

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4322283号
(P4322283)

(45) 発行日 平成21年8月26日(2009.8.26)

(24) 登録日 平成21年6月12日(2009.6.12)

(51) Int. Cl.		F I			
G 1 0 H	1/00	(2006.01)	G 1 0 H	1/00	Z
G 1 0 G	1/00	(2006.01)	G 1 0 G	1/00	
G 1 0 G	3/04	(2006.01)	G 1 0 G	3/04	
			G 1 0 H	1/00	1 0 2 Z
			G 1 0 H	1/00	B

請求項の数 4 (全 22 頁)

(21) 出願番号 特願2007-45193 (P2007-45193)
 (22) 出願日 平成19年2月26日(2007.2.26)
 (65) 公開番号 特開2008-209572 (P2008-209572A)
 (43) 公開日 平成20年9月11日(2008.9.11)
 審査請求日 平成20年9月29日(2008.9.29)

(73) 特許権者 301021533
 独立行政法人産業技術総合研究所
 東京都千代田区霞が関1-3-1
 (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 100111763
 弁理士 松本 隆
 (72) 発明者 後藤 真孝
 茨城県つくば市東1-1-1 独立行政法
 人産業技術総合研究所つくばセンター内
 (72) 発明者 藤島 琢哉
 静岡県浜松市中沢町10番1号 ヤマハ株
 式会社内

最終頁に続く

(54) 【発明の名称】 演奏判定装置およびプログラム

(57) 【特許請求の範囲】

【請求項1】

演奏音を外部から取得する演奏音取得手段と、
 前記演奏音取得手段により取得される演奏音が本来有しているべき基本周波数である目標基本周波数を取得する目標基本周波数取得手段と、

楽器が発生する音の高調波構造を模した音モデルを用い、前記演奏音取得手段により取得された演奏音の周波数成分が各種の基本周波数に対応した音モデルを重み付け加算した混合分布となるように各音モデルに対する重み値を逐次更新して最適化し、前記演奏音取得手段により取得された演奏音の基本周波数を推定する手段であって、前記各種の基本周波数に対応した音モデルに対する重み値の最適化に際して、各音モデルに対する重み値に
対し、前記目標基本周波数取得手段により取得された目標基本周波数における重み値を強調した事前分布を付与する基本周波数推定手段と、

前記目標基本周波数取得手段により取得された目標基本周波数と前記基本周波数推定手段における基本周波数の推定結果とに基づいて、前記目標基本周波数取得手段により取得された目標基本周波数と前記演奏音取得手段により取得された演奏音の基本周波数との一致判定を行う判定手段と

を具備することを特徴とする演奏判定装置。

【請求項2】

前記基本周波数推定手段は、前記音モデルの高調波構造における各高調波成分の大きさの比率を変化させつつ、前記各音モデルに対する重み値を逐次更新して最適化する処理を

行うことを特徴とする請求項 1 に記載の演奏判定装置。

【請求項 3】

前記判定手段は、前記目標基本周波数取得手段により取得された目標基本周波数においてピークとなる確率密度関数を発生し、この目標基本周波数の確率密度関数と前記基本周波数推定手段により最適化された各種の基本周波数に対応した音モデルの重み値により与えられる演奏音の基本周波数の確率密度関数とに基づいて、前記一致判定を行うことを特徴とする請求項 1 または 2 に記載の演奏判定装置。

【請求項 4】

コンピュータを、

演奏音を外部から取得する演奏音取得手段と、

前記演奏音取得手段により取得される演奏音が本来有しているべき基本周波数である目標基本周波数を取得する目標基本周波数取得手段と、

楽器が発生する音の高調波構造を模した音モデルを用い、前記演奏音取得手段により取得された演奏音の周波数成分が各種の基本周波数に対応した音モデルを重み付け加算した混合分布となるように各音モデルに対する重み値を逐次更新して最適化し、前記演奏音取得手段により取得された演奏音の基本周波数を推定する手段であって、前記各種の基本周波数に対応した音モデルに対する重み値の最適化に際して、各音モデルに対する重み値に対し、前記目標基本周波数取得手段により取得された目標基本周波数における重み値を強調した事前分布を付与する基本周波数推定手段と、

前記目標基本周波数取得手段により取得された目標基本周波数と前記基本周波数推定手段における基本周波数の推定結果とに基づいて、前記目標基本周波数取得手段により取得された目標基本周波数と前記演奏音取得手段により取得された演奏音の基本周波数との一致判定を行う判定手段として機能させることを特徴とするコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、楽譜等において指定された通りの音高の演奏音が発音されているか否かの判定を行う演奏判定装置および演奏判定プログラムに関する。

【背景技術】

【0002】

従来から演奏の独習機能を備えた楽器が各種提供されている。鍵盤楽器を例に挙げると、この種の独習機能を備えた楽器は、表示装置への表示等により、押鍵すべき鍵をユーザに指示し、ユーザによって押鍵された鍵を検知して、正しい押鍵がなされているかをユーザに知らせ、ユーザに鍵盤演奏の独習を行わせる。このような独習機能を実現するためには、ユーザによって押鍵された鍵を検知する必要がある。従って、キースキャン機構を備えていない鍵盤楽器には、このような独習機能を持たせることができないという問題があった。そこで、鍵盤操作等を検知するのではなく、演奏音を收音して、その周波数解析を行い、楽譜通りの正しい音高の演奏音が発生しているか否かを判定する技術が提案されるに至った。例えば特許文献 1 に開示された技術では、各種の音高のピアノ音を收音して、それらの周波数解析を行い、各音高のピアノ音のパワースペクトルを求めて予め記憶しておく。そして、ピアノ演奏が行われる際には、演奏音を收音し、その周波数解析を行ってパワースペクトルを求め、この演奏音のパワースペクトルと予め記憶した各種の音高のピアノ音の各パワースペクトルとの類似度を求め、類似度の高さにより楽譜通りの演奏がなされているか否かを判定する。

【特許文献 1】特開 2004 - 341026 号公報

【特許文献 2】特許第 3413634 号

【非特許文献 1】後藤 真孝："リアルタイム音楽情景記述システム：全体構想と音高推定手法の拡張", 情報処理学会 音楽情報科学研究会 研究報告 2000-MUS-37-2, Vol.2000, No.94, pp.9-16, 2000 年 10 月 16 日

【発明の開示】

【発明が解決しようとする課題】

【0003】

しかしながら、一般に楽音のパワースペクトルは多くの周波数位置に倍音成分を持っており、各成分の割合も多様であるため、対比される2つの楽音がある場合に、それらの基本周波数は異なっているがパワースペクトルの形状は類似している、ということが起こりうる。このため、上述した特許文献1の技術では、ある基本周波数の演奏音が收音された場合に、この演奏音とは基本周波数が異なるがパワースペクトルの形状が類似したピアノ音が選択され、演奏音の音高が誤って判定される可能性がある、という問題があった。また、特許文献1の技術では、收音した演奏音の基本周波数を求めないので、「本来、この音高であるべきものをこの音高で演奏した。」といった形式で演奏の誤りを指摘することができない、という問題があった。

10

【0004】

この発明は、以上説明した事情に鑑みてなされたものであり、演奏音の基本周波数を正確に判定することができる演奏判定装置を提供することを目的としている。

【課題を解決するための手段】

【0005】

この発明は、演奏音を外部から取得する演奏音取得手段と、前記演奏音取得手段により取得される演奏音が本来有しているべき基本周波数である目標基本周波数を取得する目標基本周波数取得手段と、楽器が発生する音の高調波構造を模した音モデルを用い、前記演奏音取得手段により取得された演奏音の周波数成分が各種の基本周波数に対応した音モデルを重み付け加算した混合分布となるように各音モデルに対する重み値を逐次更新して最適化し、前記演奏音取得手段により取得された演奏音の基本周波数を推定する基本周波数推定手段と、前記目標周波数取得手段により取得された目標基本周波数と前記基本周波数推定手段における基本周波数の推定結果とに基づいて、前記演奏音取得手段により取得された演奏音の基本周波数に関する判定を行う判定手段とを具備することを特徴とする演奏判定装置およびコンピュータを該演奏判定装置として機能させるコンピュータプログラムを提供する。

20

【0006】

かかる発明によれば、楽器が発生する音の高調波構造を模した音モデルを用い、演奏音取得手段により取得された演奏音の周波数成分が各種の基本周波数に対応した音モデルを重み付け加算した混合分布となるように各音モデルに対する重み値を逐次更新して最適化し、演奏音取得手段により取得された演奏音の基本周波数を推定するようにしたので、演奏音の基本周波数を精度良く推定し、演奏音の基本周波数に関する判定を正確に行うことができる。

30

【発明を実施するための最良の形態】

【0007】

以下、図面を参照し、この発明の実施の形態を説明する。

【0008】

<全体構成>

図1は、この発明による演奏判定装置の一実施形態である教習伴奏システムの構成を示すブロック図である。この教習伴奏システムは、例えば鍵盤楽器等の楽器において稼動するシステムであり、ユーザに当該楽器の演奏の独習を行わせるシステムである。図1において、制御部101は、各種のプログラムを実行するCPUと、このCPUにワークエリアとして用いられるRAM等により構成されている。図1において、制御部101を表すボックス内には、制御部101のCPUによって実行されるプログラムのうち、本実施形態による教習伴奏システムとしての機能を実現するためのプログラムの処理内容が図示されている。操作部102は、各種のコマンドや情報をユーザから受け取るための装置であり、楽器の本体に配備されたパネルスイッチ等の操作子群によって構成されている。HDD(ハードディスク装置)103は、各種のプログラムやデータベースを記憶した記憶装置である。本実施形態による教習伴奏システムとしての機能を実現するためのプログラム

40

50

もこのHDD103に記憶されている。制御部101のCPUは、操作部102の操作により、教習伴奏システムとしての機能を起動すべき旨のコマンドが与えられたとき、この教習伴奏システムとしての機能を実現するためのプログラムをRAMにロードして実行する。

【0009】

收音部104は、外界の音を收音してアナログ音響信号を出力するマイクロフォンと、このアナログ音声信号をデジタル音響信号に変換するA/D変換器等により構成されている。本実施形態において、この收音部104は、演奏音を外部から取得する演奏音取得手段として用いられる。

【0010】

楽曲記憶部105は、楽曲データを記憶するための記憶装置であり、例えばRAMにより構成されている。ここで、楽曲データは、楽曲を構成するメロディパート、ベースパート等の各種のパートに対応した演奏データの集合体である。1つのパートに対応した演奏データは、演奏音の発生を指示するイベントデータと、演奏音の発生タイミングを指示するタイミングデータとにより構成された時系列データである。データ入力部106は、各種の楽曲の楽曲データを外部から取り込むための手段である。例えばFDやICメモリ等の記憶媒体から楽曲データを読み出す装置やネットワークを介してサーバから楽曲データをダウンロードする通信装置等がこのデータ入力部106として用いられる。

【0011】

サウンドシステム107は、デジタル音響信号をアナログ音響信号に変換するD/A変換器と、アナログ音響信号を音として出力するスピーカ等により構成されている。表示部108は、例えば液晶パネルディスプレイであり、本実施形態では、演奏すべき楽曲の表示、押鍵すべき鍵をユーザに示すための鍵盤の画像の表示、ユーザにより行われた演奏が適切なものであったか否かの判定結果の表示等を行うための手段として用いられる。なお、判定結果は、表示に限らず、報知音や振動による提示などによりユーザに示してもよい。

【0012】

次に、本実施形態における教習伴奏システムとしての機能を実現するためのプログラムの処理内容について説明する。まず、楽曲入力処理111は、操作部102を介して与えられるコマンドに従って、データ入力部106により楽曲データ105aを取得し、楽曲記憶部105に格納する処理である。演奏位置制御処理112は、ユーザが演奏すべき位置を制御するとともに、その演奏位置に対応した演奏データを楽曲記憶部105内の楽曲データ105aから抽出して出力し、かつ、この抽出した演奏データに基づき、ユーザが演奏すべき音の基本周波数である目標基本周波数を求めて出力する処理である。演奏位置制御処理112における演奏位置の制御の態様には2通りある。第1の態様は、ユーザがあるパートの楽器演奏をしており、この楽器演奏によりある演奏音が発生したとき、その演奏音が楽曲データ中の該当パートの演奏データにより指定された正しい音高を持った演奏音である場合に、その演奏音の次の演奏音の位置に演奏位置を進める態様である。第2の態様は、自動演奏、すなわち、パート毎にタイミングデータにより指定されたタイミングで各イベントデータを順次読み出し、この読み出しに連動させて演奏位置を進める態様である。演奏位置制御処理112にいずれの態様による演奏位置の制御を行わせるかは、操作部102を介して与えられるコマンドにより決定される。また、ユーザが楽曲データ105aにおけるいずれのパートの演奏を行うかも、操作部102を介して与えられるコマンドにより決定される。

【0013】

楽曲再生処理113は、演奏位置制御処理112により出力された演奏位置における演奏データの中からユーザの演奏パート以外のパートの演奏データを選択し、その演奏データが示す演奏音(すなわち、伴奏音)の波形のサンプルデータを生成してサウンドシステム107に供給する処理である。楽曲表示処理114は、ユーザが演奏を行うべき演奏位置および演奏音を示す情報を表示部108に表示させる処理である。この楽曲表示処理1

10

20

30

40

50

14の態様にも各種のものがあり得る。ある態様において、楽曲表示処理114では、楽曲データ105aに従って、演奏すべき楽曲の楽譜を表示部108に表示させるとともに、演奏位置における演奏データに基づいて、ユーザが演奏すべき演奏位置を示すマークをその楽譜上に表示させる。他の態様における楽曲表示処理114では、例えば鍵盤の画像を表示部108に表示させるとともに、演奏位置における演奏データに基づいて、ユーザが押鍵すべき鍵を示すマークを表示させる。

【0014】

基本周波数推定処理115は、楽器が発生する音の高調波構造を模した音モデル115Mを用い、收音部104によって收音された演奏音の周波数成分が各種の基本周波数に対応した音モデル115Mを重み付け加算した混合分布となるように各音モデル115Mに対する重み値を最適化し、最適化された各音モデルに対する重み値に基づいて、收音部104によって收音された演奏音の基本周波数を推定する処理である。本実施形態における基本周波数推定処理115では、演奏位置制御処理112から出力される目標基本周波数を事前知識として使用し、基本周波数の推定を行う。類似度評価処理116は、基本周波数推定処理115により推定された基本周波数と、演奏位置制御処理112により得られた目標基本周波数との類似度を算出する処理である。一致判定処理117は、類似度評価処理116により得られた類似度に基づき、基本周波数推定処理115により推定された基本周波数と演奏位置制御処理112により得られた目標基本周波数とが一致しているといえるか否かを判定する処理である。一致判定処理117の判定結果は、判定結果表示処理118に引き渡される他、上述した演奏位置制御処理112に引き渡される。演奏位置制御処理112では、上述した第1の態様が操作部102の操作により選択されている場合、一致判定処理117の判定結果が肯定的である場合に限り、演奏位置を次の演奏音の位置に進める制御が行われる。判定結果表示処理118は、一致判定処理117の判定結果、すなわち、ユーザが演奏データ通りの音高で演奏音を発生させたか否かを表示部108に表示させる処理である。

【0015】

<基本周波数推定処理115の内容>

次に本実施形態における基本周波数推定処理115の処理内容の詳細について説明する。この基本周波数推定処理115は、特許文献2に開示された技術を基盤とし、これに非特許文献1に開示された改良を加えたものである。

【0016】

特許文献2の技術では、メロディ音のものと考えられる帯域に属する周波数成分と、ベース音のものと考えられる帯域に属する周波数成分とを入力音響信号からBPFにより別々に取り出し、それらの各帯域の周波数成分に基づき、メロディ音およびベース音の各々の基本周波数の推定を行う。

【0017】

さらに詳述すると、特許文献2の技術では、音の高調波構造に対応した確率分布を持った音モデルを用意し、メロディ音の帯域の各周波数成分、ベース音の帯域の各周波数成分が、様々な基本周波数に対応した各音モデルを重み付け加算した混合分布であると考えられる。そして、各音モデルの重みの値をEM(Expectation-Maximization)アルゴリズムを用いて推定する。

【0018】

このEMアルゴリズムは、隠れ変数を含む確率モデルに対して最尤推定を行うための反復アルゴリズムであり、局所最適解を求めることができる。ここで、最も大きな重みの値を持つ確率分布は、その時点で最も優勢な高調波構造であるとみなすことができるため、あとはその優勢な高調波構造における基本周波数を音高として求めればよい。この手法は基本周波数成分の存在に依存しないため、ミッシングファンダメンタル現象も適切に扱うことができ、基本周波数成分の存在に依存せずに、最も優勢な高調波構造を求めることができる。

【0019】

非特許文献 1 は、特許文献 2 の技術に対して次の拡張を行ったものである。

< 拡張 1 : 音モデルの多重化 >

特許文献 2 の技術では、同一基本周波数には 1 つの音モデルしか用意されていなかったが、実際には、ある基本周波数に、異なる高調波構造を持つ音が入れ替わり立ち替わり現れることがある。そこで、同一基本周波数に対して複数の音モデルを用意し、入力音響信号をそれらの混合分布としてモデル化した。

< 拡張 2 : 音モデルのパラメータの推定 >

特許文献 2 の技術では、音モデルにおいて各高調波成分の大きさの比を固定していた（ある理想的な音モデルを仮定していた）。これは実世界の混合音中の高調波構造とは必ずしも一致しておらず、精度向上のためには洗練される余地が残されていた。そこで、音モデルの高調波成分の比率もモデルパラメータに加え、各時刻において EM アルゴリズムにより推定するようにした。

< 拡張 3 : モデルパラメータに関する事前分布の導入 >

特許文献 2 の技術では、音モデルの重み（基本周波数の確率密度関数）に関する事前知識は仮定していなかった。しかし、この基本周波数の推定技術の用途によっては、たとえ事前に基本周波数がどの周波数の近傍にあるかを与えてでも、より誤検出の少ない基本周波数を求めたいというような要求も発生し得る。例えば、演奏分析やビブラート分析等の目的では、楽曲をヘッドホン聴取しながらの歌唱や楽器演奏によって、各時刻におけるおおよその基本周波数を事前知識として用意しておき、実際の楽曲中のより正確な基本周波数を求めることが求められている。そこで、特許文献 2 におけるモデルパラメータ（音モデルに対する重み値）の最尤推定の枠組みを拡張し、モデルパラメータに関する事前分布に基づいて最大事後確率推定（MAP 推定；Maximum A Posteriori Probability Estimation）を行うようにした。その際、< 拡張 2 > においてモデルパラメータに加えた音モデルの高調波成分の大きさの比率に関する事前分布も導入した。

【 0 0 2 0 】

図 2 は、特許文献 2 の技術と非特許文献 1 の技術を組み合わせることにより構成された本実施形態における基本周波数推定処理 1 1 5 の処理内容を示す図である。基本周波数推定処理 1 1 5 ではメロディラインとベースラインを推定する。メロディは他よりも際立って聞こえる単音の系列、ベースはアンサンブル中で最も低い単音の系列であり、その時間的な変化の軌跡をそれぞれメロディライン $D_m(t)$ 、ベースライン $D_b(t)$ と呼ぶ。時刻 t における基本周波数 F_0 を $F_i(t)$ ($i = m, b$)、振幅を $A_i(t)$ とすると、これらは以下のように表される。

【数 1】

$$D_m(t) = \{F_m(t), A_m(t)\} \quad (1)$$

【数 2】

$$D_b(t) = \{F_b(t), A_b(t)\} \quad (2)$$

【 0 0 2 1 】

このメロディライン $D_m(t)$ およびベースライン $D_b(t)$ を收音部 1 0 4 により收音された演奏音を示す入力音響信号から得るための手段として、基本周波数推定処理 1 1 5 は、瞬時周波数の算出 1、周波数成分の候補の抽出 2、周波数帯域の制限 3、メロディラインの推定 4 a およびベースラインの推定 4 b の各処理を含む。また、メロディラインの推定 4 a およびベースラインの推定 4 b の各処理は、基本周波数の確率密度関数の推定

4 1 およびマルチエージェントモデルによる基本周波数の継時的な追跡 4 2 の各処理を各々含む。本実施形態では、ユーザの演奏パートがメロディパートである場合にはメロディラインの推定 4 a が実行され、ユーザの演奏パートがベースパートである場合にはベースラインの推定 4 b が実行されるようになっている。

【 0 0 2 2 】

< < 瞬時周波数の算出 1 > >

この処理では、入力音響信号を複数の B P F からなるフィルタバンクに与え、フィルタバンクの各 B P F の出力信号について、位相の時間微分である瞬時周波数 (Flanagan, J. L. and Golden, R.M.: Phase Vocoder, The Bell System

Technical J., Vol.45, pp.1493-1509 (1966)参照) を計算する。ここでは、上記 Flanagan の手法を用い、短時間フーリエ変換 (STFT) の出力をフィルタバンク出力と解釈して、効率良く瞬時周波数を計算する。入力音響信号 $x(t)$ に対する窓関数 $h(t)$ を用いた STFT が式 (3) および (4) により与えられるとき、瞬時周波数 (ω, t) は式 (5) により求めることができる。

10

【 0 0 2 3 】

【数 3】

$$X(\omega, t) = \int_{-\infty}^{+\infty} x(\tau) h(t-\tau) e^{-j\omega\tau} d\tau \quad (3)$$

20

【数 4】

$$= a + jb \quad (4)$$

【数 5】

$$\lambda(\omega, t) = \omega + \frac{a \frac{\partial b}{\partial t} - b \frac{\partial a}{\partial t}}{a^2 + b^2} \quad (5)$$

30

【 0 0 2 4 】

ここで、 $h(t)$ は時間周波数の局所化を与える窓関数である (例えば、最適な時間周波数の局所化を与えるガウス関数に 2 階のカーディナル B - スプライン関数を畳み込んで作成した時間窓など)。

40

【 0 0 2 5 】

この瞬時周波数を計算するのに、ウェーブレット変換を用いても良い。ここでは、計算量を減らすために STFT を用いるが、単一の STFT のみを用いたのでは、ある周波数帯域における時間分解能や周波数分解能が悪くなってしまう。そこで、マルチレートフィルタバンク (Vetterli, M.: A Theory of Multirate Filter Banks, IEEE Trans. on ASSP, Vol. ASSP-35, No.3, pp. 356-372 (1987)、参照) を構成し、リアルタイムに実行可能という制約のもとで、ある程度妥当な時間周波数分解能を得る。

【 0 0 2 6 】

< < 周波数成分の候補の抽出 2 > >

この処理では、フィルタの中心周波数からその瞬時周波数への写像に基づいて、周波数

50

成分の候補を抽出する (Charpentier, F.J.: Pitch detection using the short-term phase

spectrum, Proc. of ICASSP 86, pp.113-116 (1986)参照)。あるSTFTフィルタの中心周波数からその出力の瞬時周波数 $(\lambda(\phi, t))$ への写像を考える。すると、もし周波数の周波数成分があるときには、 $\lambda(\phi, t)$ がこの写像の不動点に位置し、その周辺の瞬時周波数の値はほぼ一定となる。つまり、全周波数成分の瞬時周波数 $\lambda_f^{(t)}$ は、次式によって抽出することができる。

【0027】

【数6】

$$\Psi_f^{(t)} = \left\{ \Psi \mid \lambda(\phi, t) - \phi = 0, \frac{\partial}{\partial \phi} (\lambda(\phi, t) - \phi) < 0 \right\} \quad (6)$$

10

これらの周波数成分のパワーは、 $\lambda_f^{(t)}$ の各周波数におけるSTFTパワースペクトルの値として得られるため、周波数成分のパワー分布関数 $\Psi_p^{(t)}(\omega)$ を次のように定義することができる。

【数7】

$$\Psi_p^{(t)}(\omega) = \begin{cases} |X(\omega, t)| & \text{if } \omega \in \Psi_f^{(t)} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

20

【0028】

<<周波数帯域の制限3>>

この処理では、抽出した周波数成分に重み付けすることで、周波数帯域を制限する。ここでは、メロディラインとベースライン用に、2種類のBPFを用意する。メロディライン用BPFは、典型的なメロディラインの主要な基本波成分および高調波成分の多くを通過させることができ、かつ、基本周波数付近の重複が頻繁に起きる周波数帯域をある程度遮断する。一方、ベースライン用BPFは、典型的なベースラインの主要な基本周波数成分および高調波成分の多くを通過させることができ、かつ、他の演奏パートがベースラインよりも優勢になるような周波数帯域をある程度遮断する。

30

【0029】

本実施形態では、以下、対数スケールの周波数をcentの単位(本来は音高差(音程)を表す尺度)で表し、Hzで表された周波数 f_{Hz} を、次のようにcentで表された周波数 f_{cent} に変換する。

40

【数8】

$$f_{cent} = 1200 \log_2 \frac{f_{Hz}}{REF_{Hz}} \quad (8)$$

【数 9】

$$\text{REF}_{\text{Hz}} = 440 \times 2^{\frac{3}{12} - 5} \quad (9)$$

平均律の半音は 100 cent に、1 オクターブは 1200 cent に相当する。

【0030】

10

周波数 x cent での BPF の周波数応答を $\text{BPF}_i(x)$ ($i = m, b$) とし、周波数成分のパワー分布関数を $\Psi_p^{(t)}(x)$ とすると、BPF を通過した周波数成分は $\text{BPF}_i(x) \Psi_p^{(t)}(x)$ と表すことができる。ただし、 $\Psi_p^{(t)}(x)$ は、周波数軸が cent で表されていることを除けば $\Psi_p^{(t)}(\)$ と同じ関数である。ここで、次の段階の準備として、BPF を通過した周波数成分の確率密度関数 $p^{(t)}(x)$ を定義する。

【数 10】

$$p_{\Psi}^{(t)}(x) = \frac{\text{BPF}_i(x) \Psi_p^{(t)}(x)}{\text{Pow}^{(t)}} \quad (10) \quad 20$$

【0031】

ここで、 $\text{Pow}^{(t)}$ は次式に示すように BPF を通過した周波数成分のパワーの合計である。

【数 11】

$$\text{Pow}^{(t)} = \int_{-\infty}^{+\infty} \text{BPF}_i(x) \Psi_p^{(t)}(x) dx \quad (11) \quad 30$$

【0032】

<< 基本周波数の確率密度関数の推定 41 >>

この基本周波数の確率密度関数の推定 41 では、BPF を通過した周波数成分の候補に対し、各高調波構造が相対的にどれくらい優勢かを表す基本周波数の確率密度関数を求める。この基本周波数の確率密度関数の推定 41 の処理内容を、非特許文献 1 に開示されたものに改良を加えた内容となっている。

40

【0033】

基本周波数の確率密度関数の推定 41 では、上述した「拡張 1」と「拡張 2」を実現するために、同一基本周波数に対して M_i 種類の音モデルがあるものとし (i はメロディ用 ($i = m$) かベース用 ($i = b$) かを示す)、基本周波数が F であり、音モデルの種類が m 番目の種類であり、モデルパラメータ $\mu^{(t)}(F, m)$ を持った音モデル $p(x | F, m, \mu^{(t)}(F, m))$ を次のように定義する。

【数 1 2】

$$\begin{aligned}
 & p(x | F, m, \mu^{(t)}(F, m)) \\
 &= \sum_{h=1}^{H_i} p(x, h | F, m, \mu^{(t)}(F, m))
 \end{aligned}$$

(1 2)

10

【数 1 3】

$$\begin{aligned}
 & p(x, h | F, m, \mu^{(t)}(F, m)) \\
 &= c^{(t)}(h | F, m) G(x; F + 1200 \log_2 h, W_i)
 \end{aligned}$$

(1 3)

20

【数 1 4】

$$\mu^{(t)}(F, m) = \{c^{(t)}(h | F, m) | h = 1 \sim H_i\}$$

(1 4)

【数 1 5】

$$G(x; x_0, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

(1 5)

30

【0 0 3 4】

この音モデルは、基本周波数がFのときに、その高調波成分がどの周波数にどれくらい現れるかをモデル化したものである。H_iは基本周波数成分も含めた高調波成分の数、W_i²はガウス分布G(x; x₀,)の分散を表す。c^(t)(h | F, m)は、基本周波数がFであるm番目の音モデルの第h次調波成分の大きさを表し、次式を満たす。

40

【数 1 6】

$$\sum_{h=1}^{H_i} c^{(t)}(h | F, m) = 1$$

(1 6)

【0 0 3 5】

50

上記式(16)に示すように、基本周波数Fの音モデルにおける重み $c^{(t)}(h|F, m)$ は、総和が1となるような予め定義された重みを用いられる。

【0036】

基本周波数の確率密度関数の推定41では、以上のような音モデルを使用し、周波数成分の確率密度関数 $p^{(t)}(x)$ が、次式で定義されるような $p(x|F, m, \mu^{(t)}(F, m))$ の混合分布モデル $p(x|\theta^{(t)})$ から生成されたと考える。

【数17】

$$p(x|\theta^{(t)}) = \int_{F_{li}}^{F_{hi}} \sum_{m=1}^{M_i} w^{(t)}(F, m) p(x|F, m, \mu^{(t)}(F, m)) dF \quad (17) \quad 10$$

【数18】

$$\theta^{(t)} = \{w^{(t)}, \mu^{(t)}\} \quad (18) \quad 20$$

【数19】

$$w^{(t)} = \{w^{(t)}(F, m) | F_{li} \leq F \leq F_{hi}, m = 1, \dots, M_i\} \quad (19)$$

【数20】

$$\mu^{(t)} = \{\mu^{(t)}(F, m) | F_{li} \leq F \leq F_{hi}, m = 1, \dots, M_i\} \quad (20) \quad 30$$

ここで、 F_{hi} と F_{li} は、許容される基本周波数の上限と下限であり、 $w^{(t)}(F, m)$ は次式を満たすような音モデルの重みである。

【数21】

$$\int_{F_{li}}^{F_{hi}} \sum_{m=1}^{M_i} w^{(t)}(F, m) dF = 1 \quad (21) \quad 40$$

【0037】

実世界の混合音に対して事前に音源数を仮定することは不可能なため、式(17)のように、あらゆる基本周波数の可能性を同時に考慮してモデル化することが重要となる。最終的に、モデル $p(x|\theta^{(t)})$ から、観測した確率密度関数 $p^{(t)}(x)$ が生成されたかのようにモデルパラメータ $\theta^{(t)}$ を推定できれば、その重み $w^{(t)}(F, m)$

)は各高調波構造が相対的にどれくらい優勢かを表すため、次式のように基本周波数の確率密度関数 $p_{F0}^{(t)}(F)$ と解釈することができる。

【数 2 2】

$$p_{F0}^{(t)}(F) = \sum_{m=1}^{Mi} w^{(t)}(F, m) \quad (F_{li} \leq F \leq F_{hi}) \quad (22)$$

10

【0038】

次に、上述した「拡張3」を実現するために、 (t) の事前分布 $p_{0i}^{(t)}$ を、式(23)のように式(24)と式(25)の積で与える。

【数 2 3】

$$p_{0i}(\theta^{(t)}) = p_{0i}(w^{(t)}) p_{0i}(\mu^{(t)}) \quad (23)$$

【数 2 4】

20

$$\begin{aligned} p_{0i}(w^{(t)}) \\ = \frac{1}{Z_w} e^{-\beta_{\mu i}^{(t)} D_w(w_{0i}^{(t)}; w^{(t)})} \end{aligned} \quad (24)$$

30

【数 2 5】

$$\begin{aligned} p_{0i}(\mu^{(t)}) \\ = \frac{1}{Z_{\mu}} e^{-\int_{F_{li}}^{F_{hi}} \sum_{m=1}^{Mi} \beta_{\mu i}^{(t)}(F, m) D_{\mu}(\mu_{0i}^{(t)}(F, m); \mu^{(t)}(F, m)) dF} \end{aligned} \quad (25)$$

40

【0039】

ここで、 $p_{0i}(w^{(t)})$ と $p_{0i}(\mu^{(t)})$ は、最も起こりやすいパラメータを $w_{0i}^{(t)}(F, m)$ と $\mu_{0i}^{(t)}(F, m)$ としたときに、そこで最大値を取るよ

50

うな単峰性の事前分布である。ただし、 Z_w 、 Z_μ は正規化係数、 $w_{0i}^{(t)}$ 、 $\mu_{0i}^{(t)}$ (F, m) は、最大値をどれくらい重視した事前分布とするかを定めるパラメータで、0 のときに無情報事前分布(一様分布)となる。また、 $D_w(w_{0i}^{(t)}; w^{(t)})$ 、 $D_\mu(\mu_{0i}^{(t)}; \mu^{(t)})$ (F, m) は、次のような K - L 情報量 (Kullback - Leibler's information) である。

【数 26】

$$D_w(W_{0i}^{(t)}; w^{(t)}) = \int_{F_{li}}^{F_{hi}} \sum_{m=1}^{M_i} w_{0i}^{(t)}(F, m) \log \frac{w_{0i}^{(t)}(F, m)}{w^{(t)}(F, m)} dF \quad (26) \quad 10$$

【数 27】

$$D_\mu(\mu_{0i}^{(t)}(F, m); \mu^{(t)}(F, m)) = \sum_{h=1}^{H_i} c_{0i}^{(t)}(h | F, m) \log \frac{c_{0i}^{(t)}(h | F, m)}{c^{(t)}(h | F, m)} \quad (27) \quad 20$$

【0040】

以上から、確率密度関数 $p^{(t)}(x)$ を観測したときに、そのモデル $p(x | \theta^{(t)})$ のパラメータ $\theta^{(t)}$ を、事前分布 $p_{0i}(\theta^{(t)})$ に基づいて推定する問題を解けばよいことがわかる。この事前分布に基づく $\theta^{(t)}$ の最大事後確率推定量 (MAP 推定量) は、次式を最大化することで得られる。

【数 28】

$$\int_{-\infty}^{+\infty} p_{\Psi}^{(t)}(x) (\log p(x | \theta^{(t)}) + \log p_{0i}(\theta^{(t)})) dx \quad (28) \quad 30$$

【0041】

この最大化問題は解析的に解くことが困難なため、前述の EM (Expectation - Maximization) アルゴリズムを用いて $\theta^{(t)}$ を推定する。EM アルゴリズムは、E ステップ (expectation step) と M ステップ (maximization step) を交互に繰返し適用することで、不完全な観測データ (この場合、 $p^{(t)}(x)$) から最尤推定をおこなうための反復アルゴリズムである。本実施形態では、EM アルゴリズムを繰り返すことにより、BPF を通過した周波数成分の確率密度関数 $p^{(t)}(x)$ を、各種の基本周波数 F に対応した複数の音モデル $p(x | F, m, \mu^{(t)}(F, m))$ を重み付け加算した混合分布と考える場合において、最も尤もらしい重みのパラメータ $\theta^{(t)} (= \{w^{(t)}(F, m), \mu^{(t)}(F, m)\})$ を求める。ここで、EM アルゴリズムの各繰返しでは、パラメータ $\theta^{(t)} (= \{w^{(t)}(F, m), \mu^{(t)}(F, m)\})$ に関して、古いパラメータ推定値 $\theta_{old}^{(t)} (= \{w_{old}^{(t)}(F, m), \mu_{old}^{(t)}(F, m)\})$ を更新して新しい (よ

り尤もらしい)パラメータ推定値 $w_{new}^{(t)}$ ($= \{w_{new}^{(t)}(F, m), \mu_{new}^{(t)}(F, m)\}$) を求めていく。 $w_{old}^{(t)}$ の初期値には、1つ前の時刻 $t-1$ における最終的な推定値を用いる。この古いパラメータ推定値 $w_{old}^{(t)}$ から新しいパラメータ推定値 $w_{new}^{(t)}$ を求める漸化式は、次のようになる。なお、この漸化式の導出過程は非特許文献 1 に詳細に説明されているので、そちらを参照されたい。

【数 29】

$$\begin{aligned} & w_{new}^{(t)}(F, m) \\ = & \frac{w_{ML}^{(t)}(F, m) + \beta_{wi}^{(t)} w_{oi}^{(t)}(F, m)}{1 + \beta_{wi}^{(t)}} \end{aligned} \quad (29) \quad 10$$

【数 30】

$$\begin{aligned} & c_{new}^{(t)}(h | F, m) \\ = & \frac{w_{ML}^{(t)}(F, m) c_{ML}^{(t)}(h | F, m) + \beta_{\mu i}^{(t)}(F, m) c_{oi}^{(t)}(h | F, m)}{w_{ML}^{(t)}(F, m) + \beta_{\mu i}^{(t)}(F, m)} \end{aligned} \quad (30) \quad 20$$

【0042】

上記式(29)および(30)における $w_{ML}^{(t)}(F, m)$ と $c_{ML}^{(t)}(h | F, m)$ は、 $w_i^{(t)} = 0$ 、 $\mu_i^{(t)}(F, m) = 0$ の無情報事前分布のとき、つまり、最尤推定の場合の推定値であり、次式により与えられる。

30

【数 31】

$$\begin{aligned} & w_{ML}^{(t)}(F, m) \\ = & \int_{-\infty}^{+\infty} p_{\Psi}^{(t)}(x) \\ & \frac{w_{old}^{(t)}(F, m) p(x | F, m, \mu_{old}^{(t)}(F, m))}{\int_{F_{li}}^{F_{hi}} \sum_{\nu=1}^{M_i} w_{old}^{(t)}(\eta, \nu) p(x | \eta, \nu, \mu_{old}^{(t)}(\eta, \nu)) d\eta} dx \end{aligned} \quad (31) \quad 40$$

【数 3 2】

$$\begin{aligned}
 & c_{ML}^{(t)}(h | F, m) \\
 &= \frac{1}{w_{ML}^{(t)}(F, m)} \int_{-\infty}^{\infty} p_{\Psi}^{(t)}(x) \\
 & \quad \frac{w_{old}^{(t)}(F, m) p(x, h | F, m, \mu_{old}^{(t)}(F, m))}{\int_{F_{li}}^{F_{hi}} \sum_{\nu=1}^{M_i} w_{old}^{(t)}(\eta, \nu) p(x | \eta, \nu, \mu_{old}^{(t)}(\eta, \nu)) d\eta} dx
 \end{aligned} \tag{32}$$

【0043】

これらの反復計算により、事前分布を考慮した基本周波数の確率密度関数 $p_{F_0}^{(t)}(F)$ が、式(23)によって $w^{(t)}(F, m)$ から求まる。さらに、すべての音モデル $p(x | F, m, \mu^{(t)}(F, m))$ の各高調波成分の大きさの比率 $c^{(t)}(h | F, m)$ も求まり、「拡張1」～「拡張3」が実現される。

【0044】

最も優勢な基本周波数 $F_i(t)$ を決定するには、次式に示すように、基本周波数の確率密度関数 $p_{F_0}^{(t)}(F)$ (式(22)より、式(29)～(32)を反復計算した最終的な推定値として得られる) を最大にする周波数を求めればよい。

【数 3 3】

$$F_i(t) = \operatorname{argmax}_F p_{F_0}^{(t)}(F) \tag{33}$$

こうして得られた周波数を音高とする。

【0045】

<< マルチエージェントモデルによる基本周波数の継時的な追跡 4.2 >>

基本周波数の確率密度関数において、同時に鳴っている音の基本周波数に対応する複数のピークが拮抗すると、それらのピークが確率密度関数の最大値として次々に選ばれてしまうことがあるため、このように単純に求めた結果は安定しないことがある。そこで、本実施形態では、大局的な観点から基本周波数を推定するために、基本周波数の確率密度関数の時間変化において複数のピークの軌跡を継時的に追跡し、その中で最も優勢で安定した基本周波数の軌跡を選択する。このような追跡処理を動的で柔軟に制御するために、マルチエージェントモデルを導入する。

【0046】

マルチエージェントモデルは、1つの特徴検出器と複数のエージェントにより構成される(図3参照)。特徴検出器は、基本周波数の確率密度関数の中で目立つピークを拾い上げる。エージェントは基本的に、それらのピークに駆動されて軌跡を追跡していく。つまり、マルチエージェントモデルは、入力中で目立つ特徴を時間的に追跡する汎用の枠組みである。具体的には、各時刻において以下の処理がおこなわれる。

【0047】

10

20

30

40

50

(1) 基本周波数の確率密度関数が求めた後、特徴検出器は目立つピーク(最大ピークに応じて動的に変化する閾値を越えたピーク)を複数検出する。そして、目立つピークのそれぞれについて、周波数成分のパワーの合計 $Pow(t)$ も考慮しながら、どれくらい将来有望なピークかを評価する。これは、現在時刻を数フレーム先の時刻とみなして、ピークの軌跡をその時刻まで先読みして追跡することで実現する。

【0048】

(2) 既に生成されたエージェントがあるときは、それらが相互作用しながら、目立つピークをそれに近い軌跡を持つエージェントへと排他的に割り当てる。複数のエージェントが割り当て候補に上がる場合には、最も信頼度の高いエージェントへと割り当てる。

【0049】

(3) 最も有望で目立つピークがまだ割り当てられていないときは、そのピークを追跡する新たなエージェントを生成する。

【0050】

(4) 各エージェントは累積ペナルティを持っており、それが一定の閾値を越えると消滅する。

【0051】

(5) 目立つピークが割り当てられなかったエージェントは、一定のペナルティを受け、基本周波数の確率密度関数の中から自分の追跡する次のピークを直接見つけようとする。もしそのピークも見つからないときは、さらにペナルティを受ける。さもなければ、ペナルティはリセットされる。

【0052】

(6) 各エージェントは、今割り当てられたピークがどれくらい有望で目立つかを表す度合いと、1つ前の時刻の信頼度との重み付き和によって、信頼度を自己評価する。

【0053】

(7) 時刻 t における基本周波数 $F_i(t)$ は、信頼度が高く、追跡しているピークの軌跡に沿ったパワーの合計が大きいエージェントに基づいて決定する。振幅 $A_i(t)$ は、基本周波数 $F_i(t)$ の高調波成分等を $p(t)$ () から抽出して決定する。

以上が本実施形態における基本周波数推定処理 115 の詳細である。

【0054】

< 本実施形態の動作 >

次に本実施形態の動作を説明する。本実施形態において演奏位置制御処理 112 では、楽曲上においてユーザが演奏すべき位置を常時監視し、その演奏位置に対応した演奏データを楽曲記憶部 105 内の楽曲データ 105a から抽出して出力し、楽曲再生処理 113 および楽曲表示処理 114 に引き渡す。また、演奏位置制御処理 112 では、演奏位置に対応した演奏データに基づいてユーザの演奏パートの演奏音の目標基本周波数を求め、基本周波数推定処理 115 に引き渡す。

【0055】

楽曲再生処理 113 では、演奏位置制御処理 112 から引き渡された演奏データに基づいて、ユーザの演奏パート以外のパートの演奏音(すなわち、伴奏音)の音響信号を生成し、サウンドシステム 107 にその再生を行わせる。また、楽曲表示処理 114 では、演奏位置制御処理 112 から引き渡された演奏データに基づいて、ユーザが演奏すべき演奏音を示す画像(例えば鍵盤において押鍵すべき鍵を示す画像)またはユーザが演奏すべき演奏位置を示す画像(楽譜上における演奏位置を示す画像)を表示部 108 に表示させる。

【0056】

ユーザが楽器演奏を行って、その演奏音が収音部 104 により収音されると、その演奏音を示す入力音響信号が基本周波数推定処理 115 に引き渡される。基本周波数推定処理 115 では、楽器が発生する音の高調波構造を模した音モデル 115M を用い、入力音響信号の周波数成分が各種の基本周波数に対応した音モデル 115M を重み付け加算した混合分布となるように各音モデル 115M に対する重み値を最適化し、最適化された各音モ

10

20

30

40

50

デルに対する重み値に基づいて、入力音響信号が示す1または複数の演奏音の基本周波数を推定する。その際、本実施形態における基本周波数推定処理115では、演奏位置制御処理112から引き渡された目標基本周波数における重みが強調された事前分布 $p_{0i}(t)$ を作成し、この事前分布 $p_{0i}(t)$ を用い、かつ、音モデルにおける各高調波成分の大きさの比を変化させつつEMアルゴリズムを実行し、演奏音の基本周波数を推定する。

【0057】

そして、類似度評価処理116では、このようにして基本周波数推定処理115により推定された基本周波数と、演奏位置制御処理112により得られた目標基本周波数との類似度を算出する。類似度として何を用いるかについては各種の様態が考えられるが、例えば基本周波数推定処理115により推定された基本周波数の目標基本周波数に対する比（すなわち、対数表現された両周波数のずれであるセント値）を所定の値（例えば一音階分のセント値）によって割り算したものを類似度としてもよい。一致判定処理117では、この類似度評価処理116により得られた類似度に基づき、基本周波数推定処理115により推定された基本周波数と演奏位置制御処理112により得られた目標基本周波数とが一致しているといえるか否かを判定する。判定結果表示処理118では、一致判定処理117の判定結果、すなわち、ユーザが演奏データ通りの音高で演奏音を発生させたか否かを表示部108に表示させる。好ましい態様では、表示部108に楽譜を表示させ、判定結果表示処理118では、ユーザに演奏の誤りを適確に知らせるため、その楽譜上において、演奏位置に対応した演奏データにより指定された演奏音の音符（すなわち、目標基本周波数を示す音符）と、実際にユーザが発生させた演奏音の基本周波数を示す音符とを例えば異なる色で表示させる。

本実施形態では、演奏位置を進めつつ、以上の処理が繰り返される。

【0058】

以上説明したように、本実施形態では、楽器が発生する音の高調波構造を模した音モデルを用い、收音部104により取得された演奏音の周波数成分が各種の基本周波数に対応した音モデルを重み付け加算した混合分布となるように各音モデルに対する重み値を最適化し、最適化された各音モデルに対する重み値に基づいて、演奏音の基本周波数を推定するようにしたので、演奏音の基本周波数を精度良く推定し、演奏音の基本周波数に関する判定を正確に行うことができる。また、本実施形態では、ユーザが発生させた演奏音の基本周波数を求めるので、「本来、この音高であるべきものをこの音高で演奏した。」といった形式で演奏の誤りをユーザに示すことができる。また、本実施形態では、演奏データから得られる目標基本周波数の位置において重みが強調された事前分布を作成し、この事前分布を用いて、基本周波数の推定を行うようにしているので、基本周波数の推定の精度を向上させることができる。また、本実施形態によれば、音モデルにおける高調波成分の大きさの比率を変えつつEMアルゴリズムを実行して演奏音の基本周波数の推定を行うようにしている。従って、ユーザが発生させる演奏音のスペクトル形状が演奏の強弱、タッチにより大きく変動する状況においても、そのスペクトル形状の変化に合わせて音モデルの高調波成分の大きさの比率を変化させ、精度よく演奏音の基本周波数の推定を行うことができる。

【0059】

<他の実施形態>

以上、この発明の一実施形態について説明したが、この発明には他にも実施形態があり得る。例えば次の通りである。

【0060】

(1) 上記実施形態では、基本周波数推定処理115において、1または複数の基本周波数を推定結果として出力するようにしたが、演奏音の基本周波数の確率密度関数を推定結果として出力するようにしてもよい。この場合、類似度評価処理116では、目標基本周波数においてピークとなるガウス分布などの確率密度関数を生成し、この目標基本周波数の確率密度関数と基本周波数推定処理115により得られた基本周波数の確率密度関数の

10

20

30

40

50

類似度を算出する。ここで、演奏位置において和音を演奏する場合には目標基本周波数は複数発生する。そのような場合には、複数の目標基本周波数において各々ピークとなる複数の確率密度関数を合成して目標基本周波数の確率密度関数を求める。演奏音の確率密度関数と目標基本周波数の確率密度関数とを用いた類似度の算出方法としては、例えば次のような各種の態様が考えられる。

【 0 0 6 1 】

(1 - 1) 2つの確率密度関数の平均2乗誤差RMS、すなわち、図4に示すように、演奏音の基本周波数の確率密度関数における確率密度と目標基本周波数の確率密度関数における確率密度との差分を2乗したものを全周波数帯域に亘って積分し、所定の定数Cにより除算したものの平方根の逆数を類似度とする。なお、この平方根の逆数の代わりに、所定の最大値からこの平方根を減算したものを類似度としてもよい。

10

【 0 0 6 2 】

(1 - 2) 図5に示すように、周波数帯域を目標基本周波数の確率密度が高い有ピッチ領域と目標基本周波数の確率密度がほぼ0である無ピッチ領域とに分け、基本周波数推定処理115により得られた演奏音の基本周波数の確率密度関数において、有ピッチ領域に属する部分の確率密度の総和と、無ピッチ領域に属する部分の確率密度の総和とを求め、前者から後者を差し引いたものを類似度としてもよい。

【 0 0 6 3 】

(1 - 3) 図6に示すように、演奏音の基本周波数の確率密度関数について、目標基本周波数を中心とした所定幅の周波数範囲における積分値を算出する。図示の例では、演奏位置において演奏すべき音が3個有り、その基本周波数がF1、F2、F3であるため、 $F1 \pm F$ 、 $F2 \pm F$ 、 $F3 \pm F$ の各範囲について演奏音の確率密度関数の積分値(図示の斜線部分)を求めている。そして、各目標基本周波数を中心とした各範囲の各積分値を類似度として算出し、これが閾値を越えているか否かにより、各目標基本周波数の音を正しく演奏したか否かを判定する。この場合、演奏位置において演奏すべき音数が多いと、演奏音の確率密度関数において目標基本周波数の確率密度関数との類似度の低いピークが多数発生するため、実際に正しい演奏を行っていたとしても、正しい演奏が行われていないという誤った判定がなされるおそれがある。そこで、このような誤判定を防止するために、演奏位置において演奏すべき音の数がkである場合に、目標基本周波数を中心とした範囲の積分値をk倍したものを類似度としてもよい。

20

30

【 0 0 6 4 】

(1 - 4) その他、演奏音の基本周波数の確率密度関数と目標基本周波数の確率密度関数から何らかの特徴量を抽出し、特徴量同士の積・べき乗・各種の数学関数などを用いた量を類似度とし、演奏音の基本周波数の確率密度関数と目標基本周波数の確率密度関数との差別化を容易にするようにしてもよい。

【 0 0 6 5 】

(1 - 5) 上記のうちの例えば2種類の方法により2種類の類似度(第1および第2の類似度とする)を求め、第1および第2の類似度を線形結合した第3の類似度を演奏音が正しい音高であるか否かの判断のための類似度として採用してもよい。この場合、目標基本周波数通りの演奏音を発生させ、あるいは目標基本周波数からずれた基本周波数の演奏音を発生させる等、各種の条件で演奏音の発生および演奏音の基本周波数の推定を行わせる。そして、それら各種の条件において、第1の類似度と第2の類似度の重みを変えつつ、基本周波数の確率密度関数と目標基本周波数の確率密度関数とを用いた第3の類似度の算出を行う。そして、周知の判別分析の手法を用いることにより、演奏音の基本周波数と目標基本周波数とが一致しているのかずれているのかの差別化が容易な第3の類似度が得られるように、第1の類似度と第2の類似度の重みのバランスを定めればよい。なお、周知の判別分析の手法として知られる手法の他、ニューラルネットワークやSVM(サポートベクターマシン)として知られる手法も用いることができる。

40

【 0 0 6 6 】

(2) 上記実施形態において、類似度評価処理116および一致判定処理117を実行す

50

る代わりに、基本周波数推定処理 1 1 5 において得られる基本周波数の確率密度関数の中から顕著なピークを選択し、そのピークが発生している基本周波数と目標基本周波数との一致度により、正しい音高で演奏が行われたか否かの判定を行うようにしてもよい。

【 0 0 6 7 】

(3) 模範となる楽器演奏を録音することにより得られた音響信号のサンプルデータを楽曲データとして用い、この楽曲データに基本周波数推定処理を施すことにより、ユーザが発生すべき演奏音の目標基本周波数を求めるようにしてもよい。すなわち、図 1 において、收音部 1 0 4 により得られる演奏音の基本周波数を推定する基本周波数推定処理 1 1 5 の他に、演奏位置制御処理 1 1 2 により抽出される演奏位置の楽曲データ(模範演奏音のサンプルデータ)から模範演奏音の基本周波数を推定する基本周波数推定処理を設け、この基本周波数推定処理により推定された模範演奏音の基本周波数を目標基本周波数として用いるのである。この態様において、模範演奏の演奏音を收音部 1 0 4 により收音し、收音部 1 0 4 から得られる音響信号を模範演奏の楽曲データとして楽曲記憶部 1 0 5 に格納するようにしてもよい。

10

【図面の簡単な説明】

【 0 0 6 8 】

【図 1】この発明による演奏判定装置の一実施形態である教習伴奏システムの構成を示すブロック図である。

【図 2】同実施形態における基本周波数推定処理 1 1 5 の処理内容を示す図である。

【図 3】同基本周波数推定処理 1 1 5 において行われるマルチエージェントモデルによる基本周波数の経時的な追跡を示す図である。

20

【図 4】同実施形態における基本周波数の類似度の算出方法の変形例を示す図である。

【図 5】同実施形態における基本周波数の類似度の算出方法の他の変形例を示す図である。

。

【図 6】同実施形態における基本周波数の類似度の算出方法の他の変形例を示す図である。

。

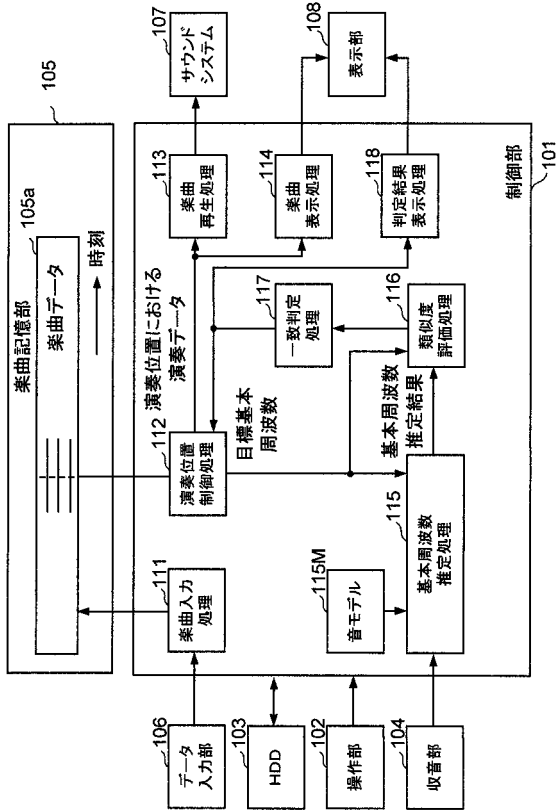
【符号の説明】

【 0 0 6 9 】

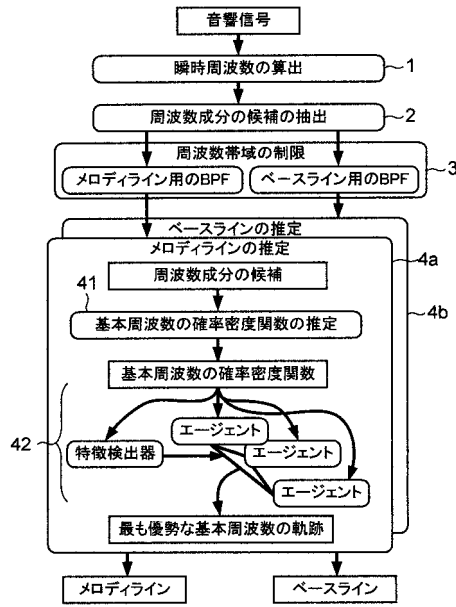
1 0 1 ……制御部、1 0 2 ……操作部、1 0 3 ……HDD、1 0 4 ……收音部、1 0 5 ……楽曲記憶部、1 0 6 ……データ入力部、1 1 1 ……楽曲入力処理、1 1 2 ……演奏位置制御処理、1 1 3 ……楽曲再生処理、1 1 4 ……楽曲表示処理、1 1 5 ……基本周波数推定処理、1 1 5 M ……音モデル、1 1 6 ……類似度評価処理、1 1 7 ……一致判定処理、1 1 8 ……判定結果表示処理、1 ……瞬時周波数の算出、2 ……周波数成分の候補の抽出、3 ……周波数帯域の制限、4 a ……メロディラインの推定、4 b ……ベースラインの推定、4 1 ……基本周波数の確率密度関数の推定、4 2 ……マルチエージェントモデルによる基本周波数の経時的な追跡。

30

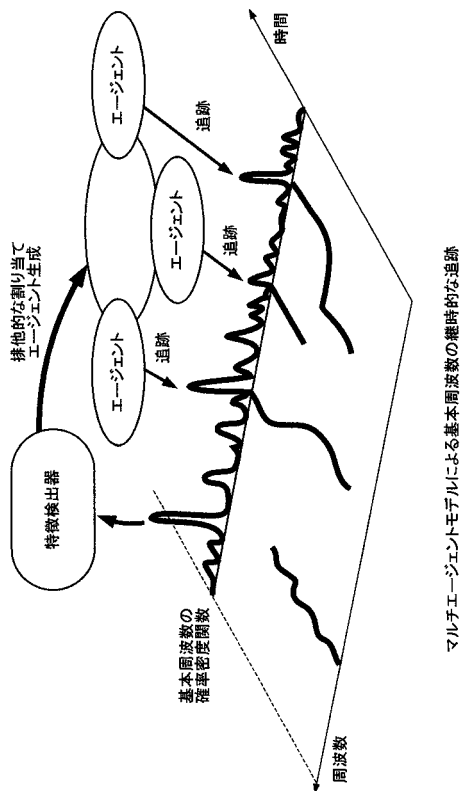
【図1】



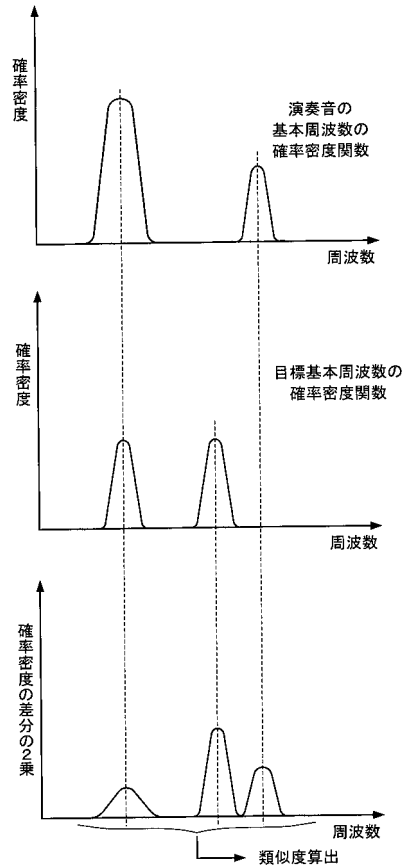
【図2】



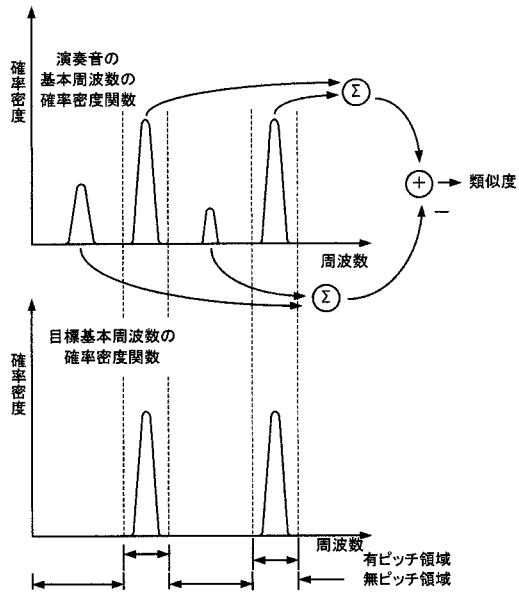
【図3】



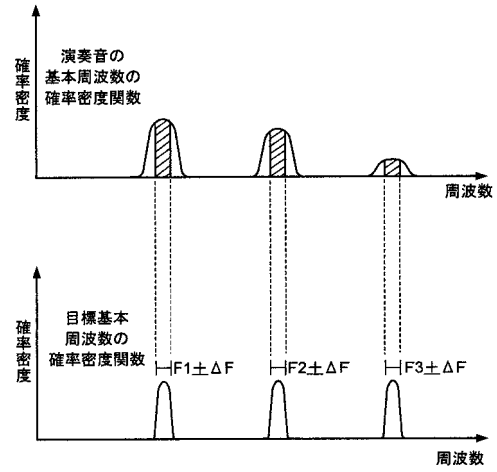
【図4】



【図5】



【図6】



フロントページの続き

(72)発明者 有元 慶太
静岡県浜松市中沢町10番1号 ヤマハ株式会社内

審査官 日下 善之

(56)参考文献 特表2004-515808(JP,A)
特開2001-128959(JP,A)
特開2003-256552(JP,A)
特開2006-030692(JP,A)
特開2004-004872(JP,A)
特開2004-341026(JP,A)
特開2001-125562(JP,A)
特開平08-286684(JP,A)
後藤真孝,リアルタイム音楽情景記述システム:全体構想と音高推定手法の拡張,情報処理学会
研究報告,日本,社団法人情報処理学会,2000年10月12日,Vol.2000,No.
94,p.9-16

(58)調査した分野(Int.Cl.,DB名)

G10H	1/00
G10G	1/00
G10G	3/04