

(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2012-37726

(P2012-37726A)

(43) 公開日 平成24年2月23日(2012.2.23)

(51) Int.Cl.

G 1 O L 21/04

(2006.01)

F 1

G 1 O L 21/04

G 1 O L 21/04

1 1 O Z

2 0 O C

テーマコード (参考)

審査請求 未請求 請求項の数 6 O L (全 22 頁)

(21) 出願番号

特願2010-177776 (P2010-177776)

(22) 出願日

平成22年8月6日 (2010.8.6)

(71) 出願人 000004352

日本放送協会

東京都渋谷区神南2丁目2番1号

591053926

財団法人エヌエイチケイエンジニアリング
サービス

東京都世田谷区砧一丁目10番11号

100064908

弁理士 志賀 正武

100108578

弁理士 高橋 詔男

齋藤 礼子

東京都世田谷区砧一丁目10番11号 日
本放送協会放送技術研究所内

最終頁に続く

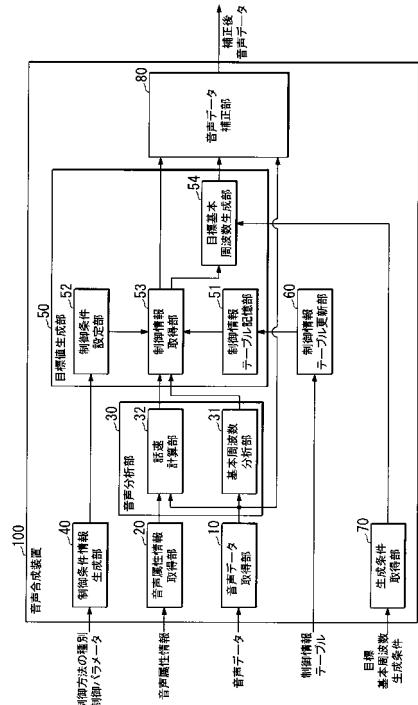
(54) 【発明の名称】音声合成装置およびコンピュータプログラム

(57) 【要約】

【課題】容易に且つ効率的に音声データを補正する。

【解決手段】制御情報テーブルを記憶する制御情報テーブル記憶部51と、音声データの基本周波数の時間変化情報を生成する基本周波数分析部31と、話速値を計算する話速計算部32と、制御方法の種別と話速の制御量または基本周波数の変化幅の制御量を指定するキー情報を設定する制御条件設定部52と、制御方法の種別に関連付けられた制御情報テーブルを読み込み、キー情報と話速値と基本周波数の時間変化情報とに基づいて話速の制御量と基本周波数の変化幅の制御量とを取得する制御情報取得部53と、基本周波数の変化幅の制御量に基づいて基本周波数の時間変化情報を生成する目標基本周波数生成部54と、話速の制御量と目標基本周波数の時間変化情報とに基づいて音声データを補正する音声データ補正部80とを備えた。

【選択図】図1



【特許請求の範囲】**【請求項 1】**

話速の制御量と基本周波数の変化幅の制御量とを対応付けた制御情報テーブルを、制御方法の種別に関連付けて記憶する制御情報テーブル記憶部と、

音声データを取得する音声データ取得部と、

前記音声データ取得部が取得した前記音声データに基づいて基本周波数の時間変化情報を生成する基本周波数分析部と、

前記音声データに基づいて話速値を計算する話速計算部と、

所望の制御方法の種別と制御情報テーブルにおける話速の制御量または基本周波数の変化幅の制御量を指定するキー情報とを設定する制御条件設定部と、

前記制御条件設定部が設定した前記制御方法の種別に関連付けられた制御情報テーブルを前記制御情報テーブル記憶部から読み込み、前記制御条件設定部が設定した前記キー情報と前記話速計算部が計算した前記話速値と前記基本周波数分析部が生成した前記基本周波数の時間変化情報をとに基づいて、話速の制御量と基本周波数の変化幅の制御量とを取得する制御情報取得部と、

前記制御情報取得部が取得した前記基本周波数の変化幅の制御量に基づいて、前記基本周波数の時間変化情報を目標基本周波数の時間変化情報を生成する目標基本周波数生成部と、

前記制御情報取得部が取得した前記話速の制御量と、前記目標基本周波数生成部が生成した前記目標基本周波数の時間変化情報をとに基づいて、前記音声データを補正して補正後音声データを出力する音声データ補正部と、

を備えることを特徴とする音声合成装置。

【請求項 2】

話速の絶対値を話速の制御量とし、基本周波数の変化幅の絶対値を基本周波数の変化幅の制御量とし、前記制御情報テーブル記憶部は、複数の話速値の絶対値それぞれに基本周波数の変化幅の絶対値を対応付けた制御情報テーブルを制御方法の種別に関連付けて記憶する

ことを特徴とする請求項 1 記載の音声合成装置。

【請求項 3】

話速の倍率を話速の制御量とし、基本周波数の変化幅の倍率を基本周波数の変化幅の制御量とし、前記制御情報テーブル記憶部は、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた制御情報テーブルを制御方法の種別に関連付けて記憶し、

前記制御条件設定部は、前記キー情報をについての所望の倍率をさらに設定し、

前記制御情報取得部は、前記制御条件設定部が設定した前記キー情報を前記倍率と前記話速計算部が計算した前記話速値と前記基本周波数分析部が生成した前記基本周波数の時間変化情報をとに基づいて、話速の制御量と基本周波数の変化幅の制御量とを取得する

ことを特徴とする請求項 1 記載の音声合成装置。

【請求項 4】

前記制御情報テーブル記憶部は、複数の指定話速値それぞれに対応して、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた制御情報テーブルを制御方法の種別に関連付けて記憶し、

前記制御情報取得部は、前記制御条件設定部が設定した前記制御方法の種別に関連し且つ前記話速計算部が計算した前記話速値と同一である指定話速値に対応する制御情報テーブルを前記制御情報テーブル記憶部から読み込み、前記制御条件設定部が設定した前記キー情報を前記倍率と前記話速値と前記基本周波数分析部が生成した前記基本周波数の時間変化情報をとに基づいて、話速の制御量と基本周波数の変化幅の制御量とを取得する

ことを特徴とする請求項 3 記載の音声合成装置。

【請求項 5】

前記制御情報テーブル記憶部は、話速の制御量である話速の倍率に基本周波数の変化幅の制御量である基本周波数の変化幅の倍率を対応付けた制御情報テーブルを、制御方法の

10

20

30

40

50

種別に関連付けて記憶し、

前記制御条件設定部は、所望の制御方法の種別を設定し、

前記制御情報取得部は、前記制御条件設定部が設定した前記制御方法の種別に関連付けられた制御情報テーブルを前記制御情報テーブル記憶部から読み込み、前記制御情報テーブルから話速の制御量と基本周波数の変化幅の制御量とを取得する

ことを特徴とする請求項1記載の音声合成装置。

【請求項6】

話速の制御量と基本周波数の変化幅の制御量とを対応付けた制御情報テーブルを、制御方法の種別に関連付けて記憶する制御情報テーブル記憶部を備えるコンピュータを、

音声データを取得する音声データ取得手段と、

前記音声データ取得手段が取得した前記音声データに基づいて基本周波数の時間変化情報を生成する基本周波数分析手段と、

前記音声データに基づいて話速値を計算する話速計算手段と、

所望の制御方法の種別と制御情報テーブルにおける話速の制御量または基本周波数の変化幅の制御量を示すキー情報を設定する制御条件設定手段と、

前記制御条件設定手段が設定した前記制御方法の種別に関連付けられた制御情報テーブルを前記制御情報テーブル記憶部から読み込み、前記制御条件設定手段が設定した前記キー情報を前記話速計算手段が計算した前記話速値と前記基本周波数分析手段が生成した前記基本周波数の時間変化情報を基づいて、話速の制御量と基本周波数の変化幅の制御量とを取得する制御情報取得手段と、

前記制御情報取得手段が取得した前記基本周波数の変化幅の制御量に基づいて、前記基本周波数の時間変化情報を目標基本周波数の時間変化情報を生成する目標基本周波数生成手段と、

前記制御情報取得手段が取得した前記話速の制御量と、前記目標基本周波数生成手段が生成した前記目標基本周波数の時間変化情報を基づいて、前記音声データを補正して補正後音声データを出力する音声データ補正手段と、

として機能させるためのコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力音声の韻律を補正する音声合成装置およびコンピュータプログラムに関する。

【背景技術】

【0002】

合成音声等の音声データを聞き易い音声データに変換するために、韻律に関する音響特徴量を変更することは有効である。そして、韻律に関する音響特徴量として例えばピッチ、パワー、および継続長を制御対象とし、ピッチもしくはパワー、またはピッチパターンの時間変動のダイナミックレンジを変更したりポーズ長を変えたりして、音声の音質を変化させる技術が知られている（例えば、特許文献1参照）。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開平10-97267号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

音声データの音声を聞き易い音声に変換するためには、音声データの韻律に関する複数種類の音響特徴量を補正することが有効である。しかしながら、複数種類の音響特徴量を調整して所望の韻律の音声に近づけるには、音響特徴量相互の関係や音響特徴量と聞き易さとの関係等の音響に関する専門知識や豊かな調整経験が必要である。また、このような

10

20

30

40

50

調整作業を、人手を伴って行うことは非常に複雑でかつ煩雑である。

そこで、本発明は上記事情に鑑みてなされたものであり、音響に関する専門知識や豊かな調整経験がなくとも音声データを補正でき、聴き易い音声データを容易に且つ効果的に得ることができる、音声合成装置およびコンピュータ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】

【0005】

[1] 上記の課題を解決するため、本発明の一態様である音声合成装置は、話速の制御量と基本周波数の変化幅の制御量とを対応付けた制御情報テーブルを、制御方法の種別に関連付けて記憶する制御情報テーブル記憶部と、音声データを取得する音声データ取得部と、前記音声データ取得部が取得した前記音声データに基づいて基本周波数の時間変化情報を生成する基本周波数分析部と、前記音声データに基づいて話速値を計算する話速計算部と、所望の制御方法の種別と制御情報テーブルにおける話速の制御量または基本周波数の変化幅の制御量を指定するキー情報を設定する制御条件設定部と、前記制御条件設定部が設定した前記制御方法の種別に関連付けられた制御情報テーブルを前記制御情報テーブル記憶部から読み込み、前記制御条件設定部が設定した前記キー情報を前記話速計算部が計算した前記話速値と前記基本周波数分析部が生成した前記基本周波数の時間変化情報をに基づいて、話速の制御量と基本周波数の変化幅の制御量とを取得する制御情報取得部と、前記制御情報取得部が取得した前記基本周波数の変化幅の制御量に基づいて、前記基本周波数の時間変化情報をから目標基本周波数の時間変化情報を生成する目標基本周波数生成部と、前記制御情報取得部が取得した前記話速の制御量と、前記目標基本周波数生成部が生成した前記目標基本周波数の時間変化情報をに基づいて、前記音声データを補正して補正後音声データを出力する音声データ補正部と、を備えることを特徴とする。

ここで、話速の制御量は、音声データ補正部が韻律の音響特徴量である話速を補正するための補正データまたは補正量を表すデータである。また、基本周波数の変化幅の制御量は、音声データ補正部が韻律の音響特徴量である基本周波数の変化幅、言い換えると音域を補正するための補正データまたは補正量を表すデータである。

この構成によれば、本発明の一態様である音声合成装置は、話速の制御量と基本周波数の変化幅の制御量とのいずれか一方を指定することにより、この一方の制御量に対応する他方の制御量を制御情報テーブル記憶部から取得する。そして、この音声合成装置は、取得した基本周波数の変化幅の制御量に基づいて目標基本周波数の時間変化情報を生成し、取得した話速の制御量と生成した目標基本周波数の時間変化情報をに基づいて、音声データの話速および基本周波数の変化幅に関する韻律を補正する。

よって、本発明の一態様である音声合成装置によれば、音声データが容易に且つ効率的に補正後音声データに補正される。

【0006】

[2] 上記[1]記載の音声合成装置において、話速の絶対値を話速の制御量とし、基本周波数の変化幅の絶対値を基本周波数の変化幅の制御量とし、前記制御情報テーブル記憶部は、複数の話速値の絶対値それぞれに基本周波数の変化幅の絶対値を対応付けた制御情報テーブルを制御方法の種別に関連付けて記憶することを特徴とする。

[3] 上記[1]記載の音声合成装置において、話速の倍率を話速の制御量とし、基本周波数の変化幅の倍率を基本周波数の変化幅の制御量とし、前記制御情報テーブル記憶部は、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた制御情報テーブルを制御方法の種別に関連付けて記憶し、前記制御条件設定部は、前記キー情報をについての所望の倍率をさらに設定し、前記制御情報取得部は、前記制御条件設定部が設定した前記キー情報を前記倍率と前記話速計算部が計算した前記話速値と前記基本周波数分析部が生成した前記基本周波数の時間変化情報をに基づいて、話速の制御量と基本周波数の変化幅の制御量とを取得することを特徴とする。

[4] 上記[3]記載の音声合成装置において、前記制御情報テーブル記憶部は、複数の指定話速値それぞれに対応して、複数の話速の倍率それぞれに基本周波数の変化幅の倍

10

20

30

40

50

率を対応付けた制御情報テーブルを制御方法の種別に関連付けて記憶し、前記制御情報取得部は、前記制御条件設定部が設定した前記制御方法の種別に関連し且つ前記話速計算部が計算した前記話速値と同一である指定話速値に対応する制御情報テーブルを前記制御情報テーブル記憶部から読み込み、前記制御条件設定部が設定した前記キー情報と前記倍率と前記話速値と前記基本周波数分析部が生成した前記基本周波数の時間変化情報をに基づいて、話速の制御量と基本周波数の変化幅の制御量とを取得することを特徴とする。

[5] 上記[1]記載の音声合成装置において、前記制御情報テーブル記憶部は、話速の制御量である話速の倍率に基本周波数の変化幅の制御量である基本周波数の変化幅の倍率を対応付けた制御情報テーブルを、制御方法の種別に関連付けて記憶し、前記制御条件設定部は、所望の制御方法の種別を設定し、前記制御情報取得部は、前記制御条件設定部が設定した前記制御方法の種別に関連付けられた制御情報テーブルを前記制御情報テーブル記憶部から読み込み、前記制御情報テーブルから話速の制御量と基本周波数の変化幅の制御量とを取得することを特徴とする。
10

【0007】

[6] 上記の課題を解決するため、本発明の一態様であるコンピュータプログラムは、話速の制御量と基本周波数の変化幅の制御量とを対応付けた制御情報テーブルを、制御方法の種別に関連付けて記憶する制御情報テーブル記憶部を備えるコンピュータを、音声データを取得する音声データ取得手段と、前記音声データ取得手段が取得した前記音声データに基づいて基本周波数の時間変化情報を生成する基本周波数分析手段と、前記音声データに基づいて話速値を計算する話速計算手段と、所望の制御方法の種別と制御情報テーブルにおける話速の制御量または基本周波数の変化幅の制御量を示すキー情報を設定する制御条件設定手段と、前記制御条件設定手段が設定した前記制御方法の種別に関連付けられた制御情報テーブルを前記制御情報テーブル記憶部から読み込み、前記制御条件設定手段が設定した前記キー情報を前記話速計算手段が計算した前記話速値と前記基本周波数分析手段が生成した前記基本周波数の時間変化情報をに基づいて、話速の制御量と基本周波数の変化幅の制御量とを取得する制御情報取得手段と、前記制御情報取得手段が取得した前記基本周波数の変化幅の制御量に基づいて、前記基本周波数の時間変化情報から目標基本周波数の時間変化情報を生成する目標基本周波数生成手段と、前記制御情報取得手段が取得した前記話速の制御量と、前記目標基本周波数生成手段が生成した前記目標基本周波数の時間変化情報をに基づいて、前記音声データを補正して補正後音声データを出力する音声データ補正手段と、として機能させる。
20

【発明の効果】

【0008】

本発明によれば、音響に関する専門知識や豊かな調整経験がなくとも音声データを補正でき、聴き易い音声データを容易且つ効率的に得ることができる。

【図面の簡単な説明】

【0009】

【図1】本発明の第1実施形態であ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】音素区切時間情報を表すテキストデータの例を示す図である。

【図3】話速値の絶対値に基本周波数の変化幅の絶対値を対応付けた制御情報テーブルの例を示す図である。
40

【図4】複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた制御情報テーブルの例を示す図である。

【図5】三つの指定話速値それぞれに対応して、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた制御情報テーブルの例を示す図である。

【図6】目標基本周波数生成部が、基本周波数の変化幅の制御量に基づいて生成する目標基本周波数の時間変化情報を模式的に示す図である。

【図7】同実施形態において、音声合成装置の韻律補正処理の処理手順を示すフローチャートである。

【図8】本発明の第2実施形態である音声合成装置の機能構成を示すブロック図である。
50

【図9】難聴者向けの制御情報が格納された制御情報テーブルの例を示す図である。

【図10】速聴向けの制御情報が格納された制御情報テーブルの例を示す図である。

【図11】早口向けの制御情報が格納された制御情報テーブルの例を示す図である。

【図12】雑音環境下向けの制御情報テーブルの例を示す図である。

【図13】同実施形態において、音声合成装置の韻律補正処理の処理手順を示すフローチャートである。

【発明を実施するための形態】

【0010】

以下、本発明を実施するための形態について、図面を参照して詳細に説明する。

10

[第1の実施の形態]

図1は、本発明の第1実施形態である音声合成装置の機能構成を示すブロック図である。同図に示すように、音声合成装置100は、音声データ取得部10と、音声属性情報取得部20と、音声分析部30と、制御条件情報生成部40と、目標値生成部50と、制御情報テーブル更新部60と、生成条件取得部70と、音声データ補正部80とを備える。

【0011】

音声データ取得部10は、外部から供給される音声データを取得し、この取得した音声データを音声分析部30に供給する。音声データは、アナログ音声信号を標本化周波数で標本化し量子化して得られるデジタル音声データであり、例えば不図示の録音装置によって録音されて得られた録音音声データや不図示の音声合成装置によって音声合成処理された合成音声データである。音声データは、例えばPCM(Pulse Code Modulation)データであり、フォーマットは、例えばWAVEである。

20

【0012】

音声属性情報取得部20は、音声データ取得部10に供給される音声データに関連する音声属性情報の供給を受けて取得し、その音声属性情報を音声分析部30に供給する。音声属性情報は、音声データの発話内容を文章化したテキスト情報、または音声データの音素区切時間情報である。

音声データの発話内容が「おはようございます（空白）ごきげんいかがですか（空白）」である場合、テキスト情報は、例えば“おはようございます。ごきげんいかがですか。”である。なお、“（空白）”は、その前後の音声の間に、発声されない区間があることを表す。また、テキスト情報に含まれる文章は、ひらがな表記の他、カタカナ表記、ローマ字表記、または漢字混じり表記のいずれであってもよい。また、本実施形態では、音声データの発話内容が日本語である場合について説明するが、本実施形態は外国語についても適用可能である。

30

音素区切時間情報は、音声データの音声の先頭に対応する時点を基準時としたときの、当該音声データの各音素の発声時間を示す情報である。音素とは、音韻論上の音の最小単位であり、母音および子音それぞれが1音素に対応する。また、撥音、長音、および促音それぞれも1音素に対応する。この音素区切時間情報の具体例については後述する。

【0013】

音声分析部30は、音声データ取得部10から供給される音声データと音声属性情報取得部20から供給される音声属性情報をそれぞれ取り込み、これら音声データと音声属性情報を分析して音声の基本周波数（以下、単に「基本周波数」と呼ぶ。）の時間変化情報と話速値とをそれぞれ求めて目標値生成部50に供給する。

40

【0014】

音声分析部30は、その機能構成として基本周波数分析部31と、話速計算部32とを含んで構成される。

基本周波数分析部31は、音声データに基づいて、基本周波数の時間変化情報を有声区間（声帯の振動を伴う音声である有聲音声の区間）において生成し、その生成した基本周波数の値に対し、有声区間の分析値を用いてスプライン関数等によりスムージング処理を行うことによって、変化が滑らかな基本周波数の時間変化情報を生成する。このとき、基本周波数分析部31は、基本周波数を抽出できない無声区間（声帯の振動を伴わない音声

50

である無聲音声の区間)については、この無声区間の前後の有声区間における基本周波数の値から内挿した値を用いて、無声区間における基本周波数の値を補間する。

なお、基本周波数分析部31は、スムージング処理を行わずに、生成した基本周波数の値そのものを基本周波数の時間変化情報としてもよい。

【0015】

音声属性情報がテキスト情報である場合、話速計算部32は、次の二通りの方法のいずれかによって音声データの話速値を計算する。

第1の話速値を計算する方法は、音声データとテキスト情報とに基づいて音素区切時間情報を生成し、生成した音素区切時間情報に基づいて話速を計算する方法である。具体的には、話速計算部32は、音声データから音声区間の総時間長を求め、この総時間長とテキスト情報とに基づき公知の音声認識処理技術によって音素区切時間情報を生成する。そして、話速計算部32は、音素区切時間情報から単位時間当たりのモーラ(Mora)数を求ることによって話速値(単位は、「モーラ/秒」)を得る。モーラとは、音の長さについての音韻論上の単位である。日本語では、おおむね拗音については仮名二文字が1モーラに対応し、直音については仮名一文字が1モーラに対応する。そして、1モーラは、一つまたは複数の音素により構成される。

第2の話速値を計算する方法は、音声データから音声区間の総時間長を求めるとともにテキスト情報からモーラ数を求めて、これら音声区間の総時間長とモーラ数とに基づいて話速値を計算する方法である。

【0016】

また、音声属性情報が音素区切時間情報である場合、話速計算部32は、その音素区切時間情報から単位時間当たりのモーラ数を求ることによって話速値を得る。

【0017】

制御条件情報生成部40は、ユーザによるキーボードやマウス等の入力装置の操作によって所望の制御方法の種別と所望の制御パラメータと所望の補正目的の種別との供給を受け、これら制御方法の種別と制御パラメータと補正目的の種別とを含めた制御条件情報を生成して目標値生成部50に供給する。

制御方法の種別は、話速の制御量と基本周波数の変化幅の制御量との制御方法を区別するデータである。話速の制御量は、音声データ補正部80が韻律の第1の音響特徴量である話速を補正するための補正データまたは補正量を表すデータであり、目標となる話速値の絶対値、または話速の倍率で表される。基本周波数の変化幅の制御量は、音声データ補正部80が韻律の第2の音響特徴量である基本周波数の変化幅、言い換えると音域を補正するための補正データまたは補正量を表すデータであり、基本周波数の変化幅の絶対値、または基本周波数の変化幅の倍率で表される。

制御方法の種別について具体的に説明する。本実施形態における制御方法の種別は定量的に制御量を指定する制御方法であり、下記の表1のように分類される。

【0018】

【表1】

制御対象	制御方法1	制御方法2 (汎用)	制御方法3 (指定話速値)
話速	絶対値の指定	倍率の指定	倍率の指定
基本周波数の変化幅	絶対値の指定	倍率の指定	倍率の指定

【0019】

表1において、制御方法1は、話速の制御量と基本周波数の変化幅の制御量とを数値の

10

20

30

40

50

絶対値で指定する方法である。制御方法2は、話速の制御量と基本周波数の変化幅の制御量とを倍率で指定する方法である。制御方法3は、音声分析部30から供給される話速値(指定話速値)に対応して、話速の制御量と基本周波数の変化幅の制御量とを倍率で指定する方法である。制御方法の種別は、制御方法1～制御方法3それぞれに対応付けられたデータである。

【0020】

制御パラメータは、目標値生成部50が話速の制御量と基本周波数の変化幅の制御量とを取得するために用いられるキー情報であり、具体的には次のとおりである。制御方法1に対応する制御パラメータは、話速の制御量と基本周波数の変化幅の制御量とのいずれかを指定する情報である。制御方法2および制御方法3に対応する制御パラメータは、話速の制御量と基本周波数の変化幅の制御量とのいずれかを指定する情報と所望の倍率である。

10

【0021】

補正目的の種別は、韻律制御の目的を指定する情報、言い換えると、目標値生成部50が取得する話速の制御量と基本周波数の変化幅の制御量との補正傾向を指定する情報である。具体的には、補正目的の種別は、例えば、「難聴者向け」、「速聴向け」、「早口向け」を指定する情報である。

【0022】

目標値生成部50は、音声分析部30から供給される話速値と基本周波数の時間変化情報を取り込み、制御条件情報生成部40から供給される制御条件情報にしたがって、話速の制御量と基本周波数の変化幅の制御量とを求める。さらに、目標値生成部50は、基本周波数の変化幅の制御量に基づいて目標基本周波数の時間変化情報を生成する。そして、目標値生成部50は、話速の制御量と目標基本周波数の時間変化情報を音声データ補正部80に供給する。

20

【0023】

次に、詳細に目標値生成部50を説明する。目標値生成部50は、その機能構成として、制御情報テーブル記憶部51と、制御条件設定部52と、制御情報取得部53と、目標基本周波数生成部54とを含んで構成される。

【0024】

制御情報テーブル記憶部51は、話速の制御量と基本周波数の変化幅の制御量とを対応付けた複数の制御情報テーブルを、制御方法の種別と補正目的の種別とに関連付けて記憶する。制御情報テーブルは、話速値の絶対値に基本周波数の変化幅の絶対値を対応付けたテーブルと、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けたテーブルと、複数の指定話速値それぞれに対応して、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けたテーブルとの3種類である。各種制御情報テーブルの具体例については後述する。

30

制御情報テーブル記憶部51は、読み出しおよび書き込み可能な記憶装置であり、例えば、磁気ハードディスク装置、半導体記憶装置等である。また、制御情報テーブル記憶部51は、光磁気ディスク等の可搬型記録媒体であってもよい。

40

【0025】

制御条件設定部52は、制御条件情報生成部40から供給される制御条件情報を取り込み、その制御条件情報に含まれる制御方法の種別と制御パラメータと補正目的の種別とを制御情報取得部53に設定する。

【0026】

制御情報取得部53は、制御条件設定部52により設定された制御方法の種別と補正目的の種別とに対応付けられた制御情報テーブルを制御情報テーブル記憶部51から読み込む。そして、制御情報取得部53は、読み込んだ制御情報テーブルと、制御条件設定部52により設定された制御パラメータと、音声分析部30からそれぞれ供給された話速値および基本周波数の時間変化情報を基づいて、話速の制御量と基本周波数の変化幅の制御量とを求める。さらに、制御情報取得部53は、話速の制御量を音声データ補正部80に

50

供給するとともに、基本周波数の変化幅の制御量を目標基本周波数生成部 54 に供給する。

【 0 0 2 7 】

具体的には、制御条件設定部 52 により制御方法の種別（例えば、制御方法 1）と制御パラメータ（例えば、キー情報は話速の制御量を指定する情報である。）と補正目的の種別（例えば、難聴者向け）とが設定された場合、制御情報取得部 53 は、制御方法 1 と難聴者向けとに対応付けられた制御情報テーブル（話速値の絶対値に基本周波数の変化幅の絶対値を対応付けた難聴者向けテーブル）を制御情報テーブル記憶部 51 から読み込む。そして、制御情報取得部 53 は、制御パラメータにしたがい、音声分析部 30 から供給された話速値の絶対値をキーとして、そのキーに対応付けられた制御量である基本周波数の変化幅の絶対値を制御情報テーブルから抽出する。そして、制御情報取得部 53 は、キーである話速値の絶対値を話速の制御量とし、制御情報テーブルから抽出した基本周波数の変化幅の絶対値を基本周波数の変化幅の制御量とする。

10

【 0 0 2 8 】

また、制御条件設定部 52 により制御方法の種別（例えば、制御方法 1）と制御パラメータ（例えば、キー情報は基本周波数の変化幅の制御量を指定する情報である。）と補正目的の種別（例えば、難聴者向け）とが設定された場合、制御情報取得部 53 は、制御方法 1 と難聴者向けとに対応付けられた制御情報テーブル（話速値の絶対値に基本周波数の変化幅の絶対値を対応付けた難聴者向けテーブル）を制御情報テーブル記憶部 51 から読み込む。そして、制御情報取得部 53 は、制御パラメータにしたがい、音声分析部 30 から供給された基本周波数の時間変化情報に基づき得られる基本周波数の変化幅の絶対値をキーとして、そのキーに対応付けられた制御量である話速値の絶対値を制御情報テーブルから抽出する。そして、制御情報取得部 53 は、キーである基本周波数の変化幅の絶対値を基本周波数の変化幅の制御量とし、制御情報テーブルから抽出した話速値の絶対値を話速の制御量とする。

20

【 0 0 2 9 】

また、制御条件設定部 52 により制御方法の種別（例えば、制御方法 2）と制御パラメータ（例えば、キー情報は話速の制御量を指定する情報であり、倍率も指定されている。）と補正目的の種別（例えば、速聴向け）とが設定された場合、制御情報取得部 53 は、制御方法 2 と速聴向けとに対応付けられた制御情報テーブル（複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた速聴向けテーブル）を制御情報テーブル記憶部 51 から読み込む。そして、制御情報取得部 53 は、制御パラメータによる話速の倍率をキーとして、そのキーに対応付けられた制御量である基本周波数の変化幅の倍率を制御情報テーブルから抽出する。そして、制御情報取得部 53 は、制御パラメータによる話速の倍率を話速の制御量とし、制御情報テーブルから抽出した基本周波数の変化幅の倍率を基本周波数の変化幅の制御量とする。

30

【 0 0 3 0 】

また、制御条件設定部 52 により制御方法の種別（例えば、制御方法 2）と制御パラメータ（例えば、キー情報は基本周波数の変化幅の制御量を指定する情報であり、倍率も指定されている。）と補正目的の種別（例えば、速聴向け）とが設定された場合、制御情報取得部 53 は、制御方法 2 と速聴向けとに対応付けられた制御情報テーブル（複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた速聴向けテーブル）を制御情報テーブル記憶部 51 から読み込む。そして、制御情報取得部 53 は、制御パラメータによる基本周波数の変化幅の倍率をキーとして、そのキーに対応付けられた制御量である話速の倍率を制御情報テーブルから抽出する。そして、制御情報取得部 53 は、制御パラメータによる基本周波数の変化幅の倍率を基本周波数の変化幅の制御量とし、制御情報テーブルから抽出した話速の倍率を話速の制御量とする。

40

【 0 0 3 1 】

また、制御条件設定部 52 により制御方法の種別（例えば、制御方法 3）と制御パラメータ（例えば、キー情報は話速の制御量を指定する情報であり、倍率も指定されている。）

50

)と補正目的の種別(例えば、早口向け)とが設定された場合、制御情報取得部53は、制御方法3と早口向けとに対応付けられて且つ音声分析部30から供給された話速値と同一である指定話速値の制御情報テーブル(指定話速値に対応して、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた早口向けテーブル)を、制御情報テーブル記憶部51から読み込む。そして、制御情報取得部53は、制御パラメータによる話速の倍率をキーとして、そのキーに対応付けられた制御量である基本周波数の変化幅の倍率を制御情報テーブルから抽出する。そして、制御情報取得部53は、制御パラメータによる話速の倍率を話速の制御量とし、制御情報テーブルから抽出した基本周波数の変化幅の倍率を基本周波数の変化幅の制御量とする。

【0032】

また、制御条件設定部52により制御方法の種別(例えば、制御方法3)と制御パラメータ(例えば、キー情報は基本周波数の変化幅の制御量を指定する情報であり、倍率も指定されている。)と補正目的の種別(例えば、早口向け)とが設定された場合、制御情報取得部53は、制御方法3と早口向けとに対応付けられて且つ音声分析部30から供給された話速値と同一である指定話速値の制御情報テーブル(指定話速値に対応して、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた早口向けテーブル)を、制御情報テーブル記憶部51から読み込む。そして、制御情報取得部53は、制御パラメータによる基本周波数の変化幅の倍率をキーとして、そのキーに対応付けられた制御量である話速値の倍率を制御情報テーブルから抽出する。そして、制御情報取得部53は、制御パラメータによる基本周波数の変化幅の倍率を基本周波数の変化幅の制御量とし、制御情報テーブルから抽出した話速の倍率を話速の制御量とする。

【0033】

目標基本周波数生成部54は、制御情報取得部53から供給された基本周波数の変化幅の制御量に基づいて、基本周波数分析部31で生成された基本周波数の時間変化情報から目標基本周波数の時間変化情報を生成して音声データ補正部80に供給する。このとき、目標基本周波数生成部54は、内部に記憶している生成条件情報にしたがって目標基本周波数の時間変化情報を生成する。

生成条件情報は、目標基本周波数生成部54が目標基本周波数の時間変化情報を生成するときの条件を指定する情報であり、平均値固定指定と、最小値固定指定と、指定値固定指定とのいずれかが指定される。

【0034】

平均値固定指定は、基本周波数分析部31が生成した基本周波数の時間変化情報の所定区間ににおける周波数値の平均値と同一の平均値となるようにして、当該区間ににおける目標基本周波数の時間変化情報を生成することを指定するものである。

最小値固定指定は、基本周波数分析部31が生成した基本周波数の時間変化情報の所定区間ににおける周波数値の最小値を基準とし、その最小値以上の周波数値となるように、当該区間ににおける目標基本周波数の時間変化情報を生成することを指定するものである。

指定値固定指定には次の二通りの指定がある。第1の指定値固定指定は、基本周波数分析部31が生成した基本周波数の時間変化情報の所定区間ににおける任意の一つの周波数値を基準として、当該区間ににおける目標基本周波数の時間変化情報を生成することを指定するものである。

また、第2の指定値固定指定は、所定区間の始点と終点とのそれぞれに対応する二つの周波数値を基準として、当該区間ににおける目標基本周波数の時間変化情報を生成することを指定するものである。

【0035】

目標基本周波数生成部54は、生成条件取得部70から生成条件情報の供給を受けると、その生成条件情報で内部に記憶した生成条件情報を書き換える。

【0036】

目標基本周波数生成部54は、制御情報取得部53から供給される基本周波数の変化幅の制御量が「基本周波数の変化幅の数値の絶対値」である場合、生成条件取得部70から

10

20

30

40

50

供給される生成条件情報にしたがって、周波数の変化幅が「基本周波数の変化幅の数値の絶対値」に合致する目標基本周波数の時間変化情報を生成する。

また、目標基本周波数生成部54は、制御情報取得部53から供給される基本周波数の変化幅の制御量が「基本周波数の変化幅の倍率」である場合、生成条件取得部70から供給される生成条件情報にしたがって、周波数の変化幅が「基本周波数の変化幅の倍率」に適合する目標基本周波数の時間変化情報を生成する。

【0037】

制御情報テーブル更新部60は、ユーザによるキーボードやマウス等の入力装置の操作によって制御情報テーブルの供給を受け、その制御情報テーブルを制御情報テーブル記憶部51に新規に書き込んだり、その制御情報テーブルで、既に記憶されている制御情報テーブルを書き換えたりする。

生成条件取得部70は、ユーザによるキーボードやマウス等の入力装置の操作によって平均値固定指定と最小値固定指定と第1の指定値固定指定と第2の指定値固定指定とのいずれかの指定を受け、その指定を示す生成条件情報を生成して目標基本周波数生成部54に供給する。

【0038】

音声データ補正部80は、制御情報取得部53から供給される話速の制御量と、目標基本周波数生成部54から供給される目標基本周波数の時間変化情報をそれぞれ取り込み、音声データ取得部10が取得した音声データの韻律を補正して補正後音声データを出力する。

具体的には、例えば、音声データ補正部80は、音声データに目標基本周波数の時間変化情報を適用して基本周波数の時間変化情報を補正する。そして、音声データ補正部80は、話速の制御量（話速値の絶対値または話速の倍率）に基づいて話速計算部32が算出した話速値を補正し、音声データの音素区切時間情報をその補正した話速値に基づいて補正する。そして、音声データ補正部80は、基本周波数の時間変化情報が補正された音声データに、補正後の音素区切時間情報を適用して時間軸方向を補正した補正後音声データを生成する。

【0039】

図2は、音素区切時間情報を表すテキストデータの例を示す図である。同図に示すように、音素区切時間情報は複数の行からなるテキストデータであり、各行が音素に対応している。この音素区切時間情報の1行目から17行目までのデータは、音声データにおける“（空白）おはようございます（空白）”の部分についての音素の区切りを示す時間情報である。

また、音素区切時間情報の各行は、区切り文字（例えば、スペースやタブ等）で区切られた3つの列を有している。一列目は、音声データの先頭に対応する時点を基準時として各音素の開始時点までの時間を1万分の1秒単位で表し、二列目は上記基準時から各音素の終了時点までの時間を1万分の1秒単位で表し、三列目は音素の音素ラベルを表す。例えば、同図において、“0 2740 s i l”は、基準時から0.274秒経過するまでの間が無声区間であることを表す。また、“2740 3168 o(お)”は、基準時から0.274秒経過した時点から、基準時から0.3168秒経過するまでの間の音素が“o(お)”であることを表す。なお、音素ラベル“s i l”は音素がないことを表す。

なお、ここでは、時刻が1万分の1秒単位である場合を一例として説明したが、1千分の1秒単位（ミリ秒）等、他の単位で表すようにしてもよい。

【0040】

図3(a)～(c)のそれぞれは、話速値の絶対値に基本周波数の変化幅の絶対値を対応付けた制御情報テーブルの例を示す図である。同図(a)～(c)それぞれの制御情報テーブルは、制御方法1を示す制御方法の種別に関連付けられている。これら制御情報テーブルにおいて、話速値の絶対値は話速の制御量であり、基本周波数の変化幅の絶対値は基本周波数の変化幅の制御量である。同図(a)は難聴者向けテーブルであり、「難聴者

「向け」を示す補正目的の種別に関係付けられている。また、同図(b)は速聴向けテーブルであり、「速聴向け」を示す補正目的の種別に関係付けられている。また、同図(c)は早口向けテーブルであり、「早口向け」を示す補正目的の種別に関係付けられている。

【 0 0 4 1 】

図4(a)～(c)のそれぞれは、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた制御情報テーブルの例を示す図である。同図(a)～(c)それぞれの制御情報テーブルは、制御方法2を示す制御方法の種別に関連付けられている。これら制御情報テーブルにおいて、話速値の倍率は話速の制御量であり、基本周波数の変化幅の倍率は基本周波数の変化幅の制御量である。同図(a)は難聴者向けテーブルであり、「難聴者向け」を示す補正目的の種別に関係付けられている。また、同図(b)は速聴向けテーブルであり、「速聴向け」を示す補正目的の種別に関係付けられている。また、同図(c)は早口向けテーブルであり、「早口向け」を示す補正目的の種別に関係付けられている。

10

【 0 0 4 2 】

図5(a)～(c)のそれぞれは、三つの指定話速値それぞれに対応して、複数の話速の倍率それぞれに基本周波数の変化幅の倍率を対応付けた制御情報テーブルの例を示す図である。同図(a)～(c)それぞれの制御情報テーブルは、制御方法3を示す制御方法の種別に関連付けられている。これら制御情報テーブルにおいて、話速値の倍率は話速の制御量であり、基本周波数の変化幅の倍率は基本周波数の変化幅の制御量である。同図(a)は難聴者向けテーブルであり、「難聴者向け」を示す補正目的の種別に関係付けられている。また、同図(b)は速聴向けテーブルであり、「速聴向け」を示す補正目的の種別に関係付けられている。また、同図(c)は早口向けテーブルであり、「早口向け」を示す補正目的の種別に関係付けられている。同図(a)、(b)、(c)とも、指定話速値が6, 8, 10(モーラ/秒)である例を表す。

20

【 0 0 4 3 】

なお、図3から図5までの制御情報テーブルの例は、話速計算部32から供給される音声データの話速値が7モーラ/秒から8モーラ/秒程度である場合の例である。

20

【 0 0 4 4 】

図6は、目標基本周波数生成部54が、制御情報取得部53から供給される基本周波数の変化幅の制御量に基づいて生成する目標基本周波数の時間変化情報を模式的に示す図である。

30

同図(a)は、生成条件情報が平均値固定指定を示す情報である場合に、目標基本周波数生成部54が生成した目標基本周波数の時間変化情報の例である。同図(a)は、時刻t1から時刻t2まで($t_1 < t_2$)の時間における音声データの基本周波数の時間変化情報1と、この基本周波数の時間変化情報1を基本周波数の変化幅の制御量に基づいて、拡大した基本周波数の時間変化情報1aと、縮小した基本周波数の時間変化情報1rとを示す。同図(a)において、基本周波数の時間変化情報1の時刻t1から時刻t2までにおける周波数値の平均値は平均値 f_{ave} である。同図(a)に示すように、目標基本周波数生成部54は、時刻t1から時刻t2までの基本周波数の時間変化情報の平均値が平均値 f_{ave} になるように、基本周波数の時間変化情報2を拡大したり縮小したりする。

40

【 0 0 4 5 】

図6(b)は、生成条件情報が最小値固定指定を示す情報である場合に、目標基本周波数生成部54が生成した目標基本周波数の時間変化情報の例である。同図(b)は、時刻t1から時刻t2まで($t_1 < t_2$)の時間における音声データの基本周波数の時間変化情報2と、この基本周波数の時間変化情報2を基本周波数の変化幅の制御量に基づいて、拡大した基本周波数の時間変化情報2aと、縮小した基本周波数の時間変化情報2rとを示す。同図(b)において、基本周波数の時間変化情報2の時刻t1から時刻t2までにおける周波数値の最小値は最小値 f_{min} である。同図(b)に示すように、目標基本周波数生成部54は、時刻t1から時刻t2までの基本周波数の時間変化情報2を周波数値の最小値 f_{min} を基準として拡大したり縮小したりする。

50

【0046】

図6(c)は、生成条件情報が第1の指定値固定指定を示す情報である場合に、目標基本周波数生成部54が生成した目標基本周波数の時間変化情報の例である。同図(c)は、時刻t1から時刻t2まで($t_1 < t_2$)の時間における音声データの基本周波数の時間変化情報3と、この基本周波数の時間変化情報3を基本周波数の変化幅の制御量に基づいて、拡大した基本周波数の時間変化情報3aと、縮小した基本周波数の時間変化情報3rとを示す。同図(c)に示すように、目標基本周波数生成部54は、時刻t1から時刻t2までの基本周波数の時間変化情報3を、任意の周波数値 f_s を基準として拡大したり縮小したりする。

【0047】

10

図6(d)は、生成条件情報が第2の指定値固定指定を示す情報である場合に、目標基本周波数生成部54が生成した目標基本周波数の時間変化情報の例である。同図(d)は、時刻t1から時刻t2まで($t_1 < t_2$)の時間における音声データの基本周波数の時間変化情報4と、この基本周波数の時間変化情報4を基本周波数の変化幅の制御量に基づいて、拡大した基本周波数の時間変化情報4aと、縮小した基本周波数の時間変化情報4rとを示す。同図(d)に示すように、目標基本周波数生成部54は、時刻t1における周波数値 f_{s1} と時刻t2($t_1 < t_2$)における周波数値 f_{s2} とを基準として、時刻t1から時刻t2までの基本周波数の時間変化情報4を拡大したり縮小したりする。

【0048】

20

次に、第1実施形態である音声合成装置100の動作について説明する。図7は、音声合成装置100の韻律補正処理の処理手順を示すフローチャートである。

まず、ステップS101において、制御条件情報生成部40は、ユーザによるキーボードやマウス等の入力装置の操作によって所望の制御方法の種別と所望の制御パラメータと補正目的の種別との供給を受けると、これら制御方法の種別と制御パラメータと補正目的の種別とを含めた制御条件情報を生成して目標値生成部50に供給する。

次に、ステップS102において、目標値生成部50の制御条件設定部52は、制御条件情報生成部40から供給される制御条件情報を取り込み、その制御条件情報に含まれる制御方法の種別と制御パラメータと補正目的の種別とを制御情報取得部53に設定する。

次に、ステップS103において、音声データ取得部10は、音声データの供給を受けると、その音声データを取り込んで音声分析部30に供給する。さらに、音声属性情報取得部20は、音声データ取得部10に供給される音声データに関連する音声属性情報の供給を受けると、その音声属性情報を取り込んで音声分析部30に供給する。

30

【0049】

次に、ステップS104において、音声分析部30は、音声データ取得部10から供給される音声データと音声属性情報取得部20から供給される音声属性情報とをそれぞれ取り込むと、これら音声データと音声属性情報とを分析し、音声の基本周波数の時間変化情報と話速値とをそれぞれ求めて目標値生成部50に供給する。詳細には、音声分析部30の基本周波数分析部31は、音声データに基づいて基本周波数の時間変化情報を生成して目標値生成部52に供給する。また、音声属性情報がテキスト情報である場合、音声分析部30の話速計算部32は、音声データとテキスト情報とに基づいて音素区切時間情報を生成し、生成した音素区切時間情報に基づいて話速を計算する(第1の話速値を計算する方法)か、音声データから音声区間の総時間長を求めるとともにテキスト情報からモーラ数を求めて、これら音声区間の総時間長とモーラ数とに基づいて話速値を計算する(第2の話速値を計算する方法)。また、音声属性情報が音素区切時間情報である場合、話速計算部32は、その音素区切時間情報から単位時間当たりのモーラ数を求ることによって話速値を得る。

40

【0050】

次に、ステップS105において、制御情報取得部53は、制御条件設定部52により設定された制御方法の種別と補正目的の種別とに対応付けられた制御情報テーブルを制御情報テーブル記憶部51から読み込む。次に、制御情報取得部53は、読み込んだ制御情

50

報テーブルと、制御条件設定部 52 により設定された制御パラメータと、音声分析部 30 からそれぞれ供給された話速値および基本周波数の時間変化情報とに基づいて、話速の制御量と基本周波数の変化幅の制御量とを求める。次に、制御情報取得部 53 は、話速の制御量を音声データ補正部 80 に供給し、基本周波数の変化幅の制御量を目標基本周波数生成部 54 に供給する。

【0051】

次に、ステップ S106において、目標基本周波数生成部 54 は、制御情報取得部 53 から供給された基本周波数の変化幅の制御量に基づいて、基本周波数分析部 31 で生成された基本周波数の時間変化情報から目標基本周波数の時間変化情報を生成して音声データ補正部 80 に供給する。このとき、目標基本周波数生成部 54 は、内部に記憶している生成条件情報にしたがって目標基本周波数の時間変化情報を生成する。10

次に、ステップ S107において、音声データ補正部 80 は、制御情報取得部 53 から供給された話速の制御量と、目標基本周波数生成部 54 から供給された目標基本周波数の時間変化情報をとをそれぞれ取り込むと、音声データ取得部 10 が取得した音声データの韻律を補正して補正後音声データを出力する。

【0052】

以上説明したように、第1実施形態である音声合成装置 100 は、ユーザによる入力装置の操作により話速の制御量と基本周波数の変化幅の制御量とのいずれか一方と補正目的とを指定することによって、補正目的に応じて、指定した一方の制御量に対応する他方の制御量を容易に且つ効率的に取得することができる。そして、音声合成装置 100 は、それぞれ取得した話速の制御量と基本周波数の変化幅の制御量とに基づいて、音声データの話速および基本周波数の変化幅に関する韻律をユーザの補正目的に沿って補正することができる。20

したがって、本実施形態である音声合成装置 100 によれば、音響に関する専門知識や豊かな調整経験を必要とせず、容易に且つ効率的に音声データを聞き易い補正後音声データに補正することができる。

【0053】

[第2の実施の形態]

図 8 は、本発明の第2実施形態である音声合成装置の機能構成を示すブロック図である。なお、本実施形態において、第1実施形態である音声合成装置 100 の構成と同一の構成については、同一の符号を付してその説明を省略する。同図に示す音声合成装置 100a は、音声合成装置 100 の構成における制御条件情報生成部 40 と目標値生成部 50 とを、それぞれ、制御条件情報生成部 40a と目標値生成部 50a とに変更したものである。30

【0054】

制御条件情報生成部 40a は、ユーザによるキーボードやマウス等の入力装置の操作によって所望の制御方法の種別の供給を受け、この制御方法の種別を含めた制御条件情報を生成して目標値生成部 50a に供給する。

本実施形態における制御方法の種別について具体的に説明する。本実施形態における制御方法の種別は定性的に制御量を指定する制御方法であり、下記の表 2 のように分類される。40

【0055】

【表2】

制御対象	制御方法A (難聴者向け)	制御方法B (速聴向け)	制御方法C (早口向け)	制御方法D (雑音環境下)
話速	遅く	速く	少し速く	固定
基本周波数の変化幅	拡大	少し縮小	縮小	拡大

10

【0056】

表2において、制御方法Aは、音声データの話速を「遅く」するよう指定するとともに、音声データの基本周波数の変化幅を「拡大」するよう指定する方法である。具体的には、例えば、制御方法Aは、音声分析部30から供給される話速値を約0.5倍から約0.9倍までの間にするよう指定するとともに、音声分析部30から供給される基本周波数の時間変化情報に基づき得られる基本周波数の変化幅を約1.15倍から約1.05倍までの間にするよう指定する方法である。

この制御方法Aは、老人性難聴の症状を有する高齢者や他の難聴者等に、ゆっくりとした速度で且つ音域を拡張した聞き取り易い音声を提供する場合に好適な制御方法である。

20

【0057】

また、表2において、制御方法Bは、音声データの話速を「速く」するよう指定するとともに、音声データの基本周波数の変化幅を「少し縮小」するよう指定する方法である。具体的には、例えば、制御方法Bは、音声分析部30から供給される話速値を約1.5倍から約2.5倍までの間にするよう指定するとともに、音声分析部30から供給される基本周波数の時間変化情報に基づき得られる基本周波数の変化幅を約0.97倍から約0.9倍までの間にするよう指定する方法である。

この制御方法Bは、聞き取り可能な程度の速い速度で且つ音域を少し狭めた速聴向けの音声を提供する場合に好適な制御方法である。

30

【0058】

また、表2において、制御方法Cは、音声データの話速を「少し速く」するよう指定するとともに、音声データの基本周波数の変化幅を「縮小」するよう指定する方法である。具体的には、例えば、制御方法Cは、音声分析部30から供給される話速値を約1.1倍から約1.5倍までの間にするよう指定するとともに、音声分析部30から供給される基本周波数の時間変化情報に基づき得られる基本周波数の変化幅を約0.96倍から約0.8倍までの間にするよう指定する方法である。

この制御方法Cは、違和感を与えない程度の速い速度で且つ音域を狭めた早口音声を提供する場合に好適な制御方法である。

40

【0059】

表2において、制御方法Dは、音声データの話速を「固定」するよう指定するとともに、音声データの基本周波数の変化幅を「拡大」するよう指定する方法である。具体的には、例えば、制御方法Dは、音声分析部30から供給される話速値を変更しないよう指定するとともに、音声分析部30から供給される基本周波数の時間変化情報に基づき得られる基本周波数の変化幅を約1.1倍から約1.2倍までの間にするよう指定する方法である。

この制御方法Dは、話速を変えることなく且つ音域を拡張して、例えば騒音・雑音環境下において聞き取り易い音声を提供する場合に好適な制御方法である。

【0060】

目標値生成部50aは、音声分析部30から供給される話速値と基本周波数の時間変化情報とを取り込み、制御条件情報生成部40aから供給される制御条件情報にしたがって

50

、話速の制御量と基本周波数の変化幅の制御量とを求める。さらに、目標値生成部50aは、基本周波数の変化幅の制御量に基づいて目標基本周波数の時間変化情報を生成する。そして、目標値生成部50aは、話速の制御量と目標基本周波数の時間変化情報を音声データ補正部80に供給する。

【0061】

次に、詳細に目標値生成部50aを説明する。目標値生成部50aは、音声合成装置100の目標値生成部50における制御情報テーブル記憶部51と制御条件設定部52と制御情報取得部53とを、制御情報テーブル記憶部51aと制御条件設定部52aと制御情報取得部53aとに変更したものである。

【0062】

制御情報テーブル記憶部51aは、話速の制御量と基本周波数の変化幅の制御量とを対応付けた複数の制御情報テーブルを、制御方法の種別に対応付けて記憶する。制御情報テーブルは、補正目的（例えば、難聴者向け、速聴、早口、雑音環境下）ごとに、話速の倍率に基本周波数の変化幅の倍率を対応付けたテーブルである。補正目的ごとの種制御情報テーブルの具体例については後述する。

制御情報テーブル記憶部51aは、読み出しおよび書き込み可能な記憶装置であり、例えば、磁気ハードディスク装置、半導体記憶装置等である。また、制御情報テーブル記憶部51aは、光磁気ディスク等の可搬型記録媒体であってもよい。

【0063】

制御条件設定部52aは、制御条件情報生成部40aから供給される制御条件情報を取り込み、その制御条件情報に含まれる制御方法の種別を制御情報取得部53aに設定する。

【0064】

制御情報取得部53aは、制御条件設定部52aにより設定された制御方法の種別に対応付けられた制御情報テーブルを制御情報テーブル記憶部51aから読み込む。そして、制御情報取得部53aは、読み込んだ制御情報テーブルに格納された話速の制御量を音声データ補正部80に供給するとともに、制御情報テーブルに格納された基本周波数の変化幅の制御量を目標基本周波数生成部54に供給する。

【0065】

具体的には、例えば、制御条件設定部52aにより制御方法の種別（制御方法A）が設定された場合、制御情報取得部53aは、制御方法Aに関係付けられた制御情報テーブル（難聴者向けの制御情報が格納されたテーブル）を制御情報テーブル記憶部51aから読み込む。そして、制御情報取得部53aは、読み込んだ制御情報テーブルに格納された話速の倍率と基本周波数の変化幅の倍率とを、それぞれ話速の制御量と基本周波数の変化幅の制御量とする。

【0066】

また、例えば、制御条件設定部52aにより制御方法（制御方法B）が設定された場合、制御情報取得部53aは、制御方法Bに関連付けられた制御情報テーブル（速聴向けの制御情報が格納されたテーブル）を制御情報テーブル記憶部51aから読み込む。そして、制御情報取得部53aは、読み込んだ制御情報テーブルに格納された話速の倍率と基本周波数の変化幅の倍率とを、それぞれ話速の制御量と基本周波数の変化幅の制御量とする。

【0067】

また、例えば、制御条件設定部52aにより制御方法（制御方法C）が設定された場合、制御情報取得部53aは、制御方法Cに関連付けられた制御情報テーブル（早口向けの制御情報が格納されたテーブル）を制御情報テーブル記憶部51aから読み込む。そして、制御情報取得部53aは、読み込んだ制御情報テーブルに格納された話速の倍率と基本周波数の変化幅の倍率とを、それぞれ話速の制御量と基本周波数の変化幅の制御量とする。

【0068】

10

20

30

40

50

また、例えば、制御条件設定部 52a により制御方法（制御方法 D）が設定された場合、制御情報取得部 53a は、制御方法 D に関連付けられた制御情報テーブル（雑音環境下向けのテーブル）を制御情報テーブル記憶部 51a から読み込む。そして、制御情報取得部 53a は、話速の倍率を 1 倍として話速の制御量とし、読み込んだ制御情報テーブルに格納された基本周波数の変化幅の倍率を基本周波数の変化幅の制御量とする。

【0069】

図 9 は、難聴者向けの制御情報が格納された制御情報テーブルの例を示す図である。同図に示す難聴者向けの制御情報テーブルは、音声データの話速を 0.7 倍する指定と、音声データの基本周波数の変化幅を 1.15 倍する指定とを含んでいる。

図 10 は、速聴向けの制御情報が格納された制御情報テーブルの例を示す図である。同図に示す速聴向けの制御情報テーブルは、音声データの話速を 2.0 倍する指定と、音声データの基本周波数の変化幅を 0.93 倍する指定とを含んでいる。

図 11 は、早口向けの制御情報が格納された制御情報テーブルの例を示す図である。同図に示す早口向けの制御情報テーブルは、音声データの話速を 1.4 倍する指定と、音声データの基本周波数の変化幅を 0.84 倍する指定とを含んでいる。

図 12 は、雑音環境下向けの制御情報テーブルの例を示す図である。同図に示す雑音環境下向けの制御情報テーブルは、音声データの話速を固定（1.0 倍）する指定と、音声データの基本周波数の変化幅を 1.15 倍する指定とを含んでいる。

なお、図 9 から図 12 までの制御情報テーブルの例は、話速計算部 32 から供給される音声データの話速値が 7 モーラ / 秒から 8 モーラ / 秒程度である場合の例である。

【0070】

次に、第 2 実施形態である音声合成装置 100a の動作について説明する。図 13 は、音声合成装置 100a の韻律補正処理の処理手順を示すフローチャートである。なお、本実施形態において、第 1 実施形態である音声合成装置 100 の処理手順と同一の処理については、同一の符号を付してその説明を省略する。

まず、ステップ S101a において、制御条件情報生成部 40a は、ユーザによるキーボードやマウス等の入力装置の操作によって所望の制御方法の種別の供給を受けると、この制御方法の種別を含めた制御条件情報を生成して目標値生成部 50a に供給する。

次に、ステップ S102a において、目標値生成部 50a の制御条件設定部 52a は、制御条件情報生成部 40a から供給される制御条件情報を取り込み、その制御条件情報に含まれる制御方法の種別を制御情報取得部 53a に設定する。

【0071】

ステップ S105a において、制御情報取得部 53a は、制御条件設定部 52a により設定された制御方法の種別に関連付けられた制御情報テーブルを制御情報テーブル記憶部 51a から読み込む。次に、制御情報取得部 53a は、読み込んだ制御情報テーブルに格納された話速の制御量を音声データ補正部 80 に供給し、制御情報テーブルに格納された基本周波数の変化幅の制御量を目標基本周波数生成部 54 に供給する。

【0072】

以上説明したように、第 2 実施形態である音声合成装置 100a は、ユーザによる入力装置の操作により補正目的に応じた制御方法の種別を指定することによって、その補正目的に対応した話速の制御量と基本周波数の変化幅の制御量とを容易に且つ効率的に取得することができる。そして、音声合成装置 100a は、それぞれ取得した話速の制御量と基本周波数の変化幅の制御量とに基づいて、音声データの話速および基本周波数の変化幅に関する韻律をユーザの補正目的に沿って補正することができる。

したがって、本実施形態である音声合成装置 100a によれば、音響に関する専門知識や豊かな調整経験を必要とせず、容易に且つ効率的に音声データを聞き易い補正後音声データに補正することができる。

【0073】

なお、第 1 実施形態である音声合成装置 100 と第 2 実施形態である音声合成装置 100a とのそれぞれは、音声属性情報取得部 20 を搭載しない構成としてもよい。その場合

10

20

30

40

50

、音声分析部30の話速計算部32は、以下の処理を行って音声データの話速値を計算する。つまり、話速計算部32は、音声データから公知の音声認識処理を適用してテキスト情報を生成するか、または音声データから音素区切時間情報を生成する。そして、話速計算部32は、第1実施形態に詳述したように、音声データとテキスト情報とに基づいて音素区切時間情報を生成し、生成した音素区切時間情報に基づいて話速を計算する（第1の話速値を計算する方法）か、音声データから音声区間の総時間長を求めるとともにテキスト情報からモーラ数を求めて、これら音声区間の総時間長とモーラ数とに基づいて話速値を計算する（第2の話速値を計算する方法）。または、話速計算部32は、音素区切時間情報から単位時間当たりのモーラ数を求ることによって話速値を得る。

【0074】

10

また、第2実施形態である音声合成装置100, 100aの制御情報テーブル記憶部51aは、話速の倍率と基本周波数の変化幅の倍率とを対応付けた複数の制御情報テーブルを、制御方法の種別に対応付けて記憶するものであった。これ以外にも、制御情報テーブル記憶部51aは、話速値の絶対値と基本周波数の変化幅の絶対値とを対応付けた複数の制御情報テーブルを、制御方法の種別に対応付けて記憶するようにしてもよい。

【0075】

また、音声合成装置100, 100aの一部の機能をコンピュータで実現するようにしてもよい。この場合、その機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することによって実現してもよい。なお、ここでいう「コンピュータシステム」とは、OS(Operating System)や周辺機器のハードウェアを含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、光ディスク、メモリカード等の可搬型記録媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時間の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持するものを含んでもよい。また上記のプログラムは、前述した機能の一部を実現するためのものであってもよく、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせにより実現するものであってもよい。

20

【0076】

30

以上、本発明の実施形態について図面を参照して詳述したが、具体的な構成はこの実施形態に限られるものではなく、本発明の要旨を逸脱しない範囲の設計等も含まれる。

【符号の説明】

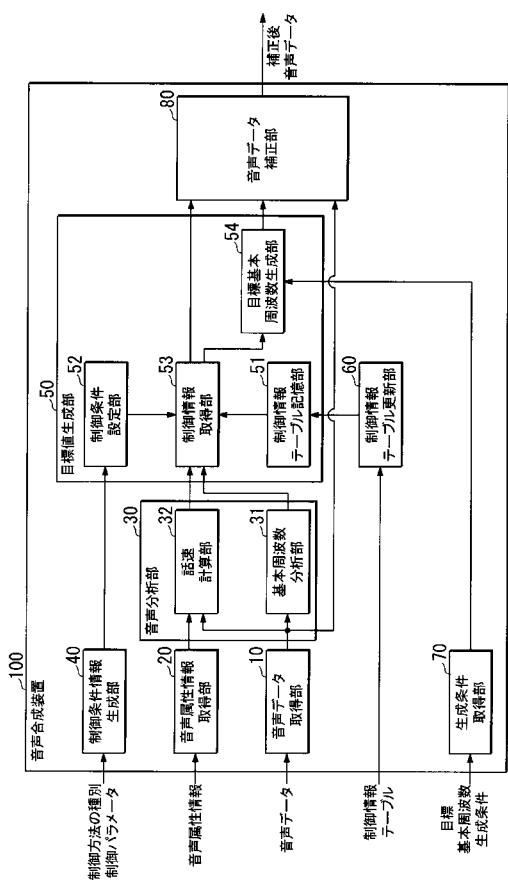
【0077】

40

- 10 音声データ取得部
- 20 音声属性情報取得部
- 30 音声分析部
- 31 基本周波数分析部
- 32 話速計算部
- 40, 40a 制御条件情報生成部
- 50, 50a 目標値生成部
- 51, 51a 制御情報テーブル記憶部
- 52, 52a 制御条件設定部
- 53, 53a 制御情報取得部
- 54 目標基本周波数生成部
- 60 制御情報テーブル更新部
- 70 生成条件取得部
- 80 音声データ補正部

50

【 図 1 】



【図2】

音素区切時間情報			
0	2740	s	i
2740	3168	o	
3168	3534	h	
3534	3944	a	
3944	4464	y	
4464	5092	o	
5092	5480	u	
5480	5886	g	
5886	6278	o	
6278	6730	z	
6730	7144	a	
7144	7652	i	
7652	8272	m	
8272	8658	a	
8658	9038	s	
9038	9476	u	
9476	15000	s	i

【 図 3 】

(a) 難聴者向けの制御情報テーブル	
話速値の絶対値 (モーラ/秒)	基本周波数の変化幅の絶対値 (Hz)
6	180

速聴向けの制御情報テーブル	
話速値の絶対値 (モーラ/秒)	基本周波数の変化幅の絶対値 (Hz)
18	170

(c) 早口向けの制御情報テーブル	
話速値の絶対値 (モーラ/秒)	基本周波数の変化幅の絶対値 (Hz)
10	150

〔 図 4 〕

(a) 難聴者向けの制御情報テーブル	
話速の倍率 (倍)	基本周波数の変化幅の倍率 (倍)
0.5	1.15
0.6	1.15
0.7	1.15
0.8	1.10
0.9	1.05
1.0	1.00

(b) 速晴向けの制御情報テーブル

速聴向けの制御情報テーブル	
話速の倍率 (倍)	基本周波数の変化幅の倍率 (倍)
1.0	1.00
1.5	0.97
2.0	0.93
2.5	0.90

(c) 早口向けの制御情報テーブル

話速の倍率 (倍)	基本周波数の変化幅の倍率 (倍)
1.0	1.00
1.1	0.96
1.2	0.91
1.3	0.87
1.4	0.84
1.5	0.80

〔 図 5 〕

指定透過率 10モード秒	
透過率 (倍)	基本周波数の変化率 (倍)
0.5	1.30
0.6	1.30
0.7	1.30
0.8	1.19
0.9	1.09
1.0	1.00

指定話速値 8モード/秒	
速さの倍率(倍)	基本周波数の変化幅の倍(倍)
0.5	1.20
0.6	1.20
0.7	1.20
0.8	1.13
0.9	1.06
1.0	1.00

指定周波数の変化幅(倍)	基本周波数の倍率(倍)
0.6	1.10
0.7	1.10
0.8	1.07
0.9	1.03
1.0	1.00

指定速度:10モード/秒	
周波数の倍率 (倍)	基本周波数の変化幅の倍率 (倍)
1.0	1.00
1.5	1.00
2.0	1.00

指定活速値：8モーラ/秒		
速度の倍率 (倍)	基本周波数の変化幅の倍率 (倍)	
1.0	1.00	
1.5	0.97	
2.0	0.93	
2.5	0.90	

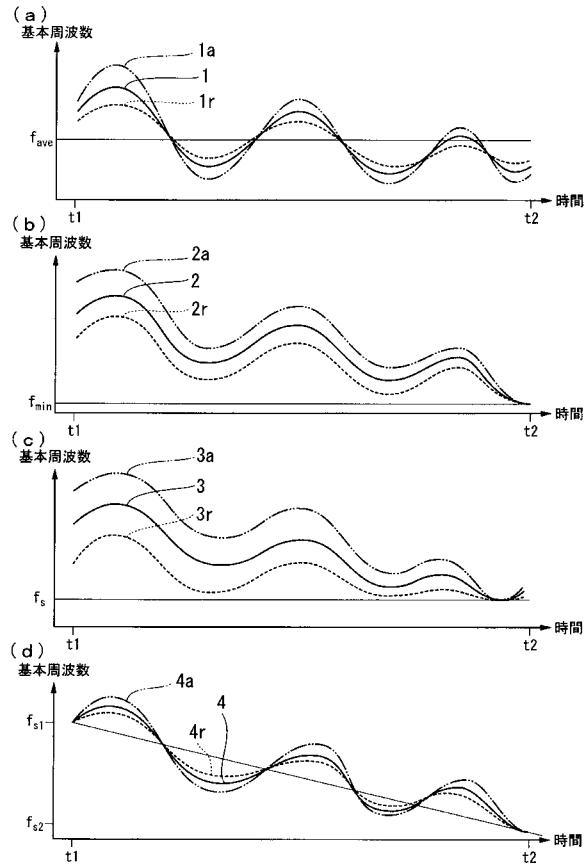
（単位：秒）	
周波数	時間
1.0	1.00
1.5	0.95
2.0	0.90
2.5	0.85
3.0	0.80

持続話速: 10モード秒	
持続話速の倍率 (倍)	基本周波数の変化幅の倍率 (倍)
1.0	1.00
1.1	0.98
1.2	0.96
1.3	0.94
1.4	0.92
1.5	0.90

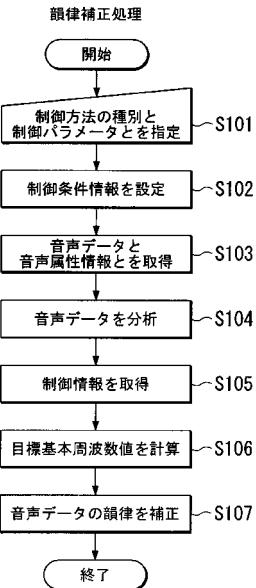
指定話速選択: 8モード/秒	
速さの倍率 (倍)	基本周波数の変化倍率 (倍)
1.0	1.00
1.1	0.96
1.2	0.93
1.3	0.89
1.4	0.85
1.5	0.85

指定時間(前回)の平均周波数 モードアーチ	基本周波数の変化幅の倍率 (倍)	基本周波数 (Hz)
1.0	1.00	1.00
1.1	0.95	0.89
1.2	0.85	0.80
1.3	0.80	0.80
1.4	0.80	0.80
1.5	0.80	0.80

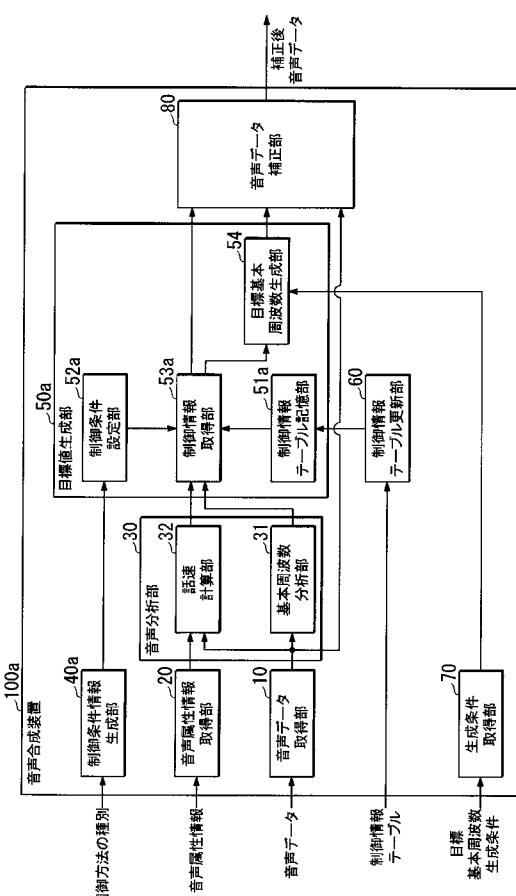
【 図 6 】



【 四 7 】



【図8】



【 図 9 】

難聴者向けの制御情報テーブル	
話速の倍率 (倍)	基本周波数の変化幅の倍率 (倍)
0.7	1.15

【 図 1 0 】

速聴向けの制御情報テーブル	
話速の倍率 (倍)	基本周波数の変化幅の倍率 (倍)
2.0	0.93

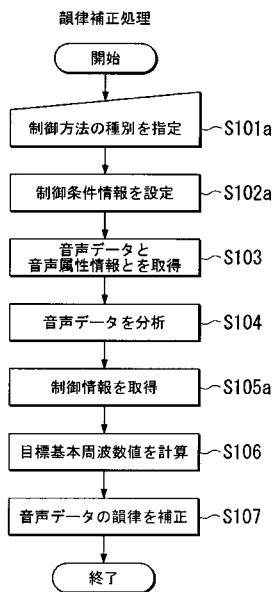
【 図 1 1 】

早口向けの制御情報テーブル	
話速の倍率 (倍)	基本周波数の変化幅の倍率 (倍)
1.4	0.84

【 図 1 2 】

雑音環境下向けの制御情報テーブル	
話速の倍率 (倍)	基本周波数の変化幅の倍率 (倍)
1.0	1.15

【図 1 3】



フロントページの続き

(72)発明者 都木 徹

東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内

(72)発明者 清山 信正

東京都世田谷区砧一丁目10番11号 財団法人工ヌエイチケインジニアリングサービス内