、合成音声の音韻を指定する音韻指定情報及び該合成音声の声質を指定する声質指定情報を取得する取得過程と、

特定話者の各音韻をあらゆる一連の音韻データを記憶する第 1 記憶手段から、前記音韻指定情報に示される音韻に対応する音韻データを抽出する第 1 抽出過程と、

前記各音韻の声質を変更するために必要な情報であって、音韻データの加工内容をあらゆる音韻データ加工情報を複数種類記憶する第 2 記憶手段から、前記声質指定情報に示される声質に対応する音韻データ加工情報を抽出する第 2 抽出過程と、

前記抽出された音韻データを、前記抽出された音韻データ加工情報に基づいて加工し、前記合成音声を生成する生成過程と

を具備することを特徴とする音声合成方法。

【請求項 6】

特定話者の各音韻をあらゆる一連の音韻データを記憶する第 1 記憶手段と、前記各音韻

10

20

30

40

50

の声質を変更するために必要な情報であって、音韻データの加工内容をあらかず音韻データ加工情報を複数種類記憶する第2記憶手段とを備えたコンピュータを、

入力されるテキスト情報から、合成音声の音韻を指定する音韻指定情報及び該合成音声の声質を指定する声質指定情報を取得する取得手段と、

前記第1記憶手段から、前記音韻指定情報に示される音韻に対応する音韻データを抽出する第1抽出手段と、

前記第2記憶手段から、前記声質指定情報に示される声質に対応する音韻データ加工情報を抽出する第2抽出手段と、

前記抽出された音韻データを、前記抽出された音韻データ加工情報に基づいて加工し、前記合成音声を生成する生成手段として機能させるための音声合成プログラム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力されるテキスト情報から合成音声を生成する音声合成装置、音声合成方法、及び音声合成プログラムに関する。

【背景技術】

【0002】

図17は、入力されるテキスト情報から合成音声を生成する従来の音声合成装置100の構成を示す図である。

入力部110は、図示せぬ操作部等から「こんにちわ」等のテキスト情報を受け取ると、このテキスト情報をテキスト解析部120に供給する。テキスト解析部120は、受け取ったテキスト情報について、単語辞書等を用いて単語解析、構文解析等を行い、「こ」、「ん」、「に」、「ち」、「わ」といったモーラ単位の各音韻をあらかず音韻情報、及び各音韻の長さ、高さ、強さをあらかず韻律情報を生成し、音声合成部130に供給する。音声合成部130は、テキスト解析部120から供給される各音韻情報に基づいて、モーラ単位の音声データ（以下、音韻データ）を音韻データベース140から取得する。そして、音声合成部130は、取得した各音韻データを韻律情報に従って適宜加工・接続等して合成音声信号を生成し、スピーカ等から合成音声として出力する。ユーザは、音声合成装置から出力される合成音声を聴取することで、入力されたテキスト情報の内容を確認することができる。

20

30

【0003】

ところが、上記音韻データベースには、特定話者（例えば、男性話者）による1種類の音韻データしか登録されていない。このため、例えば若い女性等が好んで使用するであろうテキスト情報（「ちょう・・・的」や「・・・みたいなあ」等のテキスト情報）を、上記特定話者の声質を有する合成音声で出力した場合には、ユーザは声質と音声内容との間に違和感を感じてしまう等の問題があった。

【0004】

かかる問題を解消するべく、複数種類の音韻データ（例えば、男性、女性、子供、老人のそれぞれの音韻データ）を音韻データベースに予め登録し、入力されるテキスト情報の内容等に応じて最適な音韻データを選択し、選択した音韻データを用いて合成音声を生成する技術が提案されている（例えば、特許文献1参照）。

40

【0005】

【特許文献1】特開2000-339137号公報（第3-4頁）

【発明の開示】

【発明が解決しようとする課題】

【0006】

特許文献1に開示された技術によれば、確かに適切な合成音声を得ることができるが、これを実現するためには複数種類の音韻データを音韻データベースに登録しなければならない。しかしながら、メモリやCPU等のハードウェア資源に大きな制約が課されている携帯端末等にこのような複数種類の音韻データを実装することはできず、結局、携帯端末

50

等においては様々な声質の合成音声を生成することができないという問題があった。

【0007】

本発明は、以上説明した事情を鑑みてなされたものであり、ハードウェア資源に大きな制約が課されている環境下においても、様々な種類の声質の合成音声を生成することができる音声合成装置、音声合成方法、及び音声合成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】

【0008】

上述した問題を解決するため、本発明に係る音声合成装置は、入力されるテキスト情報から、合成音声の音韻を指定する音韻指定情報及び該合成音声の声質を指定する声質指定情報を取得する取得手段と、特定話者の各音韻をあらゆる一連の音韻データを記憶する第1記憶手段と、前記各音韻の声質を変更するために必要な情報であって、音韻データの加工内容をあらゆる音韻データ加工情報を複数種類記憶する第2記憶手段と、前記第1記憶手段から、前記音韻指定情報に示される音韻に対応する音韻データを抽出する第1抽出手段と、前記第2記憶手段から、前記声質指定情報に示される声質に対応する音韻データ加工情報を抽出する第2抽出手段と、前記抽出された音韻データを、前記抽出された音韻データ加工情報に基づいて加工し、前記合成音声を生成する生成手段とを具備することを特徴とする。

10

【0009】

かかる構成によれば、第1抽出手段によって抽出された音韻データが第2抽出手段によって抽出された音韻データ加工情報に基づいて加工され、これにより、合成音声が生成される。この音韻データ加工情報に様々な加工条件（例えば、音韻のフォーマットの変更内容等）を設定することで、1種類の音韻データ（例えば、男性話者の音韻データ）から様々な音質の合成音声を得ることが可能となる。

20

【0010】

ここで、上記構成にあつては、前記各音韻データは、それぞれ対応する音韻のフォーマットを複数備え、前記音韻データ加工情報には、前記フォーマットの変更内容をあらゆるフォーマット変更情報が含まれ、前記生成手段は、前記音韻データに示される音韻の各フォーマットを、前記フォーマット変更情報に基づいて変更し、変更後の各フォーマットを加算した後の信号波形に基づいて前記合成音声を生成する態様が好ましい。

30

これに加え、前記フォーマットは、フォーマット周波数とフォーマットレベルとの対によって構成され、前記フォーマット変更情報には、前記フォーマット周波数の変更内容をあらゆるフォーマット周波数変更情報と前記フォーマットレベルの変更内容をあらゆるフォーマットレベル変更情報が含まれ、前記生成手段は、前記音韻データに示される音韻の各フォーマット周波数と各フォーマットレベルのそれぞれを、前記フォーマット周波数変更情報及び前記フォーマットレベル変更情報に基づいて変更することにより、前記変更後の各フォーマットを得る態様がより好ましい。

【0011】

また、前記取得手段は、前記テキスト情報から前記音韻指定情報及び前記声質指定情報を取得するほか、前記合成音声のピッチを指定するピッチ指定情報を取得し、前記生成手段は、前記変更後の各フォーマットを加算した後の信号波形に対して前記ピッチ指定情報に示されるピッチを与えることにより、前記合成音声を得る態様も好ましい。

40

【発明の効果】

【0012】

以上説明したように、本発明によれば、ハードウェア資源に大きな制約が課されている環境下においても、様々な種類の声質の合成音声を生成することが可能となる。

【発明を実施するための最良の形態】

【0013】

以下、本発明に係る実施の形態について図面を参照しながら説明する。

A. 本実施形態

50

図1は、本実施形態に係る音声合成装置100の機能構成を示す図である。なお、本実施形態では、携帯電話やPHS(Personal Handyphone System)、PDA(Personal Digital Assistance)などハードウェア資源に大きな制約が課される携帯端末に実装される場合を想定するが、これに限らず様々な電子機器に適用可能である。

【0014】

入力部210は、図示せぬ操作部等を介して入力されるテキスト情報をテキスト解析部220に供給する。図2は、テキスト情報を例示した図である。

テキスト本文情報は、合成音声として出力すべきテキストの内容(例えば「こんにちわ」)をあらわす情報である。なお、図2ではひらがなのみによって表されたテキスト本文情報を示しているが、ひらがなのみならず、漢字、ローマ字、カタカナ等の各種文字や各種記号によって表されたものであっても良い。

【0015】

声質データ番号(声質指定情報)は、後述する複数の声質パラメータ(音韻データ加工情報)をそれぞれ識別するためのユニークな番号(図2では、K1~Kn)である。本実施形態では、この声質パラメータを適宜選択・利用することで、特定話者(本実施形態では、「男性話者」を想定)による1種類の音韻データから、様々な声質の合成音声を得ることが可能となる(詳細は後述)。

【0016】

ピッチ情報(ピッチ指定情報)は、合成音声にピッチを与える(いいかえれば、合成音声のピッチを指定する)ための情報であり、「C(ド)」~「H(シ)」等の音階を指定する情報によって構成されている(図2参照)。

【0017】

テキスト解析部220は、入力部210から供給されるテキスト情報を解析し、解析結果を音韻データ取得部230、声質変更部250、音声信号生成部270にそれぞれ供給する。具体的には、図2に示すようなテキスト情報が供給されると、テキスト解析部220は、まず、「こんにちわ」といったテキスト本文情報を「こ」、「ん」、「に」、「ち」、「わ」といったモーラ単位の音韻に分解する。なお、モーラとは、読みの単位をあらわし、基本的には1つの子音と1つの母音から構成されるものをいう。

【0018】

テキスト解析部(取得手段)220は、このようにしてテキスト本文情報をモーラ単位の音韻に分解すると、これら合成音声の各音韻を指定する音韻情報(音韻指定情報)を生成し、音韻データ取得部230に順次供給する。さらに、テキスト解析部220は、声質データ番号(例えばK3)、ピッチ情報(例えばC(ド))をテキスト情報からそれぞれ取得し、取得した声質データ番号を声質変更部250に供給する一方、取得したピッチ情報を音声信号生成部270に供給する。

【0019】

音韻データ取得部(第1抽出手段)230は、テキスト解析部220から供給される音韻情報を検索キーとして音韻データベース240を検索することにより、音韻情報に示される音韻に対応する音韻データを取得する。図3は、音韻データベース240の登録内容を例示した図である。図3に示すように、音韻データベース(第1記憶手段)240には、1人の男性話者のモーラ単位の各音韻(「あ」、「い」、・・・「ん」等)をあらわす一連の音韻データ1~mが登録されるほか、該一連の音韻データの数(以下、登録音韻データ数)等が登録されている。

【0020】

図4は、ある音韻(例えば「こ」等)をあらわす音韻データの構成を例示した図であり、図5は、音韻データに含まれる各フレーム情報を説明するための図である。なお、図5のAは、上記男性話者がある音韻(例えば「こ」等)を発声したときの音声波形vwと各フレームFRとの関係を示しており、図5のB、図5のC、図5のDは、それぞれ第1フレームFR1、第2フレームFR2、第nフレームFRnに係るフォルマント分析結果を示している。

10

20

30

40

50

【 0 0 2 1 】

図 4 に示すように、音韻データは、第 1 フレーム情報 ~ 第 n フレーム情報によって構成されている。各フレーム情報は、対応する各フレーム F R (図 5 参照) をフォルマント分析することにより得られる第 1 フォルマント情報 ~ 第 k フォルマント情報と、各フレーム F R の音声が有声であるか無声であるかをあらかず有声 / 無声判別フラグ (例えば、 “ 1 ” = 有声、 “ 0 ” = 無声) とを備えている。

【 0 0 2 2 】

各フレーム情報を構成する第 1 フォルマント情報 ~ 第 k フォルマント情報は、対応するフォルマントをあらかずフォルマント周波数 F とフォルマントレベル A との対によって構成されている (図 5 の B ~ 図 5 の D 参照) 。例えば、第 1 フレーム情報を構成する第 1 フォルマント情報 ~ 第 k フォルマント情報は、それぞれ (F 1 1 、 A 1 1) 、 (F 1 2 、 A 1 2) 、 . . . (F 1 k 、 A 1 k) といいたフォルマント周波数とフォルマントレベルとの対によって構成され (図 5 の B 参照) 、 . . . 第 n フレーム情報を構成する第 1 フォルマント情報 ~ 第 k フォルマント情報は、それぞれ (F n 1 、 A n 1) 、 (F n 2 、 A n 2) 、 . . . (F n k 、 A n k) といいたフォルマント周波数とフォルマントレベルとの対によって構成される (図 5 の D 参照) 。

10

【 0 0 2 3 】

音韻データ取得部 2 3 0 は、テキスト解析部 2 2 0 から供給される各音韻情報 (「こ」、「ん」、「に」、「ち」、「わ」等をあらかず各音韻情報) に基づいて対応する各音韻データを取得すると、これらを声質変更部 2 5 0 に供給する。

20

【 0 0 2 4 】

声質変更部 2 5 0 は、音韻データ取得部 2 3 0 によって取得された各音韻データに示される音韻の声質を変更する。詳述すると、声質変更部 (第 2 抽出手段) 2 5 0 は、まず、テキスト解析部 2 2 0 から供給される声質データ番号を検索キーとして声質データベース (第 2 記憶手段) 2 6 0 を検索し、対応する声質パラメータを取得する。そして、声質変更部 2 5 0 は、取得した声質パラメータに基づいて、上記各音韻の声質の変更を行う。

【 0 0 2 5 】

図 6 は、声質データベース 2 6 0 の登録内容を例示した図である。

図 6 に示すように、声質データベース (第 2 記憶手段) 2 6 0 には、上記各音韻の声質を変更するために必要な情報であって、音韻データの加工内容をあらかず複数種類の声質パラメータ 1 ~ L と、該声質パラメータの数をあらかず登録数情報とが格納されている。

30

【 0 0 2 6 】

図 7 は、声質パラメータの構成の一例を示す図である。

図 7 に示すように、声質パラメータ (音韻データ加工情報) は、当該パラメータを特定するための声質データ番号と、合成音声の性別を変更するか否かをあらかず性別変更フラグと、第 1 ~ 第 k フォルマントの変更内容をあらかず第 1 ~ 第 k フォルマント変更情報とを備えている。ここで、例えば上記性別変更フラグが “ 1 ” に設定されている場合には、声質変更部 2 5 0 によって合成音声の性別を変更するための処理 (以下、性別変更処理) が行われる一方、上記性別変更フラグが “ 0 ” に設定されている場合には、上記性別変更処理は行われない (詳細は後述) 。なお、本実施形態では、男性話者による 1 種類の音韻データを想定しているため、この性別変更フラグが “ 1 ” に設定されている場合には、合成音声の特徴は男性的なものから女性的なものに変更される。一方、性別変更フラグが “ 0 ” に設定されている場合には、合成音声の特徴は男性的なもののまま変更されない。

40

【 0 0 2 7 】

一方、各フォルマント変更情報は、後述する各フォルマントの基本波形 (正弦波等) を選択するための基本波形選択情報と、各フォルマント周波数の変更内容をあらかずフォルマント周波数変更情報と、フォルマント周波数の各レベルの変更内容をあらかずフォルマントレベル変更情報とを備えている。

【 0 0 2 8 】

各フォルマント周波数変更情報及び各フォルマントレベル変更情報には、それぞれフォ

50

ルマント周波数のシフト量、発振速度、発振レベルをあらゆる情報及びフォルマントレベルのシフト量、発振速度、発振レベルをあらゆる情報が含まれている。なお、フォルマント周波数及びフォルマントレベルのシフト量、発振速度、発振レベルに関する詳細は後述する。

【0029】

図8は、声質変更部250によって実行される声質変更処理を示すフローチャートである。

声質変更部(生成手段)250は、テキスト解析部220から声質データ番号を受け取ると、該声質データ番号を検索キーとして声質データベース260を検索し、対応する声質パラメータを取得する(ステップS1)。そして、声質変更部250は、取得した声質パラメータに含まれる性別変更フラグを参照し、合成音声の性別を変更すべきか(すなわち、性別変更処理を実行すべきか)否かを判断する(ステップS2)。声質変更部250は、例えば性別変更フラグが“0”に設定され、性別変更すべきでない判断すると、ステップS3をスキップしてステップS4に進む一方、例えば性別変更フラグが“1”に設定され、性別変更すべきと判断すると、ステップS3に進み、性別変更処理を実行する。

10

【0030】

図9は、記憶手段(図示略)に格納されている性別変更処理用のマッピング関数 m_f を例示した図であり、図10及び図11は、男性及び女性がそれぞれ同一の音韻(例えば、「あ」等)を発声したときの分析結果を示す図である。なお、図9に示すマッピング関数 m_f の横軸は入力周波数(声質変更部250に入力されるフォルマント周波数)、縦軸は出力周波数(声質変更部250から出力される周波数変更後のフォルマント周波数)をあらわし、 f_{max} は入力可能なフォルマント周波数の最大値をあらわす。また、図10及び図11に示す分析グラフ g_1 、 g_2 の横軸は周波数、縦軸はレベルをあらわす。

20

【0031】

図10と図11に示す分析グラフ g_1 、 g_2 を比較して明らかなように、男性の音韻の第1フォルマント周波数 f_{m1} ~第4フォルマント周波数 f_{m4} は、女性の音韻の第1フォルマント周波数 f_{f1} ~第4フォルマント周波数 f_{f4} と比較して低い。そこで、本実施形態では、図9に示すように、直線 n_1 (入力周波数=出力周波数;破線部分参照)よりも上側に位置するマッピング関数 m_f (実線部分参照)を用いることにより、男性的な特徴を有する音韻を女性的な特徴を有する音韻に変更する。

30

【0032】

具体的には、声質変更部250は、入力される音韻データの各フォルマント周波数を、図9に示すマッピング関数 m_f を用いて周波数の高い方向にシフトする。これにより、入力される男性の音韻の各フォルマント周波数は、女性的な特徴を有するものに変更される。なお、女性の音韻のフォルマント周波数が入力される場合には、上記とは逆に、直線 n_1 よりも下側に位置するマッピング関数 $m_{f'}$ (図9に一点鎖線で示す部分参照)を利用すれば良い。

【0033】

声質変更部250は、上記性別変更処理を実行し、ステップS4に進むと、各フォルマント周波数変更情報に示される各フォルマント周波数のシフト量に従って、各フォルマント周波数をシフトする。さらに、声質変更部250は、シフトした各フォルマント周波数を発振させるべく、周波数発振処理を実行する(ステップS5)。

40

【0034】

図12は、記憶手段(図示略)に格納されている周波数発振処理に用いられる発振テーブル T_A を例示した図であり、図13は、該発振テーブル T_A から読み出される発振値と時間の関係を例示した図である。なお、本実施形態では、便宜上、同一の発振テーブル T_A を用いて上記各フォルマント周波数を発振させる場合を想定するが、各フォルマント周波数毎に発振値等が異なる発振テーブルを用いても良い。

【0035】

発振テーブル T_A は、発振値を時系列順に登録したテーブルである。声質変更部250

50

は、各フォルマント周波数変更情報に示されるフォルマント周波数の発振速度に従って、発振テーブルT Aに登録されている発振値の読み出し速度（あるいは発振値を読み飛ばす数）を制御する一方、読み出した各発振値に対して各フォルマント周波数変更情報に示されるフォルマント周波数の発振レベルを乗するといった周波数発振処理を実行する。これにより、図14に示すようなフォルマント周波数 f_m を発振速度 s_p 、発振レベル l_v で発振させた波形を得ることができる。なお、本実施形態では、フォルマント周波数の発振レベルの演算量を削減するために、上記発振テーブルT Aを利用する態様を例示したが、発振テーブルT Aを利用することなく所定の関数等を利用してフォルマント周波数の発振レベルを求めても良い。

【0036】

10

声質変更部250は、周波数発振処理を実行すると、ステップ6に進み、各フォルマント周波数変更情報に示される各フォルマントレベルのシフト量に従って、各フォルマントレベルをシフトする。さらに、声質変更部250は、シフトした各フォルマントレベルを発振させるべく、レベル発振処理を実行し（ステップS7）、処理を終了する。なお、レベル発振処理に用いられる発振テーブルや該発振テーブルを用いて各フォルマントレベルを発振させる場合の動作については、上記各フォルマント周波数を発振させる場合とほぼ同様に説明することができるため割愛する。また、フォルマントレベルの発振について、フォルマント周波数の発振と同一の発振テーブルを用いて発振させても良いが、フォルマント周波数の発振とは異なる発振テーブルを用いて発振させても良い。

【0037】

20

声質変更部（生成手段）250は、取得した声質パラメータ（音韻データ加工情報）に基づいて各音韻の声質を変更（すなわち、音韻データを加工）すると、各フォルマント毎の基本波形選択情報、各フォルマント周波数、及び各フォルマントレベルを音声信号生成部270に供給する。

音声信号生成部270は、声質変更部250から供給される基本波形選択情報を受け取ると、この基本波形選択情報に示される波形データを波形データベース280から取得する。なお、この基本波形選択情報に示される基本波形は、各フォルマント毎に異なっても良く、例えば低い周波数のフォルマントの基本波形を正弦波とする一方、個性を表す高い周波数のフォルマントの基本波形を正弦波以外の波形（例えば、矩形波やのこぎり波など）等としても良い。もちろん、複数種類の基本波形を利用することなく、単一の基本波形（例えば、正弦波）のみを利用しても良い。

30

【0038】

音声信号生成部（生成手段）270は、このようにして各波形データを選択すると、選択した各波形データ、各フォルマント周波数、各フォルマントレベルを用いて各フォルマント毎のフォルマント波形を生成する。そして、音声信号生成部（生成手段）270は、各フォルマント波形を加算し、合成音声信号を生成する。さらに、音声信号生成部270は、生成した合成音声信号に、テキスト解析部220から供給されるピッチ情報（ピッチ指定情報）に示されるピッチを与える処理（以下、ピッチ付与処理）を施す。

【0039】

図15は、ピッチ付与処理を説明するための図である。なお、図15では、説明の理解を容易にするために、正弦波の合成音声信号にピッチが付与される場合を例示している。

40

音声信号生成部270は、テキスト解析部220から供給されるピッチ情報に基づいて図15に示す時間エンベロープ t_p の周期を算出する。ここで、合成音声のピッチは、時間エンベロープ t_p の周期に依存し、時間エンベロープ t_p の周期が長くなればピッチは低くなる一方、時間エンベロープ t_p の周期が短くなればピッチは高くなる。音声信号生成部270は、このようにして時間エンベロープ t_p の周期を求めると、時間エンベロープ t_p と合成音声信号との乗算を、求めた時間エンベロープ t_p の周期で繰り返し行うことにより、所定のピッチが付与された合成音声信号を得る。

【0040】

図16は、声質変更処理及びピッチ付与処理が施された特定フォルマントのフォルマン

50

ト波形を例示した図である。図 16 に示すように、声質変更に関わる処理（例えば、フォルマント周波数やフォルマントレベルの発振処理等）は、フレーム周期（フレーム単位）で制御することが可能となっている。音声信号生成部（生成手段）270 は、上記の如く所定のピッチが付与された合成音声信号を得ると、これを合成音声として外部へ出力する。これにより、ユーザは、音声合成装置 100 に入力したテキスト本文（「こんにちわ」等）の内容を、所望する声質の合成音声によって確認することが可能となる。

【0041】

以上説明したように、本実施形態に係る音声合成装置によれば、声質変更部においてフォルマントを単位とする様々な声質変更処理が施されるため、記憶される音韻データが 1 種類（すなわち、特定話者の音韻データのみ）であっても、様々な声質の音声合成が可能となる。

10

【0042】

B. その他

以上説明した本実施形態では、音声合成装置 100 に入力されるテキスト情報にピッチ情報が含まれている場合を例示したが（図 2 参照）、該テキスト情報にピッチ情報が含まれない場合もある。かかる場合を想定して音韻データベース 240 に予め代替ピッチ情報を登録し（図 3 の括弧書き参照）、テキスト情報にピッチ情報が含まれていない場合には、この代替ピッチ情報に示されるピッチ（例えば、C（ド）等）を合成音声のピッチとして利用するようにしても良い。また、代替ピッチ情報のほか、図 4 に示す各フレーム毎のフォルマント情報の数（フォルマント数情報；図 3 の括弧書き参照）を音韻データベース

20

【0043】

また、以上説明した音声合成装置 100 に係る諸機能は、CPU（或いは DSP）が ROM 等のメモリに格納されたプログラムを実行することによって実現されるため、かかるプログラムについて CD-ROM 等の記録媒体に記録して頒布したり、インターネット等の通信ネットワークを介して頒布しても良い。

【図面の簡単な説明】

【0044】

【図 1】本実施形態に係る音声合成装置の機能構成を示すブロック図である。

【図 2】同実施形態に係るテキスト情報を例示した図である。

30

【図 3】同実施形態に係る音韻データベースの登録内容を例示した図である。

【図 4】同実施形態に係る音韻データの構成を例示した図である。

【図 5】同実施形態に係る音韻データに含まれる各フレーム情報を説明するための図である。

【図 6】同実施形態に係る声質データベースの登録内容を例示した図である。

【図 7】同実施形態に係る声質パラメータの構成の一例を示す図である。

【図 8】同実施形態に係る声質変更処理を示すフローチャートである。

【図 9】同実施形態に係るマッピング関数を例示した図である。

【図 10】同実施形態に係る男性の音韻の分析結果を示す図である。

40

【図 11】同実施形態に係る女性の音韻の分析結果を示す図である。

【図 12】同実施形態に係る発振テーブルを例示した図である。

【図 13】同実施形態に係る発振テーブルから読み出される発振値と時間の関係を例示した図である。

【図 14】同実施形態に係るフォルマント周波数発振を説明するための図である。

【図 15】同実施形態に係るピッチ付与処理を説明するための図である。

【図 16】同実施形態に係る声質変更処理及びピッチ付与処理が施された特定フォルマントのフォルマント波形を例示した図である。

【図 17】従来の音声合成装置の機能構成を示す図である。

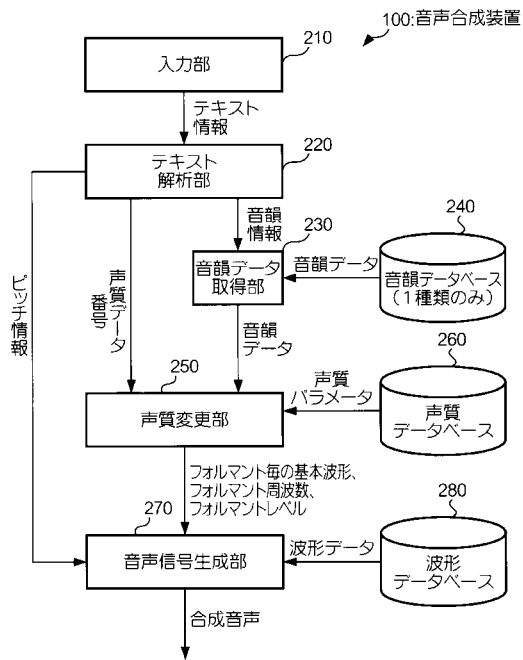
【符号の説明】

【0045】

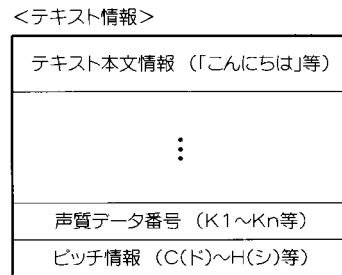
50

100・・・音声合成装置、210・・・入力部、220・・・テキスト解析部、230
 ・・・音韻データ取得部、240・・・音韻データベース、250・・・声質変更部、2
 60・・・声質データベース、270・・・音声信号生成部、280・・・波形データベ
 ース。

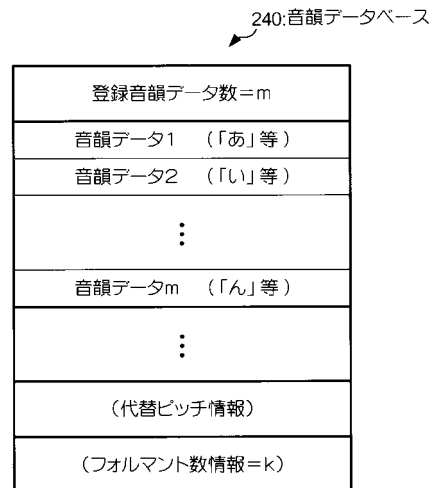
【 図 1 】



【 図 2 】



【 図 3 】

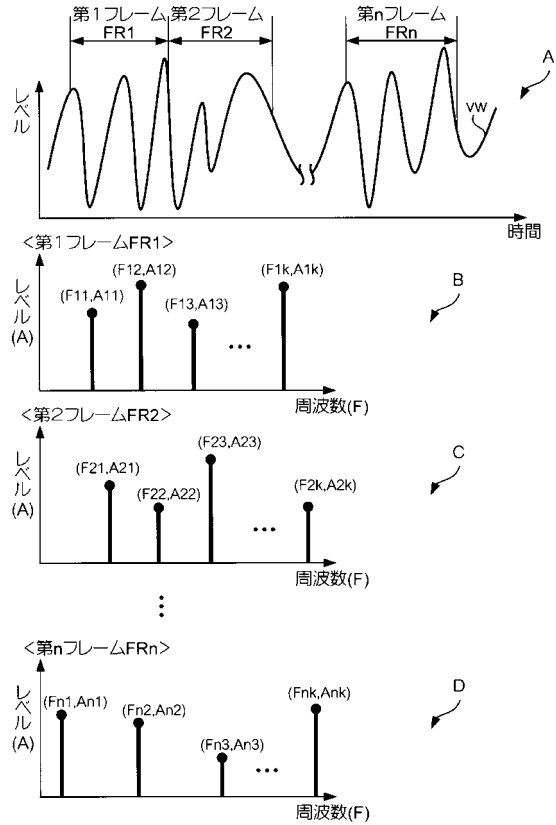


【 図 4 】

<音韻データ>

第1フレーム 情報	有声/無声判別フラグ	
	第1フォルマント情報	第1フォルマント周波数
		第1フォルマントレベル
	第2フォルマント情報	第2フォルマント周波数
		第2フォルマントレベル
⋮	⋮	
第kフォルマント情報	第kフォルマント周波数	
	第kフォルマントレベル	
第2フレーム 情報	有声/無声判別フラグ	
	第1フォルマント情報	第1フォルマント周波数
		第1フォルマントレベル
	第2フォルマント情報	第2フォルマント周波数
		第2フォルマントレベル
⋮	⋮	
第kフォルマント情報	第kフォルマント周波数	
	第kフォルマントレベル	
⋮	⋮	
第nフレーム 情報	⋮	⋮

【 図 5 】



【 図 6 】

260:声質データベース

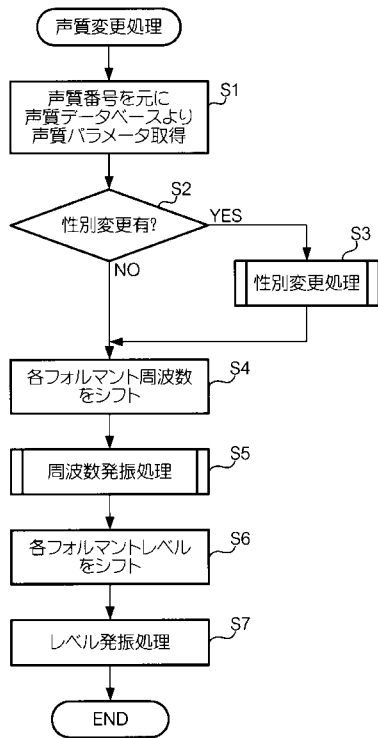
登録数情報=L
声質パラメータ1
声質パラメータ2
⋮
声質パラメータL

【 図 7 】

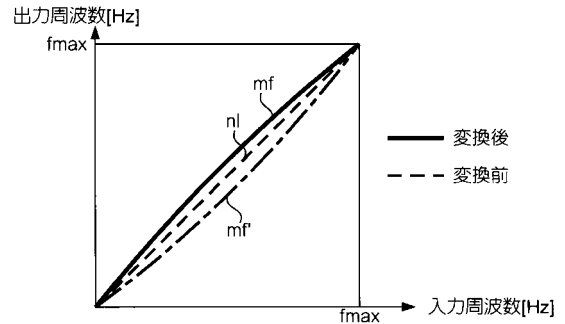
<声質パラメータ>

声質データ番号		
性別変更フラグ		
第1 フォルマント 変更情報	基本波形選択情報	
	フォルマント 周波数 変更情報	フォルマント周波数のシフト量
		フォルマント周波数の発振速度
フォルマント レベル変 更情報	フォルマント周波数の発振レベル	
第2 フォルマント 変更情報
⋮	⋮	⋮
第k フォルマント 変更情報

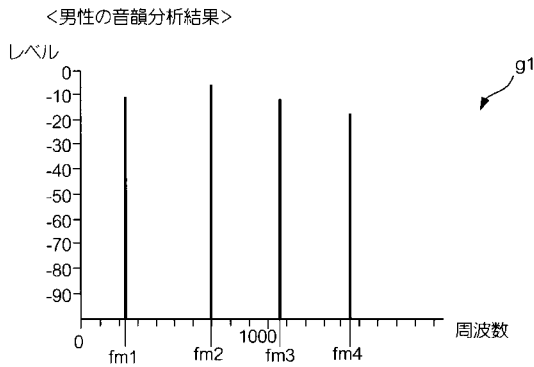
【 図 8 】



【 図 9 】



【 図 1 0 】

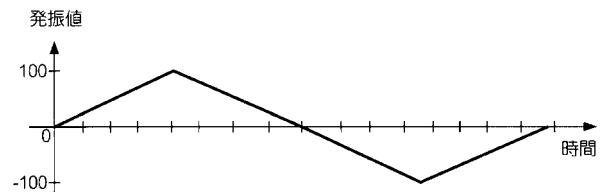


【 図 1 2 】

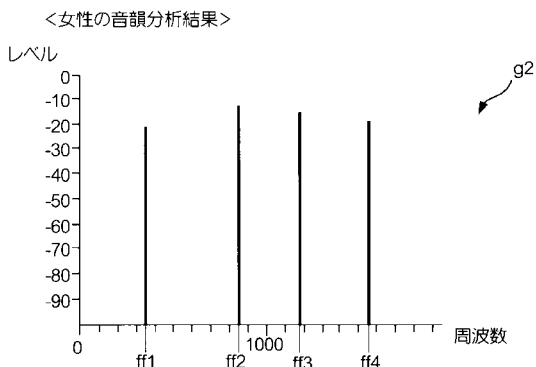
<発振テーブルTA>

0	25	50	75	100	75	50	25	0	-25	-50	-75	-100	-75	-50	-25	0
---	----	----	----	-----	----	----	----	---	-----	-----	-----	------	-----	-----	-----	---

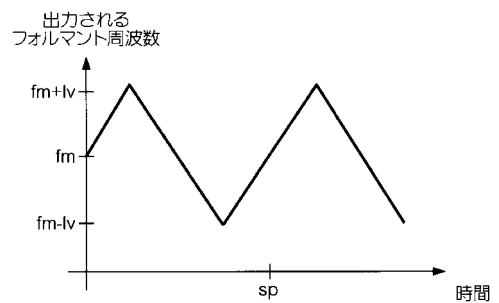
【 図 1 3 】



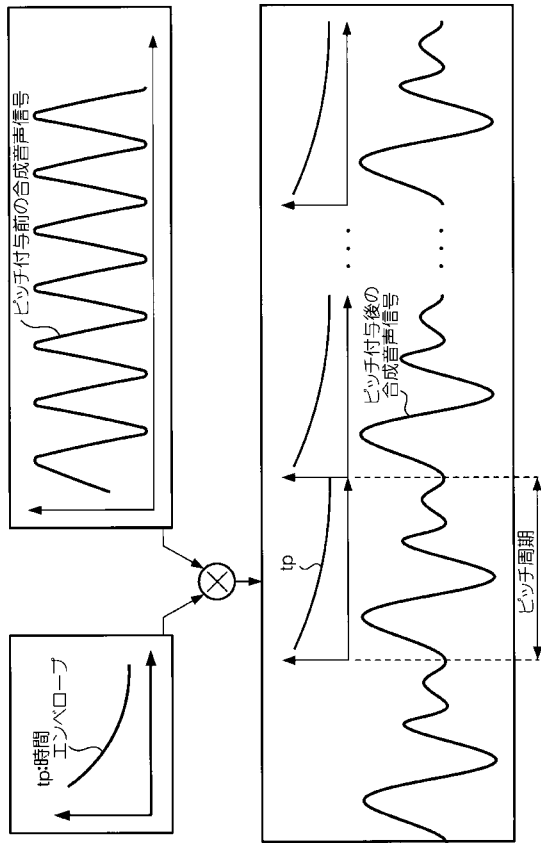
【 図 1 1 】



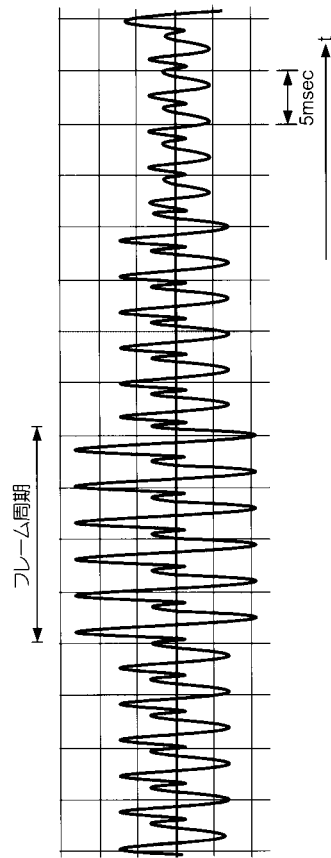
【 図 1 4 】



【 図 1 5 】



【 図 1 6 】



【 図 1 7 】

