

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3966074号

(P3966074)

(45) 発行日 平成19年8月29日(2007.8.29)

(24) 登録日 平成19年6月8日(2007.6.8)

(51) Int. Cl.		F I		
G 1 0 L	13/08	(2006.01)	G 1 0 L	13/08 1 3 2
G 1 0 L	13/02	(2006.01)	G 1 0 L	13/08 1 2 7 A
			G 1 0 L	13/02 1 2 2 B

請求項の数 7 (全 13 頁)

(21) 出願番号	特願2002-152787 (P2002-152787)	(73) 特許権者	000004075
(22) 出願日	平成14年5月27日(2002.5.27)		ヤマハ株式会社
(65) 公開番号	特開2003-345400 (P2003-345400A)		静岡県浜松市中区中沢町10番1号
(43) 公開日	平成15年12月3日(2003.12.3)	(74) 代理人	100125689
審査請求日	平成16年11月24日(2004.11.24)		弁理士 大林 章
		(74) 代理人	100125335
			弁理士 矢代 仁
		(74) 代理人	100121108
			弁理士 高橋 太郎
		(72) 発明者	久湊 裕司
			静岡県浜松市中沢町10番1号ヤマハ株式会社内
		審査官	新川 圭二

最終頁に続く

(54) 【発明の名称】 ピッチ変換装置、ピッチ変換方法及びプログラム

(57) 【特許請求の範囲】

【請求項1】

ピッチデータの示すピッチを有する歌唱音声信号を合成する歌唱合成手段を備えた歌唱合成装置において使用されるピッチ変換装置であって、
合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力する入力手段と、
複数の入力ピッチをそれぞれ複数の音声ピッチに変換するためのピッチ変換関数であって、
入力ピッチが所定の下限ピッチよりも低い場合には入力ピッチより高くなるように、
入力ピッチが所定の上限ピッチよりも高い場合には入力ピッチより低くなるように、
入力ピッチが所定の下限ピッチと所定の上限ピッチとの間である場合には入力ピッチと等しくなるように変換するピッチ変換関数を記憶する記憶手段と、
前記入力手段から入力されるピッチ毎に該ピッチを前記ピッチ変換関数に基づいて音声ピッチに変換し、該音声ピッチを示すデータを前記ピッチデータとして前記歌唱合成手段に供給する変換手段とを備えたピッチ変換装置。

【請求項2】

前記入力手段は、歌唱者を示す歌唱者データを入力し、前記記憶手段は、前記ピッチ変換関数を歌唱者毎に記憶し、前記変換手段は、前記歌唱者データの示す歌唱者に対応するピッチ変換関数に基づいてピッチ変換を行なう請求項1記載のピッチ変換装置。

【請求項3】

前記変換手段は、ピッチ変換の際に入力ピッチに依存する乱数的なピッチ変動を音声ピッチに付与する請求項1又は2記載のピッチ変換装置。

10

20

【請求項 4】

前記変換手段は、ピッチ変換の際に経時的なピッチ変動を音声ピッチに付与する請求項 1 又は 2 記載のピッチ変換装置。

【請求項 5】

複数の入力ピッチをそれぞれ複数の音声ピッチに変換するためのピッチ変換関数であって、入力ピッチが所定の下限ピッチよりも低い場合には入力ピッチより高くなるように、入力ピッチが所定の上限ピッチよりも高い場合には入力ピッチより低くなるように、入力ピッチが所定の下限ピッチと所定の上限ピッチとの間である場合には入力ピッチと等しくなるように変換するピッチ変換関数を記憶する記憶手段と、ピッチデータの示すピッチを有する歌唱音声信号を合成する歌唱合成手段とを備えた歌唱合成装置において使用される ピッチ変換方法であって、

10

合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力するステップと、このステップで入力されるピッチ毎に該ピッチを前記ピッチ変換関数に基づいて音声ピッチに変換し、該音声ピッチを示すデータを前記ピッチデータとして前記歌唱合成手段に供給するステップとを含むピッチ変換方法。

【請求項 6】

コンピュータと、ピッチデータの示すピッチを有する歌唱音声信号を合成する歌唱合成手段とを備えた歌唱合成装置において使用されるプログラムであって、前記コンピュータを、

合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力する入力手段と、複数の入力ピッチをそれぞれ複数の音声ピッチに変換するためのピッチ変換関数であって、入力ピッチが所定の下限ピッチよりも低い場合には入力ピッチより高くなるように、入力ピッチが所定の上限ピッチよりも高い場合には入力ピッチより低くなるように、入力ピッチが所定の下限ピッチと所定の上限ピッチとの間である場合には入力ピッチと等しくなるように変換するピッチ変換関数を記憶する記憶手段と、

20

前記入力手段から入力されるピッチ毎に該ピッチを前記ピッチ変換関数に基づいて音声ピッチに変換し、該音声ピッチを示すデータを前記ピッチデータとして前記歌唱合成手段に供給する変換手段として機能させるプログラム。

【請求項 7】

コンピュータを備えた歌唱合成装置において使用されるプログラムであって、前記コンピュータを、

30

合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力する入力手段と、複数の入力ピッチをそれぞれ複数の音声ピッチに変換するためのピッチ変換関数であって、入力ピッチが所定の下限ピッチよりも低い場合には入力ピッチより高くなるように、入力ピッチが所定の上限ピッチよりも高い場合には入力ピッチより低くなるように、入力ピッチが所定の下限ピッチと所定の上限ピッチとの間である場合には入力ピッチと等しくなるように変換するピッチ変換関数を記憶する記憶手段と、

前記入力手段から入力されるピッチ毎に該ピッチを前記ピッチ変換関数に基づいて音声ピッチに変換し、該音声ピッチを示すピッチデータを送出する変換手段と、

この変換手段から送出されるピッチデータの示す音声ピッチを有する歌唱音声信号を合成する歌唱合成手段として機能させるプログラム。

40

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、歌唱合成に用いるに好適なピッチ変換装置、ピッチ変換方法及びプログラムに関するものである。

【0002】

【従来の技術】

従来、音声合成装置としては、合成音声のピッチにゆらぎを付与するようにしたものが知られている（例えば、特開平 9 - 281994 号公報参照）。

50

【0003】

この従来技術では、デジタル音声波形データをアナログ音声信号に変換するD/A変換器に供給するクロック信号として、格納部から読出したクロック間隔ゆらぎデータに応じてクロック周期にゆらぎをもたせたクロック信号を用いることによりD/A変換出力(アナログ音声信号)のピッチにゆらぎを付与している。

【0004】

【発明が解決しようとする課題】

人間がある音符に対応する音声を発生するとき、物理的に一定の高さ(ピッチ)で発生するのは歌唱を職業とする人でも困難であり、一般的に発声ピッチは音符ピッチから多少ずれ、加えて経時的なピッチ変動も生ずる。特に、歌唱を職業としない一般の人が歌唱した場合には、上記のようなピッチずれやピッチ変動の傾向が強く、歌唱の上手さ(又は下手さ)を評価するための1つの要素となる。また、ピッチのずれ方に歌唱者の特徴が見られる場合もある。その上、人が発声できる上限又は下限に近いピッチの音声を発生しようとすると、声の発生機構に物理的な負担がかかるため、発生したいピッチと、実際に発声したピッチとが異なる(上限近くの高音ではピッチが下がり易く、下限近くの低音ではピッチが上がり易い)という現象がある。

10

【0005】

上記した従来技術によれば、クロック間隔ゆらぎデータの値をピッチ上昇方向又はピッチ下降方向に変化させることによりピッチ変動の方向及び量を变化させることができるが、平均ピッチで見た場合にピッチ変動を加える前のピッチ(基準ピッチ)を变化させることはできず、ピッチ変動の時間的なパターンを变化させることもできない。換言すれば、上記したような歌唱者の発声ピッチや経時的なピッチ変動を再現することはできない。

20

【0006】

この発明の目的は、歌唱合成の際に歌唱者の発声ピッチや経時的なピッチ変動を再現することができる新規なピッチ変換装置、ピッチ変換方法及びプログラムを提供することにある。

【0007】

【課題を解決するための手段】

この発明に係る第1のピッチ変換装置は、ピッチデータの示すピッチを有する歌唱音声信号を合成する歌唱合成手段を備えた歌唱合成装置において使用されるピッチ変換装置であって、合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力する入力手段と、複数の入力ピッチをそれぞれ複数の音声ピッチに変換するためのピッチ変換関数であって、入力ピッチが所定の下限ピッチよりも低い場合には入力ピッチより高くなるように、入力ピッチが所定の上限ピッチよりも高い場合には入力ピッチより低くなるように、入力ピッチが所定の下限ピッチと所定の上限ピッチとの間である場合には入力ピッチと等しくなるように変換するピッチ変換関数を記憶する記憶手段と、前記入力手段から入力されるピッチ毎に該ピッチを前記ピッチ変換関数に基づいて音声ピッチに変換し、該音声ピッチを示すデータを前記ピッチデータとして前記歌唱合成手段に供給する変換手段とを備えたものである。

30

【0008】

第1のピッチ変換装置によれば、複数の入力ピッチをそれぞれ複数の音声ピッチに変換するためのピッチ変換関数が記憶手段に記憶され、このピッチ変換関数に基づいて入力に係る各ピッチが歌唱音声合成用の音声ピッチに変換される。ピッチ変換関数において、複数の音声ピッチとして歌唱者の複数の発声ピッチをそれぞれ用いると、合成歌唱音声において歌唱者の発声ピッチやピッチ特徴を再現することができ、例えば発声可能な上限ピッチの近くではピッチを若干低くすると共に発声可能な下限ピッチの近くではピッチを若干高くすることができる。

40

【0009】

第1のピッチ変換装置において、前記入力手段は、歌唱者を示す歌唱者データを入力し、前記記憶手段は、前記ピッチ変換関数を歌唱者毎に記憶し、前記変換手段は、前記歌唱者

50

データの示す歌唱者に対応するピッチ変換関数に基づいてピッチ変換を行なうようにしてもよい。このようにすると、歌唱者毎に発声ピッチやピッチ特徴を再現することができる。

【0010】

第1のピッチ変換装置においては、ピッチ変換の際に入力ピッチに依存する乱数的な(ランダムな)ピッチ変動を音声ピッチに付与するようにしてもよい。このようにすると、合成歌唱音声に一層自然なピッチ変化を付与することができる。また、ピッチ変換の際に歌唱者の実際の音声に含まれる経時的なピッチ変動を音声ピッチに付与するようにしてもよい。このようにすると、歌唱者の経時的に不安定なピッチ変動を再現することができる。

【0014】

この発明に係る第1のピッチ変換方法は、複数の入力ピッチをそれぞれ複数の音声ピッチに変換するためのピッチ変換関数であって、入力ピッチが所定の下限ピッチよりも低い場合には入力ピッチより高くなるように、入力ピッチが所定の上限ピッチよりも高い場合には入力ピッチより低くなるように、入力ピッチが所定の下限ピッチと所定の上限ピッチとの間である場合には入力ピッチと等しくなるように変換するピッチ変換関数を記憶する記憶手段と、ピッチデータの示すピッチを有する歌唱音声信号を合成する歌唱合成手段とを備えた歌唱合成装置において使用されるピッチ変換方法であって、合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力するステップと、このステップで入力されるピッチ毎に該ピッチを前記ピッチ変換関数に基づいて音声ピッチに変換し、該音声ピッチを示すデータを前記ピッチデータとして前記歌唱合成手段に供給するステップとを含むものである。

【0015】

第1のピッチ変換方法によれば、第1のピッチ変換装置に関して前述したと同様にピッチ変換を行なうことができる。

【0018】

この発明に係る第1のプログラムは、コンピュータと、ピッチデータの示すピッチを有する歌唱音声信号を合成する歌唱合成手段とを備えた歌唱合成装置において使用されるプログラムであって、前記コンピュータを、合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力する入力手段と、複数の入力ピッチをそれぞれ複数の音声ピッチに変換するためのピッチ変換関数であって、入力ピッチが所定の下限ピッチよりも低い場合には入力ピッチより高くなるように、入力ピッチが所定の上限ピッチよりも高い場合には入力ピッチより低くなるように、入力ピッチが所定の下限ピッチと所定の上限ピッチとの間である場合には入力ピッチと等しくなるように変換するピッチ変換関数を記憶する記憶手段と、前記入力手段から入力されるピッチ毎に該ピッチを前記ピッチ変換関数に基づいて音声ピッチに変換し、該音声ピッチを示すデータを前記ピッチデータとして前記歌唱合成手段に供給する変換手段として機能させるものである。

【0019】

この発明に係る第2のプログラムは、コンピュータを備えた歌唱合成装置において使用されるプログラムであって、前記コンピュータを、合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力する入力手段と、複数の入力ピッチをそれぞれ複数の音声ピッチに変換するためのピッチ変換関数であって、入力ピッチが所定の下限ピッチよりも低い場合には入力ピッチより高くなるように、入力ピッチが所定の上限ピッチよりも高い場合には入力ピッチより低くなるように、入力ピッチが所定の下限ピッチと所定の上限ピッチとの間である場合には入力ピッチと等しくなるように変換するピッチ変換関数を記憶する記憶手段と、前記入力手段から入力されるピッチ毎に該ピッチを前記ピッチ変換関数に基づいて音声ピッチに変換し、該音声ピッチを示すピッチデータを送出する変換手段と、この変換手段から送られるピッチデータの示す音声ピッチを有する歌唱音声信号を合成する歌唱合成手段として機能させるものである。

【0020】

第1又は第2のプログラムによれば、第1のピッチ変換装置に関して前述したと同様にピ

10

20

30

40

50

ッチ変換を行なうことができる。

【 0 0 2 1 】

この発明に係る第 3 のプログラムは、コンピュータと、ピッチデータの示すピッチを有する歌唱音声信号を合成する歌唱合成手段とを備えた歌唱合成装置において使用されるプログラムであって、前記コンピュータを、

合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力する入力手段と、複数の入力ピッチのうちの各入力ピッチ毎に該入力ピッチに対する音声ピッチの経時的変動分を示すピッチ差分データを記憶する記憶手段と、前記入力手段から入力されるピッチ毎に該ピッチに対応するピッチ差分データを前記記憶手段から読出すと共に入力に係るピッチに対して読出しに係るピッチ差分データの示す音声ピッチの経時的変動分を加算してピッチ変換を行ない、このピッチ変換後のピッチを示すデータを前記ピッチデータとして前記歌唱合成手段に供給する変換手段として機能させるものである。

10

【 0 0 2 2 】

この発明に係る第 4 のプログラムは、コンピュータを備えた歌唱合成装置において使用されるプログラムであって、前記コンピュータを、

合成すべき順次の歌唱音声にそれぞれ対応して順次にピッチを入力する入力手段と、複数の入力ピッチのうちの各入力ピッチ毎に該入力ピッチに対する音声ピッチの経時的変動分を示すピッチ差分データを記憶する記憶手段と、

前記入力手段から入力されるピッチ毎に該ピッチに対応するピッチ差分データを前記記憶手段から読出すと共に入力に係るピッチに対して読出しに係るピッチ差分データの示す音声ピッチの経時的変動分を加算してピッチ変換を行ない、このピッチ変換後のピッチを示すピッチデータを送出する変換手段と、

20

この変換手段から送出されるピッチデータの示す音声ピッチを有する歌唱音声信号を合成する歌唱合成手段として機能させるものである。

【 0 0 2 3 】

第 3 又は第 4 のプログラムによれば、第 2 のピッチ変換装置に関して前述したと同様にピッチ変換を行なうことができる。

【 0 0 2 4 】

30

【 発明の実施の形態 】

図 1 は、この発明の一実施形態に係る歌唱合成装置を示すものである。

【 0 0 2 5 】

図 1 の歌唱合成装置は、入力部 1 0、ピッチ変換装置 1 2 及び歌唱合成器 1 8 を含むもので、ピッチ変換装置 1 2 は、ピッチ変換器 1 4 及びデータベース 1 6 を備えている。

【 0 0 2 6 】

入力部 1 0 は、歌唱者を示す歌唱者データ、音声素片（単一の音素 [音韻] 又は音素連鎖）を示す音声素片データ、音符のピッチ及び長さを示す音符データ、合成音声の音強度を示す音強度データ等を入力するもので、入力に係る音符ピッチ P_i を示す音符ピッチデータ及び入力に係る歌唱者 S を示す歌唱者データは、ピッチ変換器 1 4 に供給される。

40

【 0 0 2 7 】

データベース 1 6 には、複数の入力ピッチ（音符ピッチ）をそれぞれ複数の音声ピッチ（出力ピッチ）に変換するためのピッチ変換データがピッチ変換関数 $[F T (S , p)]$ 又はピッチ変換表の形で歌唱者毎に記憶されている。

【 0 0 2 8 】

図 2 には、歌唱者 S_1 , S_2 , S_3 にそれぞれ対応する 3 つのピッチ変換関数 $F T (S_1 , p)$, $F T (S_2 , p)$, $F T (S_3 , p)$ をデータベース 1 6 に記憶した例を示す。ここで、 p は、入力ピッチを表わす。

【 0 0 2 9 】

図 2 に示すピッチ変換装置 1 2 において、ピッチ変換器 1 4 は、入力部 1 0 からの歌唱者

50

データの示す歌唱者 S に対応するピッチ変換関数をデータベース 16 にて参照すると共に、入力部 10 からの音符ピッチデータの示す音符ピッチ P_i に対応する音声ピッチ P_o を参照に係るピッチ変換関数に基づいて算出する。そして、算出された音声ピッチ P_o を示すピッチデータを歌唱合成器 18 に出力する。

【0030】

データベース 16 がピッチ変換データをピッチ変換表の形で記憶している場合、ピッチ変換器 14 は、入力部 10 からの歌唱者 S に対応するピッチ変換表を参照すると共に、入力部 10 からの音符ピッチデータの示す音符ピッチ P_i に対応する音声ピッチ P_o を参照に係るピッチ変換表から読出す。そして、読出された音声ピッチ P_o を示すピッチデータを歌唱合成器 18 に供給する。

10

【0031】

歌唱合成器 18 は、入力部 10 からの歌唱者データ、音声素片データ、音符長データ及び音強度データと、ピッチ変換器 14 からのピッチデータとに基づいて歌唱音声信号を合成するものである。歌唱合成方式としては、種々のものが公知であり、そのうちから適切なものを選択して歌唱合成器 18 を構成することができる。

【0032】

歌唱合成器 18 では、一例として、歌唱者データの示す歌唱者と、音声素片データの示す音声素片とに対応した音声成分データを用いて歌唱音声信号を合成する。このとき、歌唱音声信号のピッチ、音長及び音強度は、それぞれピッチデータ、音符長データ及び音強度データに応じて決定される。

20

【0033】

図 3 は、ピッチ変換関数の一例を示すものである。図 3 において、横軸の入力ピッチ [cent] は、ピッチ変換器 14 に入力される音符ピッチに相当し、縦軸の出力ピッチ [cent] は、ピッチ変換器 14 から出力される音声ピッチに相当する。

【0034】

図 3 に示すピッチ変換関数 $FT(S, p)$ は、所定の下限ピッチ PL と所定の上限ピッチ PH との間では出力ピッチが入力ピッチと等しいが、入力ピッチが上限ピッチ PH より高いときは人の発声可能な上限ピッチに近づくにつれて徐々に出力ピッチが入力ピッチより低くなると共に、入力ピッチが下限ピッチ PL より低いときは人の発声可能な下限ピッチに近づくにつれて徐々に出力ピッチが入力ピッチより高くなるような形状になっている。

30

【0035】

【数 1】

$$FT(S, p) > p \quad \text{if } p < PL$$

$$FT(S, p) = p \quad \text{if } PL \leq p \leq PH$$

$$FT(S, p) < p \quad \text{if } PH < p$$

具体例としては、 $PH < p$ の領域では出力ピッチが入力ピッチより最大で数 10 セント程度低くなると共に $p < PL$ の領域では出力ピッチが入力ピッチより最大で数 10 セント程度高くなるようなピッチ変換関数を用いることができる。

【0036】

図 3 に示したようなピッチ変換関数は、歌唱者毎に適切な形状のものが用意され、図 2 に関して前述したように歌唱者毎にデータベース 16 に記憶される。ピッチ変換器 14 は、入力に係る歌唱者 S に対応するピッチ変換関数を参照して入力ピッチ P_i を出力ピッチ P_o に変換する。このようなピッチ変換を数式的に表現すると、次の数 2 の通りである。

40

【0037】

【数 2】

$$P_o = FT(S, P_i)$$

図 4 は、図 3 のピッチ変換関数を用いたピッチ変換の一例を示すもので、(A) は、変換前のピッチ変化(入力ピッチの変化)を示し、(B) は、変換後のピッチ変化(出力ピッチの変化)を示す。図 4 (A) において、順次の入力ピッチは、合成すべき順次の歌唱音

50

声にそれぞれ対応するものである。図4によれば、PLより低音域では、出力ピッチが入力ピッチより高くなると共にPHより高音域では出力ピッチが入力ピッチより低くなり、PL以上でPH以下の中音域では出力ピッチが入力ピッチに等しくなっているのがわかる。図4の例では、入力ピッチを離散的に与えたが、そうである必要はなく、連続的に与えても構わない。

【0038】

図3に示したピッチ変換関数は、直線に近似したものであるが、歌唱者やピッチに依存する乱数的な(ランダムな)ピッチ変動分 $rand(S, p)$ を加えた次の数3の式に示すようなピッチ変換関数を用いてもよい。

【0039】

【数3】

$$FT(S, p) + rand(S, p)$$

このようなピッチ変換関数を用いると、ピッチ変換の際に図4(A)に示すような順次の入力ピッチにそれぞれ応答して順次の出力ピッチにランダムなピッチ変化が加わるようになり、合成音声に一層自然な変化を付与することができる。

【0040】

上記した実施形態において、データベース16には、時間に依存しないピッチ変換関数 $FT(S, p)$ を記憶する例を示したが、データベース16には、時間に依存するピッチ変換関数を記憶し、このピッチ変換関数を参照してピッチ変換を行なうようにしてもよい。一例としてデータベース16には、ピッチ差分 $FT(S, p, t)$ を示すピッチ差分データをピッチ変換データとして歌唱者毎に記憶する。ピッチ差分 $FT(S, p, t)$ は、歌唱者Sが音符ピッチpに対応する音声を発生したときに時間tの進行に従って音符ピッチpに対する音声ピッチの差分を表わすものである。

【0041】

データベース16にピッチ差分データをピッチ変換関数 $FT(S, p, t)$ の形で歌唱者毎に記憶しておいた場合、ピッチ変換器14は、入力に係る歌唱者Sに対応するピッチ変換関数 $FT(S, p, t)$ を参照して入力ピッチ P_i を出力ピッチ P_o に変換する。このようなピッチ変換を数式的に表現すると、次の数4に示す通りである。

【0042】

【数4】

$$P_o = P_i + FT(S, P_i, t)$$

この場合のピッチ変換は、入力ピッチ P_i に対して入力ピッチ P_i 対応のピッチ差分 $FT(S, P_i, t)$ を加算することにより行なわれる。

【0043】

データベース16には、上記のようにピッチ変換関数 $FT(S, p, t)$ を記憶する代わりに、ピッチ差分 $FT(S, p, t)$ の経時的な変化波形を表わすピッチ差分データを記憶するようにしてもよい。図5は、このようなピッチ差分データを歌唱者 $S_1 \dots S_n$ (n は2以上の整数)のうちの各歌唱者毎に $p_1 \sim p_{25}$ の25ピッチ分記憶した例を示す。ピッチ $p_1 \sim p_{25}$ は、100セント(半音)刻みで1200~3600[cent]となっている。データベース16にピッチ差分データを記憶すると、後述のピッチ波形データを記憶する場合に比べてデータ量が少なくて済む。

【0044】

図5の例において、各ピッチ毎のピッチ差分データとしては、実際の歌唱に基づくものを用いるとよい。一例を示すと、歌唱者 S_1 にピッチ p_1 に対応する音声を実際に発生させると共に、ピッチ p_1 に対する発生音声のピッチの差分の経時的変化波形を求め、この変化波形を表わすピッチ差分データを用いる。このようにすると、歌唱者の特性を反映したピッチ変化を再現可能になると共に、より人間的な微細なピッチ変化を表現可能になる。

【0045】

ピッチ変換器14は、入力に係る歌唱者Sに対応するピッチ差分データのうち入力ピッチ P_i に対応するピッチ差分データを参照して前述の数4の式に従って入力ピッチ P_i を出

10

20

30

40

50

力ピッチ P_0 に変換する。図 6 は、図 5 のピッチ差分データを用いたピッチ変換の一例を示すもので、(A) は、図 4 (A) と同様に変換前のピッチ変化 (入力ピッチの変化) を示し、(B) は、変換後のピッチ変化 (出力ピッチの変化) を示す。図 6 によれば、人の発声可能な上限ピッチ又は下限ピッチの近傍では図 4 に関して前述したと同様に出力が入力ピッチよりそれぞれ低く又は高くなると共に、人の発声可能な上限ピッチ又は下限ピッチの近傍ではピッチの変動量 (ゆらぎ量) が大きくなることからわかる。従って、人間的な発声ピッチやピッチ変動の再現が可能となる。

【0046】

ピッチは、離散値ではなく連続値であるので、図 5 の例においてすべてのピッチに対応可能とするには全ピッチ分のピッチ差分データを記憶することになり、記憶するデータ量が膨大なものになってしまう。また、ピッチ差分 $FT(S, p, t)$ の変化が長く継続するピッチ差分データについても、記憶するデータ量が膨大なものになってしまう。このような記憶データ量の増大を回避するためには、次の (イ) 又は (ロ) のような対策を適宜採用することができる。

【0047】

(イ) 複数の離散的なピッチについてそれぞれピッチ差分 $FT(S, p, t)$ を表わすピッチ差分データを記憶した場合において、入力ピッチとピッチが丁度一致するピッチ差分データを検出できないときは、入力ピッチとピッチが最も近いピッチ差分データを参照してピッチ変換を行なう。また、入力ピッチとピッチが近い 2 つのピッチ差分データから補間により新たなピッチ差分データを求めてピッチ変換を行なってもよい。

【0048】

(ロ) ピッチ差分データとしては、ピッチ差分の変化継続時間が所定値以内のものを記憶しておき、入力ピッチの時間長がピッチ差分 $FT(S, p, t)$ の変化継続時間を越えたときは、ピッチ差分の変化波形において時間 0 等の適当な位置に戻って再びピッチ差分データを読み出す。

【0049】

図 5 の例において、データベース 16 には、歌唱者毎に複数のピッチにそれぞれ対応してピッチ差分の変化波形を表わすピッチ差分データを記憶したが、歌唱者毎に複数のピッチにそれぞれ対応してピッチ変化波形を表わすピッチ波形データをピッチ変換データとして記憶するようにしてもよい。この場合、入力に係る歌唱者 S に対応する複数のピッチ波形データのうち入力ピッチ P_i に対応するピッチ波形データを読み出して出力ピッチ P_0 とすることによりピッチ変換を行なう。ピッチ波形データを実際の歌唱に基づいて作成すると、歌唱者の発声ピッチや経時的なピッチ変動を再現することができる。

【0050】

上記したようなピッチ変換処理は、パーソナルコンピュータ等の小型コンピュータにおいてソフトウェア処理として実行するようにしてもよい。すなわち、ROM 又は RAM 等の記憶手段にストアしたプログラムに従って CPU (中央処理装置) にピッチ変換処理を実行させるようにしてもよい。

【0051】

図 7 は、この発明の他の実施形態に係る歌唱合成装置を示すもので、この装置は、例えば特許第 2906970 号に示される SMS (Spectral Modeling Synthesis) 技術を用いて歌唱合成を行なうものである。

【0052】

ステップ S1 では、歌唱音声信号を入力し、ステップ S2 では、入力された歌唱音声信号に SMS 分析処理及び区間切出し処理を施す。

【0053】

SMS 分析処理では、入力音声信号を一連の時間フレームに区分し、各フレーム毎に FFT (Fast Fourier Transform) 等により 1 組の強度 (マグニチュード) スペクトルデータを生成し、各フレーム毎に 1 組の強度スペクトルデータから複数のピークに対応する線スペクトルを抽出する。これらの線スペクトルの振幅値及び周波数を表わすデータを調和

10

20

30

40

50

成分 (Deterministic Component) のデータと称する。次に入力音声波形のスペクトルから調和成分のスペクトルを差引いて残差スペクトルを得る。この残差スペクトルを非調和成分 (Stochastic Component) と称する。

【 0 0 5 4 】

区間切出し処理では、SMS 分析処理で得られた調和成分のデータ及び非調和成分のデータを音声素片に対応して区分する。音声素片とは、歌詞の構成要素であり、例えば [a] , [i] のような単一の音素 (又は音韻: Phoneme) 又は例えば「 a_i 」, 「 a_p 」のような音素連鎖 (複数音素の連鎖) からなるものである。

【 0 0 5 5 】

データベース 20 には、音声素片毎に調和成分のデータ及び非調和成分のデータが記憶される。また、データベース 20 には、データベース 16 に関して前述したと同様にピッチ変換データ (ピッチ差分データ又はピッチ波形データである場合も含む) が記憶されている。

10

【 0 0 5 6 】

歌唱合成に際しては、ステップ S 3 で歌詞データ及びメロディデータを入力する。そして、ステップ S 4 では、歌詞データが表わす音素列に音素列 / 音声素片変換処理を施して音素列を音声素片に区分し、音声素片毎にそれに対応する調和成分のデータ及び非調和成分のデータを音声素片データとしてデータベース 20 から読出す。

【 0 0 5 7 】

ステップ S 5 では、データベース 20 から読出された音声素片データ (調和成分のデータ及び非調和成分のデータ) に音声素片接続処理を施して音声素片データ同士を発音順に接続する。

20

【 0 0 5 8 】

ステップ S 6 では、ピッチ変換処理を行なう。すなわち、ステップ S 3 で入力されたメロディデータの示す音符ピッチを前述したと同様にしてデータベース 20 のピッチ変換データ (ピッチ差分データ又はピッチ波形データである場合も含む) に基づいて音声ピッチに変換し、該音声ピッチを示すピッチデータを生成する。

【 0 0 5 9 】

ステップ 7 では、音声素片毎に調和成分のデータとステップ S 6 で生成されたピッチデータの示す音声ピッチとに基づいて該音声ピッチに適合した新たな調和成分のデータを生成する。このとき、新たな調和成分のデータでは、ステップ S 5 の処理を受けた調和成分のデータが表わすスペクトル包絡の形状をそのまま引継ぐようにスペクトル強度を調整すると、ステップ S 1 で入力した音声信号の音色を再現することができる。

30

【 0 0 6 0 】

ステップ S 8 では、ステップ S 7 で生成した調和成分のデータとステップ S 5 の処理を受けた非調和成分のデータとを音声素片毎に加算する。そして、ステップ S 9 では、ステップ S 8 で加算処理を受けたデータを音声素片毎に逆 FFT 等により時間領域の歌唱音声信号に変換する。ステップ 9 の処理の結果として得られる歌唱音声信号は、デジタル形式の信号であり、D/A 変換器 22 によりアナログ形式の歌唱音声信号に変換される。

【 0 0 6 1 】

一例として、「サイタ」(s a i t a) という歌唱音声合成するには、データベース 20 から音声素片「# s」, 「s_a」, 「a」, 「a_i」, 「i」, 「i_t」, 「t_a」, 「a」, 「a#」(#は無音を表わす) にそれぞれ対応する音声素片データを読み出してステップ S 5 で接続する。そして、ステップ S 7 では、音声素片毎にステップ S 6 での変換に係るピッチを有する調和成分のデータを生成し、ステップ S 8 の加算処理、ステップ S 9 の変換処理及び変換器 22 の D/A 変換処理を経ると、「サイタ」の歌唱音声信号が得られる。

40

【 0 0 6 2 】

上記したステップ S 1 ~ S 9 の処理は、パーソナルコンピュータ等の小型コンピュータにおいてソフトウェア処理として実行してもよく、あるいは電子回路等のハードウェアを用

50

いて実行してもよい。

【0063】

【発明の効果】

以上のように、この発明によれば、記憶手段に記憶したピッチ変換データを用いて入力音符ピッチを歌唱合成用の音声ピッチに変換する構成にしたので、歌唱者の発声ピッチや経時的なピッチ変動を再現できる効果が得られる。また、人間の実際の発声における経時的に不安定なピッチ変動を忠実に再現したり、音域による発声ピッチの精度の違いを表現したり、歌唱者によるピッチ変化の違いを表現したりすることも可能となる。

【0064】

その上、入力ピッチに対してピッチ差分データの示す音声ピッチの変動分を加算してピッチ変換を行なう構成にしたので、記憶するデータ量が少なく済む利点もある。 10

【図面の簡単な説明】

【図1】 この発明の一実施形態に係る歌唱合成装置を示すブロック図である。

【図2】 ピッチ変換装置の動作を説明するためのブロック図である。

【図3】 ピッチ変換関数の一例を示すグラフである。

【図4】 図3のピッチ変換関数を用いたピッチ変換の一例を示す図であり、(A)は、変換前のピッチ変化を示すグラフ、(B)は、変換後のピッチ変化を示すグラフである。

【図5】 データベースにおけるピッチ差分データの記憶状況を示すグラフである。

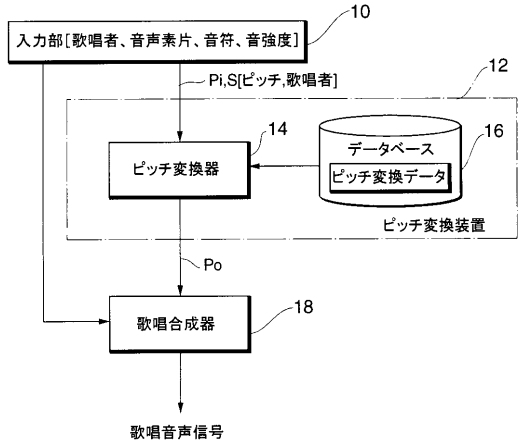
【図6】 図5のピッチ差分データを用いたピッチ変換の一例を示す図であり、(A)は、変換前のピッチ変化を示すグラフ、(B)は、変換後のピッチ変化を示すグラフである 20

。【図7】 この発明の他の実施形態に係る歌唱合成装置を示すブロック図である。

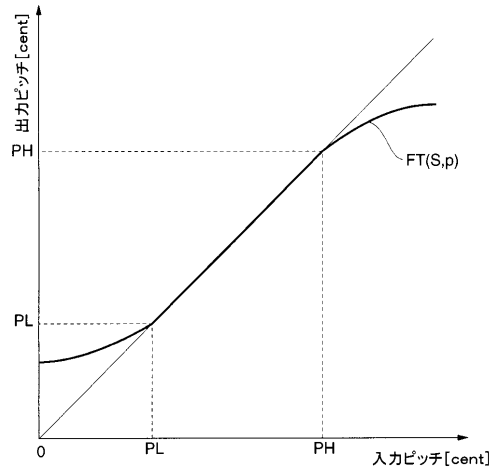
【符号の説明】

10：入力部、12：ピッチ変換装置、14：ピッチ変換器、16, 20：データベース、18：歌唱合成器、22：D/A変換器、S1：歌唱音声信号入力処理、S2：SMS分析及び区間切出し処理、S3：歌詞データ及びメロディデータ入力処理、S4：音素列-音声素片変換処理、S5：音声素片接続処理、S6：ピッチ変換処理、S7：調和成分生成処理、S8：加算処理、S9：時間領域の歌唱音声信号に変換する処理。

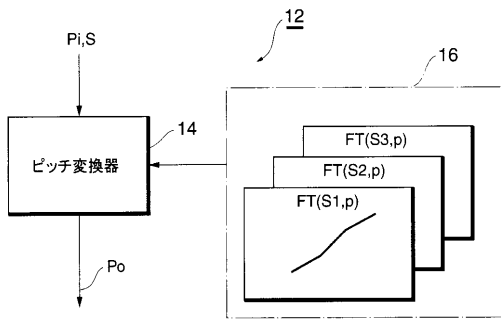
【 図 1 】



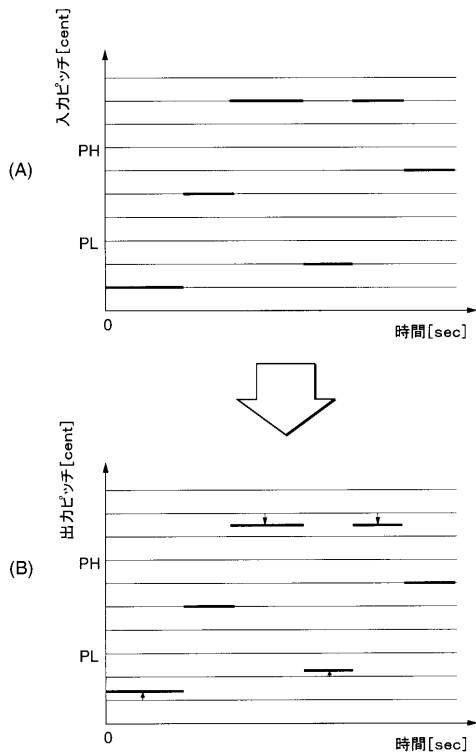
【 図 3 】



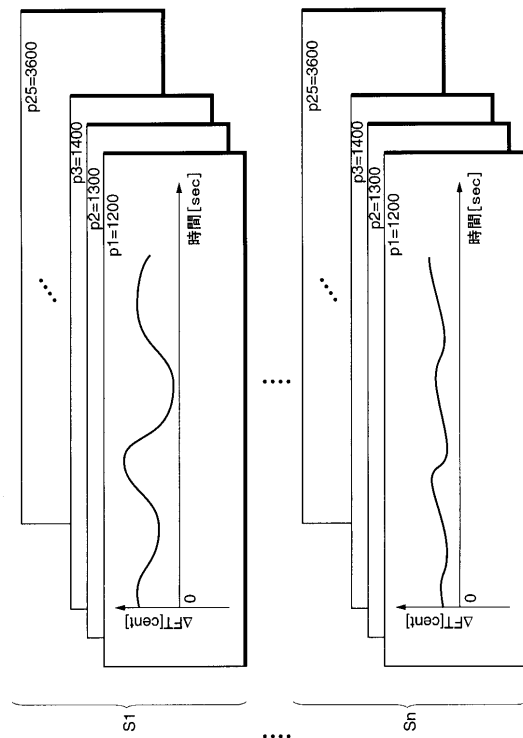
【 図 2 】



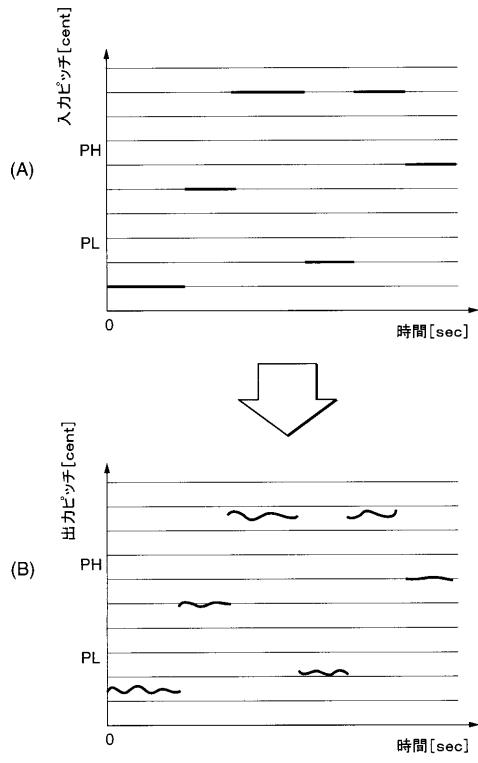
【 図 4 】



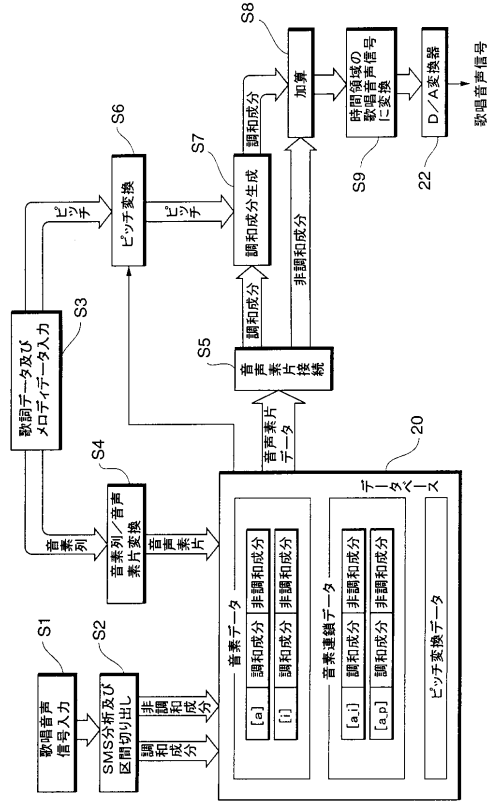
【 図 5 】



【 図 6 】



【 図 7 】



フロントページの続き

- (56)参考文献 特開平7 - 146695 (JP, A)
特開2002 - 73064 (JP, A)
特開平1 - 112296 (JP, A)
特開2002 - 149180 (JP, A)
特開平2 - 129700 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 13/00-13/08

G10K 15/04