

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第3941611号  
(P3941611)

(45) 発行日 平成19年7月4日(2007.7.4)

(24) 登録日 平成19年4月13日(2007.4.13)

(51) Int.C1.

F 1

G 1 O L 21/04 (2006.01)  
G 1 O L 13/02 (2006.01)G 1 O L 21/04 120 D  
G 1 O L 13/02 122 B

請求項の数 4 (全 12 頁)

(21) 出願番号 特願2002-198486 (P2002-198486)  
 (22) 出願日 平成14年7月8日 (2002.7.8)  
 (65) 公開番号 特開2004-38071 (P2004-38071A)  
 (43) 公開日 平成16年2月5日 (2004.2.5)  
 審査請求日 平成16年5月27日 (2004.5.27)

(73) 特許権者 000004075  
 ヤマハ株式会社  
 静岡県浜松市中区中沢町10番1号  
 (74) 代理人 100092820  
 弁理士 伊丹 勝  
 (72) 発明者 篠持 秀紀  
 静岡県浜松市中沢町10番1号 ヤマハ株  
 式会社内  
 審査官 横本 剛

最終頁に続く

(54) 【発明の名称】歌唱合成装置、歌唱合成方法及び歌唱合成用プログラム

## (57) 【特許請求の範囲】

## 【請求項1】

合成すべき歌唱演奏データを示す歌唱情報を入力する歌唱情報入力部と、  
 音声素片データを保持する音韻データベースと、  
 前記歌唱情報に基づいて前記音韻データベースに記憶された音声素片データを選択する  
 選択部と、

歌唱音声の声質を変換するための声質変換パラメータであって、歌唱音声を男性的にするか又は女性的にするかを表す係数を入力する声質変換パラメータ入力部と、

前記声質変換パラメータに含まれる係数に基づき(1)式で定義されるマッピング関  
 数を生成するマッピング関数生成部と、

$f_{out} = (f_s / 2) \times (2 \times f_{in} / f_s) \dots \dots (1)$

(但し、 $f_{out}$ は出力周波数、 $f_s$ はサンプリング周波数、 $f_{in}$ は入力周波数)

選択された前記音声素片データに従ってスペクトル包絡を生成し該スペクトル包絡の口  
 一カルピークの周波数を入力周波数として前記生成したマッピング関数に基づき変換し、  
 該変換後のスペクトル包絡を用いて歌唱音声を生成する歌唱合成部とを備え、

前記マッピング関数は、前記スペクトル包絡の変換前後において、最低周波数と最高周  
 波数の値を変化させず一定とするように設定されている  
 ことを特徴とする歌唱合成装置。

## 【請求項2】

前記声質変換パラメータ入力部は、前記声質変換パラメータを時間的に変化させる声質

10

20

変換パラメータ調整手段を備えたものである請求項1に記載の歌唱合成装置。

【請求項3】

歌唱を合成するための歌唱情報を入力する歌唱情報入力ステップと、  
音声素片データを予め音韻データベースに保持させておくと共に、前記歌唱情報に基づいて前記音声素片データを選択する選択ステップと、  
歌唱音声の声質を変換するための声質変換パラメータであって、歌唱音声を男性的にするか又は女性的にするかを表す係数を入力する声質変換パラメータ入力ステップと、  
入力された前記前記声質変換パラメータに含まれる係数に基づき(1)式で定義されるマッピング関数を生成するマッピング関数生成ステップと、

$$f_{out} = (f_s / 2) \times (2 \times f_{in} / f_s) \quad \dots \quad (1)$$

(但し、 $f_{out}$ は出力周波数、 $f_s$ はサンプリング周波数、 $f_{in}$ は入力周波数)

選択された前記音声素片データに従ってスペクトル包絡を生成し該スペクトル包絡のローカルピークの周波数を入力周波数として前記生成したマッピング関数に基づき変換し、該変換後のスペクトル包絡を用いて歌唱音声を生成する歌唱合成ステップと  
を備え、

前記マッピング関数は、前記スペクトル包絡の変換前後において、最低周波数と最高周波数の値を変化させず一定とするように設定されたものである  
ことを特徴とする歌唱合成方法。

【請求項4】

歌唱を合成するための歌唱情報を入力する歌唱情報入力ステップと、  
音声素片データを予め音韻データベースに保持させておくと共に、前記歌唱情報に基づいて前記音声素片データを選択する選択ステップと、  
歌唱音声の声質を変換するための声質変換パラメータであって、歌唱音声を男性的にするか又は女性的にするかを表す係数を入力する声質変換パラメータ入力ステップと、  
入力された前記声質変換パラメータに含まれる係数に基づき(1)式で定義されるマッピング関数を生成するマッピング関数生成ステップと、

$$f_{out} = (f_s / 2) \times (2 \times f_{in} / f_s) \quad \dots \quad (1)$$

(但し、 $f_{out}$ は出力周波数、 $f_s$ はサンプリング周波数、 $f_{in}$ は入力周波数)

選択された音声素片データに従ってスペクトル包絡を生成し該スペクトル包絡のローカルピークの周波数を入力周波数として前記生成したマッピング関数に基づき変換し、該変換後のスペクトル包絡を用いて歌唱音声を生成する歌唱合成ステップと  
をコンピュータに実行させるように構成され、

前記マッピング関数は、前記スペクトル包絡の変換前後において、最低周波数と最高周波数の値を変化させず一定とするように設定されている  
ことを特徴とする歌唱合成用プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、人間の歌唱音声を合成する歌唱合成装置、歌唱合成方法及び歌唱合成用プログラムに関する。

【0002】

【従来の技術】

従来の歌唱合成装置においては、人間の実際の歌声から取得したデータをデータベースとして保存しておき、入力された演奏データ（音符、歌詞、表情等）の内容に合致したデータをデータベースより選択する。そして、この選択された演奏データに基づいて、本物の人の歌声に近い歌唱音声を合成している。

【0003】

【発明が解決しようとする課題】

ところで、人間が歌を歌うときには、音楽的なコンテキスト（楽曲中の位置、表情付けなど）により声の音色を変化させて歌うのが普通である。例えば、同じ歌詞であっても楽曲

10

20

30

40

50

の前半部分では普通に歌うが、後半部分では感情を込めて歌う、といった具合である。従って、歌唱合成装置により自然な歌声を合成するには、歌唱中の声の音色を音楽的なコンテキストに従って変化させることが必要となってくる。

#### 【0004】

しかし、従来の歌唱合成装置では、歌唱者データを入力して、歌唱者の異同に応じて歌わせ方を変化させることは行われていたが、同じ歌唱者の場合、同じ音韻コンテキストに対しては、基本的には1つの音韻テンプレートだけを使用しており、音色のバリエーションを付けることは行われていなかった。このため、合成される歌唱音声は音色の変化が乏しいものであった。

本発明は、この点に鑑みてなされたものであり、合成される歌唱音声を表情豊かなものとすることのできる歌唱合成装置を提供することを目的とする。 10

#### 【0005】

##### 【課題を解決するための手段】

上記目的達成のため、本出願の第1の発明に係る歌唱合成装置は、合成すべき歌唱演奏データを示す歌唱情報を入力する歌唱情報入力部と、音声素片データを保持する音韻データベースと、前記歌唱情報に基づいて前記音韻データベースに記憶された音声素片データを選択する選択部と、歌唱音声の声質を変換するための声質変換パラメータであって、歌唱音声を男性的にするか又は女性的にするかを表す係数を入力する声質変換パラメータ入力部と、前記声質変換パラメータに含まれる係数に基づき(1)式で定義されるマッピング関数を生成するマッピング関数生成部と、 20

$$f_{out} = (fs / 2) \times (2 \times f_{in} / fs) \quad \dots \quad (1)$$

(但し、f<sub>out</sub>は出力周波数、f<sub>s</sub>はサンプリング周波数、f<sub>in</sub>は入力周波数)

選択された前記音声素片データに従ってスペクトル包絡を生成し該スペクトル包絡のローカルピークの周波数を入力周波数として前記生成したマッピング関数に基づき変換し、該変換後のスペクトル包絡を用いて歌唱音声を生成する歌唱合成部とを備え、前記マッピング関数は、前記スペクトル包絡の変換前後において、最低周波数と最高周波数の値を変化させず一定とするように設定されていることを特徴とする。 20

#### 【0006】

この第1の発明に係る音声合成装置によれば、声質変換パラメータの変化により、合成される歌唱音声の声質を変化させることができる。このため、時間的に前後して現れる同一特徴パラメータ、すなわち同一の歌唱部分であっても、それぞれ任意の異なる声質に変換することができ、合成歌唱音声を変化に富みアリティに溢れたものとすることができます。 30

#### 【0007】

上記目的達成のため、本出願の第2の発明に係る歌唱合成方法は、歌唱を合成するための歌唱情報を入力する歌唱情報入力ステップと、音声素片データを予め音韻データベースに保持させておくと共に、前記歌唱情報に基づいて前記音声素片データを選択する選択ステップと、歌唱音声の声質を変換するための声質変換パラメータであって、歌唱音声を男性的にするか又は女性的にするかを表す係数を入力する声質変換パラメータ入力ステップと、入力された前記声質変換パラメータに含まれる係数に基づき(1)式で定義されるマッピング関数を生成するマッピング関数生成ステップと、 40

$$f_{out} = (fs / 2) \times (2 \times f_{in} / fs) \quad \dots \quad (1)$$

(但し、f<sub>out</sub>は出力周波数、f<sub>s</sub>はサンプリング周波数、f<sub>in</sub>は入力周波数)

選択された前記音声素片データに従ってスペクトル包絡を生成し該スペクトル包絡のローカルピークの周波数を入力周波数として前記生成したマッピング関数に基づき変換し、該変換後のスペクトル包絡を用いて歌唱音声を生成する歌唱合成ステップとを備え、前記マッピング関数は、前記スペクトル包絡の変換前後において、最低周波数と最高周波数の値を変化させず一定とするように設定されたものであることを特徴とする。

#### 【0008】

上記目的達成のため、本出願の第2の発明に係る歌唱合成用プログラムは、歌唱を合成 50

するための歌唱情報を入力する歌唱情報入力ステップと、音声素片データを予め音韻データベースに保持させておくと共に、前記歌唱情報に基づいて前記音声素片データを選択する選択ステップと、歌唱音声の声質を変換するための声質変換パラメータであって、歌唱音声を男性的にするか又は女性的にするかを表す係数を入力する声質変換パラメータ入力ステップと、入力された前記声質変換パラメータに含まれる係数に基づき(1)式で定義されるマッピング関数を生成するマッピング関数生成ステップと、

$$f_{out} = (f_s / 2) \times (2 \times f_{in} / f_s) \quad \dots \quad (1)$$

(但し、 $f_{out}$ は出力周波数、 $f_s$ はサンプリング周波数、 $f_{in}$ は入力周波数)

選択された音声素片データに従ってスペクトル包絡を生成し該スペクトル包絡のローカルピークの周波数を入力周波数として前記生成したマッピング関数に基づき変換し、該変換後のスペクトル包絡を用いて歌唱音声を生成する歌唱合成ステップとをコンピュータに実行させるように構成され、前記マッピング関数は、前記スペクトル包絡の変換前後において、最低周波数と最高周波数の値を変化させず一定とするように設定されていることを特徴とする。

#### 【0009】

##### 【発明の実施の形態】

以下、本発明の実施の形態を図面に基づいて詳細に説明する。

図1は、本発明の実施の形態に係る歌唱合成装置の構成を示す機能ブロック図である。まずこの図1の歌唱合成装置の概要を説明すると、音韻データベース10は、得られた歌唱データから切り出された音素連鎖データと定常部分データとを保持している。演奏データ保持部11の演奏データは、遷移部分と伸ばし音部分とに区切られた後、音素連鎖データはその遷移部分において基本的にはそのまま用いられる。このため、歌唱の重要な部分を占める遷移部分の歌唱が自然に聞こえ、合成歌唱の品質が高められている。この歌唱合成装置は、例えば一般のパーソナルコンピュータにより実現することができ、図1に示す各ブロックの機能は、パーソナルコンピュータ内部のCPUやRAM、ROMなどにより達成され得る。DSPやロジック回路によって構成することも可能である。

#### 【0010】

以下、この図1の歌唱合成装置の詳細を説明する。

前述のように、音韻データベース10は、演奏データに基づいて合成音を合成するためのデータを保持している。この音韻データベース10の作成例を図2により説明する。まず図2に示すように、実際に録音或いは取得した歌唱データ等の音声信号をSMS(spectral modeling synthesis)分析手段31により、調和成分(正弦波成分)と非調和成分に分離する。SMS分析の代わりに、LPC(Lineare Predictive Coding)等の他の分析手法を用いてもよい。

次に、音素切り分け手段32により、音素切り分け情報に基づき、音声信号を音素ごとに切り分ける。音素切り分け情報は、例えば人間が音声信号の波形を見ながら所定のスイッチ動作を行うことにより与えるのが通常である。

#### 【0011】

そして、音素ごとに切り分けられた音声信号の調和成分から、特徴パラメータ抽出手段33により特徴パラメータが抽出される。特徴パラメータには、励起波形エンベロープ、励起レゾナンス、フォルマント、差分スペクトルなどがある。

#### 【0012】

励起波形エンベロープ(ExcitationCurve)は、声帯波形の大きさ(dB)を表わすEGain、声帯波形のスペクトルエンベロープの傾きを表わすESlopeDepth、声帯波形のスペクトルエンベロープの最大値から最小値への深さ(dB)を表わすESlopeの3つのパラメータによって構成されており、以下の式[数1]で表わすことが出来る。

#### 【0013】

##### 【数1】

$$\text{Excitation Curve } (f) = \text{EGain} + \text{ESlopeDepth} \times (\exp(-\text{ESlope} \times f) - 1)$$

#### 【0014】

10

20

30

40

50

励起レゾナンスは、胸部による共鳴を表わす特徴パラメータで、中心周波数 (ERFreq)、バンド幅 (ERBW)、アンプリチュード (ERAmpl) の 3 つのパラメータにより構成され、2 次フィルター特性を有している。

#### 【0015】

フォルマントは、1から12個程度のレゾナンスを組み合わせることにより声道による共鳴を表わす特徴パラメータで、中心周波数 (FormantFreq<sub>i</sub>、<sub>i</sub>はレゾナンスの番号)、バンド幅 (FormantBWi、<sub>i</sub>はレゾナンスの番号)、強度 (FormantAmpl<sub>i</sub>、<sub>i</sub>はレゾナンスの番号) の 3 つのパラメータにより構成される。

#### 【0016】

差分スペクトルは、上記の励起波形エンベロープ、励起レゾナンス、フォルマントの 3 つで表現することの出来ない元の調和成分との差分のスペクトルを持つ特徴パラメータである。

10

#### 【0017】

この抽出された特徴パラメータを、音韻名と対応させて音韻データベース 10 に記憶させる。非調和成分も、同様にして音韻名と対応させてそのスペクトルデータを音韻データベース 10 に記憶させる。この音韻データベース 10 では、図 2 に示すように、音素連鎖データと定常部分データとに分けて記憶される。以下では、この音素連鎖データと定常部分データとを総称して「音声素片データ」と称する。

#### 【0018】

音素連鎖データは、先頭音素名、後続音素名、特徴パラメータ及び非調和成分を対応付けたデータ列である。

20

一方、定常部分データは、1つの音韻名と特徴パラメータ列と非調和成分とを対応付けたデータ列である。

#### 【0019】

図 1 に戻って、11は演奏データを保持するための演奏データ保持部である。演奏データは、例えば音符、歌詞、ピッチバンド、ダイナミクス等の情報を含んだMIDI情報である。

音声素片選択部 12 は、演奏データ保持部 11 に保持される演奏データの入力をフレーム単位で受け付けるとともに（以下、この 1 単位をフレームデータという）、入力されたフレームデータ中の歌詞データに対応する音声素片データを音韻データベース 10 から選択して読み出す機能を有する。

30

#### 【0020】

先行音素連鎖データ保持部 13、後方音素連鎖データ保持部 14 は、定常部分データを処理するために使用されるものである。先行音素連鎖データ保持部 13 は、処理すべき定常部分データより先行する音素連鎖データを保持するものであり、一方、後方音素連鎖データ保持部 14 は、処理すべき定常部分データより後方の音素連鎖データを保持するものである。

#### 【0021】

特徴パラメータ補間部 15 は、先行音素連鎖データ保持部 13 に保持された音素連鎖データの最終フレームの特徴パラメータと、後方音素連鎖データ保持部 14 に保持された音素連鎖データの最初のフレームの特徴パラメータとを読み出し、タイマ 29 の示す時刻に対応するように特徴パラメータを時間的に補間する。

40

#### 【0022】

定常部分データ保持部 16 は、音声素片選択部 12 により読み出された音声素片データのうち、定常部分データを一時保持する。一方、音素連鎖データ保持部 17 は、音素連鎖データを一時保持する。

#### 【0023】

特徴パラメータ変動抽出部 18 は、定常部分データ保持部 16 に保持された定常部分データを読み出してその特徴パラメータの変動（ゆらぎ）を抽出し、変動成分として出力する機能を有する。

50

加算部 K 1 は、特徴パラメータ補間部 1 5 の出力と特徴パラメータ変動抽出部 1 8 の出力を加算して、伸ばし音部分の特徴パラメータを出力する部分である。

フレーム読出し部 1 9 は、音素連鎖データ保持部 1 7 に保持された音素連鎖データを、タイム 2 9 に示す時刻に従ってフレームデータとして読出し、特徴パラメータと非調和成分とに分けて出力する部分である。

#### 【 0 0 2 4 】

ピッチ決定部 2 0 は、フレームデータ中の音符データ、ピッチベンドデータに基づき、最終的に合成する合成音のピッチを決定する部分である。また特徴パラメータ補正部 2 1 は、加算部 K 1 から出力された伸ばし音部分の特徴パラメータ、及びフレーム読出し部 1 9 から出力された遷移部分の特徴パラメータを、ピッチ決定部 2 0 で決定したピッチや演奏データ中に含まれるダイナミクス情報等に基づいて補正する部分である。特徴パラメータ補正部 2 1 の前段にはスイッチ SW 1 が設けられ、伸ばし音部分の特徴パラメータと遷移部分の特徴パラメータとを選択的に特徴パラメータ補正部 2 1 に入力するようになっている。この特徴パラメータ補正部 2 1 での詳しい処理内容は後述する。スイッチ SW 2 は、定常部分データ保持部 1 6 から読み出された伸ばし音部分の非調和成分と、フレーム読出し部 1 9 から読み出された遷移部分の非調和成分を切り替えて出力する。

#### 【 0 0 2 5 】

倍音列生成部 2 2 は、決定したピッチに従い、フォルマント合成を行うための倍音列を周波数軸上に生成する部分である。

スペクトル包絡生成部 2 3 は、特徴パラメータ補正部 2 1 で補正された補正後の特徴パラメータに従って、スペクトル包絡を生成する部分である。

#### 【 0 0 2 6 】

倍音振幅・位相計算部 2 4 は、倍音列生成部 2 2 で生成された各倍音の振幅及び位相を、スペクトル包絡生成部 2 3 で生成したスペクトル包絡に付加する部分である。

#### 【 0 0 2 7 】

声質変換部 2 5 は、外部より入力される声質変換パラメータに基づき、倍音振幅・位相計算部 2 4 を介して入力される調和成分のスペクトル包絡に変更を加え、これにより合成歌唱音声の声質を変換する機能を有する。

声質変換部 2 5 は、入力される声質変換パラメータに基づき、図 3 ( a ) に示すように、入力されたスペクトル包絡 S e のローカルピークの位置をシフトさせ、これにより声質の変換を実行する。図 3 ( a ) の場合、ローカルピークが全体的に高い方にシフトしているので、変更後の出力音声は、変更前に比べ女性的な又は子供っぽい声に変化することになる。

#### 【 0 0 2 8 】

本実施の形態では、声質変換パラメータ調整部 2 5 C から出力される声質変換パラメータに基づき、マッピング関数生成部 2 5 M において図 3 ( b ) に示すようなマッピング関数 M f を生成する。声質変換部 2 5 は、このマッピング関数 M f に基づきスペクトル包絡のローカルピークの位置をシフトさせる。このマッピング関数 M f は、横軸を入力周波数( 声質変換部 2 5 に入力されるスペクトル包絡のローカルピークの周波数 ) とし、縦軸を出力周波数( 声質変換部 2 5 から出力されるスペクトル包絡のローカルピークの周波数 ) をとっている。このため、マッピング関数 M f が、入力周波数 = 出力周波数を示す直線 N L よりも上側に位置する部分においては、その入力周波数のローカルピークは、マッピング関数 M f による変換後、周波数の高い方向にシフトする。逆に、マッピング関数 M f が、直線 N L よりも下側に位置する部分においては、その入力周波数のローカルピークは、マッピング関数 M f による変換後、周波数の低い方向にシフトする。

#### 【 0 0 2 9 】

そして、このマッピング関数 M f の形状は、声質変換パラメータ調整部 2 5 C を使用して時間的に変化させることができるようになっている。例えば、ある時間帯では、マッピング関数を直線 N L と一致させ、別の時間帯では図 3 ( b ) に示す直線 N L に関しマッピング関数 M f と対称な曲線を生成する、というような変化をさせることが可能である。これ

10

20

30

40

50

により、楽曲の音楽的なコンテキストなどに合わせて出力される歌唱音声の声質が時間的に変化し、変化の多い表情豊かな歌唱音声とすることができる。声質変換パラメータ調整部 25C としては、例えばパーソナルコンピュータのマウス、キーボード等を採用することができる。

なお、マッピング関数  $M_f$  は、その形状をどのように変化させる場合であっても、最低周波数（図3の例では 0 Hz）と最高周波数の値は変化させないようにし、これにより周波数帯域が声質変換の前後で変わらないようにするのが好適である。

#### 【0030】

図4は、マッピング関数  $M_f$  の他の例を示している。図4(a)は、低域側では周波数を高い側へシフトさせ、高域側では周波数を低い方へシフトさせるマッピング関数  $M_f$  の例を示している。この場合、聴感上重要な低域側で周波数の高い方へのシフトが行われているため、出力歌唱音声は、全体に子供のような、又はいわゆるダックボイスのような声となる。図4(b)に示すマッピング関数  $M_f$  は、全体に出力周波数を低い側へシフトさせ、中心周波数付近でそのシフト量を最大としている。この例では、聴感上重要な低域側で周波数の低い方へのシフトが行われているため、出力歌唱音声は、太い男性的な声となる。

10

この図4(a)(b)の場合にも、マッピング関数  $M_f$  の形状は、声質変換パラメータ調整部 25C により、時間的に変化させることが可能である。

#### 【0031】

声質変換部 26 は、定常部分データ保持部 16 と、フレーム読み出し部 19 より出力される非調和成分の入力を受けるとともに、声質変換部 25 と同様に、声質変換パラメータに基づいてマッピング関数生成部 26M で生成したマッピング関数  $M_f'$  を使用して非調和成分のスペクトル包絡に変更を加える。マッピング関数  $M_f'$  の形状は、声質変換パラメータ調整部 26C により変更することができる。

20

加算部 K2 は、声質変換部 25 の出力としての調和成分と、声質変換部 26 から出力された非調和成分とを加算する。

逆FFT部 27 は、加算部 K2 の出力値を逆高速フーリエ変換して、周波数軸表現であった信号を時間軸表現の信号に変換するものである。

重ね合せ部 28 は、時系列順に処理される歌詞データについて次々に得られる信号をその時系列に沿った形で重ね合わせることにより、合成歌唱音声を出力するものである。

30

#### 【0032】

次に、特徴パラメータ補正部 21 の詳細について図5に基づいて説明する。特徴パラメータ補正部 21 は、振幅決定手段 41 を備えている。この振幅決定手段 41 は、ダイナミクス - 振幅変換テーブル  $T_{da}$  を参照して演奏データ保持部 11 から入力されるダイナミクス情報に相当する所望の振幅値  $A_1$  を出力する。

また、スペクトル包絡生成手段 42 は、スイッチ SW1 から出力された特徴パラメータに基づき、スペクトル包絡を生成する部分である。

#### 【0033】

倍音列生成手段 43 は、ピッチ決定部 20 で決定されたピッチに基づいて倍音列を生成する。振幅計算手段 44 は、生成されたスペクトル包絡及び倍音に対応する振幅値  $A_2$  を計算する。振幅の計算は、例えば逆FFT 等により実行することができる。

40

加算部 K3 は、振幅決定手段 41 で決定された所望の振幅値  $A_1$  と、振幅計算手段 44 で計算された振幅値  $A_2$  との差を出力する。ゲイン補正手段 45 は、この差に基づき、振幅値の補正量を計算するとともに、この補正量に従って特徴パラメータを補正する。これにより、所望の振幅に合致する新たな特徴パラメータが得られる。

#### 【0034】

なお、図5では、テーブル  $T_{da}$  に基づき、ダイナミクスのみに基づいて振幅を決定しているが、これに加えて、音素の種類も考慮して振幅を決定するようなテーブルを採用してもよい。すなわち、同じダイナミクスであっても音素が異なる場合には、異なる振幅値を与えるようなテーブルを採用してもよい。同様に、ダイナミクスに加えて周波数を考慮し

50

て振幅を決定するようなテーブルを採用してもよい。

#### 【0035】

次に、この実施の形態に係る歌唱合成装置の作用を、図6に示すフローチャートを参照しつつ説明する。

演奏データ保持部11は、時系列順にフレームデータを出力する。遷移部分と伸ばし音部分とが交互に現れ、遷移部分と伸ばし音部分とでは処理のされ方が異なる。

#### 【0036】

演奏データ保持部11よりフレームデータが入力されると(S1)、音声素片選択部12において、フレームデータ中の歌詞データに基づき、そのフレームデータが伸ばし音部分に関するものか、遷移部分に関するものかが判断される(S2)。伸ばし音部分である場合には(YES)、先行音素連鎖データ保持部13、後方音素連鎖データ保持部14、定常部分データ保持部16に、それぞれ先行音素連鎖データ、後方音素連鎖データ、定常部分データが転送される(S3)。

#### 【0037】

続いて、特徴パラメータ補間部15が、先行音素連鎖データ保持部13に保持された先行音素連鎖データの最終フレームの特徴パラメータを取り出すと共に、後方音素連鎖データ保持部14に保持された後方音素連鎖データの最初のフレームの特徴パラメータを取り出し、この2つの特徴パラメータを補間することにより、処理中の伸ばし音部分の特徴パラメータを生成する(S4)。

#### 【0038】

また、定常部分データ保持部16に保持された定常部分データの特徴パラメータが、特徴パラメータ変動抽出部18に供給され、該定常部分の特徴パラメータの変動成分が抽出される(S5)。この変動成分が、加算部K1において特徴パラメータ補間部15から出力された特徴パラメータと加算される(S6)。この加算値が伸ばし音部分の特徴パラメータとしてスイッチSW1を介して特徴パラメータ補正部21に出力され、特徴パラメータの補正が実行される(S9)。一方、定常部分データ保持部16に保持された定常部分データの非調和成分は、スイッチSW2を介して声質変換部26に供給される。

#### 【0039】

スペクトル包絡生成部23は、この補正後の特徴パラメータについてのスペクトル包絡を生成する。倍音振幅・位相計算部24は、スペクトル包絡生成部23で生成したスペクトル包絡に従い、倍音列生成部22で生成された各倍音の振幅及び位相を計算する。声質変換部25では、声質変換パラメータに応じて、スペクトル包絡生成部23で生成したスペクトル包絡のローカルピークの位置を変更し、この変更後のスペクトル包絡を加算部K2に出力する。

#### 【0040】

一方、S2において、取得されたフレームデータが遷移部分のものである(NO)と判定された場合には、その遷移部分の音素連鎖データが、音素連鎖データ保持部17により保持される(S7)。次に、フレーム読出し部19が、音素連鎖データ保持部17に保持された音素連鎖データを、タイム29に示す時刻に従ってフレームデータとして読出し、特徴パラメータと非調和成分とに分けて出力する(S8)。特徴パラメータの方は特徴パラメータ補正部21に向けて出力され、非調和成分はスイッチSW2を介して声質変換部26に向けて出力される。声質変換部26では、声質変換パラメータ調整部26Cからの声質変換パラメータに応じて生成されたマッピング関数Mf'により、この非調和成分の変更が行われ、この変更後の非調和成分が加算部K2に向けて出力される。この遷移部分の特徴パラメータは、特徴パラメータ補正部21、スペクトル包絡生成部23、倍音振幅・位相計算部24等で上述の伸ばし音部分の特徴パラメータと同様の処理を受ける。

#### 【0041】

なお、スイッチSW1、SW2は、処理中のデータの種類によって切り替わるようになっているので、スイッチSW1については、伸ばし音部分を処理している間は、加算部K1の方に特徴パラメータ補正部21を接続するようにされ、遷移部分を処理している間は、

10

20

30

40

50

フレーム読み出し部 19 の方に特徴パラメータ補正部 21 を接続するようにされている。また、スイッチ SW 2 については、伸ばし音部分を処理している間は、定常部分データ保持部 16 の方に声質変換部 26 を接続するようにされ、遷移部分を処理している間は、フレーム読み出し部 19 の方に声質変換部 26 を接続するようにされている。

こうして遷移部分、伸ばし音部分の特徴パラメータ及び非調和成分が演算されると、その加算値が逆 FFT 部 27 で処理され、重ね合せ部 28 により重ね合わせられ、最終的な合成波形が出力される (S 10)。

#### 【0042】

以上、本発明の実施の形態について説明したが、本発明はこれに限定されるものではない。例えば、上記実施の形態では、声質変換パラメータをマッピング関数という形で表現しているが、声質変換パラメータを、演奏データ保持部 11 内に MIDI データとして含ませるようにしてよい。

また、上記実施の形態では、スペクトル包絡生成部 23 からの出力としてのスペクトル包絡のローカルピーク周波数をマッピング関数による調整の対象としているが、調整の対象はスペクトル包絡の全体、又は任意の部分でも良く、またローカルピークの周波数だけに限らず振幅などスペクトル包絡を表わす他のパラメータを調整の対象としてもよい。また、音韻データベース 10 から読み出された特徴パラメータ ( 例えば EGain、ESlope、ESlopeDepth 等 ) を調整の対象にするようにしてもよい。

または、特徴パラメータ補正部 21 からの出力である特徴パラメータに変更を加えることも可能である。この際、各特徴パラメータの種類ごとにマッピング関数を持てばよい。

または、加算部 K 2 での手前において調和成分と非調和成分のいずれか一方を声質変換パラメータに基づいて増幅又は減衰させ、その割合を変更させた上で加算部 K 2 で加算せることによってもよい。また、調和成分だけを調整の対象としてもよい。また、逆 FFT 部 27 から出力される時間軸の信号を調整の対象としてもよい。

#### 【0043】

また、マッピング関数を、次の式で表わしてもよい。

#### 【数 2】

$$f_{out} = ( f_s / 2 ) \times ( 2 \times f_{in} / f_s )$$

#### 【0044】

ただし、 $f_s$  はサンプリング周波数、 $f_{in}$  は入力周波数、 $f_{out}$  は出力周波数である。また、 $\alpha$  は、出力歌唱音声を男性的にするか、女性的にするかを決定する因子であり、 $\alpha$  が正の値であれば [ 数 2 ] で表わされるマッピング関数は下に凸な関数となり、出力歌唱音声は男性的なものとなる。また、 $\alpha$  が負の値であれば、女性又は子供らしい声となる ( 図 7 参照 )。

#### 【0045】

また、マッピング関数を表現する座標系上にいくつかの点 ( ブレークポイント ) を指定し、それらを結ぶ直線としてマッピング関数を定義することもできる。この場合、声質変換パラメータは座標値によるベクトルとして表現される。

#### 【0046】

#### 【発明の効果】

以上説明したように、本発明によれば、声質変換パラメータを時間的に変化させることができ、これにより、時間的に前後して現れる同一特徴パラメータ、すなわち同一の歌唱部分であっても、それぞれ任意の異なる声質に変換することができ、合成歌唱音声を変化に富みアリティに溢れたものとすることができます。

#### 【図面の簡単な説明】

【図 1】 本発明の実施の形態に係る歌唱合成装置の機能プロック図である。

【図 2】 図 1 に示す音韻データベース 10 の作成例を示す。

【図 3】 声質変換部 25 による入出力の変換のされ方、及びマッピング関数生成部 25 M で生成されるマッピング関数 M\_f の一例を示す。

【図 4】 マッピング関数 M\_f の他の例を示す。

10

20

30

40

50

【図5】 図1に示す特徴パラメータ補正部21の詳細を示す。

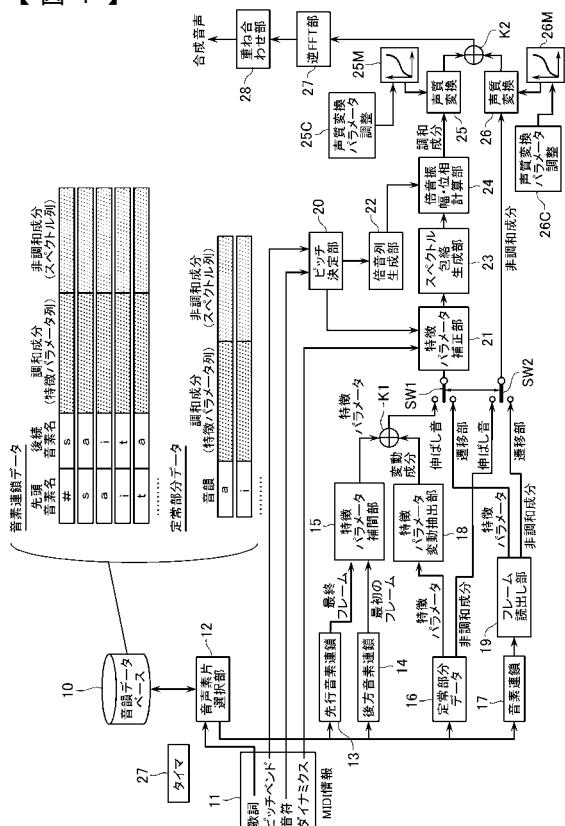
【図6】 第1の実施の形態に係る歌唱合成装置におけるデータ処理の手順を示すフローチャートである。

【図7】 マッピング関数Mfの他の例を示す。

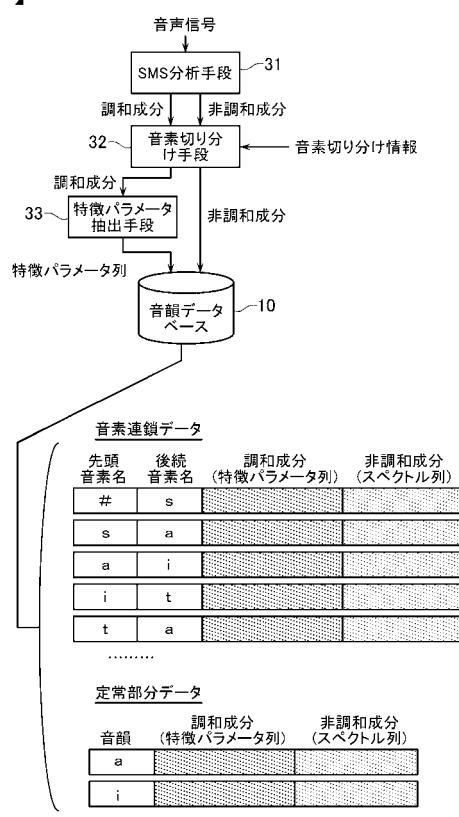
【符号の説明】

10...音韻データベース、11...演奏データ保持部、12...音素片選択部、13...先行音素連鎖データ保持部、14...後方音素連鎖データ保持部、15...特徴パラメータ補間部、16...定常部分データ保持部、17...音素連鎖データ保持部、18...特徴パラメータ変動抽出部、19...フレーム読み出し部、K1、K2...加算部、20...ピッチ決定部、21...特徴パラメータ補正部、22...倍音列生成部、23...スペクトル包絡生成部、24...倍音振幅・位相計算部、25, 26...声質変換部、25M、26M...マッピング関数生成部、25C、26C...声質変換パラメータ調整部、27...逆FFT部、28...重ね合せ部、29...タイマ、31...SMS分析手段、32...音素切り分け手段、33...特徴パラメータ抽出手段、41...振幅決定手段、43...倍音列生成手段、44...振幅計算手段、K3...加算部、45...ゲイン補正部  
10

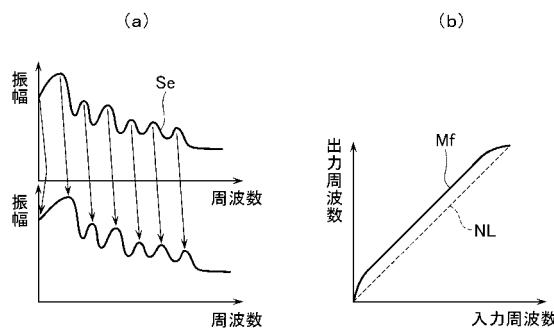
【図1】



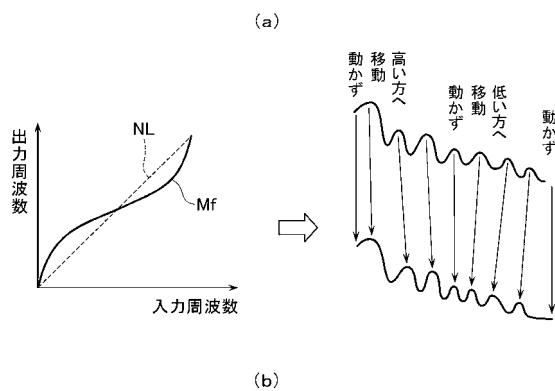
【図2】



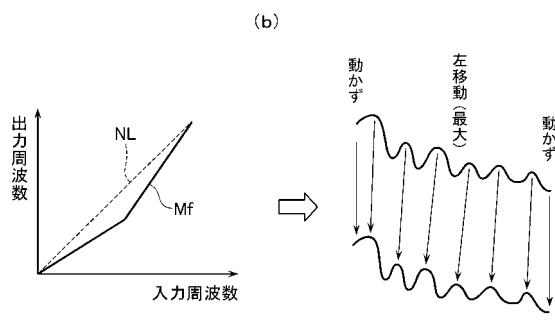
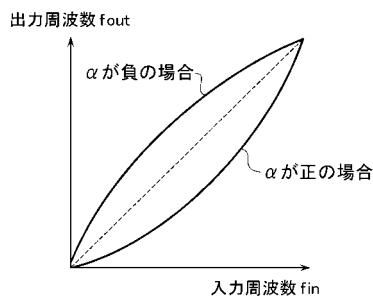
【図3】



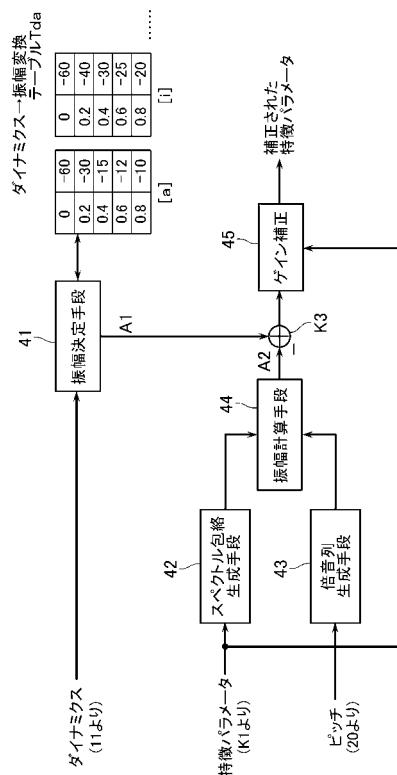
【図4】



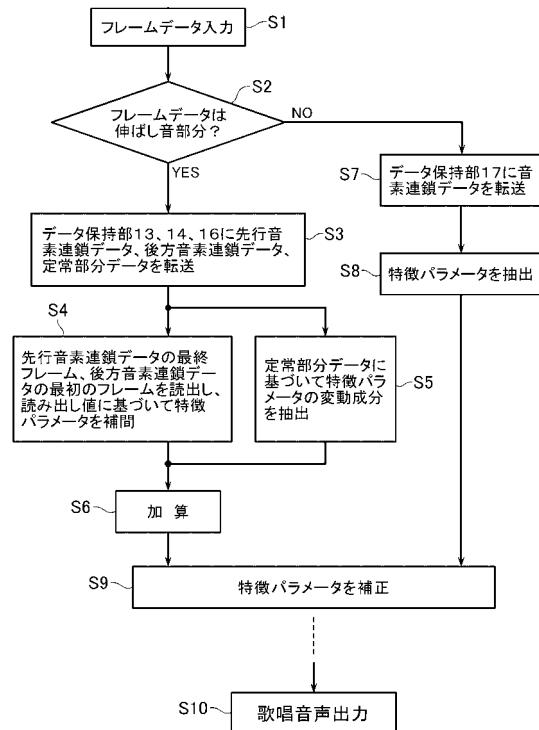
【図7】



【図5】



【図6】



---

フロントページの続き

- (56)参考文献 特開2000-250572(JP,A)  
特表2001-522471(JP,A)  
特開2001-013963(JP,A)  
特開平09-179576(JP,A)  
特開平10-260815(JP,A)  
特許第3732793(JP,B2)  
特許第3703394(JP,B2)  
特開2003-066982(JP,A)  
特開平05-260082(JP,A)  
特開2003-087437(JP,A)  
特開平07-104792(JP,A)  
特開2003-223178(JP,A)  
特開平11-015489(JP,A)  
特開平10-049192(JP,A)  
蓑田他, ホルマント分析・合成による声質変換, 電子情報通信学会技術研究報告[音声], 日本  
1992年 5月22日, Vol.92、No.35、SP92-7, p.1-8  
三木他, スペクトル包絡及びピッチ変換による声質変換, 日本音響学会2001年春季研究発表  
会講演論文集, 日本, 2001年 3月14日, Vol.1、3-Q-5, p.351-352  
阿部, 波形処理による声質変換装置(VarioVoice), 日本音響学会平成9年度春季研究発表会講  
演論文集, 日本, 1997年 3月17日, Vol.1、3-7-6, p.269-270

(58)調査した分野(Int.Cl., DB名)

G10L 21/04

G10L 13/02