

(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2015-114361

(P2015-114361A)

(43) 公開日 平成27年6月22日(2015.6.22)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 O G 3/04 (2006.01)	G 1 O G 3/04	5 D 1 8 2
G 1 O L 25/51 (2013.01)	G 1 O L 25/51 3 0 0	

審査請求 未請求 請求項の数 4 O L (全 14 頁)

(21) 出願番号	特願2013-253993 (P2013-253993)	(71) 出願人	000004075
(22) 出願日	平成25年12月9日 (2013.12.9)		ヤマハ株式会社
		(74) 代理人	110000213
			特許業務法人プロスペック特許事務所
		(74) 代理人	100168756
			弁理士 日比野 元彦
		(72) 発明者	前澤 陽
			静岡県浜松市中区中沢町10番1号ヤマハ株式会社内
		Fターム(参考)	5D182 AC01 AD06 AD10

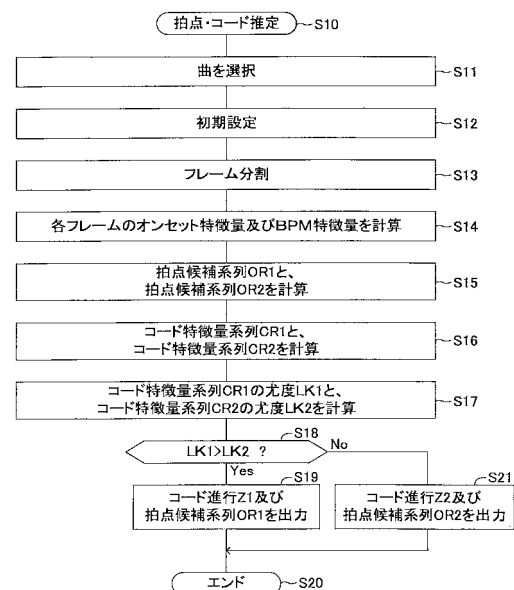
(54) 【発明の名称】 音響信号分析装置及び音響信号分析プログラム

(57) 【要約】 (修正有)

【課題】拍点、コード進行、及び小節線の位置の推定精度を向上させた音響信号分析装置及び音響信号分析プログラムを提供する。

【解決手段】楽曲の音響信号を取り込み、第1拍点候補系列OR1、及び第1拍点候補系列に対して半拍分ずれた第2拍点候補系列OR2を検出する。次に、第1拍点系列及び前記第2拍点系列の隣り合う2つの拍点候補の間のコードの特徴を表わす第1ビート同期型コード特徴量系列CR1及び第2ビート同期型コード特徴量系列CR2をそれぞれ計算する。楽曲のコード進行を表わす確率モデルのうち、第1ビート同期型コード特徴量系列CR1及び第2ビート同期型コード特徴量系列CR2が所定の基準を満たす確率モデルをそれぞれ1つずつ選択し、選択した2つの確率モデルのうち尤度が高い確率モデルに基づいて、楽曲の拍点、コード進行、及び小節線の位置を推定する。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

分析対象としての楽曲の演奏音を表わす音響信号を取り込む音響信号取得手段と、

前記取り込んだ音響信号に基づいて、前記楽曲の複数の拍点候補からなる第 1 拍点候補系列、及び前記第 1 拍点候補系列を構成する複数の拍点候補に対してそれぞれ半拍分ずれた複数の拍点候補からなる第 2 拍点候補系列を検出する拍点候補系列検出手段と、

前記取り込んだ音響信号に基づいて、前記第 1 拍点系列を構成する複数の拍点候補のうちの隣り合う 2 つの拍点候補の間のコードの特徴をそれぞれ表わすビート同期型コード特徴量から構成された第 1 ビート同期型コード特徴量系列を計算するとともに、前記第 2 拍点系列を構成する複数の拍点候補のうちの隣り合う 2 つの拍点候補の間のコードの特徴をそれぞれ表わすビート同期型コード特徴量から構成された第 2 ビート同期型コード特徴量系列を計算するビート同期型コード特徴量系列計算手段と、

10

前記楽曲のコード進行を表わす確率モデルであって、1 小節内の拍数、前記楽曲の調及び最初の拍点の拍子位置の組み合わせに応じて拍点間におけるコードの遷移確率が設定された確率モデルのうち、前記第 1 ビート同期型コード特徴量系列が所定の基準を満たす確率モデル及び前記第 2 ビート同期型コード特徴量系列が所定の基準を満たす確率モデルをそれぞれ 1 つずつ選択し、前記選択した 2 つの確率モデルのうち尤度が大きい確率モデルに基づいて、前記楽曲の拍点、コード進行、及び小節線の位置を推定する推定手段と、を備えた音響信号分析装置。

【請求項 2】

20

請求項 1 に記載の音響信号分析装置において、

前記ビート同期型コード特徴量計算手段は、

前記隣り合う 2 つの拍点候補の間に位置する複数の区間ごとにコードの特徴量を表す個度特徴量を計算するコード特徴量計算手段と、

前記隣り合う 2 つの拍点候補の間に位置する複数の区間のコード特徴量を平滑化することにより、前記ビート同期型コード特徴量を計算するコード特徴量平滑化手段と、を備えた音響信号分析装置。

【請求項 3】

請求項 1 又は 2 に記載の音響信号分析装置において、

前記拍点候補系列検出手段は、

30

前記楽曲の各区間における拍の存在に関する特徴を表わす第 1 特徴量及びテンポに関する特徴を表わす第 2 特徴量を計算する拍・テンポ特徴量計算手段と、

前記楽曲の各区間における拍の存在に関する物理量及びテンポに関する物理量の組み合わせにより分類された状態の系列として記述された複数の確率モデルのうち、前記第 1 特徴量及び前記第 2 特徴量が前記楽曲の各区間において同時に観測される確率を表わす観測尤度の系列が所定の基準を満たす確率モデルを選択することにより、前記楽曲における拍点及びテンポの推移を同時に推定する拍点・テンポ推定手段と、を備えた音響信号分析装置。

【請求項 4】

音響信号分析装置が備えるコンピュータに、

40

分析対象としての楽曲の演奏音を表わす音響信号を取り込む音響信号取得ステップと、

前記取り込んだ音響信号に基づいて、前記楽曲の複数の拍点候補からなる第 1 拍点候補系列、及び前記第 1 拍点候補系列を構成する複数の拍点候補に対してそれぞれ半拍分ずれた複数の拍点候補からなる第 2 拍点候補系列を検出する拍点候補系列検出ステップと、

前記取り込んだ音響信号に基づいて、前記第 1 拍点系列を構成する複数の拍点候補のうちの隣り合う 2 つの拍点候補の間のコードの特徴をそれぞれ表わすビート同期型コード特徴量から構成された第 1 ビート同期型コード特徴量系列を計算するとともに、前記第 2 拍点系列を構成する複数の拍点候補のうちの隣り合う 2 つの拍点候補の間のコードの特徴をそれぞれ表わすビート同期型コード特徴量から構成された第 2 ビート同期型コード特徴量系列を計算するビート同期型コード特徴量系列計算ステップと、

50

前記楽曲のコード進行を表わす確率モデルであって、1小節内の拍数、前記楽曲の調及び最初の拍点の拍子位置の組み合わせに応じて拍点間におけるコードの遷移確率が設定された確率モデルのうち、前記第1ビート同期型コード特徴量系列が所定の基準を満たす確率モデル及び前記第2ビート同期型コード特徴量系列が所定の基準を満たす確率モデルをそれぞれ1つずつ選択し、前記選択した2つの確率モデルのうち尤度が大きい確率モデルに基づいて、前記楽曲の拍点、コード進行、及び小節線の位置を推定する推定ステップと、を実行させる、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、楽曲を表わす音響信号を分析して、楽曲における拍点（拍のタイミング）及び楽曲の各区間で発音されるコード（和音）を検出する音響信号分析装置及び音響信号分析プログラムに関する。

【背景技術】

【0002】

従来から、例えば、下記非特許文献1に示されているように、楽曲における拍点、及び楽曲の各区間で発音されるコード（和音）を検出する音響信号分析装置は知られている。この音響信号分析装置においては、まず、音響信号を分析して、楽曲の拍点を決定している。そして、前記検出した拍点においてコード変化が生起し、かつ小節の先頭でコード変化が生起するという仮定の下で、楽曲の各区間のコード及び小節線の位置を検出している。

【先行技術文献】

【非特許文献】

【0003】

【非特許文献1】M. Goto et al. , “SONGLE: A WEB SERVICE FOR ACTIVE MUSIC LISTENING IMPROVED BY USER CONTRIBUTIONS”, ISMIR, 2011, p. 311 - 316

【発明の概要】

【0004】

上記非特許文献1には、誤って裏拍を拍点（表拍）として選択してしまう可能性や、楽曲のテンポが真のテンポの倍のテンポとなるような拍点を選択してしまう可能性を考慮して、尤もらしい拍点を選択すると記載されているが、その選択手段については具体的には開示されていない。また、裏拍を拍点（表拍）として選択した場合や、楽曲のテンポが真のテンポの倍のテンポとなるような拍点を選択した場合には、コードの検出精度及び小節線の位置の検出精度が低下する。

【0005】

本発明は上記問題に対処するためになされたもので、その目的は、拍点、コード進行、及び小節線の位置の推定精度を向上させた音響信号分析装置及び音響信号分析プログラムを提供することにある。なお、下記本発明の各構成要件の記載においては、本発明の理解を容易にするために、実施形態の対応箇所の符号を括弧内に記載しているが、本発明の各構成要件は、実施形態の符号によって示された対応箇所の構成に限定解釈されるべきものではない。

【0006】

上記目的を達成するために、本発明の特徴は、分析対象としての楽曲の演奏音を表わす音響信号を取り込む音響信号取得手段（S11）と、前記取り込んだ音響信号に基づいて、前記楽曲の複数の拍点候補からなる第1拍点候補系列（OR1）、及び前記第1拍点候補系列を構成する複数の拍点候補に対してそれぞれ半拍分ずれた複数の拍点候補からなる第2拍点候補系列（OR2）を検出する拍点候補系列検出手段（S15）と、前記取り込んだ音響信号に基づいて、前記第1拍点系列を構成する複数の拍点候補のうちの隣り合う

2つの拍点候補の間のコードの特徴をそれぞれ表わすビート同期型コード特徴量から構成された第1ビート同期型コード特徴量系列(CR1)を計算するとともに、前記第2拍点系列を構成する複数の拍点候補のうちの隣り合う2つの拍点候補の間のコードの特徴をそれぞれ表わすビート同期型コード特徴量から構成された第2ビート同期型コード特徴量系列(CR2)を計算するビート同期型コード特徴量系列計算手段(S16)と、前記楽曲のコード進行を表わす確率モデルであって、1小節内の拍数、前記楽曲の調及び最初の拍点の拍子位置の組み合わせに応じて拍点間におけるコードの遷移確率が設定された確率モデルのうち、前記第1ビート同期型コード特徴量系列が所定の基準を満たす確率モデル及び前記第2ビート同期型コード特徴量系列が所定の基準を満たす確率モデルをそれぞれ1つずつ選択し、前記選択した2つの確率モデルのうち尤度が大きい確率モデルに基づいて、前記楽曲の拍点、コード進行、及び小節線の位置を推定する推定手段(S17~S21)と、を備えた音響信号分析装置としたことにある。

10

【0007】

この場合、前記ビート同期型コード特徴量計算手段は、前記隣り合う2つの拍点候補の間に位置する複数の区間(t_i)ごとにコードの特徴量を表すコード特徴量(XC)を計算するコード特徴量計算手段と、前記隣り合う2つの拍点候補の間に位置する複数の区間のコード特徴量を平滑化することにより、前記ビート同期型コード特徴量を計算するコード特徴量平滑化手段と、を備えるといふ。

【0008】

また、この場合、拍点候補系列検出手段は、前記楽曲の各区間における拍の存在に関する特徴を表わす第1特徴量及びテンポに関する特徴を表わす第2特徴量を計算する拍・テンポ特徴量計算手段と、前記楽曲の各区間における拍の存在に関する物理量及びテンポに関する物理量の組み合わせにより分類された状態の系列として記述された複数の確率モデルのうち、前記第1特徴量及び前記第2特徴量が前記楽曲の各区間において同時に観測される確率を表わす観測尤度の系列が所定の基準を満たす確率モデルを選択することにより、前記楽曲における拍点及びテンポの推移を同時に推定する拍点・テンポ推定手段と、を備えるといふ。

20

【0009】

一般に、コードの変化は、表拍で生起する可能性が高い。そのため、誤って裏拍を表拍として選択してしまった場合には、拍点と拍点の間においてコードが変化する可能性が高い。そのため、この場合、ビート同期型コード特徴量は、コードの特徴を的確に表現できていない。つまり、裏拍を表拍として選択してしまった場合には、真の拍点(つまり表拍)を選択した場合に比べて、ビート同期型コード特徴量系列の尤度が低くなる。そこで、本発明に係る音響信号分析装置は、第1拍点候補系列と、第1拍点候補系列を構成する各拍点候補に対して半拍分ずれた拍点候補からなる第2拍点候補系列を検出し、第1拍点候補系列及び第2拍点候補系列に関するビート同期型コード特徴量系列の尤度をそれぞれ計算する。そして、両尤度を比較して、尤度の高い拍点候補系列を選択する。これにより、裏拍を誤って表拍として選択してしまうことを抑制できる。また、ビート同期型コード特徴量系列の尤度が最も高くなるような拍子位置及び調の組み合わせが選択されて、楽曲における拍点、コード進行及び小節線の位置が同時に(一体的)に推定される。したがって、本発明に係る音響信号分析装置によれば、拍点、コード進行、及び小節線の位置の推定精度を従来よりも向上させることができる。

30

40

【0010】

また、本発明は、音響信号分析装置が備えるコンピュータに適用されるコンピュータプログラムとしても実施可能である。

【図面の簡単な説明】

【0011】

【図1】本発明の一実施形態に係る音響信号分析装置の構成を表わすブロック図である。

【図2】拍点・コード推定処理を表わすフローチャートである。

【図3】分析対象の音響信号の波形を表わすグラフである。

50

【図４】拍点候補を計算するための確率モデルの概念図である。

【図５】コムフィルタのブロック図である。

【図６】ＢＰＭ特徴量の計算結果を示すグラフである。

【図７】テンプレートの構成を示す表である。

【図８】コード特徴量の概念図である。

【図９】ビート同期型コード特徴量の概念図である。

【発明を実施するための形態】

【００１２】

本発明の一実施形態に係る音響信号分析装置１０について説明する。音響信号分析装置１０は、以下説明するように、楽曲を表わす音響信号を取り込んで、その楽曲における拍点及びテンポの推移を検出する。音響信号分析装置１０は、図１に示すように、入力操作子１１、コンピュータ部１２、表示器１３、記憶装置１４、外部インターフェース回路１５及びサウンドシステム１６を備えており、これらがバスＢＳを介して接続されている。

10

【００１３】

入力操作子１１は、オン・オフ操作に対応したスイッチ（例えば数値を入力するためのテンキー）、回転操作に対応したボリューム又はロータリーエンコーダ、スライド操作に対応したボリューム又はリニアエンコーダ、マウス、タッチパネルなどから構成される。これらの操作子は、演奏者の手によって操作されて、分析対象の楽曲の選択、音響信号の分析開始又は停止、楽曲の再生又は停止（後述するサウンドシステム１６からの出力又は停止）、音響信号の分析に関する各種パラメータの設定などに用いられる。入力操作子１１を操作すると、その操作内容を表す操作情報が、バスＢＳを介して、後述するコンピュータ部１２に供給される。

20

【００１４】

コンピュータ部１２は、バスＢＳにそれぞれ接続されたＣＰＵ１２ａ、ＲＯＭ１２ｂ及びＲＡＭ１２ｃからなる。ＣＰＵ１２ａは、詳しくは後述する音響信号分析プログラム及びそのサブルーチンをＲＯＭ１２ｂから読み出して実行する。ＲＯＭ１２ｂには、音響信号分析プログラム及びそのサブルーチンに加えて、初期設定パラメータ、表示器１３に表示される画像を表わす表示データを生成するための図形データ及び文字データなどの各種データが記憶されている。ＲＡＭ１２ｃには、音響信号分析プログラムの実行時に必要なデータが一時的に記憶される。

30

【００１５】

表示器１３は、液晶ディスプレイ（ＬＣＤ）によって構成される。コンピュータ部１２は、図形データ、文字データなどを用いて表示すべき内容を表わす表示データを生成して表示器１３に供給する。表示器１３は、コンピュータ部１２から供給された表示データに基づいて画像を表示する。例えば分析対象の楽曲の選択時には、楽曲のタイトルリストが表示される。また、例えば分析終了時には、拍点及び小節線を表わすグラフやコード進行を表わすコード名の系列が表示される。

【００１６】

また、記憶装置１４は、ＨＤＤ、ＦＤＤ、ＣＤ－ＲＯＭ、ＭＯ、ＤＶＤなどの大容量の不揮発性記録媒体と、同各記録媒体に対応するドライブユニットから構成されている。記憶装置１４には、複数の楽曲をそれぞれ表わす複数の楽曲データが記憶されている。楽曲データは、楽曲を所定のサンプリング周期（例えば１／４４１００秒）でサンプリングして得られた複数のサンプル値からなり、各サンプル値が記憶装置１４における連続するアドレスに順に記録されている。楽曲のタイトルを表わすタイトル情報、楽曲データの容量を表わすデータサイズ情報なども楽曲データに含まれている。楽曲データは予め記憶装置１４に記憶されていてもよいし、後述する外部インターフェース回路１５を介して外部機器から取り込んでもよい。記憶装置１４に記憶されている楽曲データは、ＣＰＵ１２ａによって読み込まれ、楽曲における拍点及びテンポの推移が分析される。

40

【００１７】

外部インターフェース回路１５は、音響信号分析装置１０を電子音楽装置、パーソナル

50

コンピュータなどの外部機器に接続可能とする接続端子を備えている。音響信号分析装置 10 は、外部インターフェース回路 15 を介して、LAN (Local Area Network)、インターネットなどの通信ネットワークにも接続可能である。

【0018】

サウンドシステム 16 は、楽曲データをアナログ音信号に変換する D/A 変換器、変換したアナログ音信号を増幅するアンプ、及び増幅されたアナログ音信号を音響信号に変換して出力する左右一対のスピーカを備えている。ユーザが入力操作子 11 を用いて分析対象の楽曲の再生を指示すると、CPU 12a は、分析対象の楽曲データをサウンドシステム 16 に供給する。これにより、ユーザは分析対象の楽曲を試聴できる。

【0019】

つぎに、音響信号分析装置 10 の動作について具体的に説明する。ユーザが音響信号分析装置 10 の図示しない電源スイッチをオンにすると、CPU 12a は、図 2 に示す拍点・コード推定プログラムを ROM 12b から読み出して実行する。なお、図 2 においては、「判断」のステップを六角形で示す。

【0020】

CPU 12a は、ステップ S10 にて拍点・コード推定処理を開始し、ステップ S11 にて、記憶装置 14 に記憶されている複数の楽曲データにそれぞれ含まれるタイトル情報を読み込んで、楽曲のタイトルをリスト形式で表示器 13 に表示する。ユーザは、入力操作子 11 を用いて、表示器 13 に表示された楽曲の中から分析対象の楽曲データを選択する。なお、ステップ S11 にて分析対象の楽曲データを選択する際、選択しようとする楽曲データが表す楽曲の一部又は全部を再生して楽曲データの内容を確認できるように構成してもよい。

【0021】

つぎに、CPU 12a は、ステップ S12 にて、音響信号分析のための初期設定を実行する。具体的には、前記選択された楽曲データのデータサイズ情報に応じた記憶領域を RAM 12c 内に確保し、前記確保した記憶領域に前記選択された楽曲データを読み込む。また、後述するオンセット特徴量 X_O、BPM 特徴量 X_Bなどを一時的に記憶する領域を RAM 12c 内に確保する。また、ユーザは、前記選択した楽曲の拍子（又は 1 小節内に含まれる拍数）を、入力操作子 11 を用いて入力する。つまり、本実施形態においては、前記選択した楽曲の拍子（又は 1 小節内に含まれる拍数）が既知であると仮定する。

【0022】

CPU 12a は、ステップ S13 にて、図 3 に示すように、前記選択された楽曲を所定の時間間隔を以て区切り、複数のフレーム t_i ($i = 0, 1, \dots, I$) に分割する。各フレームの長さは共通である。説明を簡単にするために、本実施形態では各フレームの長さを 125ms とする。上記のように、各楽曲のサンプリング周期は 1/44100 秒であるので、各フレームは、約 5000 個のサンプル値から構成されている。

【0023】

次に、CPU 12a は、複数の拍点候補から構成される拍点候補系列 OR1 及び拍点候補系列 OR2 を計算する。ここで、拍点候補系列 OR1 及び拍点候補系列 OR2 の計算手順の概略を説明する。まず、拍点候補系列 OR1 が次のようにして計算される。拍の存在に関する特徴を表すオンセット特徴量 X_O 及びテンポに関する特徴を表す BPM (beats per minute (1 分間あたりの拍数)) 特徴量 X_B をフレーム t_i ごとに計算する。そして、各フレーム t_i における拍周期 b の値 (テンポの逆数に比例する値) 及び次の拍までのフレーム数 n の値の組み合わせに応じて分類された状態 $q_{b, n}$ の系列として記述された確率モデル (隠れマルコフモデル) のうち、観測値としてのオンセット特徴量 X_O 及び BPM 特徴量 X_B が同時に観測される確率を表わす観測尤度の系列が最も尤もらしい確率モデルを選択する (図 4 参照)。これにより、分析対象の楽曲における拍点候補系列 OR1 が検出される。そして、前記検出された拍点候補系列 OR1 を用いて、拍点候補系列 OR1 を構成する複数の拍点に対して半拍ずれた複数の拍点候補から構成された拍点候補系列 OR2 が計算される (図 9 参照)。なお、拍周期 b は、フレームの数に

10

20

30

40

50

よって表わされる。したがって、拍周期 b の値は「 $1 \leq b \leq b_{max}$ 」を満たす整数であり、拍周期 b の値が「 b 」である状態では、フレーム数 n の値は「 $0 \leq n < \infty$ 」を満たす整数である。

【0024】

次に、拍点候補系列 $OR1$ 及び拍点候補系列 $OR2$ の計算手順について具体的に説明する。まず、CPU12aは、ステップS14にて、フレームごとに、オンセット特徴量 XO 及びBPM特徴量 XB を計算する。

【0025】

フレーム t_i のオンセット特徴量 $XO(t_i)$ は、次のようにして計算される。CPU12aは、まず、フレームごとに短時間フーリエ変換を実行し、各周波数ビンの信号強度を計算する。次に、CPU12aは、メルフィルタバンクを用いて、各周波数帯域 fb_x (例えば、 $x = 1, 2, \dots, 20$) の信号強度 $M(fb_x, t_i)$ を計算する。次に、CPU12aは、フレーム間における各周波数帯域の信号強度の増加量 $R(fb_x, t_i)$ を計算する。下記の式(1)に示すように、フレーム間における前記各周波数帯域の信号強度の増加量の総和がオンセット特徴量 $XO(t_i)$ である。

【数1】

$$XO(t_i) = \sum_{w_k} \max(0, R(fb_x, t_i)) \quad \dots(1)$$

【0026】

フレーム t_i のBPM特徴量 $XB(t_i)$ は、次のようにして計算される。CPU12aは、まず、オンセット特徴量 $XO(t_0), XO(t_1), \dots$ をこの順にフィルタバンク FB (図5参照)に入力する。フィルタバンク FB は、拍周期 b の値に応じてそれぞれ設けられた複数のコムフィルタ CF_b からなる。コムフィルタ CF_b は、1つのデータが入力される度に1つのデータを出力する。コムフィルタ CF_b は、過去の出力データを拍周期 b の値に応じた個数だけ記憶する $FIFO (= First\ In\ First\ Out)$ メモリを有しており、入力されたデータと前記記憶手段に記憶されているデータのうちの最古のデータを所定の比率 (例えば、 $1:1$ (すなわち、 $\alpha = 0.5$)) で加算して出力する。オンセット特徴量 XO の系列 $XO(t) \{ = XO(t_0), XO(t_1), \dots \}$ をフィルタバンク FB に入力することにより得られたデータ XD_b の系列 $XD_b(t) \{ = XD_b(t_0), XD_b(t_1), \dots \}$ を時系列的に逆にして、フィルタバンク FB に再度入力することにより、拍周期 b に関するBPM特徴量の系列 $XB_b(t) \{ = XB_b(t_0), XB_b(t_1), \dots \}$ が得られる。フレーム t_i のBPM特徴量 $XB(t_i)$ は、拍周期 b ごとに計算されたBPM特徴量 $XB_{b=1, 2, \dots}(t_i)$ の集合として表わされる (図6参照)。

【0027】

次に、CPU12aは、ステップS15にて、ビタビアルゴリズムを用いて、最尤の状態系列を推定する。これにより、拍点候補系列 $OR1$ が推定される。なお、拍点候補系列 $OR1$ が推定される際、拍周期 b の値の系列 (つまり、テンポの推移) も同時 (一体的) に推定される。

【0028】

具体的には、CPU12aは、まず、オンセット特徴量 $XO(t_i)$ 及びBPM特徴量 $XB(t_i)$ の観測尤度 $LO(t_i)$ 及び観測尤度 $LB(t_i)$ をそれぞれ計算する。ここで、オンセット特徴量 $XO(t_i)$ は、次の拍点までのフレーム数 n の値に応じて設定された正規分布に従うものとする。つまり、オンセット特徴量 XO の観測尤度 $LO(t_i)$ は、次の拍点までのフレーム数 n の値に応じて設定された正規分布の確率変数としてオンセット特徴量 XO を代入することにより計算される。例えば、フレーム数 n の値が「0」であるときは、平均値が「3」であって、且つ分散が「1」である正規分布が用いられる。また、拍周期 b の値が「 b 」であって、フレーム数 n の値が「 $n/2$ 」であるときは、平均値が「1」であって、且つ分散が「1」である正規分布が用いられる。また、フレ

10

20

30

40

50

ーム数 n の値が「0」及び「 $\frac{n}{2}$ 」のいずれの値とも異なるとき、平均値が「0」であって、且つ分散が「1」である正規分布が用いられる。

【0029】

また、BPM特徴量 $XB(t_i)$ の観測尤度 LB は、拍周期 b ごとに設けられたテンプレート TMP_b に対する BPM特徴量 XB の適合度に相当する。つまり、下記の式(2)に示すように、テンプレート TMP_b と BPM特徴量 $XB(t_i)$ の内積が、観測尤度 $LB(t_i)$ である。なお、この演算式における「 κ_b 」は、オンセット特徴量 $XO(t_i)$ に対する BPM特徴量 $XB(t_i)$ の重みを決定する係数である。つまり、「 κ_b 」を大きく設定するほど、結果的に、BPM特徴量 $XB(t_i)$ が重視される。また、この演算式における $Z(\kappa_b)$ は、「 κ_b 」に依存する正規化係数である。つまり、前記テンプレート TMP_b は、BPM特徴量 $XB(t_i)$ を構成する BPM特徴量 $XB_b(t_i)$ にそれぞれ乗算される係数 κ_b , $\{\kappa_1, \kappa_2, \dots, \kappa_b\}$ の系列からなる(図7参照)。テンプレート TMP_b を構成する係数 κ_b のうち、インデックス b が拍周期 b に等しい係数及び拍周期 b の整数倍に等しい係数が極大となるように、テンプレート TMP_b が設定されている。

10

【数2】

$$LB(t_i) = \frac{1}{Z(\kappa_b)} \exp \left(\kappa_b \frac{\sum_b XB_b(t_i) \cdot \delta_{\gamma,b}}{\sqrt{\sum_b (XB_b(t_i))^2} \sqrt{\sum_b (\delta_{\gamma,b})^2}} \right) \quad \dots (2)$$

20

【0030】

次に、CPU12aは、観測尤度 $LO(t_i)$ と観測尤度 $LB(t_i)$ との積の対数である対数観測尤度 $LOB(t_i)$ (下記の式(3)参照)を用いて、尤度が最大となる状態系列を計算する。この最尤の状態系列の計算においては、ビタビアルゴリズムを用いる。

【数3】

$$LOB(t_i) = \text{Log}(LO(t_i)LB(t_i)) \quad \dots (3)$$

30

【0031】

なお、本実施形態においては、(拍周期 b の値が「 s 」であり、且つフレーム数 n の値が「 s 」である状態から、拍周期 b の値が「 e 」であり、且つフレーム数 n の値が「 e 」である状態への対数遷移確率 T の値は、次のように設定されている(図4参照)。「 $e = 0$ 」、「 $e = s$ 」、かつ「 $e = e - 1$ 」のとき、対数遷移確率 T の値は、「 -0.2 」である。また、「 $s = 0$ 」、「 $e = s + 1$ 」、かつ「 $e = e - 1$ 」のとき、対数遷移確率 T の値は、「 -0.6 」である。また、「 $s = 0$ 」、「 $e = s - 1$ 」、かつ「 $e = e - 1$ 」のとき、対数遷移確率 T の値は、「 -0.6 」である。また、「 $s > 0$ 」、「 $e = s$ 」、かつ「 $e = s - 1$ 」のとき、対数遷移確率 T の値は、「0」である。上記以外の対数遷移確率 T の値は、「 $-$ 」である。すなわち、フレーム数 n の値が「0」である状態($s = 0$)から次の状態へ遷移するとき、拍周期 b の値は「1」だけ増減され得る。このとき、フレーム数 n の値は、遷移後の拍周期 b の値より「1」だけ小さい値に設定される。また、フレーム数 n の値が「0」でない状態($s > 0$)から次の状態へ遷移するとき、拍周期 b の値は変更されず、フレーム数 n の値が「1」だけ減少する。

40

【0032】

上記のようにして計算された最尤の状態系列を構成する各状態のうち、フレーム数 n が「0」である状態の系列が拍点候補系列 OR 1 である。また、前記最尤の状態系列を構成する各状態のうち、拍周期 b とフレーム数 n の値が下記の式(4)を満たす状態の系列が拍点候補系列 OR 2 である。

50

【数 4】

$$\frac{b-n}{b} = 0.5 \quad \dots(4)$$

【0033】

これにより、拍点候補系列OR1を構成する複数の拍点候補から見てそれぞれ半拍分（例えば、4分の4拍子の楽曲においては8分音符の長さ）ずれた複数の拍点候補からなる拍点候補系列OR2が得られる。

【0034】

つぎに、CPU12aは、ステップS16にて、拍点候補系列OR1を構成する各拍点候補における和音の特徴を表わすビート同期型コード特徴量XBC1の系列であるビート同期型コード特徴量系列CR1と、拍点候補系列OR2を構成する各拍点候補における和音の特徴を表わすビート同期型コード特徴量XBC2の系列であるビート同期型コード特徴量系列CR2を計算する。

【0035】

ビート同期型コード特徴量系列CR1及びビート同期型コード特徴量系列CR2は次のようにして計算される。まず、各フレーム t_i の各周波数ビンのパワーを、その周波数に最も近い音高の周波数（例えば平均律における各音高の基本周波数）にマッピングする。上記のようにして各音高にマッピングされたパワーのうち、低音域（例えば「B1」以下）に属するパワーをピッチクラス（C, C#, D, ..., B#）ごとに加算（又は積算）する。このようにして計算された各ピッチクラスのパワーからなる12次元の特徴量をベース特徴量HPCP^(B)と呼ぶ（図8参照）。また、各音高にマッピングされたパワーのうち、高音域（例えば「C2」以上）に属するパワーをピッチクラス（C, C#, D, ..., B#）ごとに加算（又は積算）する。このようにして計算された各ピッチクラスのパワーからなる12次元の特徴量をトレブル特徴量HPCP^(T)と呼ぶ。

【0036】

また、低音域のパワーのL2ノルムをベースパワー^(B)と呼び、高音域のパワーのL2ノルムをトレブルパワー^(T)と呼ぶ。

【0037】

各フレーム t_i に関するベース特徴量HPCP^(B)、トレブル特徴量HPCP^(T)、ベースパワー^(B)及びトレブルパワー^(T)からなる26次元の特徴量をコード特徴量XC(t_i)と呼ぶ。

【0038】

ビート同期型コード特徴量XBC1(m)は、拍点候補系列OR1における「m」番目の拍点候補と「m+1」番目の拍点候補の間のフレームのコード特徴量XC(t_i)を平滑化することにより得られる26次元の特徴量である。また、ビート同期型コード特徴量XBC2(m)は、拍点候補系列OR2における「m」番目の拍点候補と「m+1」番目の拍点候補の間のフレームのコード特徴量XC(t_i)を平滑化することにより得られる26次元の特徴量である。なお、上記の平滑化とは、例えば、前記拍点候補間のフレームのベース特徴量HPCP^(B)の平均を計算するとともに、前記拍点候補間のフレームのトレブル特徴量HPCP^(T)、ベースパワー^(B)及びトレブルパワー^(T)のメジアンをそれぞれ計算することを意味する。拍点候補系列OR1を構成する全ての拍点候補に関して、ビート同期型コード特徴量XBC1を計算することにより、ビート同期型コード特徴量系列CR1が計算される。また、拍点候補系列OR2を構成する全ての拍点候補に関して、ビート同期型コード特徴量XBC2を計算することにより、ビート同期型コード特徴量系列CR2が計算される。

【0039】

次に、CPU12aは、ステップS17にて、最尤のビート同期型コード特徴量系列CR1の尤度LK1及び最尤のビート同期型コード特徴量系列CR2の尤度LK2を計算する。最尤のビート同期型コード特徴量系列CR1の尤度LK1及び最尤のビート同期型コ

10

20

30

40

50

ード特徴量系列 C R 2 の尤度 L K 2 の計算手順は共通である。そこで、以下の説明においては、ビート同期型コード特徴量系列 C R 1 及びビート同期型コード特徴量系列 C R 2 を単にビート同期型コード特徴量系列 C R と表記する。また、ビート同期型コード特徴量 X B C 1 及びビート同期型コード特徴量 X B C 2 を単にビート同期型コード特徴量 X B C と表記する。

【 0 0 4 0 】

ここで、ベース特徴量 H P C P ^(B) 及びトレブル特徴量 H P C P ^(T) は、v M F (= v o n M i s e s F i s h e r) 分布に従うと仮定する。また、ベースパワー ^(B) 及びトレブルパワー ^(T) は、ガンマ分布に従うと仮定する。一般の楽曲は、和音が発生されている区間においてパワーが強い。一方、和音が発生されていない区間においてはパワーが弱い。また、和音が発生されていない区間におけるベース特徴量 H P C P ^(B) 及びトレブル特徴量 H P C P ^(T) を構成する各ピッチクラスのパワーの分布と、和音が発生されている区間における前記各ピッチクラスのパワーの分布とが異なる。そこで、和音が発生されている状態と和音が発生されていない状態とに関し、ベース特徴量 H P C P ^(B) 及びトレブル特徴量 H P C P ^(T) 並びにベースパワー ^(B) 及びトレブルパワー ^(T) を同時に学習する。また、1つの和音に対して、単一の v M F 分布及びガンマ分布を用いるのではなく、複数のモデル (混合モデル) を設定しておき、それらの重み付き線形和としてビート同期型コード特徴量の観測尤度を定義する。ビート同期型コード特徴量 X B C の観測尤度は、ビート同期型コード特徴量 X B C の構成要素としてのベース特徴量 H P C P ^(B)、トレブル特徴量 H P C P ^(T)、ベースパワー ^(B) 及びトレブルパワー ^(T)、並びに v M F 分布の平均 μ_k 、v M F 分布の分散 σ_k 、ガンマ分布の尺度母数 u 、ガンマ分布の形状母数 v 、及び前記線形和の重み w_k を用いて、下記の式 (5) のように表わされる。なお、「 k 」は、混合モデルを構成する分布を識別するためのインデックスである。また、低音域に関する変数の右上の括弧内には、「 B 」が表記されている。また、中高音域に関する変数の右上の括弧内には、「 T 」が表記されている。また、「 j 」は、和音 j に関するパラメータを表わす。例えば、「 j 」は、ベース特徴量 H P C P ^(B) 及びトレブル特徴量 H P C P ^(T) の形状 (各ピッチクラスのパワーの分布) を表わす。

【 数 5 】

$$\begin{aligned} p(HPCP^{(B)}, HPCP^{(T)}, \rho^{(B)}, \rho^{(T)} | \theta_j) \\ = \sum_k w_k vMF(HPCP^{(B)} | \mu_k^{(B)}, \kappa_k^{(B)}) vMF(HPCP^{(T)} | \mu_k^{(T)}, \kappa_k^{(T)}) \\ \times Gamma(\rho^{(B)} | u_k^{(B)}, v_k^{(B)}) Gamma(\rho^{(T)} | u_k^{(T)}, v_k^{(T)}) \quad \dots (5) \end{aligned}$$

【 0 0 4 1 】

また、一般に、和音から和音への遷移確率は、楽曲の調 k e y に依存する。例えば、和音「 C 」から和音「 F 」への遷移は八長調の楽曲において生起する可能性が高い。また、和音から和音への遷移確率は、拍点の拍子位置 s (直前の小節線から数えた拍数) に依存する。例えば、4分の4拍子の楽曲において、4拍目 (つまり $s = 4$) の和音が「 $G7$ 」であるとき、次の小節の1拍目の和音は「 C 」である可能性が高い (ドミナントモーション)。そこで、音響信号分析装置 10 は、和音の遷移確率を記憶した複数のデータベースを備える。各データベースは、拍子にそれぞれ対応している。つまり、音響信号分析装置 10 は、例えば、4分の3拍子の楽曲を分析するときに用いるデータベース、4分の4拍子の楽曲を分析するときに用いるデータベース、8分の6拍子の楽曲を分析するときに用いるデータベースなどを備える。そして、各データベースには、和音から和音への遷移確率が、調 k e y 及び拍子位置 s に関連づけて記憶されている。和音から和音への遷移確率は、種々の楽曲における和音の遷移を学習することにより決定される。これらのデータベースは R O M 12 b に記憶されている。ここで、「 $m - 1$ 」番目の拍点候補における和音が「 j 」であって、かつ「 m 」番目の拍点候補における和音が「 j 」である確率を下記

の式(6)のように表記する。なお、1小節内の拍数にそれぞれ対応したデータベースが設けられていても良い。

【数6】

$$\tau_j(m, j' | s, key) \quad \dots (6)$$

【0042】

すると、最初の拍の拍子位置 s 及び調 key が既知であるという条件下においては、ビート同期型コード特徴量系列 CR の尤度 $LK(s, key)$ は、下記の式(7)のように表される。

10

【数7】

$$LK(s, key) = \prod_{m,k} p(HPCP^{(B)}, HPCP^{(T)}, \rho^{(B)}, \rho^{(T)} | \theta_j)^{Z_j(m)} \prod_{j'} \tau_j(m, j' | s, key)^{Z_{j'}(m-1)Z_j(m)} \quad \dots (7)$$

【0043】

なお、式(7)における「 $Z_j(m)$ 」は、次に説明するような2値変数である。つまり、「 $Z_j(m)$ 」は、「 m 」番目の拍点候補の和音が「 j 」である場合に「1」であり、その他の場合に「0」である。また、「 $Z_{j'}(m-1)Z_j(m)$ 」は、「 $m-1$ 」番目の拍点候補における和音が「 j' 」であって、かつ「 m 」番目の拍点候補における和音が「 j 」である場合にのみ「1」であり、その他の場合に「0」である。

20

【0044】

ここで、拍点候補系列 $OR1$ についての尤度 $LK(s, key)$ を尤度 $LK^{(1)}(s, key)$ と表記する。最尤のビート同期型コード特徴量系列 $CR1$ の尤度 $LK1$ は、式(8)に基づいて計算される。

【数8】

$$LK1 = \operatorname{argmax}_{s, key} LK^{(1)}(s, key) \quad \dots (8)$$

30

【0045】

また、拍点候補系列 $OR2$ についての尤度 $LK(s, key)$ を尤度 $LK^{(2)}(s, key)$ と表記する。最尤のビート同期型コード特徴量系列 $CR2$ の尤度 $LK2$ は、式(9)に基づいて計算される。

【数9】

$$LK2 = \operatorname{argmax}_{s, key} LK^{(2)}(s, key) \quad \dots (9)$$

【0046】

なお、 $CPU12a$ は、ビタビアルゴリズムを用いて尤度 $LK1$ 及び尤度 $LK2$ を計算する。また、 $CPU12a$ は、尤度 $LK1$ 及び尤度 $LK2$ を計算する際、ステップ $S12$ においてユーザによって入力された拍子(又は1小節内の拍数)に応じたデータベースを参照して和音の遷移確率を決定する。

40

【0047】

次に、 $CPU12a$ は、ステップ $S18$ にて、尤度 $LK1$ と尤度 $LK2$ と比較する。尤度 $LK1$ が尤度 $LK2$ よりも大きいとき、 $CPU12a$ は、「Yes」と判定して、ステップ $S19$ にて、前記推定された拍点候補系列 $OR1$ 、最尤のビート同期型コード特徴量系列 $CR1$ 及び拍子位置 s を出力する。具体的には、拍点候補系列 $OR1$ 及び拍子位置 s に基づいて、拍点及び小節線を表わすグラフを表示器 13 に表示する。また、最尤のビート同期型コード特徴量系列 $CR1$ に基づいて、コード進行 $Z1$ (コード名の系列) を計算

50

して、表示器 13 に表示する。コード進行 Z 1 は、最尤のビート同期型コード特徴量系列 C R 1 を構成する各ビート同期型コード特徴量 X B C 1 (m) に対応するコード名の系列である。そして、C P U 12 a は、ステップ S 20 にて、拍点・コード推定処理を終了する。

【0048】

一方、ステップ S 18 において、尤度 L K 1 が尤度 L K 2 以下であるとき、C P U 12 a は、「No」と判定して、ステップ S 21 にて、前記推定された拍点候補系列 O R 2、最尤のビート同期型コード特徴量系列 C R 2 及び拍子位置 s を出力する。具体的には、拍点候補系列 O R 2 及び拍子位置 s に基づいて、拍点及び小節線を表わすグラフを表示器 13 に表示する。また、最尤のビート同期型コード特徴量系列 C R 2 に基づいて、コード進行 Z 2 (コード名の系列) を計算して、表示器 13 に表示する。コード進行 Z 2 は、最尤のビート同期型コード特徴量系列 C R 2 を構成する各ビート同期型コード特徴量 X B C 2 (m) に対応するコード名の系列である。そして、C P U 12 a は、ステップ S 20 にて、拍点・コード推定処理を終了する。

【0049】

一般に、コードの変化は、表拍で生起する可能性が高い。そのため、フレーム t_i ごとに計算されたコード特徴量 $X C (t_i)$ は、表拍において大きく変化する可能性が高い。よって、誤って裏拍を表拍として選択してしまった場合には、拍点と拍点の間においてコード特徴量 $X C (t_i)$ が大きく変化する可能性が高い。そのため、この場合、ビート同期型コード特徴量は、コードの特徴を的確に表現できていない。つまり、裏拍を表拍として選択してしまった場合には、真の拍点(つまり表拍)を選択した場合に比べて、ビート同期型コード特徴量系列の尤度が低くなる。そこで、音響信号分析装置 10 は、拍点候補系列 O R 1 と、拍点候補系列 O R 1 を構成する各拍点候補に対して半拍分ずれた拍点候補からなる拍点候補系列 O R 2 を検出し、拍点候補系列 O R 1 及び拍点候補系列 O R 2 に関するビート同期型コード特徴量系列の尤度 L K 1 及び尤度 L K 2 を計算する。そして、尤度 L K 1 及び尤度 L K 2 を比較して、尤度の高い拍点候補系列を選択する。これにより、誤って裏拍を表拍として選択してしまうことを抑制できる。また、ビート同期型コード特徴量系列の尤度が最も高くなるような拍子位置 s 及び調 k e y の組み合わせが選択されて、楽曲における拍点、コード進行及び小節線の位置が同時に(一体的)に推定される。したがって、音響信号分析装置 10 によれば、拍点、コード進行、及び小節線の位置の推定精度を従来よりも向上させることができる。

【0050】

さらに、本発明の実施にあたっては、上記実施形態に限定されるものではなく、本発明の目的を逸脱しない限りにおいて種々の変更が可能である。

【0051】

例えば、上記実施形態では、楽曲全体の演奏音が楽曲データとして記憶されており、その楽曲データを分析して、拍点、コード進行、及び小節線の位置を推定している。しかし、これに代えて、楽曲の演奏音をリアルタイムで取り込みつつ、取り込んだ演奏音を表わすデータを上記実施形態と同様に分析して、拍点、コード進行、及び小節線の位置を推定しても良い。

【符号の説明】

【0052】

10・・・音響信号分析装置、C R 1・・・ビート同期型コード特徴量系列、C R 2・・・ビート同期型コード特徴量系列、H P C P (T)・・・トレブル特徴量、H P C P (B)・・・ベース特徴量、j・・・和音、k e y・・・調、O R 1・・・拍点候補系列、O R 2・・・拍点候補系列、s・・・拍子位置、X B・・・BPM特徴量、X B C 1・・・ビート同期型コード特徴量、X B C 2・・・ビート同期型コード特徴量、X C・・・コード特徴量、X O・・・オンセット特徴量、(T)・・・トレブルパワー、(B)・・・ベースパワー

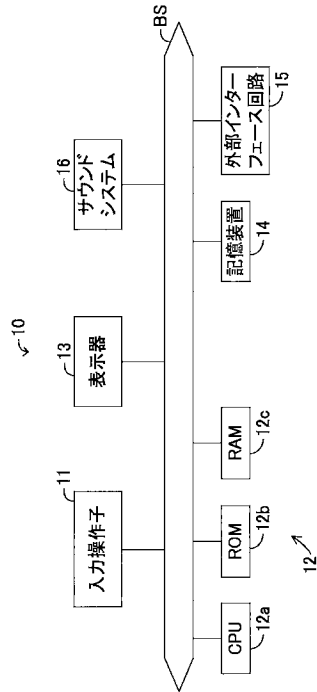
10

20

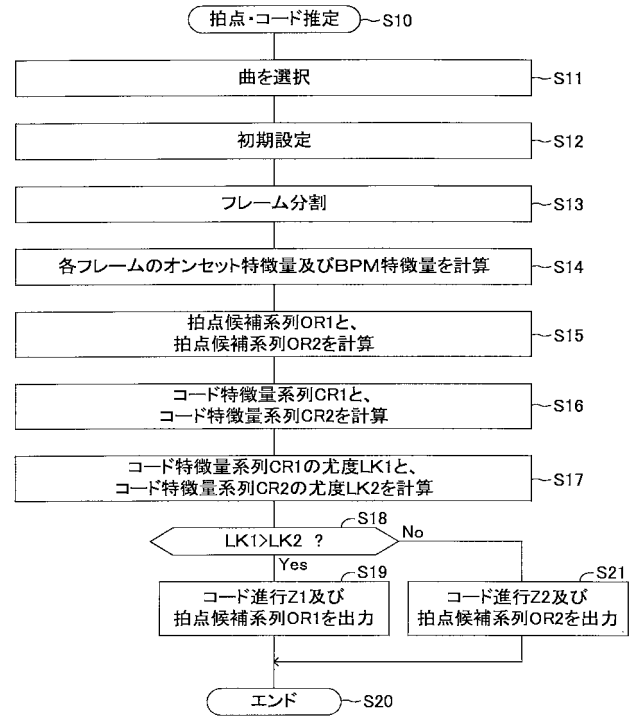
30

40

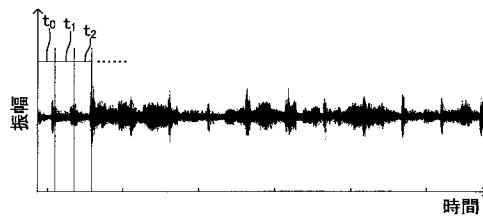
【図 1】



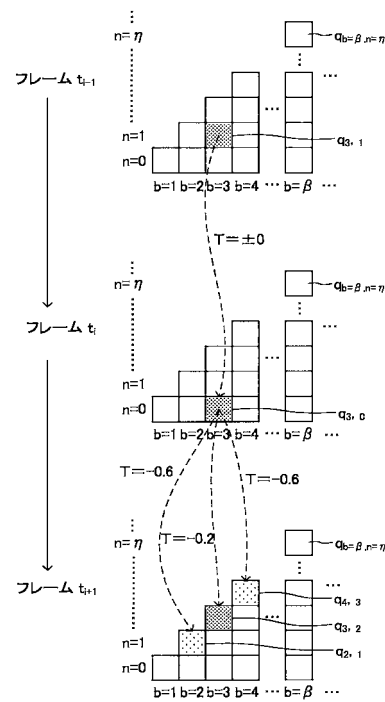
【図 2】



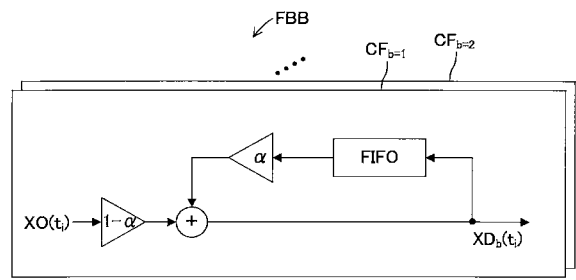
【図 3】



【図 4】



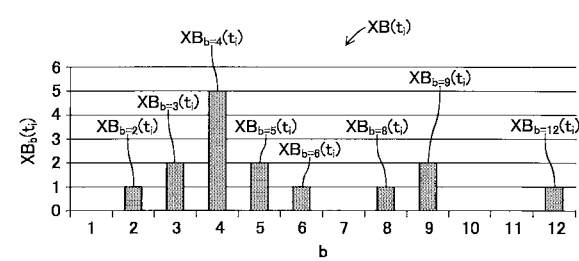
【 図 5 】



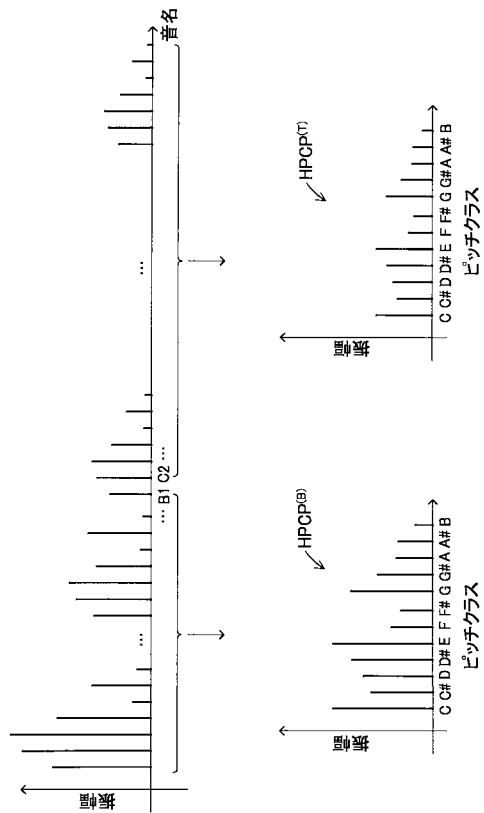
【 図 7 】

	$\delta_{2,1}$	$\delta_{2,2}$	$\delta_{2,3}$	$\delta_{2,4}$	$\delta_{2,5}$	$\delta_{2,6}$	$\delta_{2,7}$	$\delta_{2,8}$	$\delta_{2,9}$	$\delta_{2,10}$	$\delta_{2,11}$	$\delta_{2,12}$
TMP ₂												
TMP ₃	$\delta_{3,1}$	$\delta_{3,2}$	$\delta_{3,3}$	$\delta_{3,4}$	$\delta_{3,5}$	$\delta_{3,6}$	$\delta_{3,7}$	$\delta_{3,8}$	$\delta_{3,9}$	$\delta_{3,10}$	$\delta_{3,11}$	$\delta_{3,12}$
TMP ₄	$\delta_{4,1}$	$\delta_{4,2}$	$\delta_{4,3}$	$\delta_{4,4}$	$\delta_{4,5}$	$\delta_{4,6}$	$\delta_{4,7}$	$\delta_{4,8}$	$\delta_{4,9}$	$\delta_{4,10}$	$\delta_{4,11}$	$\delta_{4,12}$
TMP ₅	$\delta_{5,1}$	$\delta_{5,2}$	$\delta_{5,3}$	$\delta_{5,4}$	$\delta_{5,5}$	$\delta_{5,6}$	$\delta_{5,7}$	$\delta_{5,8}$	$\delta_{5,9}$	$\delta_{5,10}$	$\delta_{5,11}$	$\delta_{5,12}$
TMP ₆	$\delta_{6,1}$	$\delta_{6,2}$	$\delta_{6,3}$	$\delta_{6,4}$	$\delta_{6,5}$	$\delta_{6,6}$	$\delta_{6,7}$	$\delta_{6,8}$	$\delta_{6,9}$	$\delta_{6,10}$	$\delta_{6,11}$	$\delta_{6,12}$

【 図 6 】



【 図 8 】



【 図 9 】

