

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2007年5月18日 (18.05.2007)

PCT

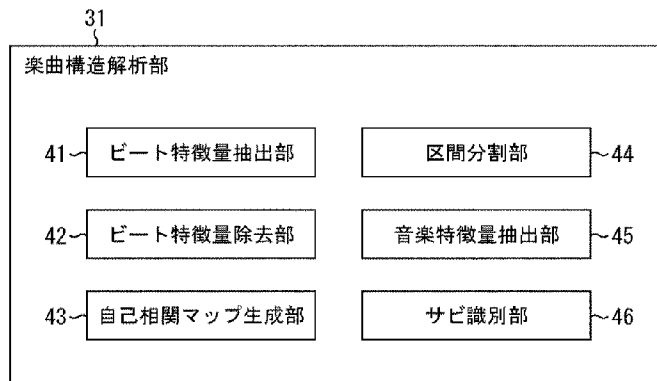
(10) 国際公開番号
WO 2007/055238 A1

- (51) 国際特許分類:
G10G 3/04 (2006.01) G10L 11/00 (2006.01)
- (21) 国際出願番号: PCT/JP2006/322273
- (22) 国際出願日: 2006年11月8日 (08.11.2006)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願2005-323153 2005年11月8日 (08.11.2005) JP
特願2006-278351 2006年10月12日 (12.10.2006) JP
- (71) 出願人 (米国を除く全ての指定国について): ソニー株式会社 (SONY CORPORATION) [JP/JP]; 東京都港区港南1丁目7番1号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 小林 由幸
- (74) 代理人: 稲本 義雄 (INAMOTO, Yoshio); 〒1600023 東京都新宿区西新宿7丁目11番18号 711ビルディング4階 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[続葉有]

(54) Title: INFORMATION PROCESSING DEVICE AND METHOD, AND PROGRAM

(54) 発明の名称: 情報処理装置および方法、並びにプログラム



- 31 MUSIC COMPOSITION STRUCTURE ANALYSIS UNIT
 41 BEAT CHARACTERISTIC AMOUNT EXTRACTION UNIT
 42 BEAT CHARACTERISTIC AMOUNT REMOVAL UNIT
 43 SELF CORRELATION MAP GENERATION UNIT
 44 SECTION DIVISION UNIT
 45 MUSIC COMPOSITION CHARACTERISTIC AMOUNT EXTRACTION UNIT
 46 DEEP VOICE IDENTIFICATION UNIT

(57) Abstract: Provided are an information processing device, method, and program capable of analyzing a waveform of a music composition having modulation by using a small calculation amount. A beat characteristic amount extraction unit (41) extracts a characteristic amount relating to a beat in the section divided for each beat. A beat characteristic amount removal unit (42) removes a phase component from a frequency component of the extracted characteristic amount. A self correlation map generation unit (43) generates self correlation information indicating correlation between a certain section and another section according to the characteristic amount from which the phase component has been removed. Thus, it is possible to analyze a waveform of a music composition having modulation with a small calculation amount. The present invention may be employed in an information processing device.

(57) 要約: 本発明は少ない演算量で、転調のある楽曲の波形を解析することができる情報処理装置および方法、並びにプログラムに関する。ビート特徴量抽出部41は、転調のある楽曲の波形をビートごとに区切った区間に於ける、ビートに関係する特徴量を抽出し、ビー

[続葉有]

WO 2007/055238 A1



(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

添付公開書類:

— 国際調査報告書

明 細 書

情報処理装置および方法、並びにプログラム

技術分野

[0001] 本発明は情報処理装置および方法、並びにプログラムに関し、特に、転調のある楽曲の波形を解析する情報処理装置および方法、並びにプログラムに関する。

背景技術

[0002] 従来より、楽曲の構造を解析するための種々の方法が提案されている。具体的には、図1の上側に示すような、楽曲の波形を解析することにより、楽曲は、図1の下側に示すように、図中左から右を時間方向とするとき、“イントロ”から開始して、“サビ”、“間奏”、“Aメロ”、“Bメロ”、“サビ”、“サビ”、“間奏”、…の順に構成されていることが解析される。そして、それらの解析の結果を基に、楽曲の中で、サビとなる区間がどこであるかを識別したりする。

[0003] なお、以下、楽曲の中ほどから曲想の変化する部分であって、特に、曲の最も盛り上がる部分をサビと称し、拍と同様に、拍子を作る単位のことをビート(beat)と称する。すなわち、ビート(拍)は、一定の時間的間隔をもって数えられるもので、小節の中でのビート(拍)の位置によりアクセントの強弱(すなわち、強拍/弱拍)が生まれ、その組み合わせにより、各種の拍子が作られる。

[0004] また、入力されてくる音声データの特徴量を抽出し、その特徴量間の類似度を計算して繰り返し区間をリストアップし、リストアップされた繰り返し区間を統合し、転調をともなう繰り返し区間の検出を行って、それらの転調をともなう繰り返し区間を統合し、統合した繰り返し区間の中から、サビ区間として適当なものを選択する方法もある(例えば、特許文献1)。

[0005] 特許文献1:特開2004-233965号公報

発明の開示

発明が解決しようとする課題

[0006] しかしながら、転調のある楽曲の波形を解析する場合、転調に対応するための演算量が増えることにより、楽曲の波形の解析に時間がかかるという問題があった。

[0007] 例えば、特開2004-233965号公報に開示されている方法は、転調をともなう繰り返し区間の検出を行うことにより、サビ区間として適当なものを選択しているが、適当なサビ区間を選択するまでに時間がかかってしまうことがある。

[0008] 本発明はこのような状況に鑑みてなされたものであり、より少ない演算量で、転調のある楽曲の波形を解析することができるようにするものである。

課題を解決するための手段

[0009] 本発明の一側面は、転調のある楽曲の波形をビートごとに区切った第1の区間における、前記ビートに関する第1の特徴量を抽出する第1の抽出手段と、抽出された前記第1の特徴量の周波数成分から位相成分を除去する除去手段と、位相成分が除去された前記第1の特徴量に基づいて、ある第1の区間と、他の第1の区間との相関を示す自己相関情報を生成する生成手段とを備える情報処理装置である。

[0010] 前記自己相関情報に基づいて、前記ビートを、複数の相関性の高いビートが繰り返されている第2の区間ごとに分割する分割手段と、分割された前記第2の区間における、前記楽曲の音楽的な特徴を示す第2の特徴量を抽出する第2の抽出手段と、前記第2の特徴量に基づいて、分割された前記第2の区間における前記楽曲の曲構成を識別する識別手段とを備えるようにすることができる。

[0011] 前記除去手段は、離散フーリエ変換により、前記第1の特徴量の周波数成分から位相成分を除去するようにすることができる。

[0012] 前記識別手段は、前記曲構成として、サビである区間を識別するようにすることができる。

[0013] 前記第1の特徴量は、前記第1の区間における、所定のオクターブ分の各音のエネルギーであるようにすることができる。

[0014] 前記第2の特徴量は、前記第2の区間における、音量、音源に関する情報、またはある周波数帯の利得であるようにすることができる。

[0015] 前記第2の特徴量は、前記第2の区間の長さ、前記第2の区間が出現した回数、または前記第2の区間が連続して繰り返された回数であるようにすることができる。

[0016] 前記生成手段は、位相成分が除去された前記第1の特徴量に基づいて、所定のビートごとに、ある第1の区間との相関を算出することにより、前記自己相関情報を生成

するようにすることができる。

[0017] 前記除去手段は、シャッフルされた前記第1の特徴量の周波数成分から位相成分を除去するようにすることができる。

[0018] 前記除去手段は、前記各音のエネルギーをシフトさせてから、前記第1の特徴量の周波数成分から位相成分を除去するようにすることができる。

[0019] 本発明の一側面は、転調のある楽曲の波形をビートごとに区切った区間における、前記ビートに関する特徴量を抽出する抽出ステップと、抽出された前記特徴量の周波数成分から位相成分を除去する除去ステップと、位相成分が除去された前記特徴量に基づいて、ある区間と、他の区間との相関を示す自己相関情報を生成する生成ステップとを含む情報処理方法である。

[0020] 本発明の一側面は、転調のある楽曲の波形をビートごとに区切った区間における、前記ビートに関する特徴量を抽出する抽出ステップと、抽出された前記特徴量の周波数成分から位相成分を除去する除去ステップと、位相成分が除去された前記特徴量に基づいて、ある区間と、他の区間との相関を示す自己相関情報を生成する生成ステップとをコンピュータに実行させるプログラムである。

[0021] 本発明の一側面においては、転調のある楽曲の波形をビートごとに区切った第1の区間における、前記ビートに関する第1の特徴量が抽出され、抽出された前記第1の特徴量の周波数成分から位相成分が除去され、位相成分が除去された前記第1の特徴量に基づいて、ある第1の区間と、他の第1の区間との相関を示す自己相関情報が生成される。

発明の効果

[0022] 以上のように、本発明の一側面によれば、少ない演算量で、転調のある楽曲の波形を解析することができる。

図面の簡単な説明

[0023] [図1]従来の楽曲の曲構造を解析する方法を説明する図である。

[図2]パーソナルコンピュータのハードウェアの構成の例を説明するブロック図である。

[図3]パーソナルコンピュータのCPUにより実行されるソフトウェアの機能的構成例を

示すブロック図である。

[図4]楽曲構造解析部による、サビ抽出の処理について説明するフローチャートである。

[図5]楽曲の波形から検出されるビートの例を示す図である。

[図6]楽曲の波形のビートごとに抽出されるビート特徴量について説明する図である。

[図7]ビート特徴量抽出部による、ビート特徴量抽出の処理の詳細について説明するフローチャートである。

[図8]ビート特徴量抽出部による、ビート特徴量抽出の処理について説明する図である。

[図9]音程と時間による各音(C1乃至B7)のエネルギーの分布の例を示す図である。

[図10]7オクターブ分の各音のエネルギーを解析する方法について説明する図である。

[図11]あるビート特徴量と、他のビート特徴量との相関について説明する図である。

[図12]ビート特徴量を除去する処理について説明する図である。

[図13]ビート特徴量を除去する処理について説明する図である。

[図14]自己相関マップ生成部による、自己相関マップ生成の処理の詳細について説明するフローチャートである。

[図15]自己相関マップ生成部により生成される、自己相関マップの例を示す図である。

[図16]自己相関マップにおける区間の分割について説明する図である。

[図17]楽曲における区間の分割について説明する図である。

[図18]楽曲の波形の分割された区間ごとに抽出される音楽特徴量について説明する図である。

[図19]楽曲の分割された区間ごとに抽出される、音楽特徴量以外のその他の特徴量について説明する図である。

[図20]サビ識別部による、サビである区間の識別について説明する図である。

[図21]サビ識別部による、サビである区間の識別の詳細について説明する図である。

[図22]自己相関マップ生成部による、自己相関マップ生成の処理の詳細について説

明するフローチャートである。

[図23]片側のNビートごとの自己相関マップの例を示す図である。

[図24]片側のNビートごとの自己相関マップの例を示す図である。

[図25]ビート特徴量のシャッフルについて説明する図である。

[図26]ビート特徴量のシフトについて説明する図である。

[図27]パーソナルコンピュータのCPUにより実行されるソフトウェアの機能的構成の他の例を示すブロック図である。

[図28]楽曲の波形から検出されるビートの例を示す図である。

[図29]ビート検出部の構成の一例を示すブロック図である。

[図30]アタック情報の抽出について説明する図である。

[図31]アタック情報の抽出について説明する図である。

[図32]基本ビート周期の検出について説明する図である。

[図33]テンポの決定について説明する図である。

[図34]テンポの修正について説明する図である。

[図35]テンポの修正について説明する図である。

[図36]楽曲特徴量の抽出について説明する図である。

[図37]拍子の決定について説明する図である。

[図38]拍子の決定について説明する図である。

[図39]小節の先頭の決定について説明する図である。

[図40]小節の先頭の決定について説明する図である。

符号の説明

- [0024] 1 パーソナルコンピュータ, 11 CPU, 12 ROM, 13 RAM, 16 入力部, 17 出力部, 18 記録部, 19 通信部, 20 ドライブ, 21 リムーバブルメディア, 31 楽曲構造解析部, 41 ビート特徴量抽出部, 42 ビート特徴量除去部, 43 自己相関マップ生成部, 44 区間分割部, 45 音楽特徴量抽出部, 46 サビ識別部, 101 ビート検出部, 111 アタック情報抽出部, 112 基本ビート周期検出部, 113 テンポ決定部, 114 楽曲特徴量抽出部, 115 テンポ修正部, 116 ビート特徴量抽出部, 117 拍子決定部, 118 小節先頭決定

部

発明を実施するための最良の形態

[0025] 以下、図面を参照しながら本発明の実施の形態について説明する。

[0026] 図2は、パーソナルコンピュータ1のハードウェアの構成の例を説明するブロック図である。

[0027] パーソナルコンピュータ1は、楽曲のデータを記録し、その楽曲のデータから再生される楽曲の波形を解析するための所定の処理を行う機器であり、本発明の情報処理装置の一例である。

[0028] 図2の例のパーソナルコンピュータ1において、CPU(Central Processing Unit)11は、ROM(Read Only Memory)12に記憶されているプログラム、または記録部18からRAM(Random Access Memory)13にロードされたプログラムに従って各種の処理を実行する。RAM13にはまた、CPU11が各種の処理を実行する上において必要なデータなども適宜記憶される。

[0029] CPU11、ROM12、およびRAM13は、バス14を介して相互に接続されている。このバス14にはまた、入出力インターフェース15も接続されている。

[0030] 入出力インターフェース15には、キーボード、マウスなどよりなる入力部16、スピーカ、LCD(Liquid Crystal Display)などのディスプレイなどよりなる出力部17、ハードディスクなどより構成される記録部18、並びにインターネットなどのネットワークを介して他の機器との通信処理を制御する通信部19が接続されている。

[0031] 入出力インターフェース15にはまた、必要に応じてドライブ20が接続され、磁気ディスク、光ディスク、光磁気ディスク、または半導体メモリなどのリムーバブルメディア21が適宜装着され、それらから読み出されたコンピュータプログラムが、必要に応じて記録部18にインストールされる。

[0032] なお、パーソナルコンピュータ1のハードウェアの構成は、図2の例に限定されず、後述する図3の機能的構成を少なくとも有していればよい。

[0033] 図3は、図2のパーソナルコンピュータ1のCPU11により実行されるソフトウェアの機能的構成例を示すブロック図である。

[0034] 楽曲構造解析部31は、(転調のある)楽曲の波形を解析するための所定の処理を

行う。

- [0035] 楽曲構造解析部31は、ビート特徴量抽出部41、自己相関マップ生成部42、区間分割部43、音楽特徴量抽出部44、およびサビ識別部45を含むようにして構成される。
- [0036] ビート特徴量抽出部41は、ビート特徴量抽出の処理を行って、ビートごとに抽出される特徴量であるビート特徴量を抽出する。ビート特徴量抽出部41は、抽出したビート特徴量をビート特徴量除去部42に供給する。なお、ビート特徴量抽出の処理の詳細は後述する。
- [0037] ビート特徴量除去部42は、ビート特徴量抽出部41から供給されるビート特徴量の位相成分を除去する。ビート特徴量除去部42は、位相成分が除去されたビート特徴量を自己相関マップ生成部43に供給する。
- [0038] 自己相関マップ生成部43は、ビート特徴量除去部42から供給される、位相成分が除去されたビート特徴量を基に、自己相関マップ生成の処理を行って、自己相関マップを生成する。自己相関マップ生成部43は、生成した自己相関マップを区間分割部44に供給する。なお、自己相関マップ生成の処理の詳細は後述する。
- [0039] 区間分割部44は、自己相関マップ生成部43から供給される自己相関マップを基に、楽曲の波形が同一のフレーズを複数回繰り返している区間ごとに分割する(セグメンテーションする)。
- [0040] 音楽特徴量抽出部45は、同一の(似ている)フレーズごとに分割されたそれぞれの区間(の楽曲の波形)から、音楽的な特徴量である音楽特徴量を抽出する。音楽特徴量抽出部45は、抽出した音楽特徴量をサビ識別部46に供給する。
- [0041] また、音楽特徴量抽出部45は、同一の(似ている)フレーズごとに分割されたそれぞれの区間(の楽曲の波形)から、音楽特徴量以外のその他の特徴量を抽出する。音楽特徴量抽出部45は、抽出したその他の特徴量をサビ識別部46に供給する。
- [0042] サビ識別部46は、音楽特徴量抽出部45から供給される音楽特徴量またはその他の特徴量を基に、同一の(似ている)フレーズごとに分割されたそれぞれの区間(の楽曲の波形)から、サビである区間を識別する。
- [0043] なお、本実施の形態では、パーソナルコンピュータ1は、上述した図2のハードウェア

ア構成を有しているので、楽曲構造解析部31は、例えば、図2のCPU11が実行するアプリケーションソフトウェアとして構成されるが、それ以外にも、パーソナルコンピュータ1のハードウェア構成を図2の構成とは異ならせることで、楽曲構造解析部31は、ハードウェア単体として構成することもできるし、ソフトウェアとハードウェアの組み合わせとして構成することもできる。

[0044] ところで、上述したように、楽曲の波形から、その楽曲の曲構成を解析することにより、その楽曲のサビとなる区間を識別することができる。したがって、次に、図4のフローチャートを参照して、楽曲構造解析部31による、サビ識別の処理について説明する。

[0045] ステップS1において、楽曲構造解析部31は、楽曲の波形からビートを検出する。例えば、ステップS1において、楽曲構造解析部31は、MP3(MPEG Audio Layer-3)やAAC(Advanced Audio Coding)などの方式で音声圧縮された音声信号より再生される楽曲の波形から、テンポ、拍子、または小節などの位置を検出する。

[0046] 図5は、楽曲の波形から検出されるビートの例を示す図である。

[0047] 図5の上側で示される楽曲の波形から検出されるビートは、図5の下側に示すように、楽曲の波形を複数に分割している、図中垂直方向の線により区切られた区間により示されている。すなわち、それらの線のうち、4つおきに示されている太い線は、小節の先頭を示し、その太線により区切られた区間の中の、さらに細かい線で区切られた4つの区間のそれぞれが、テンポ(4分音符の長さ)を示している。すなわち、図5に示される楽曲の波形は、図中下側の波形を分割する線により区切られた区間ごとに付されている数字(1乃至4)で示されるように、4ビートの楽曲となる。

[0048] すなわち、例えば、楽曲構造解析部31は、図5で示すように、図中上側に示す楽曲の波形から、図中下側に示すビートを検出する。

[0049] 図4のフローチャートに戻り、ステップS2において、ビート特徴量抽出部41は、ビート特徴量抽出の処理を行って、ビート特徴量を抽出する。

[0050] 図6は、楽曲の波形のビートごとに抽出されるビート特徴量について説明する図である。

[0051] 図6に示される楽曲の波形においては、図5の下側と同様に、楽曲の波形がビート

ごとに分割されており(図6に示す例では12区間に分割されている)、ビート特徴量抽出部41は、それらの分割されているビートごとに、それぞれ、ビート特徴量を抽出する。

[0052] ここで、ステップS2における、ビート特徴量抽出部41による、ビート特徴量抽出の処理の詳細について、図7のフローチャートを参照して説明する。

[0053] ステップS11において、ビート特徴量抽出部41は、ビートごとに分割されている楽曲の波形から、例えば、ビートごとに7オクターブ分の各音のエネルギーを算出する。

[0054] 図8は、ビート特徴量抽出部41による、ビート特徴量抽出の処理について説明する図である。

[0055] 図8で示すように、例えば、7オクターブ分の各音のエネルギーを算出する場合、オクターブ1乃至オクターブ7(以下、O1乃至O7とも称する)のそれぞれの、C, C#, D, D#, E, F, F#, G, G#, A, A#, Bであるコードのエネルギーを算出することになる。ここで、1オクターブには、周波数の低い方から、C(ド), C#, D(レ), D#, E(ミ), F(ファ), F#, G(ソ), G#, A(ラ), A#, B(シ)の12の音程(音)が含まれる。例えば、オクターブ1(O1)のC, C#, D, D#, E, F, F#, G, G#, A, A#, Bの12の音程を、それぞれ、C1, C#1, D1, D#1, E1, F1, F#1, G1, G#1, A1, A#1, B1と称し、オクターブ2(O2)のC, C#, D, D#, E, F, F#, G, G#, A, A#, Bの12の音程を、それぞれ、C2, C#2, D2, D#2, E2, F2, F#2, G2, G#2, A2, A#2, B2と称する。また、オクターブ3(O3)乃至オクターブ7(O7)についても同様である。

[0056] すなわち、ビート特徴量抽出部41は、例えば、ビートごとに分割されている楽曲の波形から、C1乃至B7のそれぞれのエネルギーを算出する。

[0057] ここで、7オクターブ分の各音(C1乃至B7)のエネルギーを算出する方法であるが、例えば、ビートごとに分割されている楽曲の波形を、音程および時間により解析することで、各音のエネルギーを算出する。次に、図9および図10を参照して、7オクターブ分の各音のエネルギーを算出する方法について説明する。

[0058] 図9は、音程と時間による各音(C1乃至B7)のエネルギーの分布の例を示す図である。

- [0059] 図9で示される例において、横軸方向は時間を示し、図中左から右方向に時間が進むことを示している。また、縦軸方向は音程を示し、図中下から上方向になるにつれて音程が高くなる(周波数が高くなる)ことを示す。
- [0060] 図中の黒で示される領域と白で示される領域とがあるが、黒で示される領域ほど、エネルギーが強いことを示し、図中の白で示される領域ほどエネルギーが弱くなることを示している。なお、図9の例では、説明を分かり易くするために、エネルギーの強さを黒と白の2色により表現しているが、実際には、さらに細かくエネルギーの強さの分布を表現するようにしてもよい。
- [0061] 図中横方向の6本の点線は、説明を分かり易くするために引かれた線であり、これらの点線により区切られている7つの区間のそれぞれが1オクターブ分を示し、すなわち、7つの区間により、O1乃至O7の7オクターブに分割されている。
- [0062] 例えば、図9で示される例において、図中下から1番目乃至3番目の区間であるO1乃至O3のそれぞれでは、各音のエネルギーは、とびとびに分布しているが、図中下から4番目乃至7番目の区間であるO4乃至O7のそれぞれでは、各音のエネルギーは、ほぼ全体的に分布していることになる。
- [0063] ビート特徴量抽出部41は、ビートごとに7オクターブ分のエネルギーを算出する場合、図10に示すように、ビート特徴量を抽出するビートの時間範囲における信号を切り出し、その切り出した信号のエネルギーを時間により平均する。その結果、ビート特徴量抽出部41は、図10の右側のグラフで示すような、各音(C1乃至B7)のエネルギーを算出することができる。
- [0064] 図7のフローチャートに戻り、ステップS12において、ビート特徴量抽出部41は、算出された各音のエネルギーに重み付けをする。例えば、ステップS12において、ビート特徴量抽出部41は、図8に示すように、算出された7オクターブ分の各音(C1乃至B7)のエネルギーに、それぞれ重み付けをする。
- [0065] ここで、各音(C1乃至B7)のエネルギーに重み付けをする方法であるが、ビート特徴量抽出部41は、例えば、ガウス分布などにより重みを算出し、算出された重みを正規化することにより、各音(C1乃至B7)における重みを特定して、重み付けをする。
- [0066] 具体的には、ガウス分布による重み $W(N)$ は、式(1)により算出される。

[0067] $W(N) = \text{Exp}(- (N - G_m) \times (N - G_m) / (2 \times G_v \times G_v)) \cdots (1)$

[0068] なお、式(1)において、 G_m は、最も重みの高い音程を示し、 G_v は、重みの広がりを示しており、それぞれ所定の値が設定される。また、 N は、音程の数を示し、すなわち、音程の数(例えば、図8の場合、C1乃至B7)だけガウス分布による重み $W(N)$ が算出されることになる。そして、算出された重み $W(N)$ は、正規化されることになる。

[0069] このように、各音(C1乃至B7)に対して、重み付けを行うことにより、例えば、高い周波数帯の音や低い周波数帯の音などの関係のない音を取り除くことができる。

[0070] 図7のフローチャートに戻り、ステップS13において、ビート特徴量抽出部41は、ビート特徴量を抽出して、ビート特徴量抽出の処理は終了する。

[0071] 例えば、ステップS13において、ビート特徴量抽出部41は、図8に示すように、重み付けがされたC1乃至B7の中から、C1, C2, C3, C4, C5, C6, C7を加算することにより、Cであるコードのエネルギーを抽出し、C#1, C#2, C#3, C#4, C#5, C#6, C#7を加算することにより、C#であるコードのエネルギーを抽出する。また、同様に、ビート特徴量抽出部41は、重み付けされたC1乃至B7の中から、D1乃至D7を加算することにより、Dであるコードのエネルギーを抽出し、D#1乃至D#7を加算することにより、D#であるコードのエネルギーを抽出し、E1乃至B7も同様に加算することにより、E乃至Bであるコードのそれぞれのエネルギーが抽出される。そして、ビート特徴量抽出部41は、図8に示すように、ビート特徴量として、C, C#, D, D#, E, F, F#, G, G#, A, A#, Bの12音分のエネルギーを抽出する。

[0072] すなわち、ビート特徴量抽出部41は、ステップS11乃至ステップS13の処理により、ビートごとに分割されている楽曲の波形のそれぞれの区間におけるビート特徴量を算出する。

[0073] なお、ビート特徴量抽出部41が抽出するビート特徴量は、上述した例に限らず、例えば、ステップS13の処理における加算の処理を行わずに、そのまま84(12×7)個のエネルギーを使用するなど、ビートごとに分割されている区間ごとの特徴量であればよい。

[0074] 図4のフローチャートに戻り、ステップS3において、ビート特徴量除去部42は、ビート特徴量抽出部41から供給されるビート特徴量の位相成分を除去し、位相成分が除

去されたビート特徴量を自己相関マップ生成部43に供給する。

- [0075] 例えば、ステップS3において、ビート特徴量除去部42は、ビートごとに分割されている楽曲の波形から、相関を算出する区間を特定する。ビート特徴量除去部42は、特定された区間における抽出されたビート特徴量に対して、離散フーリエ変換(DFT (Discrete Fourier Transform))を施すことにより、抽出されたビート特徴量の位相成分を除去する。
- [0076] 図11は、あるビート特徴量と、他のビート特徴量との相関について説明する図である。
- [0077] 図11で示される例において、ビートごとに分割されている楽曲の波形のそれぞれのビート特徴量を、それぞれ、楽曲の波形の下側のC, C#, D, D#, E, F, F#, G, G#, A, A#, Bである12音分のコード(図11の例では、C, Bのコード名のみを記載している)を示している12個の四角で表現している。
- [0078] ある楽曲の波形においては、例えば、位置Xを含んでいる区間(のビート特徴量)と、位置Yを含んでいる区間(のビート特徴量)との相関(異なる2箇所との相関)を算出する場合、位置Xを含んでいる区間から時間方向に後ろとなる、図中左下がり斜線でハッチングされた4区間(4ビート)分のビート特徴量と、位置Yを含んでいる区間から時間方向に後ろとなる、図中右下がり斜め線でハッチングされた4区間(4ビート)分のビート特徴量との相関を算出する。
- [0079] このとき、例えば、ビート特徴量除去部42は、相関を算出する、図中左下がり斜線および右下がり斜め線によりハッチングされた4区間分のビート特徴量のそれぞれに対して、離散フーリエ変換を施すことにより、それらのビート特徴量の位相成分を除去する。
- [0080] すなわち、例えば、ビート特徴量除去部42は、図12に示すように、図11の左下がり斜線でハッチングされた4区間(4ビート)のビート特徴量に対して、離散フーリエ変換を施して位相成分を除去することにより、変換前に、48(12個×4区間)個あったビート特徴量を、25($12 \times 4 / 2 + 1$)個に減少させる。
- [0081] 同様に、例えば、ビート特徴量除去部42は、図11の右下がり斜線でハッチングされた4区間(4ビート)分のビート特徴量に対して、離散フーリエ変換を施して位相成

分を除去することにより、ビート特徴量を、48個から25個に減少させる。

- [0082] このとき、離散フーリエ変換により得られたビート特徴量(例えば、25個のビート特徴量)は、図13に示すように、例えば、“ドミンド”をキーCやキーD(キーCと比較してキーが2個上に上がっている)で演奏したときなどにおいて、CやDなどの調によらず一定となる。すなわち、転調のある楽曲において、離散フーリエ変換により得られたビート特徴量は、転調とは関係なく一定であるので、この25個のビート特徴量同士の相関を1度計算することにより、転調に対応した相関の計算を行うことができる。その結果、相関に関する演算量を減らすことができるので、より少ない演算量で、転調のある楽曲の波形を解析することができる。
- [0083] なお、ビート特徴量の相関を算出する場合の区間であるが、上述した4区間(4ビート)に限らず、例えば、1, 8, または10などであってもよく、また、その区間も、基準となる位置(例えば、図11のX, Y)から時間方向に後方とは限らず、前方や前後を含むようにしてもよい。すなわち、ビート特徴量の相関を算出する区間は、例えば、図11において、位置Xを含んでいる区間から前方3区間(3ビート)と、後方6区間(6ビート)の合計8区間(8ビート)のビート特徴量としてもよい。
- [0084] また、上述した例において、ビート特徴量除去部42は、離散フーリエ変換により、ビート特徴量の位相成分を除去するとして説明したが、本発明はそれに限らず、ビート特徴量の位相成分を除去できる方法であればよい。
- [0085] 図4のフローチャートに戻り、ステップS4において、自己相関マップ生成部43は、ビート特徴量除去部42から供給される、位相成分が除去されたビート特徴量を基に、自己相関マップ生成の処理を行って、自己相関マップを生成する。
- [0086] ここで、ステップS4における、自己相関マップ生成部43による、自己相関マップ生成の処理の詳細について、図14のフローチャートを参照して説明する。
- [0087] ステップS21において、自己相関マップ生成部43は、ビート特徴量除去部42から供給される、図11に示すような、位置Xを含んでいる4区間(図中左下がり斜線でハッチングされた区間)の位相成分が除去されたビート特徴量と、位置Yを含んでいる4区間(図中右下がり斜線でハッチングされた区間)の位相成分が除去されたビート特徴量との相関を算出する。

[0088] 換言すれば、自己相関マップ生成部43は、転調のある楽曲に対応するために、数ビート分のビート特徴量を離散フーリエ変換した結果のうち、ノルム成分のみを用いて相関を算出しているとも言える。

[0089] 具体的には、ビート特徴量の相関は、例えば、式(2)により算出される。

[0090] [数1]

$$\text{Pearson}(X, Y) = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}} \dots$$

[0091] なお、式(2)において、例えば、Xは、図11の位置Xを含んでいる4区間(4ビート)の位相成分が除去されたビート特徴量を示し、Yは、図11の位置Yを含んでいる4区間(4ビート)の位相成分が除去されたビート特徴量を示す。

[0092] ステップS22において、自己相関マップ生成部43は、算出された各ビート間の相関に基づいて、自己相関マップを生成する。

[0093] 図15は、自己相関マップ生成部43により生成される、自己相関マップの例を示す図である。

[0094] 図15で示される自己相関マップの例において、水平方向の軸は、ある楽曲(の1曲中)の開始から終了までのビートを表しており、その時間の方向は、図中左から右に向かう方向とされている。また、垂直方向の軸は、水平方向の軸で示した楽曲と同一の楽曲(の1曲中)の開始から終了までのビートを表しており、その時間の方向は、図中上から下に向かう方向とされている。すなわち、水平方向の軸は、図中右に行くほど、その楽曲の終わりの方となり、垂直方向の軸は、図中下に行くほど、その楽曲の終わりの方となる。なお、水平方向の軸と、垂直方向の軸とは、同一の楽曲のビートを表しているため、その軸の長さは同一となっている。

[0095] また、図中の黒で示されている領域は、水平方向の軸と垂直方向の軸により特定される楽曲の各ビートの相関が高いことを示し、図中の白で示されている領域は、水平

方向の軸と垂直方向の軸により特定される楽曲の各ビートの相関が小さいことを示す。なお、当然ながら、図中の正方形で示す自己相関マップの左上から右下に伸びる対角線は、水平方向の軸と垂直方向の軸により特定される楽曲のビートが同一となる箇所であるので、図中の黒い線(対角線)で示すように相関が高いことになる。

[0096] 図15の自己相関マップの例は、例えば、式(2)により算出された、ビート特徴量の相関に基づいて、ビート間の相関の高い部分を黒で示し、ビート間の相関の小さい部分を白で表していることになる。すなわち、自己相関マップは、ある楽曲において同じフレーズとなる部分(相関のある部分)を、黒い線により表している図となる。

[0097] 図4のフローチャートに戻り、ステップS5において、区間分割部44は、自己相関マップ生成部43から供給される自己相関マップを基に、楽曲の波形が同一のフレーズ(似ているフレーズ)を複数回繰り返している区間ごとに分割する(セグメンテーションする)。

[0098] 具体的には、図16の例で示されるように、図15の自己相関マップの例において、上から1番目の点線Aで囲まれる領域では、ほぼ同様の(似ている)模様となる右下斜め線が4回出現しており、その出現している区間を区間Aとし、上から2番目の点線Cで囲まれる領域では、ほぼ同様の模様となる右下斜め線が2回出現しており、その出現している区間を区間Cとする。また、同様に、上から3番目の点線Bで囲まれる領域では、ほぼ同様の模様となる右下斜め線が2回出現しており、その出現している区間を区間Bとし、上から4番目の点線Dで囲まれる領域では、ほぼ同様の模様となる右下斜め線が2回出現しており、その出現している区間を区間Dとする。さらに、図示はしないが、同様に、ほぼ同様の模様となる右下斜め線が出現している区間を、それぞれ、区間E、区間F、区間G、・・・とする。

[0099] このように、同一のフレーズを複数回繰り返している区間ごとに分割された図16の自己相関マップを、図16の水平方向の軸で示す楽曲のビートに合わせて表現すると、図17のように表現することができる。すなわち、図17は、似ているフレーズとなる区間ごとに分割された楽曲(の波形)を示す。

[0100] 図17において、楽曲は、図16で分割した区間に対応して、A, B, C, D, E, B, B, C, ...である区間の順に、それぞれに分割されている。なお、その時間の方向は、

図中左から右に向かう方向となり、それぞれの区間においては、より長い区間の方が、その区間が長い時間再生されることを示し、逆に、より短い区間の方が、その区間が短い時間再生されることを示している。

- [0101] 図4のフローチャートに戻り、ステップS6において、音楽特徴量抽出部45は、同一の(似ている)フレーズごとに分割されたそれぞれの区間(の楽曲の波形)から、音楽特徴量を抽出し、抽出した音楽特徴量をサビ識別部46に供給する。
- [0102] ここで、音楽特徴量抽出部45により抽出される音楽特徴量は、例えば、音量、音源に関する情報、ある周波数帯の利得、周波数のバランス、音の多さ、ドラムやベースなどのリズム楽器の割合、または、ボーカルが存在するかなどの音楽的な特徴量となる。
- [0103] 図18は、楽曲の波形の分割された区間ごとに抽出される音楽特徴量について説明する図である。
- [0104] 図18に示される分割された区間は、図17と同様に、図16の自己相関マップを分割した区間に対応して分割されており(A, B, C, D, E, B, B, C, …である区間の順に、それぞれに分割されている)、音楽特徴量抽出部45は、それらの分割されている区間ごとに、それぞれ、例えば、音量などの音楽特徴量を抽出する。
- [0105] 図4のフローチャートに戻り、また、ステップS6において、音楽特徴量抽出部45は、同一の(似ている)フレーズごとに分割されたそれぞれの区間(の楽曲の波形)から、音楽特徴量以外のその他の特徴量を抽出し、抽出したその他の特徴量をサビ識別部46に供給する。
- [0106] ここで、音楽特徴量抽出部45により抽出されるその他の特徴量は、例えば、分割されたある区間の長さ、1曲中にある区間が何回出現するかの回数、または、ある区間が繰り返される回数などの特徴量となる。
- [0107] 図19は、楽曲の分割された区間ごとに抽出される、音楽特徴量以外のその他の特徴量について説明する図である。
- [0108] 図19で示される例において、音楽特徴量抽出部45は、その他の特徴量として、区間Bの長さなどのセグメントの長さ(ビート単位)、1曲中に区間Bが何回出現するかを表している1曲中での同じセグメントの出現回数(繰り返し回数)、区間Bが繰り返され

ている回数を表している連続繰り返し回数などを抽出する。

- [0109] また、その他の特徴量は、図19を参照して説明した以外でも、例えば、1曲中のセグメントの相対位置(例えば、5分の曲において、あるセグメントが1分から始まる場合は、 $1/5=0.2$)、または直前または直後に出現するセグメントのバリエーションの多さなどであってもよい。
- [0110] なお、詳細は後述するが、音楽特徴量およびその他の特徴量を、より多く抽出することにより、例えば、サビ識別部46は、楽曲の曲構造を識別する精度をより向上させることができる。
- [0111] なお、音楽特徴量抽出部45は、同一の(似ている)フレーズごとに分割されたそれぞれの区間(の楽曲の波形)から、音楽特徴量またはその他の特徴量のうち、いずれか一方を抽出し、抽出したいずれか一方の特徴量をサビ識別部46に供給するようにしてもよい。
- [0112] 図4のフローチャートに戻り、ステップS7において、サビ識別部46は、音楽特徴量抽出部45から供給される音楽特徴量またはその他の特徴量を基に、同一の(似ている)フレーズごとに分割されたそれぞれの区間(の楽曲の波形)から、サビである区間を識別して、サビ識別の処理は終了する。
- [0113] 例えば、ステップS7において、サビ識別部46は、図20で示すように、抽出された分割されたそれぞれの区間ごとの音楽特徴量(X1)とその他の特徴量(X2)を基に、それぞれの区間における“サビらしさ”(y(%))を算出することにより、サビである区間を識別する。
- [0114] 具体的には、例えば、サビ識別部46は、音楽特徴量(X1)、その他の特徴量(X2)と、正解となるサビデータ(例えば、ユーザにより設定される、ある楽曲のどこがサビとなるかを示すデータ)を用いて、サビを識別するための学習することにより、所定の関数(例えば、“サビらしさ”を演算する関数)を生成する。サビ識別部46は、音楽特徴量抽出部45から供給される音楽特徴量(X1)およびその他の特徴量(X2)を、その生成された所定の関数によって演算することにより、それぞれの区間における“サビらしさ”(y(%))を算出して、サビである区間を識別する。
- [0115] なお、サビ識別部46が算出する“サビらしさ”(y(%))であるが、上述した方法に限

らず、サビである確率を算出することのできる方法であればよい。

[0116] 図21は、サビ識別部46による、サビである区間の識別の詳細について説明する図である。

[0117] 図21で示される例において、楽曲(の波形)は、図17などと同様に、A, B, C, D, E, B, B, C, …である区間の順に、それぞれの区間が分割されている。サビ識別部46は、例えば、分割されたそれぞれの区間において、“区間Aのサビらしさ”が0%、“区間Bのサビらしさ”が60%、“区間Cのサビらしさ”が10%、“区間Dのサビらしさ”が30%、“区間Eのサビらしさ”が40%、“区間Bのサビらしさ”が70%、“区間Bのサビらしさ”が80%、“区間Cのサビらしさ”が15%、…を、それぞれ算出する。

[0118] そして、サビ識別部46は、例えば、算出された“サビらしさ”の平均が最も高い区間をサビである区間として識別する。例えば、サビ識別部46は、図21で示すように、区間Bのそれぞれの“サビらしさ”が、60%、70%、80%、…となり、“サビらしさ”の平均が最も高いので、区間Bをサビである区間として識別する。

[0119] 以上のようにして、楽曲構造解析部31は、より少ない演算量で、転調のある楽曲の波形から、その楽曲の曲構成を識別することができる。

[0120] 例えば、パーソナルコンピュータ1の、例えば、ハードディスクなどからなる記録部18に記録されている膨大な数の楽曲のデータより得られる楽曲の波形から、より少ない演算量で、その楽曲の曲構成からサビである区間を識別することにより、楽曲を選択するとき、曲の頭から再生させるのではなく、サビのみを再生することができる。その結果、ユーザは、例えば、曲の一番盛り上がるサビの部分を視聴して、所望の楽曲を検索したりすることができる。

[0121] また、より少ない演算量で、その楽曲の曲構成からサビである区間を識別することができるので、例えば、容量が少ない記録媒体に、楽曲のサビの区間のみを切り出して記録させることで、より多くの楽曲のデータを記録することができる。

[0122] なお、上述した例においては、楽曲の波形から、曲構成として“サビ”を識別するとして説明したが、本発明はそれに限定されず、“サビ”の代わりに、例えば、“イントロ”、“間奏”、“Aメロ”、“Bメロ”などを識別するようにしてもよい。

[0123] ところで、自己相関マップ生成部43による、自己相関マップの生成の処理であるが

、上述した処理(図14の自己相関マップ生成の処理)に限定されず、位相成分が除去されたビート特徴量を基に、自己相関マップを生成することができる方法であればよい。例えば、図4のステップS5の処理における、区間分割部44によるセグメンテーションの処理において、“サビ”である区間を識別するだけでよい場合(すなわち、“Aメロ”、“Bメロ”である区間を識別する必要がない場合)、通常、“サビ”の長さとして考えることのできる長さである、所定のビート以上(例えば、20ビート以上など)の区間のみ、分割することができればよいことになる。すなわち、自己相関マップからは、所定のビート以上の長さの繰り返しのみが検出することができればよいことになる。そこで、まず、片側(例えば、後述する図23の垂直方向の軸)において、所定のビートごとに自己相関マップを生成し、その後、相関が高い(すなわち、繰り返しの可能性がある)箇所の前後のみ、相関を計算することにより、さらに“サビ”を識別する処理の負荷を低減することができる。

[0124] したがって、次に、図22のフローチャートを参照して、“サビ”である区間を識別するだけでよい場合の、図4のステップS4における、自己相関マップ生成部43による、自己相関マップ生成の処理の詳細について説明する。

[0125] ステップS31において、自己相関マップ生成部43は、片側のNビートごとに、ビート間の相関を算出する。例えば、ステップS31において、自己相関マップ生成部43は、後述する図23の垂直方向の軸においてNビートごとに、ビート特徴量除去部42から供給される、図11に示すような、位置Xを含んでいる4区間(図中左下がり斜線でハッチングされた区間)の位相成分が除去されたビート特徴量と、位置Yを含んでいる4区間(図中右下がり斜線でハッチングされた区間)の位相成分が除去されたビート特徴量との相関を算出する。

[0126] 具体的には、後述する図23の垂直方向の軸におけるNビートごとのビート特徴量の相関は、例えば、上述した、式(2)により算出される。

[0127] ステップS32において、自己相関マップ生成部43は、算出された各ビート間の相関に基づいて、自己相関マップを生成する。

[0128] 図23は、自己相関マップ生成部43により生成される、片側のNビートごとの自己相関マップの例を示す図である。

- [0129] 図23で示される自己相関マップの例は、図15と同様に、水平方向の軸は、ある楽曲(の1曲中)の開始から終了までのビートを表し、垂直方向の軸は、水平方向の軸で示した楽曲と同一の楽曲(の1曲中)の開始から終了までのビートを表しており、水平方向の軸は、図中右に行くほど、その楽曲の終わりの方となり、垂直方向の軸は、図中下に行くほど、その楽曲の終わりの方となる。また、図中の黒で示されている領域は、水平方向の軸と垂直方向の軸により特定される楽曲の各ビートの相関が高いことを示し、図中の白で示されている領域は、水平方向の軸と垂直方向の軸により特定される楽曲の各ビートの相関が小さいことを示す。
- [0130] 図23の自己相関マップの例では、垂直方向の軸におけるNビートごとに、ビート特徴量の相関を算出しているため、相関の高い黒で示される領域は、垂直方向の軸におけるNビートごとにしか現れないことになる。逆に言えば、図23の自己相関マップの例では、あるNビートと、その次のNビートとの間では、ビート特徴量の相関を算出していないため、必ず白い領域で示されることになる。
- [0131] すなわち、図23の自己相関マップの例は、黒い点で示される箇所では、位相成分の除去されたビート特徴量同士の相関が高いことになる。
- [0132] なお、上述したように、図23の自己相関マップにおいては、通常、“サビ”として考えられる長さが、例えば、20ビート以上である場合、ユーザは、予め、 $N=20$ (ビート)を設定しておくことになる。
- [0133] 図22のフローチャートに戻り、ステップS33において、自己相関マップ生成部43は、生成された自己相関マップにおける、相関のある箇所から前方に相関を算出する。
- [0134] 図24は、片側のNビートごとの自己相関マップの例を示す図である。
- [0135] 図24で示される自己相関マップの例の軸は、図23の自己相関マップの例と同様であるので、その説明は省略する。
- [0136] 図24で示される自己相関マップの例は、図23の自己相関マップの例において黒い点で示される箇所から、前方にNビートの相関と、後述する後方にNビートの相関を算出して、相関の高い領域を黒で示し、相関の小さい領域を白で示したときの図となる。
- [0137] 具体的には、例えば、自己相関マップ生成部43は、図23の自己相関マップの例

において、垂直方向の軸における1番目のNビート上の黒い点で示される箇所から、図中の矢印a1で示す方向に、Nビートの相関を算出することにより、上述した、図15の自己相関マップの例と同様に、図24の自己相関マップを生成することになる。

[0138] 図22のフローチャートに戻り、ステップS34において、自己相関マップ生成部43は、生成された自己相関マップにおける、相関のある箇所から後方に相関を算出する。

[0139] 具体的には、例えば、自己相関マップ生成部43は、図23の自己相関マップの例において、垂直方向の軸における2番目のNビート上の黒い点で示される箇所から、図中の矢印a2で示す方向に、Nビートの相関を算出することにより、上述した、図15の自己相関マップの例と同様に、図24の自己相関マップを生成することになる。

[0140] すなわち、図24で示すように、自己相関マップ生成部43は、Nビートごとにビート特徴量の相関を算出し、算出した相関の高かった箇所の前後の最大Nビートを、高い相関の続く範囲でのみ相関を計算し続けることで、自己相関マップを生成する。

[0141] このように、まず、片側(例えば、図23の垂直方向の軸)において、所定のビートごとに自己相関マップ(図23の自己相関マップ)を生成し、その後、相関が高い箇所の前後のみ、相関を計算して自己相関マップ(図24の自己相関マップ)を生成することにより、さらにサビ識別の処理の負荷を低減し、高速にサビ識別の処理を行うことができる。

[0142] ところで、上述した、ステップS3の処理において、ビート特徴量除去部42は、ビート特徴量抽出部41から供給されるビート特徴量をシャッフルしてから、シャッフルされたビート特徴量から、位相成分を除去するようにしてもよい。

[0143] 図25は、ビート特徴量のシャッフルについて説明する図である。

[0144] 図25で示される例において、図中の4つの四角は、図12などと同様に、ビートごとに分割されている楽曲の波形のビート特徴量を表しており、4つの四角は、それぞれ、12個の四角で表している、C、C#、D、D#、E、F、F#、G、G#、A、A#、Bである12音分のコード(図25の例では、C、Bのコード名のみを記載している)からなる。

[0145] 例えば、ビート特徴量除去部42は、この4つの四角を、図中左から右方向に、最初のビート、2番目、3番目、4番目と称する場合、2番目と3番目とをシャッフルしてから

、シャッフルされたビート特徴量に対して、離散フーリエ変換などを施すことにより、抽出されたビート特徴量の位相成分を除去する。

- [0146] このように、ビート特徴量をシャッフルすることで、自己相関マップ(サビなどの曲構成を識別する処理)の精度を上げることができる。
- [0147] なお、上述した例においては、図25の2番目と3番目の四角をシャッフルするとして説明したが、本発明においてはそれに限らず、例えば、1番目乃至4番目の四角のそれぞれをシャッフルするようにしてもよい。
- [0148] また、例えば、自己相関マップ(サビなどの曲構成を識別する処理)の精度を上げるために、離散フーリエ変換などにより、ビート特徴量の位相成分を除去してから、各ビートの相関を算出した後に、自己相関マップの片方(例えば、図15の水平方向の軸または垂直方向の軸)のビート特徴量を11回シフトさせながら、再度相関を算出するようにしてもよい。
- [0149] 図26は、ビート特徴量のシフトについて説明する図である。
- [0150] 図26で示される例においては、図中左から1番目の四角(以下、1番目の四角と称し、同様に、2番目の四角、3番目の四角、・・・、12番目の四角と称する)は、図12などと同様に、ビートごとに分割されている楽曲の波形のビート特徴量を表しており、さらにその四角は、その四角の中の12個の四角で表している、C, C#, D, D#, E, F, F#, G, G#, A, A#, Bである12音分のコードからなる。
- [0151] 例えば、まず、ビート特徴量除去部42は、1番目の四角で表すビート特徴量に対して、離散フーリエ変換などを施すことにより、ビート特徴量の位相成分を除去する。自己相関マップ生成部43は、位相成分が除去された1番目の四角を基に、ビート特徴量の相関を算出する。
- [0152] 次に、1番目の四角から、12音分のコードを図中下方向に1つシフトさせることにより、ビート特徴量は、2番目の四角で表すように、図中下から、C#, D, D#, E, F, F#, G, G#, A, A#, B, Cの順にシフトすることになる。
- [0153] このとき、例えば、ビート特徴量除去部42は、2番目の四角で表すビート特徴量に対して、離散フーリエ変換などを施すことにより、ビート特徴量の位相成分を除去する。自己相関マップ生成部43は、位相成分が除去された2番目の四角を基に、ビート

特徴量の相関を算出する。

- [0154] 同様に、ビート特徴量の12音分のコードを1つずつ順に下方向にシフトさせることにより、例えば、ビート特徴量除去部42は、3番目乃至11番目の四角(図示せず)で表すビート特徴量に対して、離散フーリエ変換などを施すことにより、ビート特徴量の位相成分を除去し、自己相関マップ生成部43は、位相成分が除去された3番目乃至11番目の四角を基に、ビート特徴量の相関を算出する。
- [0155] そして最後に、11番目の四角から、12音分のコードを図中下方向に1つシフトさせることにより、12番目の四角(図中右から1番目の四角)で表すように、図中下から、B, C, C#, D, D#, E, F, F#, G, G#, A, A#の順にシフトすることになる。
- [0156] このとき、例えば、ビート特徴量除去部42は、12番目の四角で表すビート特徴量に対して、離散フーリエ変換などを施すことにより、ビート特徴量の位相成分を除去する。自己相関マップ生成部43は、位相成分が除去された12番目の四角を基に、ビート特徴量の相関を算出する。
- [0157] このように、ビート特徴量を11回シフトさせながら、相関を算出することにより、自己相関マップ生成部43は、より精度の高い自己相関マップを生成することができる。
- [0158] なお、図26では、説明を分かり易くするために、1区間(1ビート)のビート特徴量のシフトについて説明したが、例えば、4区間(4ビート)などであっても、同様にビート特徴量をシフトさせることにより、自己相関マップの精度をさらに上げることができる。
- [0159] ところで、本実施の形態では、ビートを検出する処理(図4のステップS1の処理)において、楽曲の波形(音声信号)から検出されるビート(テンポ、拍子、小節の先頭など)の検出精度を上げることで、例えば、セグメンテーションの処理(図4のステップS5の処理)を行う場合、分割される各区間の境界は小節の先頭としたほうがいいので、そのような小節の先頭を正確に求めることで、正確にセグメンテーションを行うことができる。そこで、次に、図27乃至図40を参照して、楽曲の波形からテンポ、拍子、小節の先頭(以下、ビートデータと称する)を検出する処理について説明する。
- [0160] 図27は、図2のパーソナルコンピュータ1のCPU11により実行されるソフトウェアの機能的構成の他の例を示すブロック図である。
- [0161] なお、図27では、図3と同様の箇所には、同一の符号が付してあり、処理が同じ部

分に関しては、その説明は繰り返しになるので省略する。すなわち、図27において、楽曲構造解析部31は、図3の楽曲構造解析部31と比較して、ビート検出部101が追加して設けられている。

- [0162] ビート検出部101は、楽曲の波形からビートデータを検出する。例えば、ビート検出部101は、MP3やAACなどの方式で音声圧縮された音声信号より再生される楽曲の波形から、テンポ、拍子、または小節の先頭などのビートデータを検出する。
- [0163] 次に、図28に示されるように、楽曲の波形から、ビート、すなわち、拍のそれぞれの位置を検出するビート検出部101について説明する。図28において、“1:1, 1:2, 1:3, 1:4, 2:1, 2:2, 2:3, 2:4, 3:1, 3:2, 3:3, 3:4, …” (この「1:1」などの数は、図中右下に記すように、「小節:拍」を表現している)のそれぞれの数に対する縦線は、楽曲の波形における拍の先頭のそれぞれの位置を示す。図28において、それぞれの数に対する縦線で示される位置から、その次の縦線の位置までの範囲は、楽曲の波形における拍の範囲を示す。
- [0164] なお、隣接する2つの縦線の間で示される長さは、例えば、4分音符の長さを示し、テンポに相当する。また、“1:1”, “2:1”, “3:1”, …である数に対する縦線で示される位置は、小節の先頭を示す。
- [0165] つまり、ビート検出部101は、図28に示すような音声信号による楽曲の波形から、図中の縦線で示す各拍や各小節の先頭、または図中隣接する2つの縦線の間で示される長さであるテンポなどを、ビートデータとして検出する。
- [0166] 図29は、ビート検出部101の構成の一例を示すブロック図である。ビート検出部101は、アタック情報抽出部111、基本ビート周期検出部112、テンポ決定部113、楽曲特徴量抽出部114、テンポ修正部115、ビート特徴量抽出部116、拍子決定部117、および小節先頭決定部118を含むようにして構成される。
- [0167] アタック情報抽出部111は、楽曲の波形を示す音声信号から時系列のアタック情報を抽出する。ここで、時系列のアタック情報とは、人間にビートを感じさせる音量の変化を時間に沿ってデータ化したものである。図30で示されるように、アタック情報は、時間に対する、人間の感じる音量を示す音量感で表される。
- [0168] 例えば、アタック情報抽出部111は、それぞれの時刻における、音声信号による音

声のレベルを示すアタック情報を音声信号から抽出する。

- [0169] また、例えば、図31で示されるように、アタック情報抽出部111は、音声信号による音声を複数のオクターブの成分に分けて、さらにそれぞれのオクターブにおける12平均率のそれぞれの高さの12の音のエネルギーを求めることで、オクターブごとの12の音のそれぞれのエネルギーを示す12音解析による時間－音程データを求める。アタック情報抽出部111は、各時刻における、12の音と複数のオクターブとについて音のエネルギーを積算し、その結果をアタック情報とする。
- [0170] また、例えば、アタック情報抽出部111は、音声信号による音声を複数のオクターブの成分に分けて、さらにそれぞれのオクターブにおける12平均率のそれぞれの高さの12の音の鳴り始めのタイミングを検出する。例えば、アタック情報抽出部111は、それぞれの音のエネルギーの時間方向の差分が予め定めた閾値より大きい場合、その音の鳴り始めであるとする。
- [0171] そして、アタック情報抽出部111は、音の鳴り始めを1とし、それ以外を0として、その値を複数のオクターブのそれぞれの12の音について積算し、その結果をアタック情報とする。
- [0172] 図31において、丸は、音の鳴り始めの位置を示す。音の鳴り始めを1とし、それ以外を0として、これを積算してアタック情報を求めた場合、アタック情報は、複数のオクターブのそれぞれの12の音において、鳴り始めが多い場合には、大きい値となり、鳴り始めが少ない場合には、小さい値となる。
- [0173] さらに、例えば、アタック情報抽出部111は、音声信号による音声を複数のオクターブの成分に分けて、さらにそれぞれのオクターブにおける12平均率のそれぞれの高さの12の音のエネルギーの変化を求める。例えば、音のエネルギーの変化は、音のエネルギーの時間方向の差分として計算される。アタック情報抽出部111は、各時刻における、音のエネルギーの変化を、複数のオクターブのそれぞれの12の音について積算し、その結果をアタック情報とする。
- [0174] 図29に戻り、アタック情報抽出部111は、このようなアタック情報を、基本ビート周期検出部112、テンポ修正部115、および拍子決定部117に供給する。
- [0175] 基本ビート周期検出部112は、コードの検出の対象となっている楽曲において、最

も基本となっている音の長さを検出する。例えば、楽曲において最も基本となっている音は、4分音符、8分音符、または16分音符で表される音である。

[0176] 以下、楽曲において、最も基本となる音の長さを基本ビート周期と称する。

[0177] 基本ビート周期検出部112において、時系列の情報であるアタック情報を通常の波形と見立てて基本ピッチ(音程)抽出を行うことで基本ビート周期が求められる。

[0178] 例えば、図32で示されるように、基本ビート周期検出部112は、時系列の情報であるアタック情報をショートタイムフーリエ変換(STFT(Short-Time Fourier Transform))する。アタック情報をショートタイムフーリエ変換することにより、周波数におけるエネルギーの強さを時系列に示す結果が得られる。

[0179] すなわち、基本ビート周期検出部112において、アタック情報の時間的な長さに対して十分に短い期間である窓の位置をアタック情報に対してずらしながら、その窓におけるアタック情報の部分をフーリエ変換し、そのフーリエ変換の結果を時系列に配置することで、それぞれの周波数におけるエネルギーの強さを時系列に示す結果が求められる。

[0180] ショートタイムフーリエ変換の結果、他の周波数のエネルギーに比較して強いエネルギーの周波数が、基本ビート周期の候補である周期として検出される。図32の下側において、濃度は、エネルギーの強さを示す。

[0181] 基本ビート周期検出部112は、アタック情報のショートタイムフーリエ変換の結果検出された周期のうち最も優位のものを基本ビート周期とする。

[0182] すなわち、基本ビート周期検出部112は、アタック情報のショートタイムフーリエ変換の結果検出された周期のうちから、予め用意した重みである基本ビートらしさと、アタック情報のショートタイムフーリエ変換の結果とを照らし合わせて、基本ビートらしさの高い周期を基本ビート周期とする。

[0183] より具体的には、基本ビート周期検出部112は、予め定めた周波数方向の重みである基本ビートらしさで、アタック情報のショートタイムフーリエ変換の結果得られた周波数ごとのエネルギーを重み付けして、重み付けにより得られた値のうち最も大きい値が得られた周期を基本ビート周期とする。

[0184] 周波数方向の重みである基本ビートらしさを用いることにより、基本ビート周期とは

なり得ない極めて低い周波数の周期や極めて高い周波数の周期が、基本ビート周期とされることを防止することができる。

[0185] 図29に戻り、基本ビート周期検出部112は、このように抽出した基本ビート周期をテンポ決定部113に供給する。

[0186] 楽曲特徴量抽出部114は、音声信号に所定の信号処理を適用することにより、楽曲から所定の数の特徴量(以下、楽曲特徴量と称する)を抽出する。例えば、楽曲特徴量抽出部114は、音声信号を複数のオクターブの成分に分けて、さらにそれぞれのオクターブにおける12平均率のそれぞれの高さの12の音の信号を求めて、それぞれのオクターブにおけるそれぞれの12の音の信号に所定の信号処理を適用することにより、楽曲特徴量を抽出する。

[0187] 例えば、楽曲特徴量抽出部114は、楽曲特徴量として、それぞれのオクターブにおけるそれぞれの12の音の信号の、単位時間当たりのピークの数を求める。

[0188] また、例えば、楽曲特徴量抽出部114は、楽曲特徴量として、それぞれのオクターブにおけるそれぞれの12の音の信号の、音程方向のエネルギーの分散を求める。

[0189] さらにまた、例えば、楽曲特徴量抽出部114は、楽曲特徴量として、それぞれのオクターブにおけるそれぞれの12の音の信号から、低域、中域、および高域のエネルギーのバランスを求める。

[0190] また、例えば、楽曲特徴量抽出部114は、楽曲特徴量として、それぞれのオクターブにおけるそれぞれの12の音の信号から、ステレオである音声信号の左右のチャンネルの信号の相関の大きさなどを求める。

[0191] 楽曲特徴量抽出部114は、このように抽出した楽曲特徴量をテンポ決定部113に供給する。

[0192] テンポ決定部113は、楽曲特徴量とテンポとによる学習によって予め構成され、楽曲特徴量抽出部114から供給された楽曲特徴量から、テンポを推定する。以下、推定されたテンポを推定テンポと称する。

[0193] テンポ決定部113は、推定テンポと、基本ビート周期検出部112から供給された基本ビート周期とから、基本ビート周期の2のX乗倍($\dots 1/8$ 倍、 $1/4$ 倍、 $1/2$ 倍、1倍、2倍、4倍、8倍 \dots)の中からテンポを決定する。例えば、楽曲の特徴量から回帰

解析で推定された推定テンポに対して、推定テンポ $\times 21/2$ と推定テンポ $\div 21/2$ との範囲に収まるように、基本ビート周期を2倍するかまたは $1/2$ 倍して得られた値がテンポとされる。

[0194] 例えば、図33で示されるように、テンポ決定部113は、基本ビート周期検出部112から供給された基本ビート周期と、推定テンポ $\div 21/2$ で求められる周期とを比較して、推定テンポ $\div 21/2$ で求められる周期より基本ビート周期(図33の上側の白丸で示される基本ビート周期)が長い場合、基本ビート周期を $1/2$ 倍する。

[0195] テンポ決定部113は、基本ビート周期検出部112から供給された基本ビート周期と、推定テンポ $\times 21/2$ で求められる周期とを比較して、推定テンポ $\times 21/2$ で求められる周期より基本ビート周期(図33の下側の白丸で示される基本ビート周期)が短い場合、基本ビート周期を2倍する。

[0196] テンポ決定部113は、推定テンポ $\times 21/2$ と推定テンポ $\div 21/2$ との範囲に収まるように、 $1/2$ 倍したか若しくは2倍したか、または繰り返し $1/2$ 倍したか若しくは繰り返し2倍した基本ビート周期(図33の黒丸で示される基本ビート周期)をテンポとする。

[0197] なお、テンポ決定部113は、基本ビート周期が推定テンポ $\times 21/2$ と推定テンポ $\div 21/2$ との範囲に収まる場合、そのままの基本ビート周期をテンポとする。

[0198] 図29に戻り、テンポ決定部113は、このように決定したテンポをテンポ修正部115に供給する。

[0199] テンポ修正部115は、アタック情報抽出部111から供給されるアタック情報により、テンポ決定部113において決定されたテンポをより細かく修正する。

[0200] まず、テンポ修正部115は、ビートの位相を修正する。

[0201] すなわち、図34で示されるように、テンポ修正部115は、アタック情報に対して決定されたテンポの周期で、拍の範囲ごとに、楽曲全体にわたりアタック情報を足し合わせる。

[0202] 例えば、テンポ修正部115は、楽曲全体にわたり、テンポの周期で決められる拍の範囲について、最初の拍の範囲乃至最後の拍の範囲のそれぞれにおける、最初のアタック情報のサンプルを足し合わせて、その結果を、拍の範囲の最初の合計値とす

る。次に、テンポ修正部115は、楽曲全体にわたり、テンポの周期で決められる拍の範囲について、最初の拍の範囲乃至最後の拍の範囲のそれぞれにおける、2番目のアタック情報のサンプルを足し合わせて、その結果を、拍の範囲の2番目の合計値とする。

[0203] 同様に、テンポ修正部115は、楽曲全体にわたり、テンポの周期で決められる拍の範囲について、最初の拍の範囲乃至最後の拍の範囲のそれぞれにおける、3番目乃至最後のアタック情報のサンプルを、拍の範囲における順ごとに足し合わせて、その結果のそれぞれを、拍の範囲の3番目乃至最後の合計値とする。

[0204] そして、テンポ修正部115は、アタック情報に対するテンポの周期の位相をずらして、同様に、拍の範囲ごとに、楽曲全体にわたりアタック情報を足し合わせる。

[0205] テンポ修正部115は、アタック情報に対するテンポの周期の位相を、アタック情報に対するテンポの周期の位相をずらして得られた合計値のうち、最も大きい合計値の得られた位相に修正する。すなわち、テンポ修正部115は、拍の位置を、最も大きい合計値の得られた、アタック情報に対するテンポの周期の位置に修正する。

[0206] また、テンポ修正部115は、テンポを修正する。

[0207] すなわち、図35で示されるように、テンポ修正部115は、テンポの周期を、その周期に対して十分短い所定の長さだけ縮めるか、または伸ばして、その縮められるかまたは伸ばされたテンポの周期で、テンポの周期ごとに、楽曲全体にわたりアタック情報を足し合わせる。

[0208] この場合も、テンポ修正部115は、楽曲全体にわたり、テンポの周期で決められる拍の範囲について、最初の拍の範囲乃至最後の拍の範囲のそれぞれにおける、最初乃至最後のアタック情報のサンプルを、拍の範囲における順ごとに足し合わせて、その結果のそれぞれを、拍の範囲の最初乃至最後の合計値とする。

[0209] テンポ修正部115は、テンポの周期を所定の長さだけ縮めるか、または伸ばして、その縮められるかまたは伸ばされたテンポの周期ごとに、楽曲全体にわたりアタック情報を足し合わせて、拍の範囲の最初乃至最後の合計値を求める。

[0210] テンポ修正部115は、元の長さか、縮められるか、または伸ばされたテンポの周期のうち、最も大きい合計値の得られた長さにテンポの周期を修正する。

- [0211] テンポ修正部115は、このようなビートの位相の修正とテンポの修正とを必要に応じて繰り返して、最終的なテンポを決定する。例えば、テンポ修正部115は、ビートの位相の修正と、テンポの修正とを、それぞれ2回など、予め決められた回数だけ繰り返して、最終的なテンポを決定する。
- [0212] 図29に戻り、テンポ修正部115は、最終的に決定されたテンポを示す拍データを、ビート特徴量抽出部116および小節先頭決定部118に供給する。
- [0213] ビート特徴量抽出部116は、テンポ修正部115から供給される拍データに基づいて、セグメント(1拍)ごとに、音楽特徴量やコードなどの楽曲特徴量を抽出する。
- [0214] 例えば、ビート特徴量抽出部116は、図36に示される楽曲の波形から、図中縦線で分けられているビートごと(図36に示す例では12区間に分割されている)に、楽曲特徴量を抽出する。
- [0215] なお、ビート特徴量抽出部116により抽出される楽曲特徴量は、例えば、上述した、楽曲特徴量抽出部114により抽出される楽曲特徴量と同様に、12音解析の結果により得られる信号に所定の信号処理を適用することで抽出される。
- [0216] 図29に戻り、ビート特徴量抽出部116は、このように抽出したビートごとの楽曲特徴量を小節先頭決定部118に供給する。
- [0217] 拍子決定部117は、アタック情報抽出部111から供給されるアタック情報に基づいて、例えば、4/4拍子、3/4拍子、6/8拍子などの拍子を決定する。
- [0218] ここで、拍子の決定方法であるが、例えば、ショートタイムフーリエ変換を施したアタック情報の倍音構造から拍子を求めることができる。例えば、図37で示されるように、拍子決定部117は、基本ビート周期検出部112と同様に、時系列の情報であるアタック情報をショートタイムフーリエ変換する。これにより、周波数におけるエネルギーの強さを時系列に示す結果が得られる。
- [0219] つまり、図37の例に示すように、アタック情報にショートタイムフーリエ変換を施すことで得られる結果により、ピークが現れた周波数が検出されるので、その周波数を利用して拍子を決定する。例えば、図中左側に示している4/4拍子において、8分音符、4分音符、2分音符、全音符のそれぞれの関係は、図中丸印で示すように、8分音符の周波数の半分が4分音符の周波数となり、図中三角印で示すように、4分音符

の周波数の半分が2分音符の周波数となり、図中バツ印で示すように、2分音符の周波数の半分が全音符の周波数となる。言い換えれば、4分音符の周期は、8分音符の周期の2倍であり、2分音符の周期は、4分音符の周期の2倍であり、全音符の周期は、2分音符の2倍であるとも言える。

[0220] また、同様に、例えば、図中中央に示している3/4拍子において、4分音符、1小節、2小節のそれぞれの関係は、図中丸印で示すように、4分音符の周波数の1/3が1小節の周波数となり、図中バツ印で示すように、1小節の周波数の半分が2小節の周波数となる。

[0221] さらにまた、同様に、図中右側に示している6/8拍子において、8分音符、4分音符、付点4分音符、1小節のそれぞれの関係は、図中丸印で示すように、8分音符の周波数の半分が4分音符の周波数となり、図中バツ印で示すように、4分音符の周波数の2/3が付点4分音符の周波数となり、4分音符の周波数の1/3が1小節の周波数となる。

[0222] つまり、拍子決定部117は、ピークが現れた周波数におけるエネルギーの間隔のパターン、すなわち、ショートタイムフーリエ変換を施したアタック情報の倍音構造から、拍子を決定する。

[0223] 図38は、拍子決定部117による、具体的な拍子の決定方法の例を説明する図である。

[0224] 図38においては、図中左側に示すような、横軸方向を時間とし、縦軸方向を周波数とするグラフ(以下、時間一周波数グラフと称する)、時間一周波数グラフの右側に示すような、横軸方向をエネルギーとし、縦軸方向を周波数とするグラフ(以下、エネルギー一周波数グラフと称する)、エネルギー一周波数グラフの右側に示すような、横軸方向をエネルギーとし、縦軸方向をログ周波数とするグラフ(以下、エネルギーログ周波数グラフと称する)の3つのグラフが示されている。

[0225] 図38において、まず、拍子決定部117は、アタック情報にショートタイムフーリエ変換を施すことにより得られる、時間一周波数グラフにおける周波数を時間方向に足し合わせることで、エネルギー一周波数グラフを得る。次に、拍子決定部117は、エネルギー一周波数グラフにおける周波数軸をログ周波数軸に変換して、エネルギー

ログ周波数グラフを得る。

- [0226] そして、拍子決定部117は、エネルギー—ログ周波数グラフに示されるログ周波数ごとのエネルギーに対して、高速フーリエ変換(FFT(Fast Fourier Transform))を施すことにより得られるパワー成分である特徴量と、予め用意した拍子のパターンとを照らし合わせて、拍子を決定する。
- [0227] 図29に戻り、拍子決定部117は、このように決定した拍子を小節先頭決定部118に供給する。
- [0228] 小節先頭決定部118には、テンポ修正部115から拍データが供給され、ビート特徴量抽出部116からビートごとの楽曲特徴量が供給され、拍子決定部117から拍子が供給される。小節先頭決定部118は、それらの拍データ、ビートごとの特徴量、および拍子に基づいて、小節の先頭を決定する。
- [0229] ここで、小節の先頭の決定方法であるが、例えば、小節の先頭では、その他の部分と比較して、コードが変化し易い、音楽的に安定したコード(例えば、いわゆるスリーコードなど)の出現確率が高い、音程がはっきり安定している、音量的にアクセントがあるなどといった特徴がある。
- [0230] 小節先頭決定部118は、これらの小節の先頭の特徴を利用して、各ビートについて小節の先頭らしさを推定する。なお、小節先頭決定部118は、例えば、各ビートの特徴量やコードなどによる学習によって予め構成され、拍データ、ビートごとの楽曲特徴量、および拍子から、小節の先頭らしさを推定する。
- [0231] 次に、図39の例に示されるように、小節先頭決定部118は、求めた小節の先頭らしさを、楽曲全体または特定の範囲において拍子の1小節あたりの音符の数ごと(例えば、4/4拍子なら4つごと、3/4拍子なら3つごと、6/8拍子なら6つごとなど)に足し合わせる。つまり、図39の例では、4/4拍子であるので4つごとに足し合わされる。より具体的には、1つ目のビート、5つ目のビート、9つ目のビート、…(以下、1/4拍目と称する)の小節の先頭らしさが足し合わされ、2つ目のビート、6つ目のビート、10個目のビート、…(以下、2/4拍目と称する)の小節の先頭らしさが足し合わされ、3つ目のビート、7つ目のビート、11個目のビート、…(以下、3/4拍目と称する)の小節の先頭らしさが足し合わされ、4つ目のビート、8つ目のビート、12個目のビート、・

・・(以下、4/4拍目と称する)の小節の先頭らしさが足し合わされる。

[0232] そして、図40の例に示されるように、小節先頭決定部118は、4/4拍子であるので4つごとに足し合わされた小節の先頭らしさを比較して、最も小節の先頭らしさが高くなった箇所のビートを小節の先頭であると決定する。より具体的には、図40の例では、1/4拍目の小節の先頭らしさの値を足し合わせた合計の値が35.6、2/4拍目の小節の先頭らしさの値を足し合わせた合計の値が12.6、3/4拍目の小節の先頭らしさの値を足し合わせた合計の値が21.5、4/4拍目の小節の先頭らしさの値を足し合わせた合計の値が13.3となり、1/4拍目の合計の値が最も先頭らしさの値が高いため、小節先頭決定部118は、1/4拍目を小節の先頭であると決定する。

[0233] このように、ビート検出部101は、楽曲の波形(音声信号)から、テンポ、拍子、小節の先頭などのビートデータを検出し、出力する。つまり、ビート検出部101は、例えば、ビートを検出する処理(図4のステップS1の処理)で、上述した方法によりビートデータを求めることができるので、より正確にビートを検出することが可能となる。その結果、例えば、セグメンテーションの処理(図4のステップS5の処理)において、正確にセグメンテーションを行うことができ、ひいてはサビを識別する処理(図4のステップS7の処理)において、サビである区間をより正確に識別することが可能となる。

[0234] ここで、従来の手法と本発明の手法のそれぞれにより、転調のある楽曲の波形を解析した場合の演算量(乗加算回数)を比較する。なお、比較するときの条件を同じにするために、例えば、楽曲の長さを5分、テンポを120、相関計算に用いるビートの数を8ビートとしてそれぞれの演算量を比較する。

[0235] まず、従来の手法により転調のある楽曲の波形を解析した場合、相関を算出するための平均とノルムを算出するときの演算量a1は、以下の通りになる。

[0236] 演算量a1=(ビートの数)×(8ビート分)×(12音)×(2(平均とノルム))=5×120×8×12×2=115, 200(回)

[0237] また、相関を算出する場合の演算量a2は、以下の通りとなる。

[0238] 演算量a2=(ビートの数)×(ビートの数/2+1)×(8ビート分)×(12音)×(12回シフト)=5×120×299×8×12×12=206, 668, 800(回)

[0239] すなわち、従来の手法の演算量Aは、以下の通りとなる。

[0240] 演算量 $A = \text{演算量}a_1 + \text{演算量}a_2 = 115,200 + 206,668,800 = 206,784,000$ (回)

[0241] 次に、本発明の手法により転調のある楽曲の波形を解析した場合、離散フーリエ変換により位相成分を除去して、相関を算出するための平均とノルムを算出するときの演算量 b_1 は、以下の通りになる。

[0242] 演算量 $b_1 = (\text{ビートの数}) \times ((8\text{ビート分}) \times (12\text{音}))^2$ (離散フーリエ変換の計算) $+ (\text{ビートの数}) \times (\text{離散フーリエ変換後の特徴量}) \times (2(\text{平均とノルム})) = 5 \times 120 \times (8 \times 12)^2 + 5 \times 120 \times 47 \times 2 = 5,529,600 + 56,400 = 5,586,000$ (回)

[0243] また、相関を算出する場合の演算量 b_2 は、以下の通りとなる。

[0244] 演算量 $b_2 = (\text{ビートの数}) \times (\text{ビートの数} / 2 + 1) \times (\text{離散フーリエ変換後の特徴量}) = 5 \times 120 \times 299 \times 47 = 8,431,800$ (回)

[0245] すなわち、本発明の手法の演算量 B は、以下の通りとなる。

[0246] 演算量 $B = \text{演算量}b_1 + \text{演算量}b_2 = 5,586,000 + 8,431,800 = 14,017,800$ (回)

[0247] したがって、従来手法の演算量 A と本発明の手法の演算量 B とを比較した場合、206,784,000 (回)と14,017,800 (回)となるので、本発明の手法の演算量 B は、従来手法の演算量 A の1/14以下の演算量(乗加算回数)、実際には、離散フーリエ変換に高速なアルゴリズムを用いることにより、 N^2 以下の乗加算回数により計算をすることができるので、より計算回数が少なくなり、より迅速に、自己相関マップを生成することができる。その結果、より迅速に、例えば、楽曲構造やサビの識別などの処理を行うことができる。

[0248] 以上のように、本発明によれば、少ない演算量で、転調のある楽曲の波形を解析することができる。

[0249] また、本発明によれば、より少ない演算量で、転調のある楽曲の波形から、その楽曲の曲構成を識別することができる。

[0250] 上述した一連の処理は、ハードウェアにより実行させることもできるが、ソフトウェアにより実行させることもできる。一連の処理をソフトウェアにより実行させる場合には、そのソフトウェアを構成するプログラムが、専用のハードウェアに組み込まれているコンピュータ、または、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどに、記録媒体からインス

トールされる。

- [0251] この記録媒体は、コンピュータとは別に、ユーザにプログラムを提供するために配布される、プログラムが記録されている磁気ディスク(フレキシブルディスクを含む)、光ディスク(CD-ROM(Compact Disc-Read Only Memory), DVD(Digital Versatile Disk))を含む)、光磁気ディスク(MD(Mini-Disk)(登録商標)を含む)、若しくは半導体メモリなどよりなる図2のリムーバブルメディア21により構成されるだけでなく、コンピュータに予め組み込まれた状態でユーザに提供される、プログラムが記録されている図2のROM12や記録部18などで構成される。
- [0252] また、上述した一連の処理を実行させるプログラムは、必要に応じてルータ、モデムなどのインターフェースを介して、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の通信媒体を介してコンピュータにインストールされるようにしてもよい。
- [0253] なお、本明細書において、記録媒体に格納されるプログラムを記述するステップは、記載された順序に沿って時系列的に行われる処理はもちろん、必ずしも時系列的に処理されなくとも、並列的あるいは個別に実行される処理をも含むものである。
- [0254] また、本発明の実施の形態は、上述した実施の形態に限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能である。

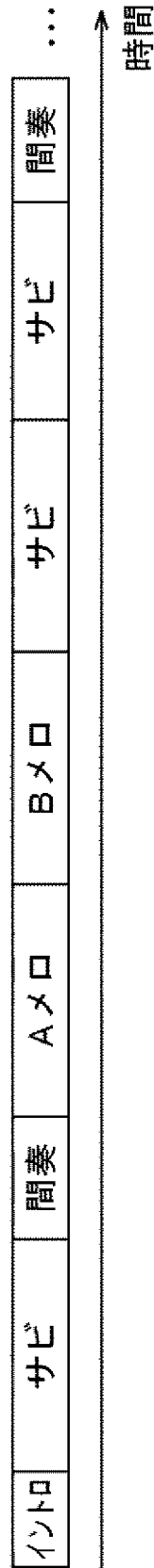
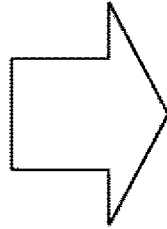
請求の範囲

- [1] 転調のある楽曲の波形をビートごとに区切った第1の区間における、前記ビートに関係する第1の特徴量を抽出する第1の抽出手段と、
抽出された前記第1の特徴量の周波数成分から位相成分を除去する除去手段と、
位相成分が除去された前記第1の特徴量に基づいて、ある第1の区間と、他の第1の区間との相関を示す自己相関情報を生成する生成手段と
を備える情報処理装置。
- [2] 前記自己相関情報に基づいて、前記ビートを、複数の相関性の高いビートが繰り返されている第2の区間ごとに分割する分割手段と、
分割された前記第2の区間における、前記楽曲の音楽的な特徴を示す第2の特徴量を抽出する第2の抽出手段と、
前記第2の特徴量に基づいて、分割された前記第2の区間における前記楽曲の曲構成を識別する識別手段と
を備える請求項1の情報処理装置。
- [3] 前記除去手段は、離散フーリエ変換により、前記第1の特徴量の周波数成分から位相成分を除去する
請求項1の情報処理装置。
- [4] 前記識別手段は、前記曲構成として、サビである区間を識別する
請求項2の情報処理装置。
- [5] 前記第1の特徴量は、前記第1の区間における、所定のオクターブ分の各音のエネルギーである
請求項1の情報処理装置。
- [6] 前記第2の特徴量は、前記第2の区間における、音量、音源に関する情報、またはある周波数帯の利得である
請求項2の情報処理装置。
- [7] 前記第2の特徴量は、前記第2の区間の長さ、前記第2の区間が出現した回数、または前記第2の区間が連続して繰り返された回数である
請求項6の情報処理装置。

- [8] 前記生成手段は、位相成分が除去された前記第1の特徴量に基づいて、所定のビートごとに、ある第1の区間との相関を算出することにより、前記自己相関情報を生成する
- 請求項4の情報処理装置。
- [9] 前記除去手段は、シャッフルされた前記第1の特徴量の周波数成分から位相成分を除去する
- 請求項1の情報処理装置。
- [10] 前記除去手段は、前記各音のエネルギーをシフトさせてから、前記第1の特徴量の周波数成分から位相成分を除去する
- 請求項5の情報処理装置。
- [11] 転調のある楽曲の波形をビートごとに区切った区間における、前記ビートに関する特徴量を抽出する抽出ステップと、
- 抽出された前記特徴量の周波数成分から位相成分を除去する除去ステップと、
- 位相成分が除去された前記特徴量に基づいて、ある区間と、他の区間との相関を示す自己相関情報を生成する生成ステップと
- を含む情報処理方法。
- [12] 転調のある楽曲の波形をビートごとに区切った区間における、前記ビートに関する特徴量を抽出する抽出ステップと、
- 抽出された前記特徴量の周波数成分から位相成分を除去する除去ステップと、
- 位相成分が除去された前記特徴量に基づいて、ある区間と、他の区間との相関を示す自己相関情報を生成する生成ステップと
- をコンピュータに実行させるプログラム。

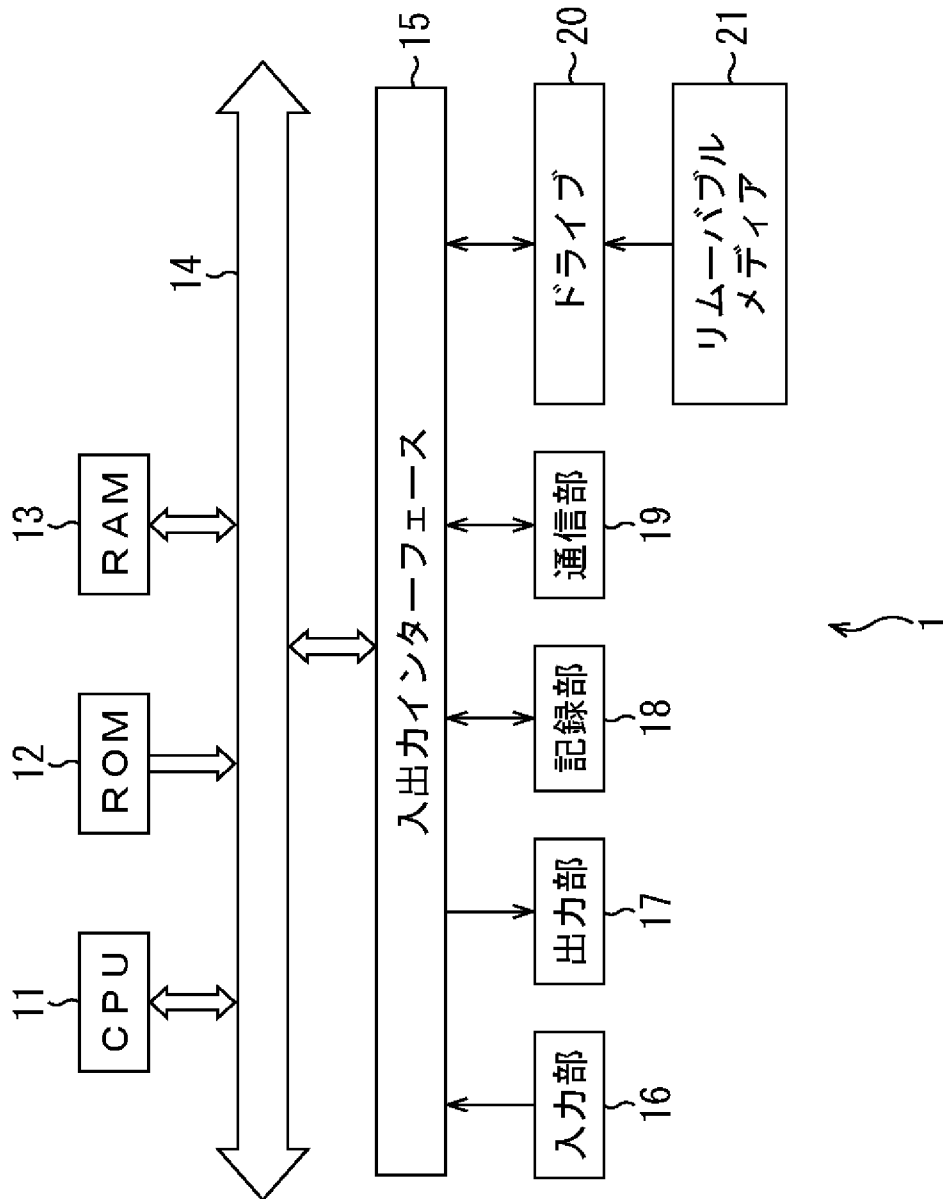
[図1]

図1

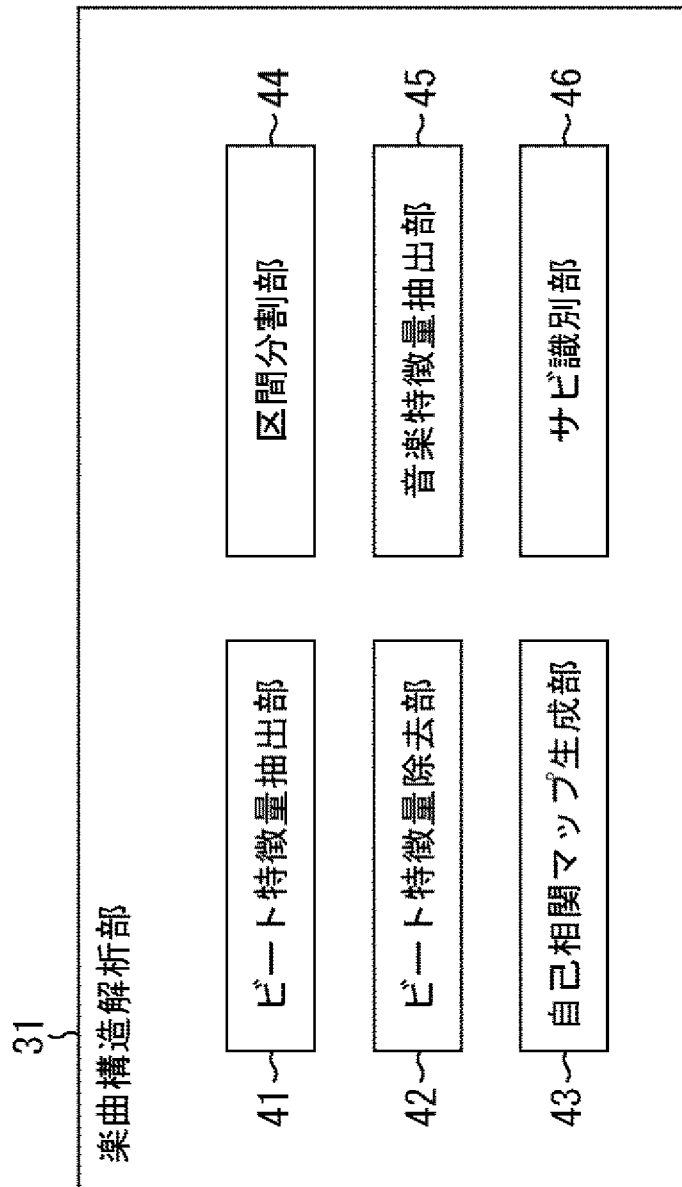


[図2]

図2

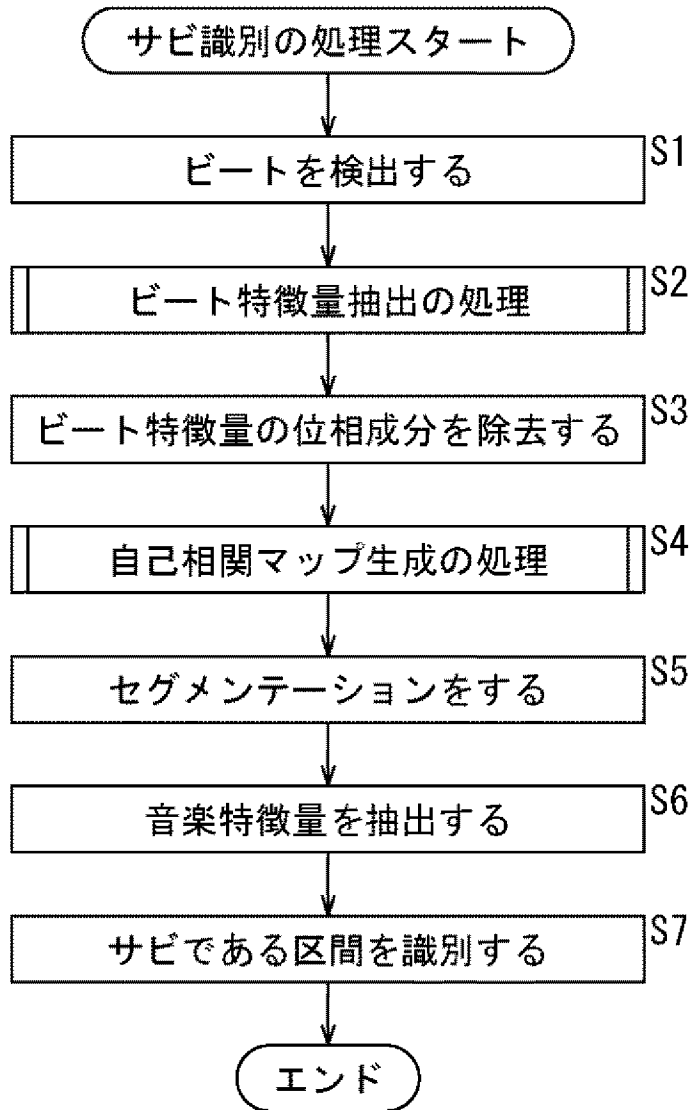


[図3]
図3

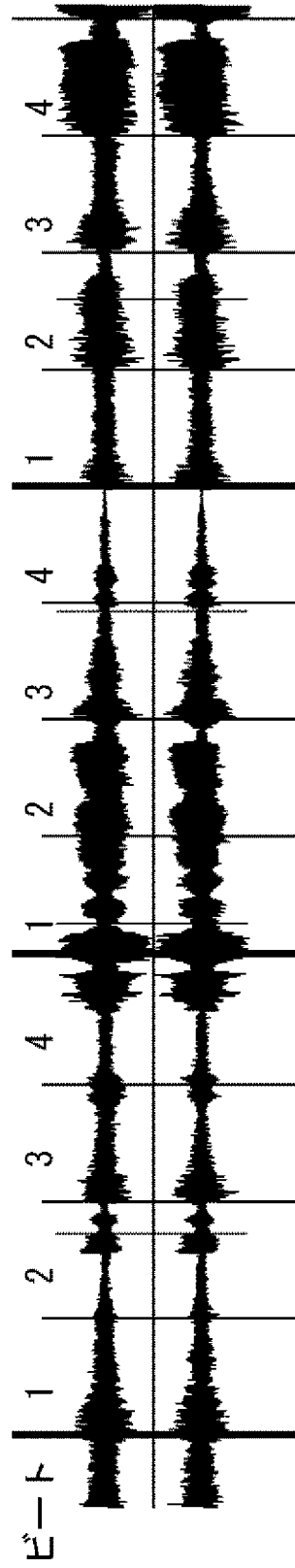
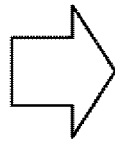
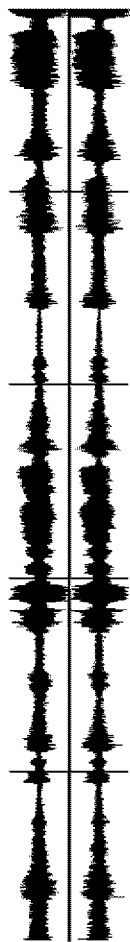


[図4]

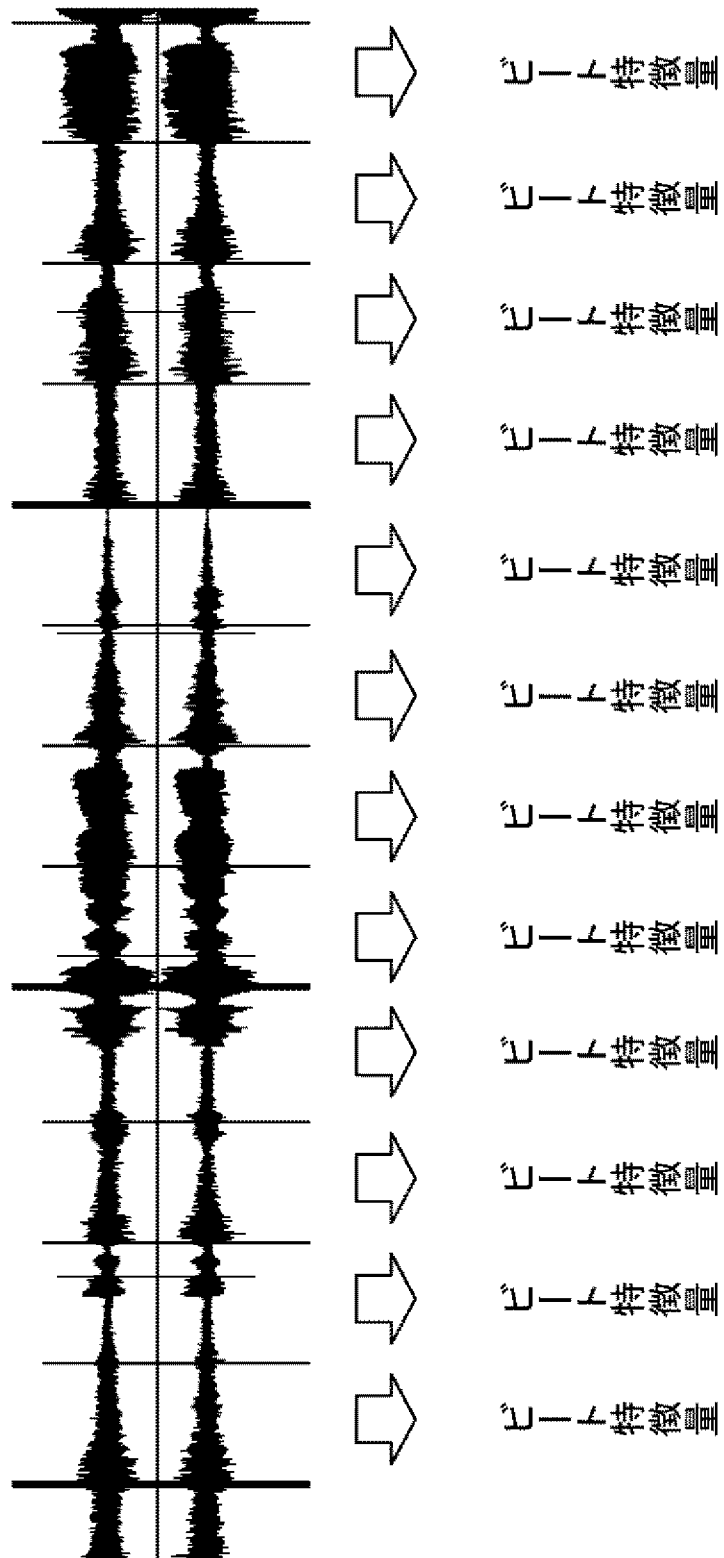
図4



[図5]
図5

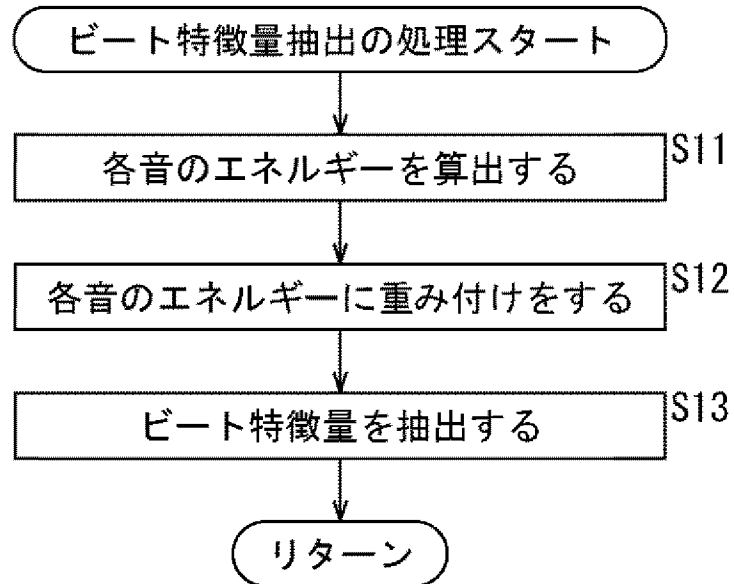


[図6]
図6



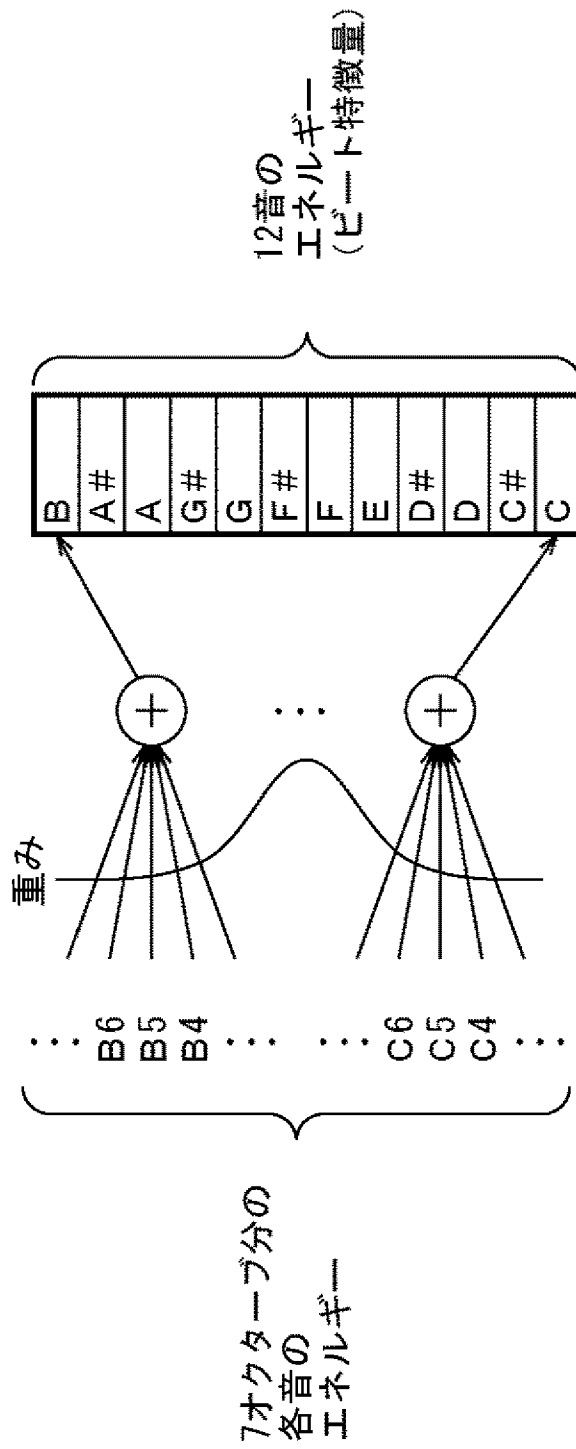
[図7]

図7



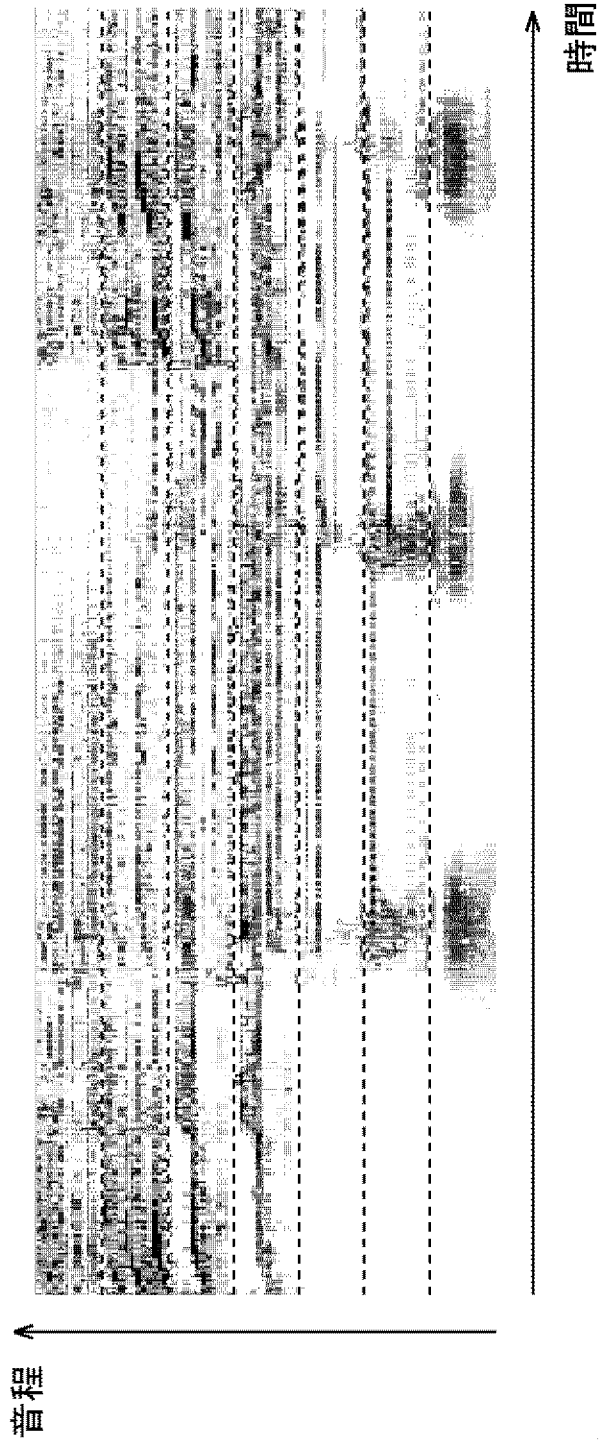
[図8]

図8



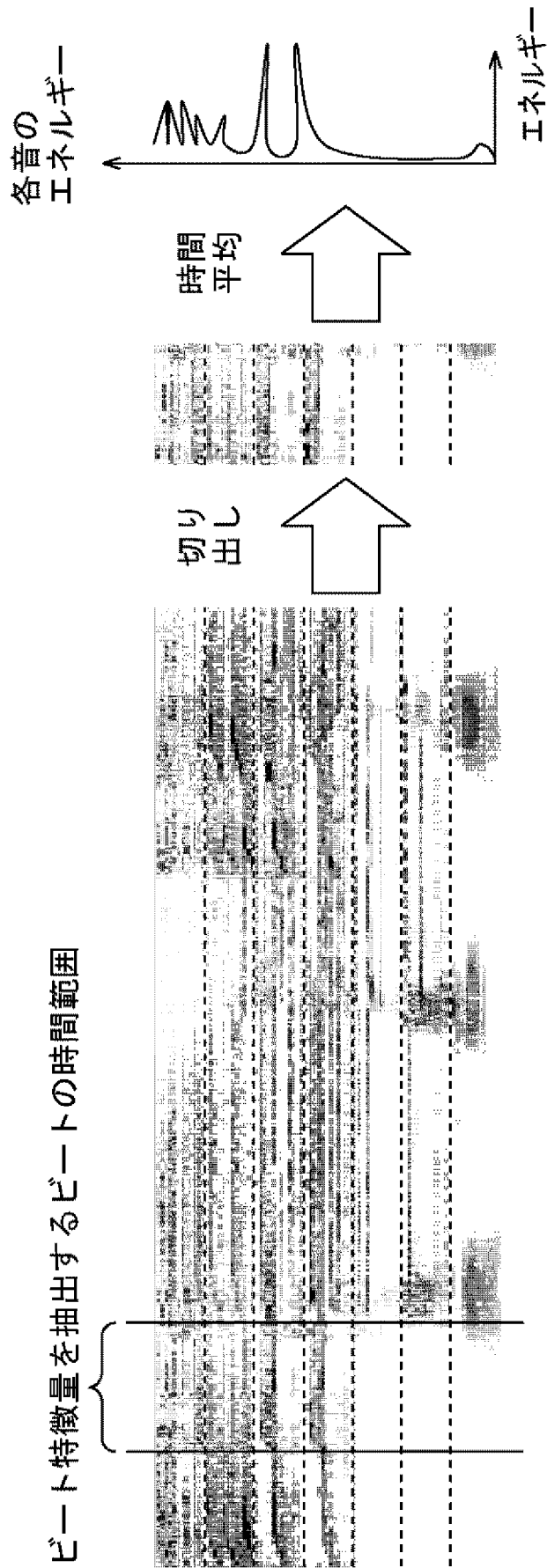
[図9]

図9



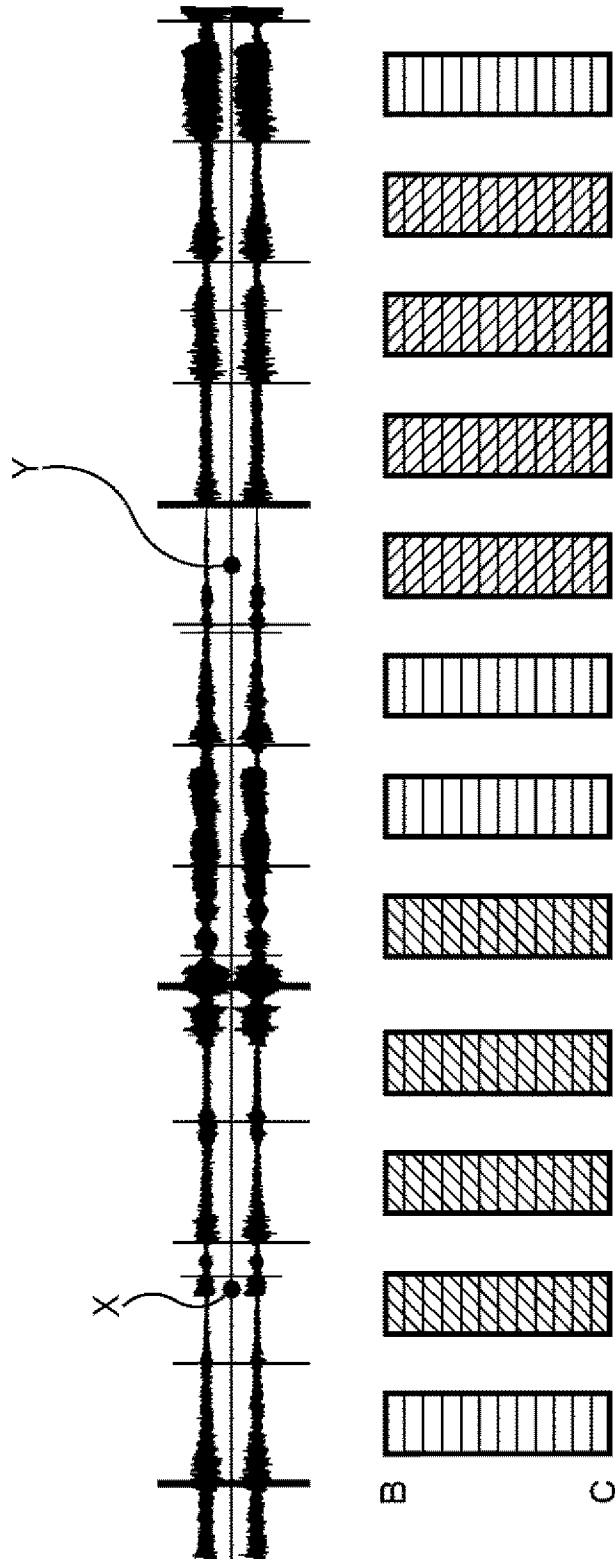
[図10]

図10



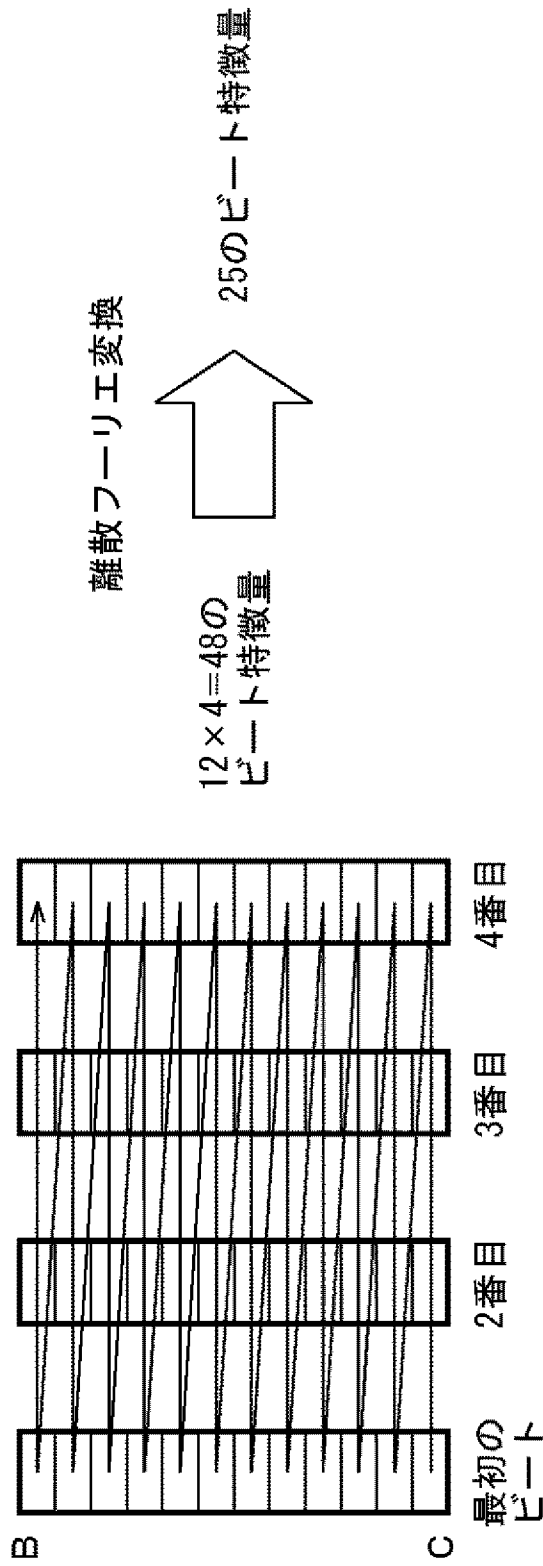
[図11]

図11

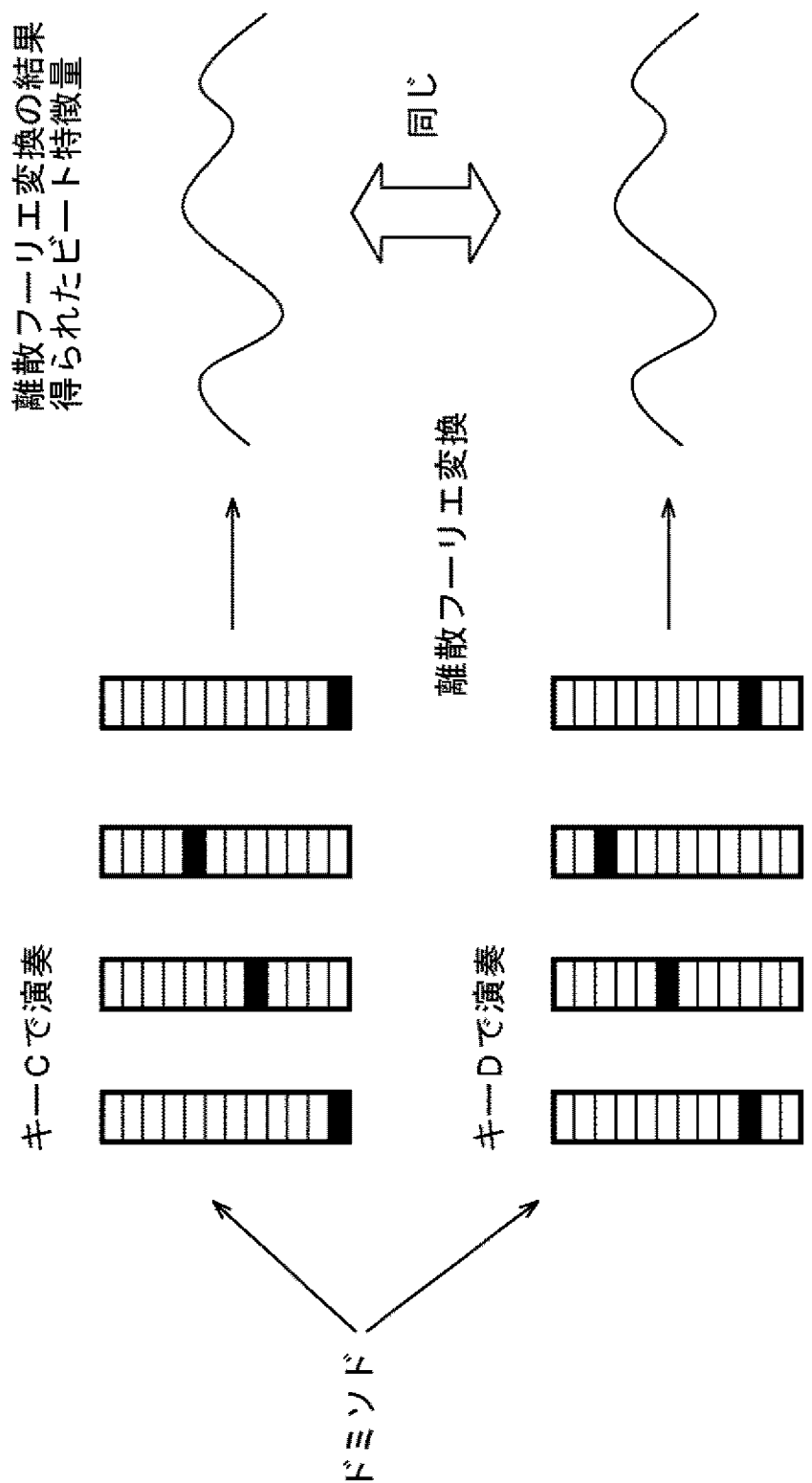


特徴量

[図12]
図12

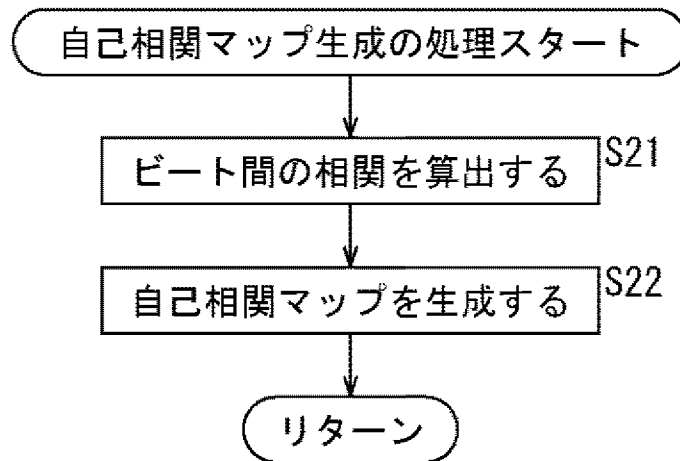


[図13]
図13



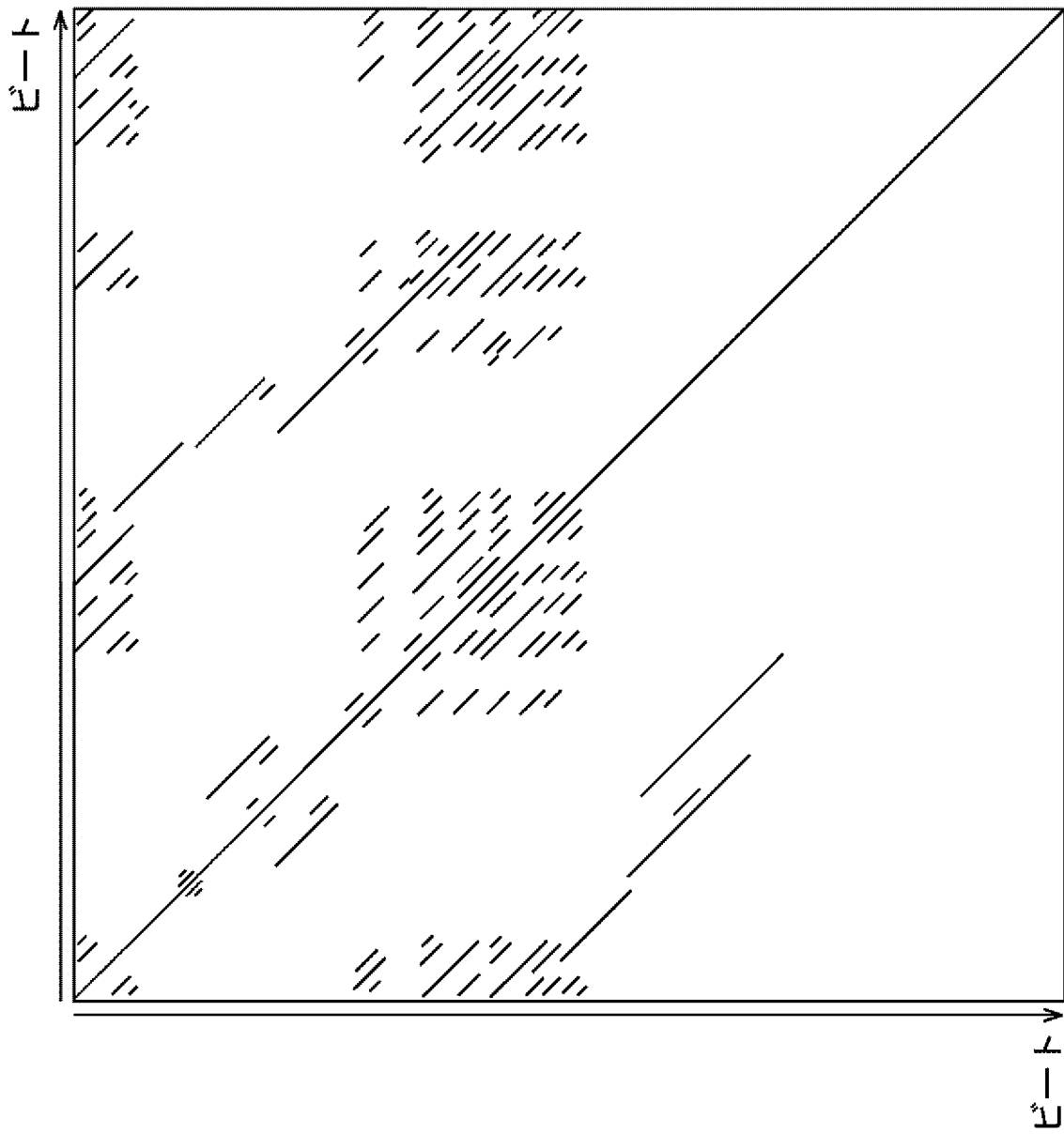
[図14]

図14



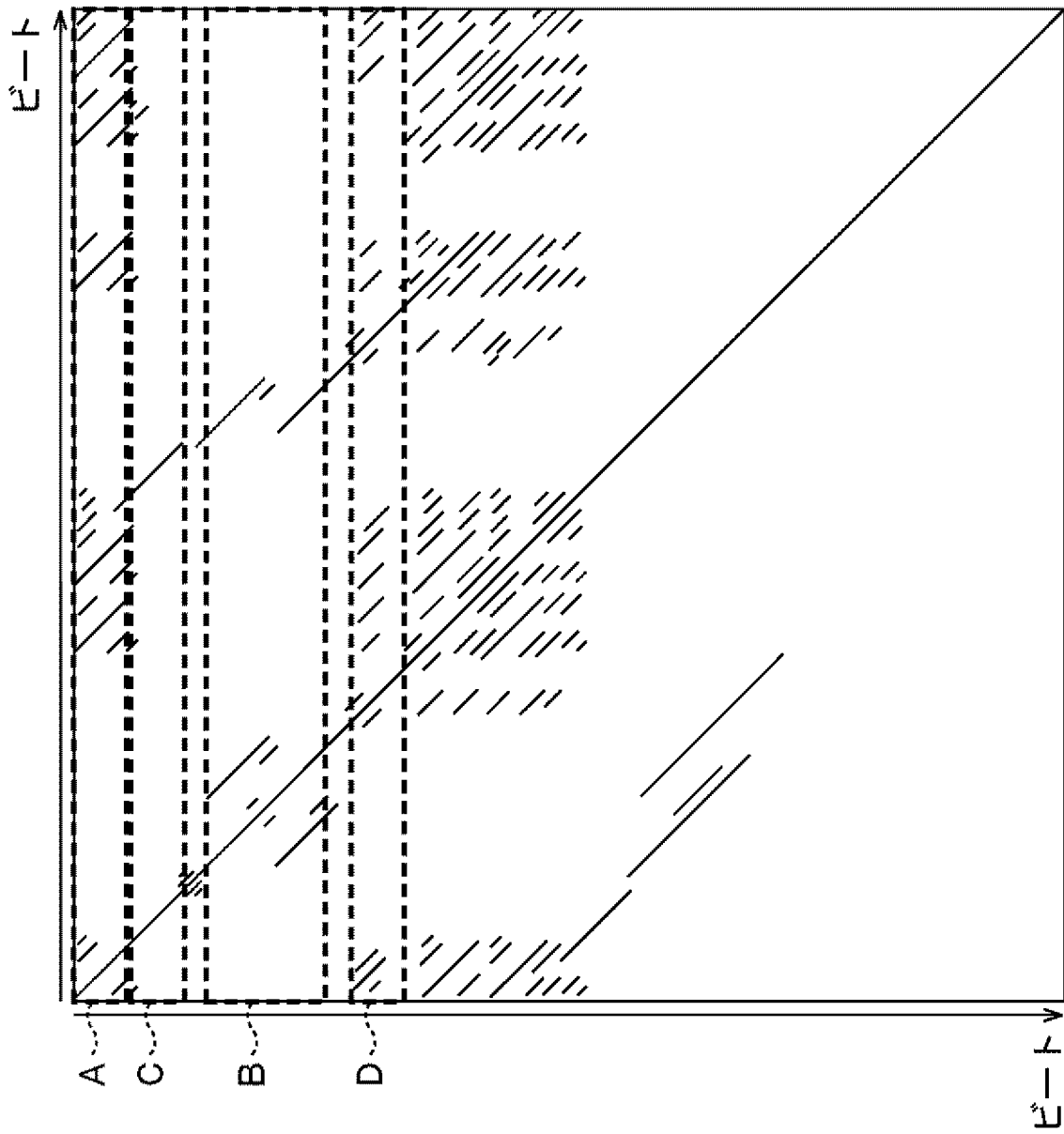
[図15]

図15



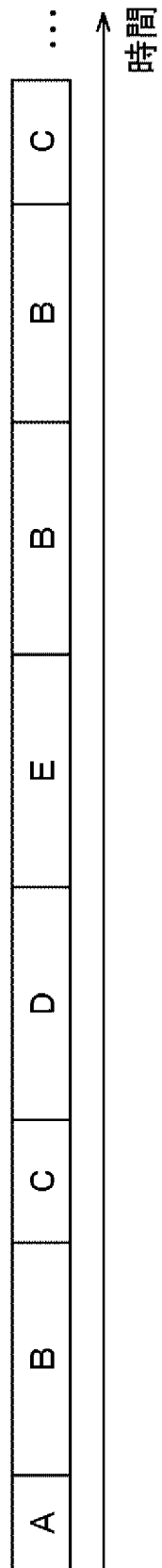
[図16]

図16



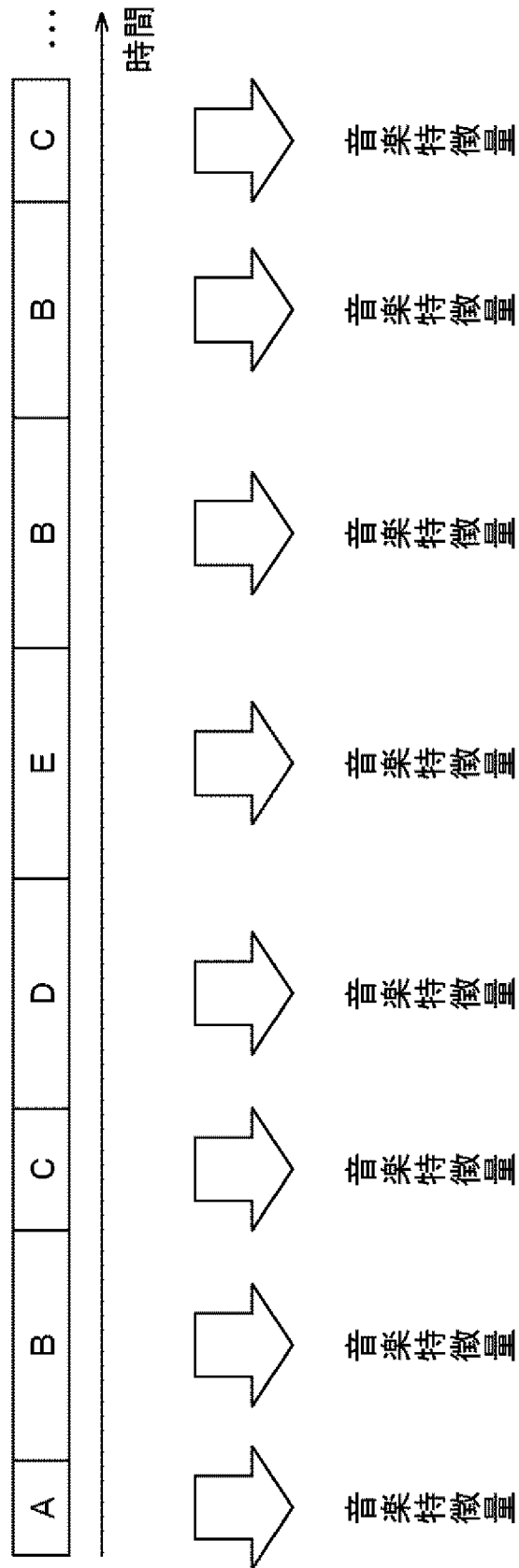
[図17]

図17



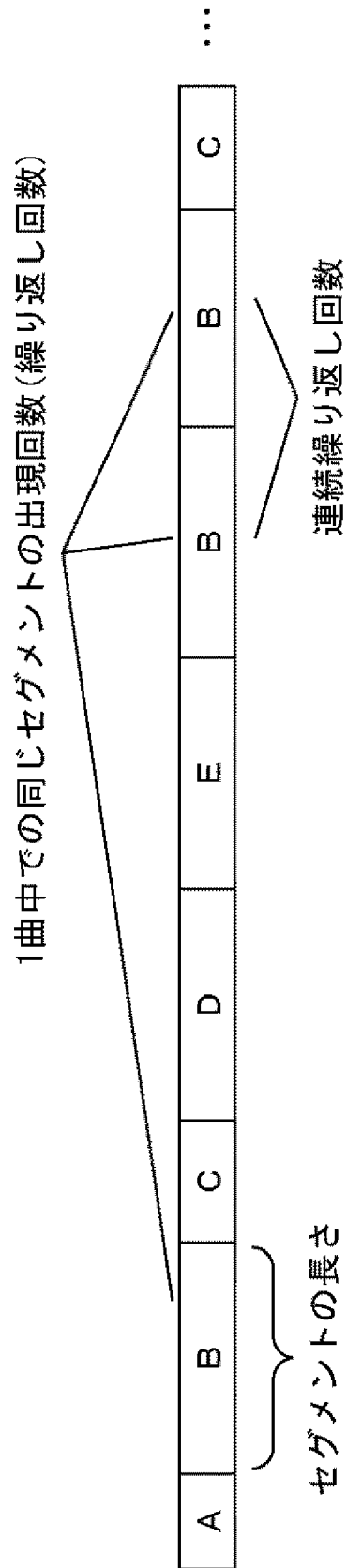
[図18]

図18



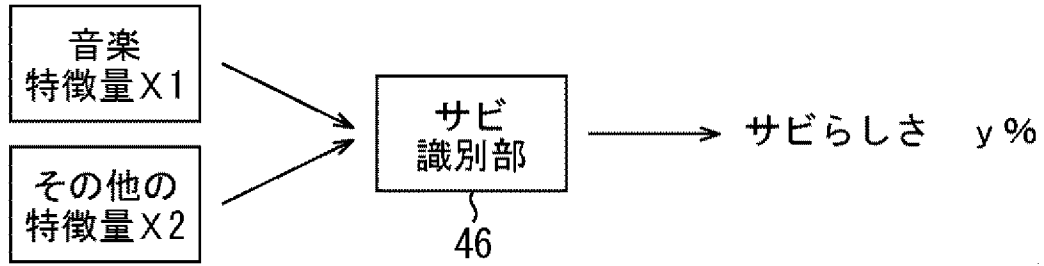
[図19]

図19



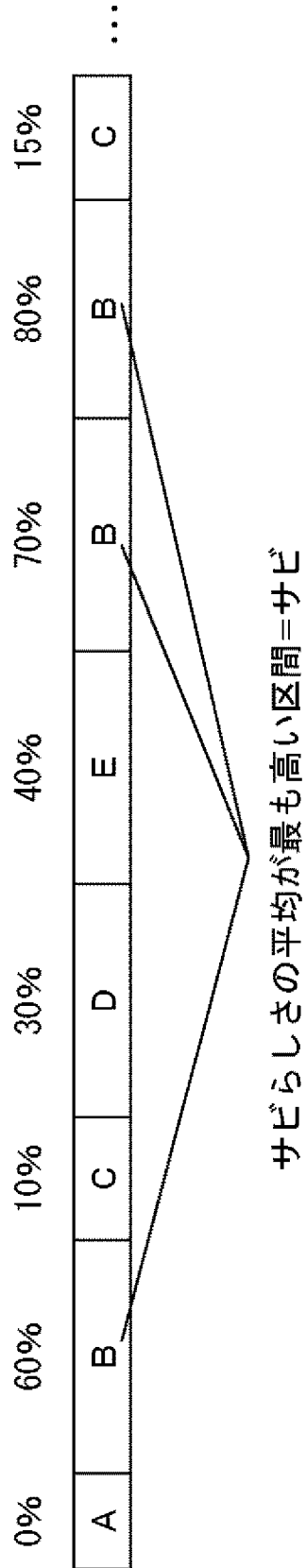
[図20]

図20



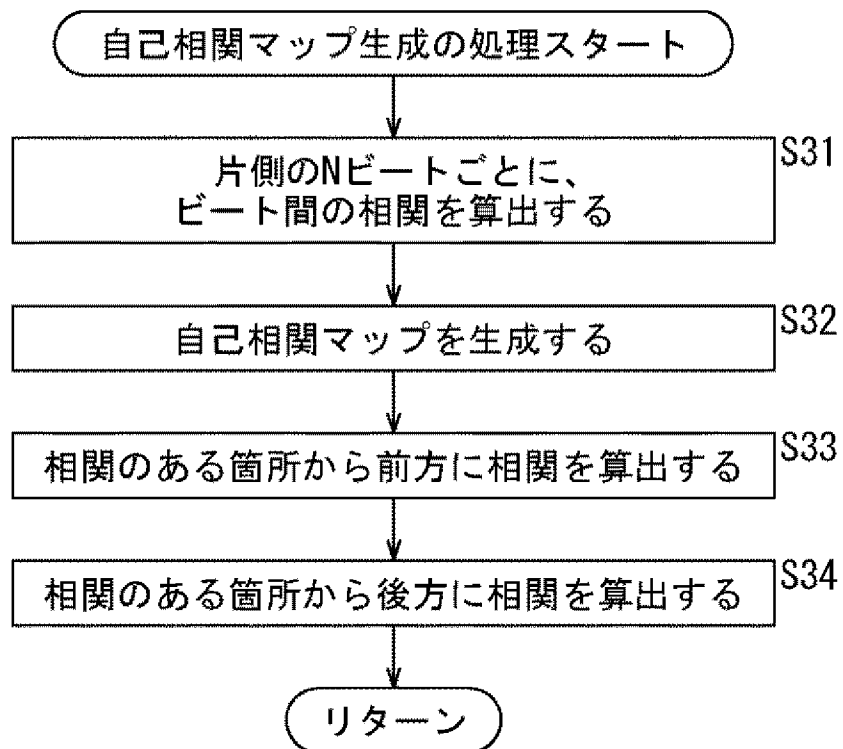
[図21]

図21



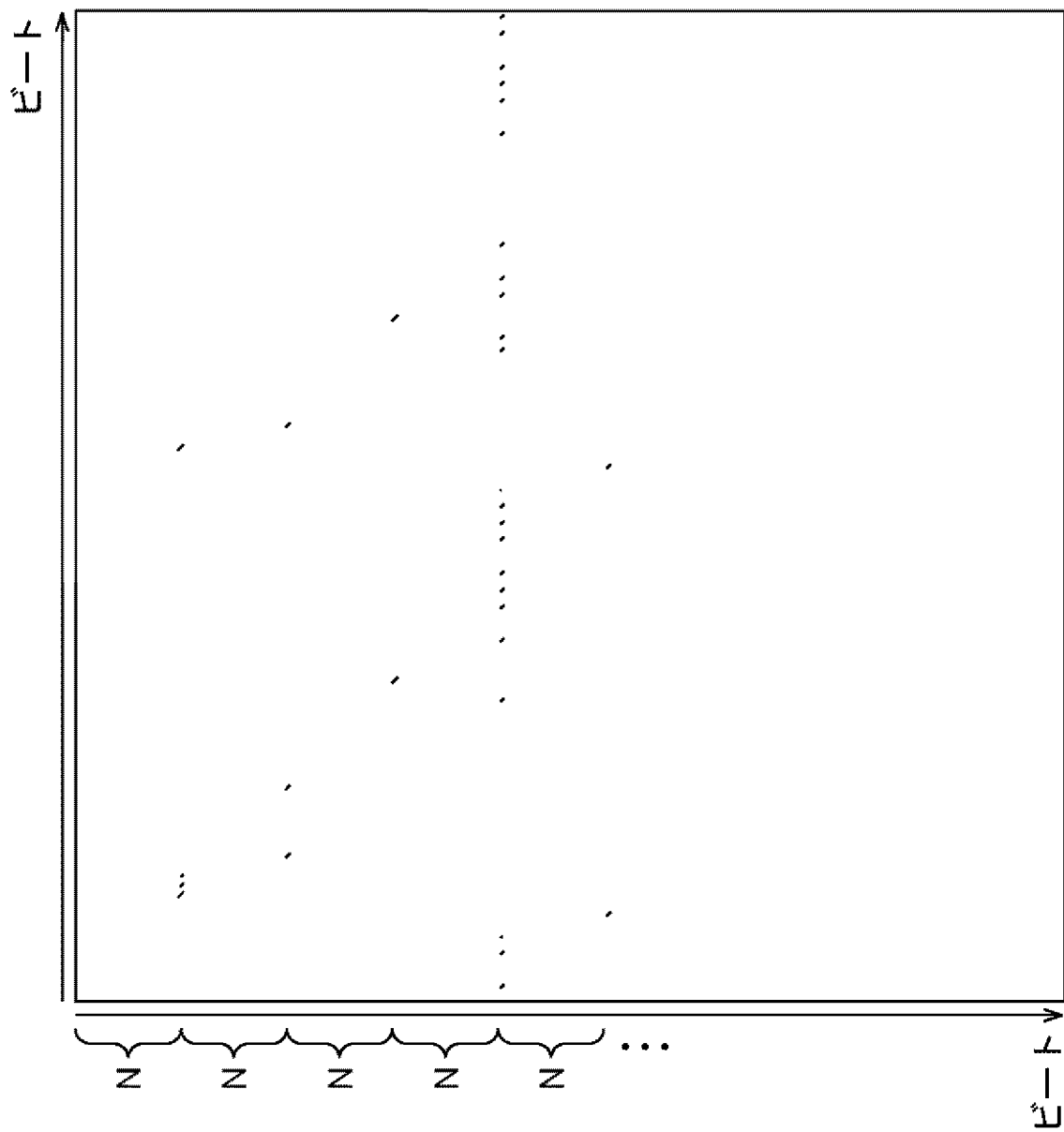
[図22]

図22



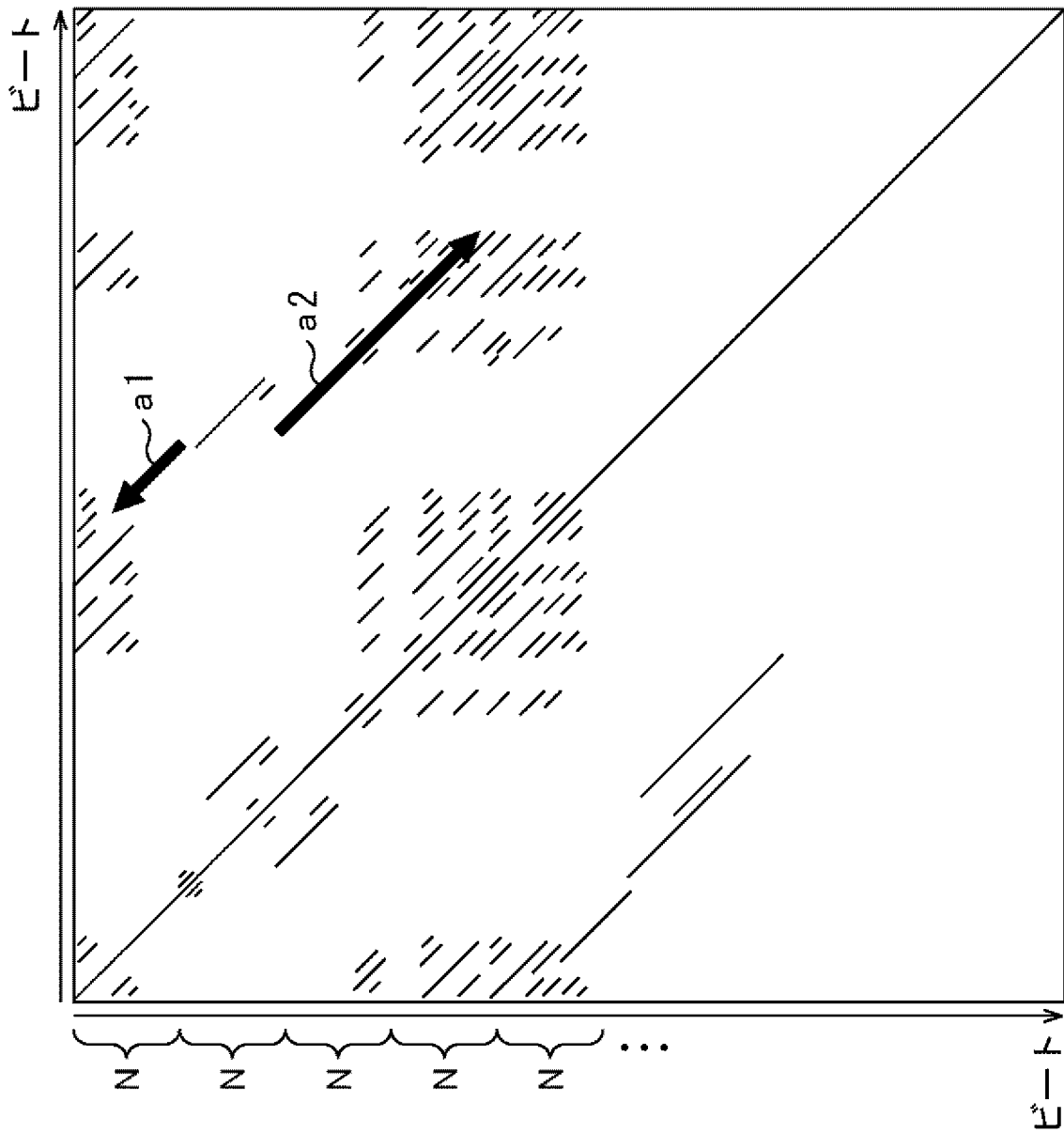
[図23]

図23



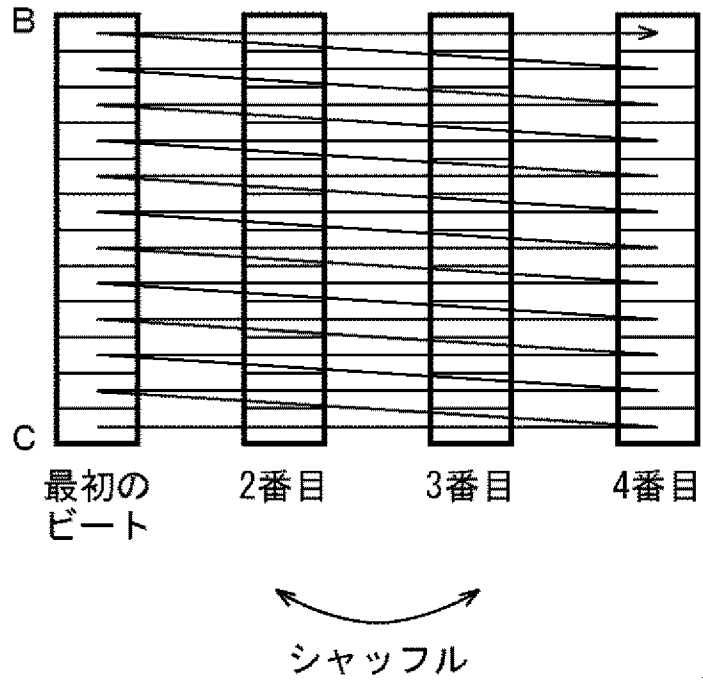
[図24]

図24



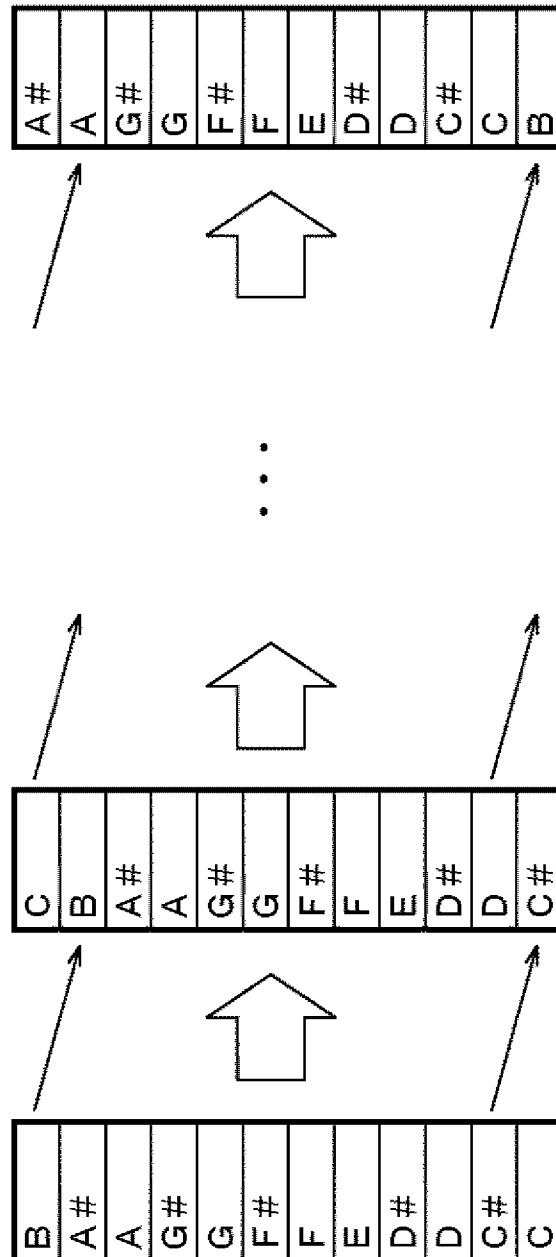
[図25]

図25

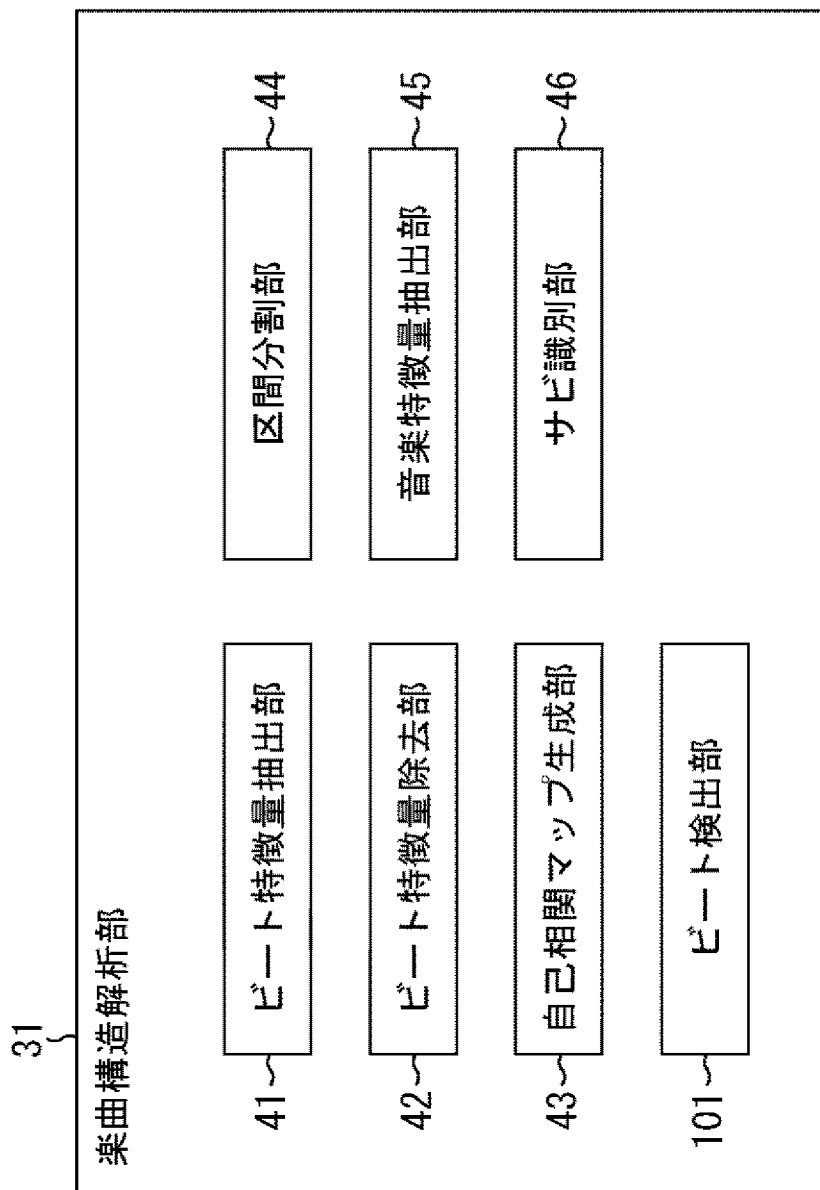


[図26]

図26

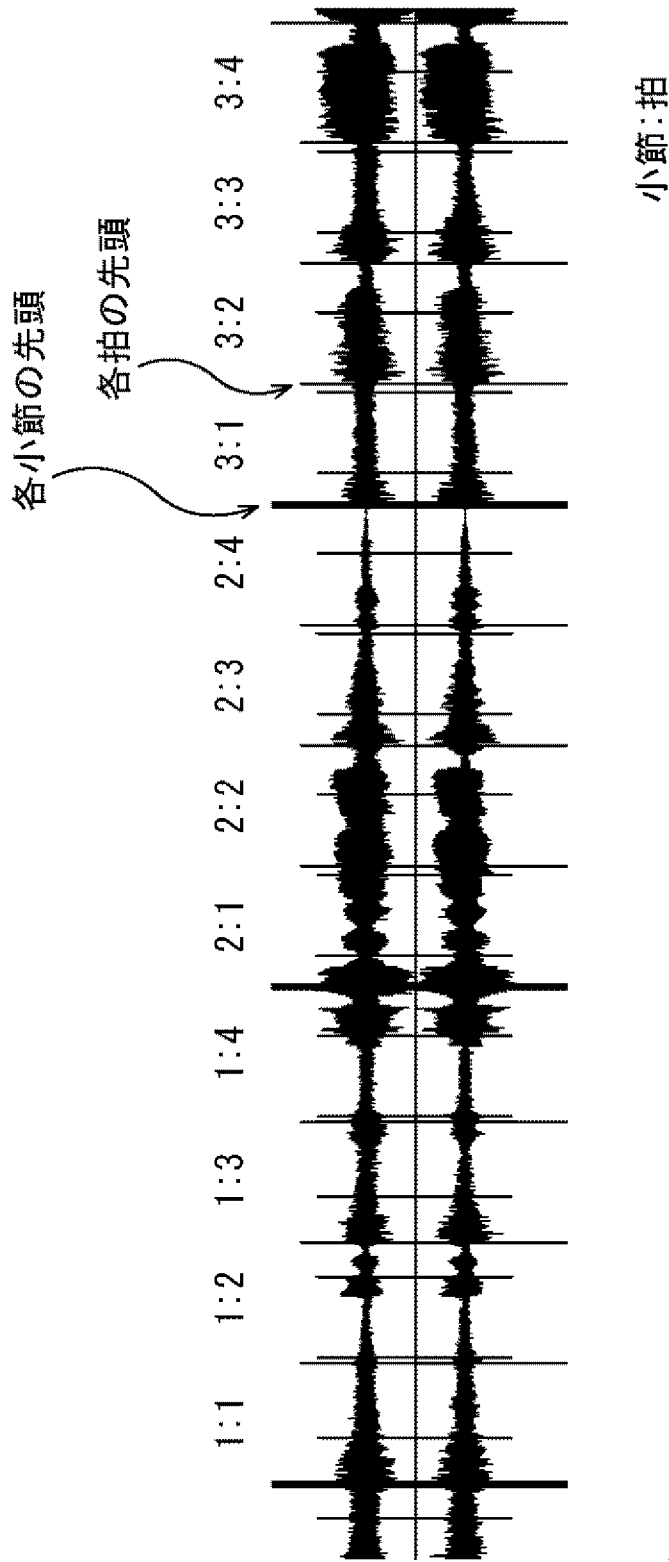


[図27]
図27

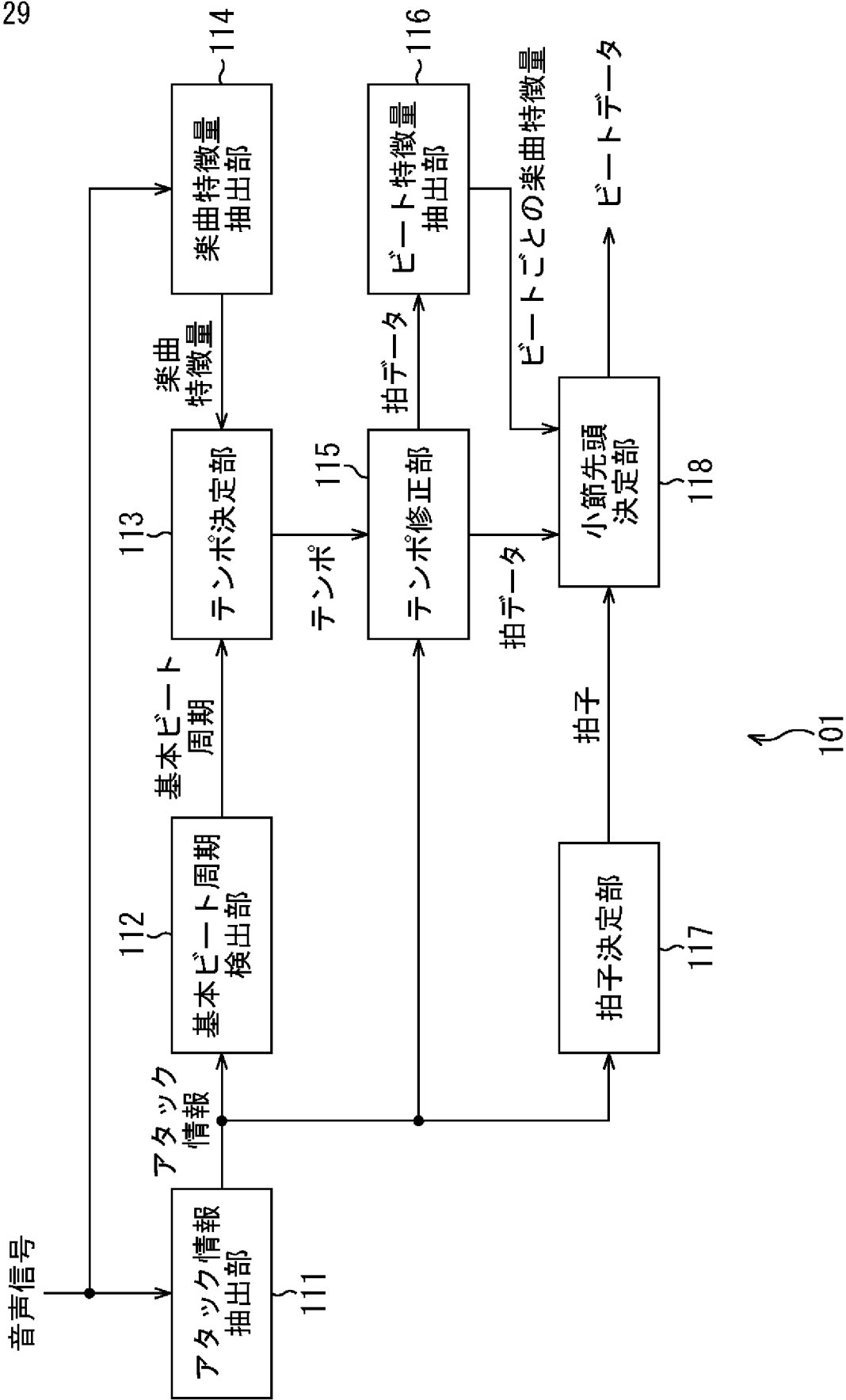


[図28]

図28

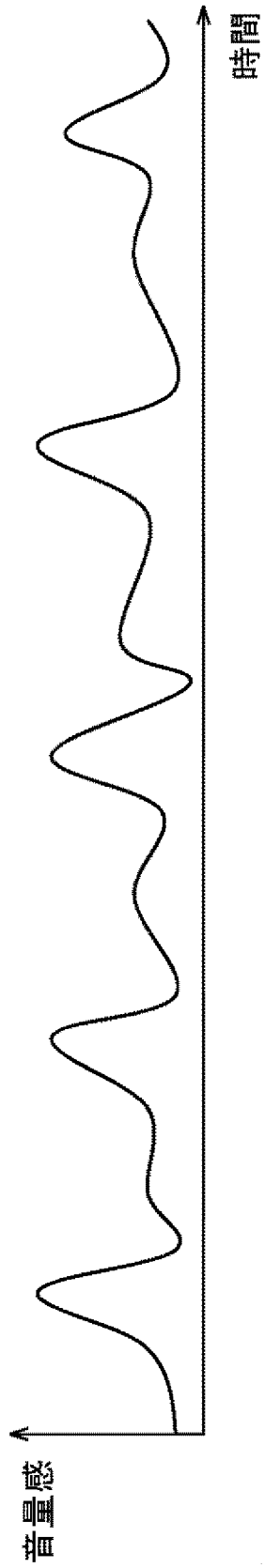


[図29]
図29



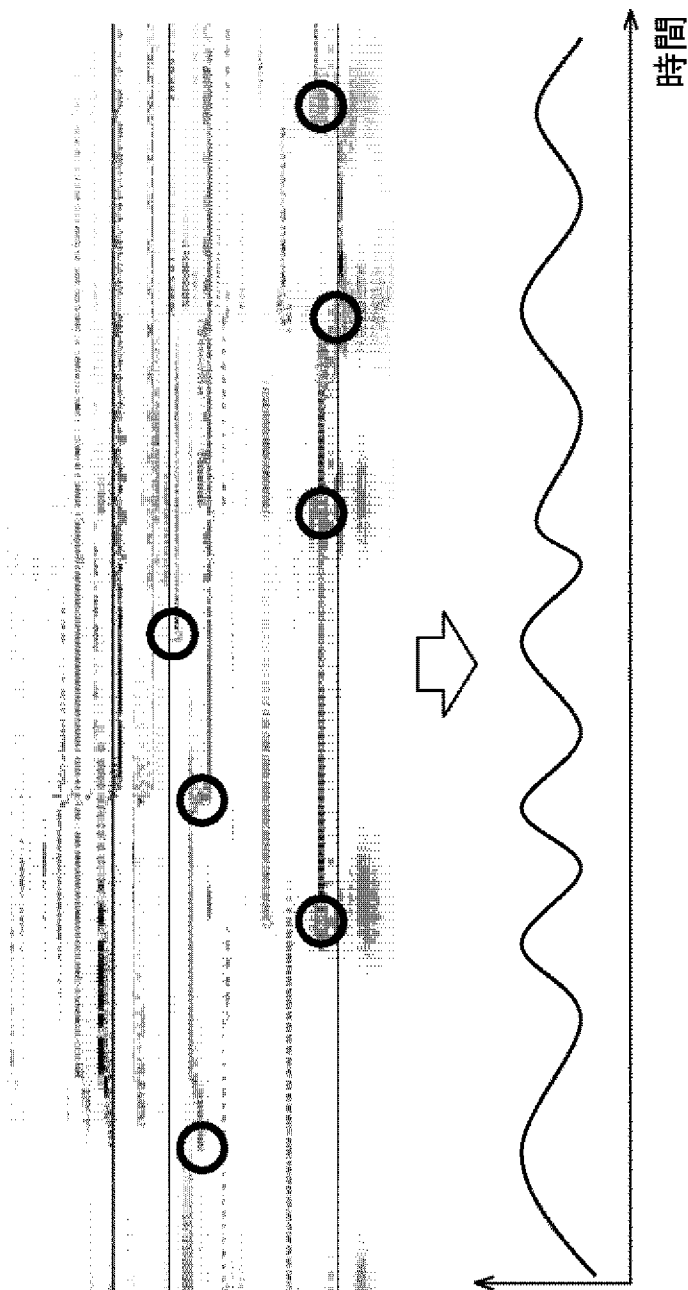
[図30]

図30



[図31]

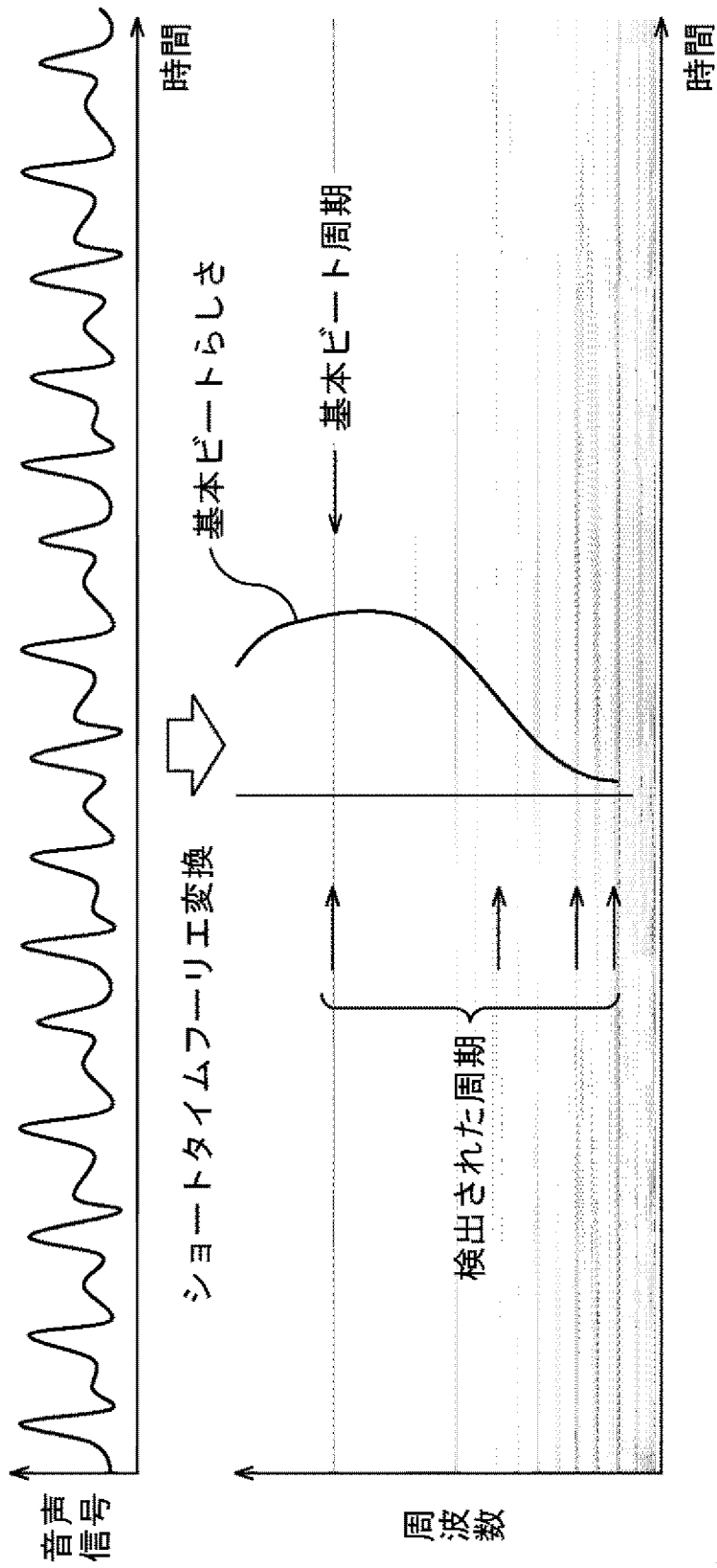
図31



12音解析による
時間-音程データ

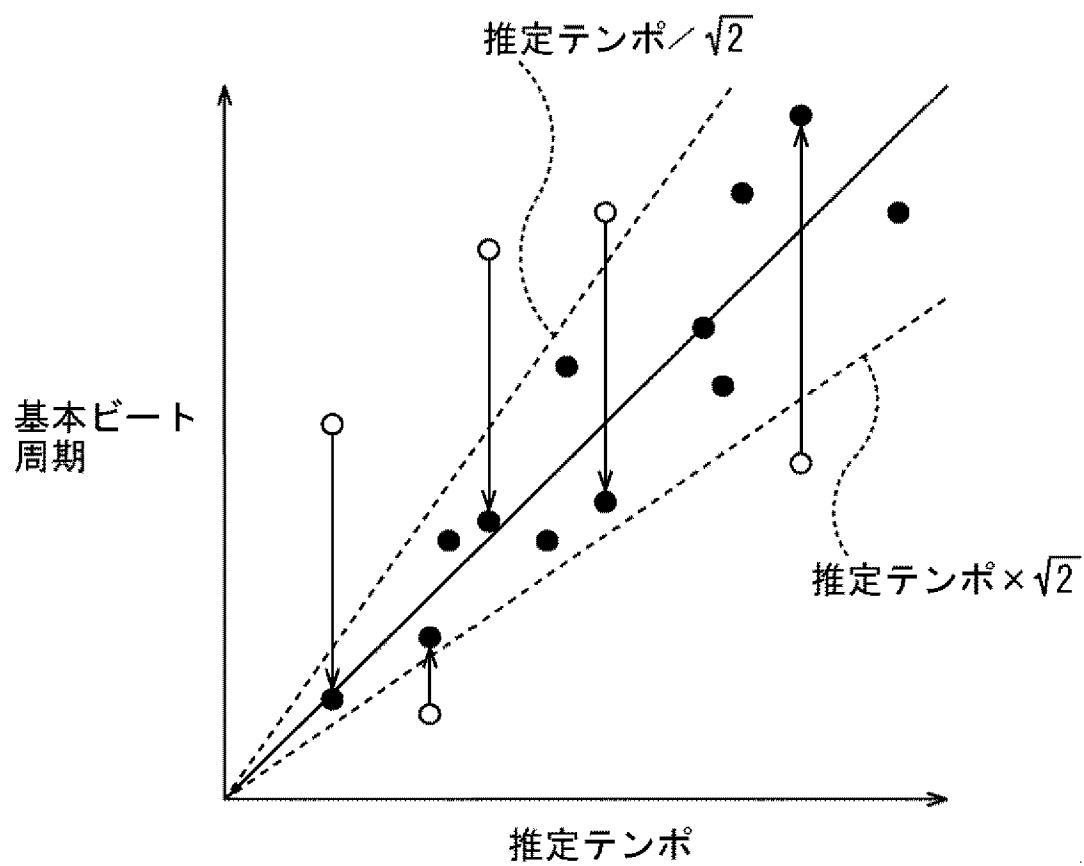
音量感
アタック情報

[図32]
図32



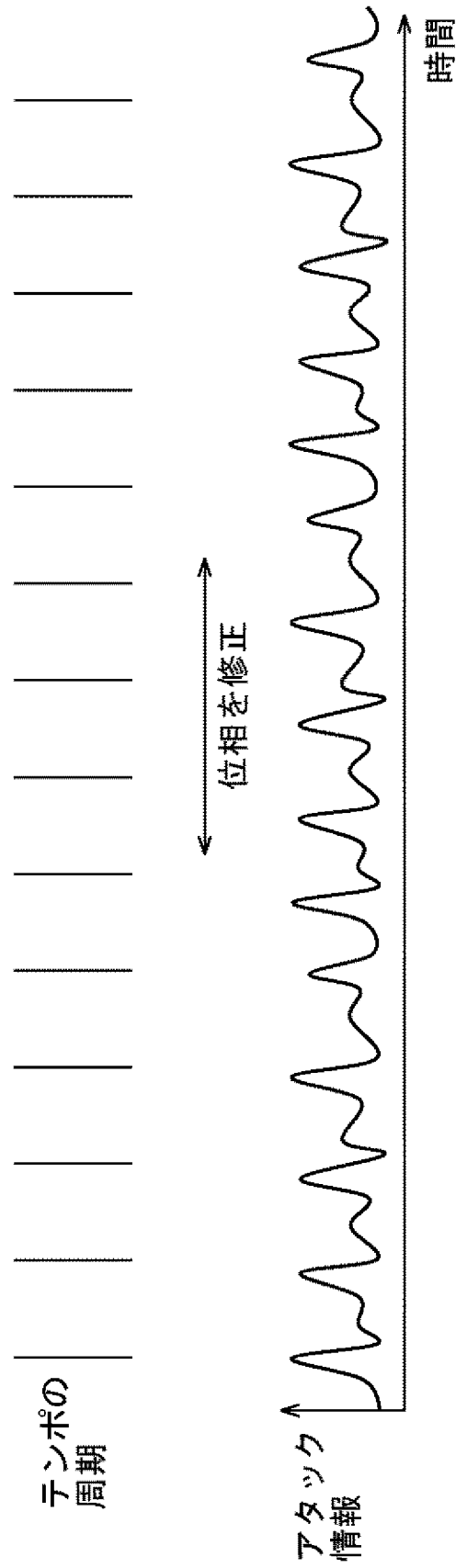
[図33]

図33

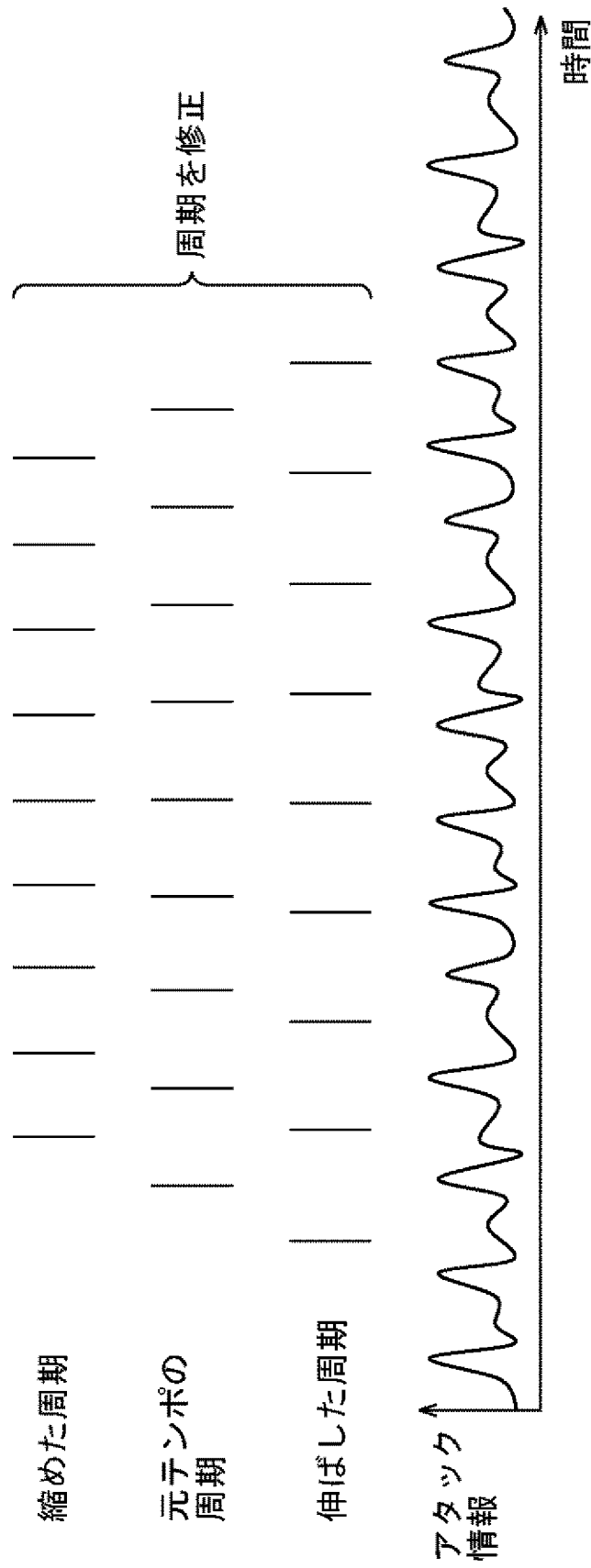


[図34]

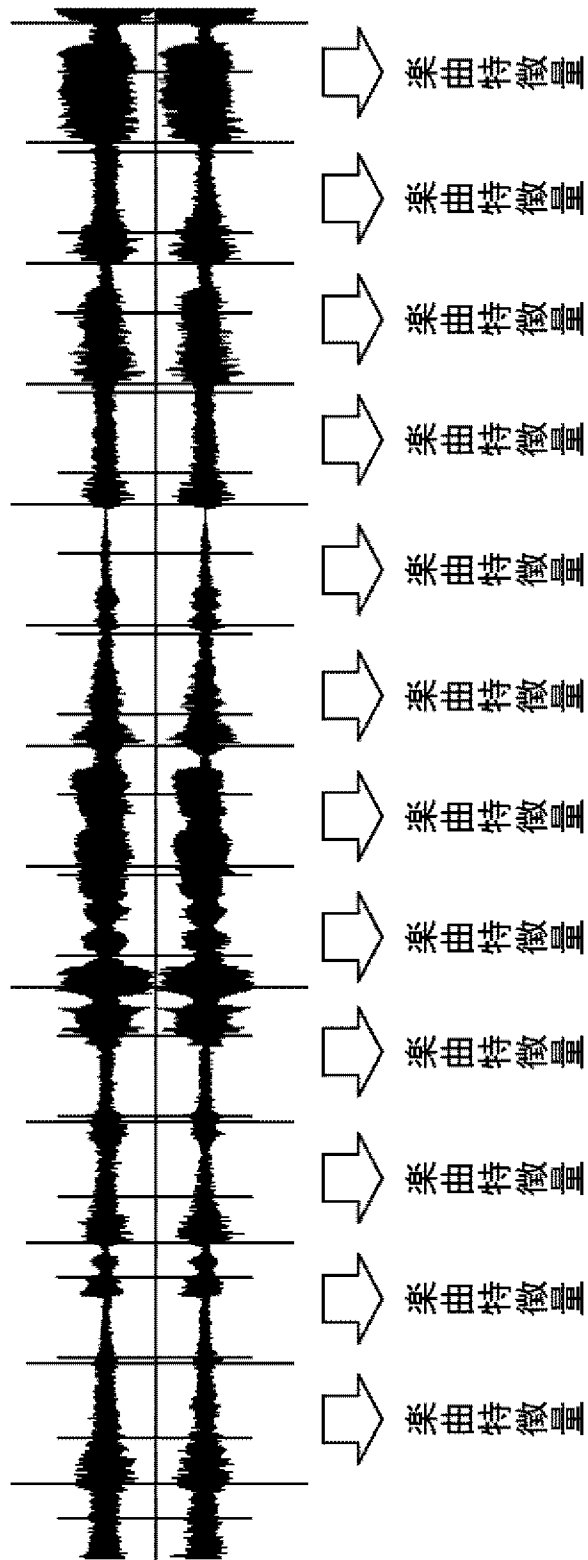
図34



[図35]
図35

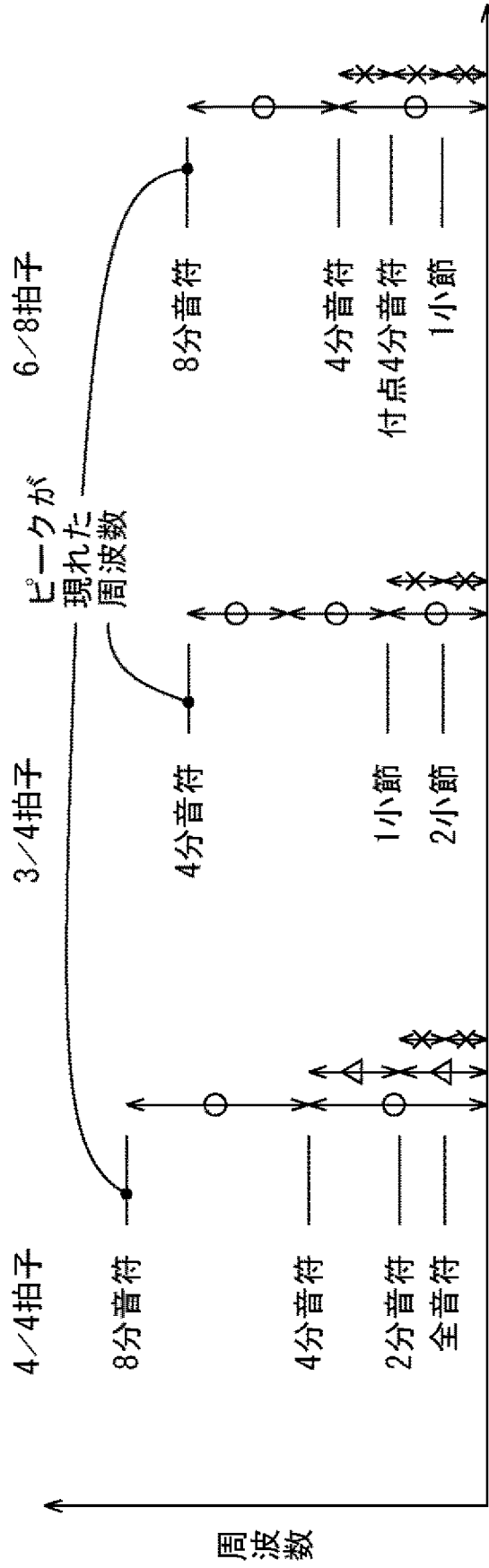


[図36]
図36



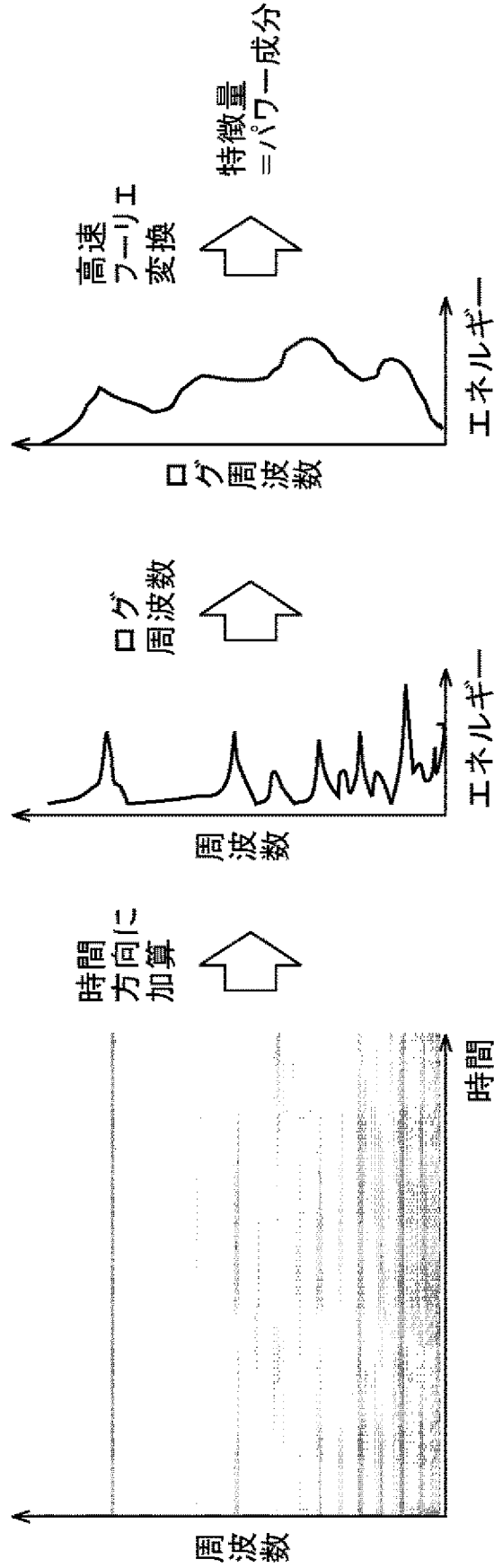
[図37]
図37

時系列のアタック情報をシヨートタイムフーリエ変換した結果



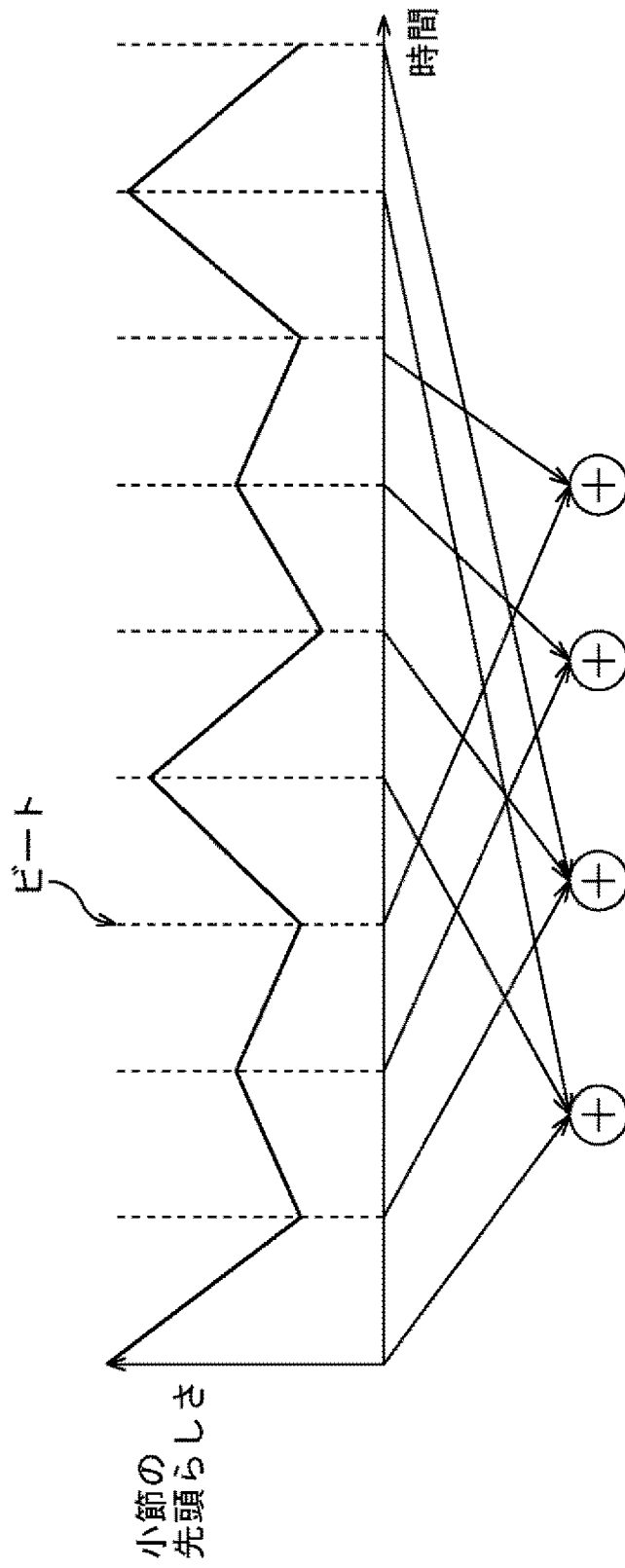
[図38]

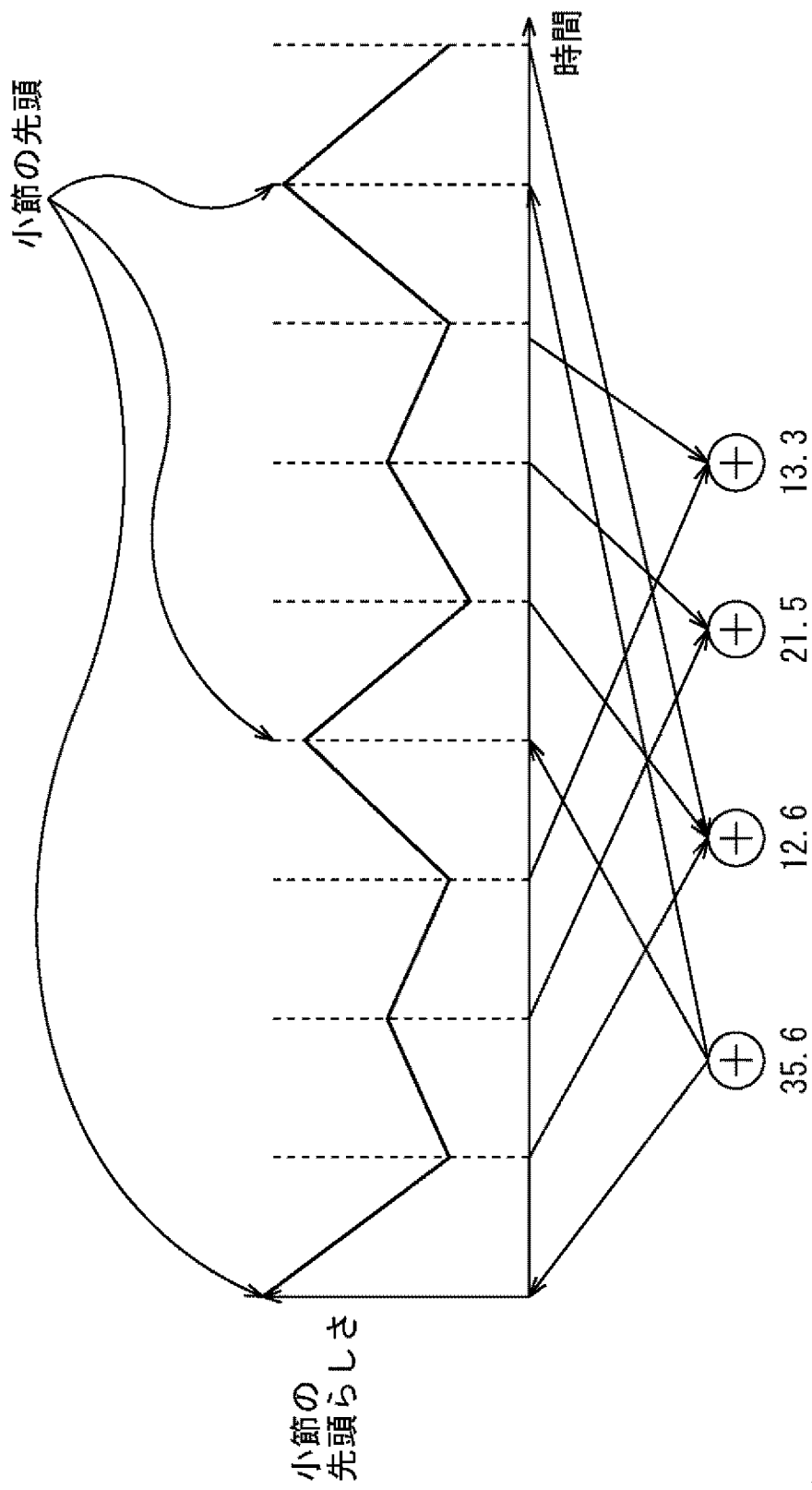
図38



[図39]

図39



[図40]
図40

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2006/322273

A. CLASSIFICATION OF SUBJECT MATTER
G10G3/04(2006.01)i, G10L11/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10G3/04, G10L11/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2006
Kokai Jitsuyo Shinan Koho	1971-2006	Toroku Jitsuyo Shinan Koho	1994-2006

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JSTPlus (JDream2)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-233965 A (National Institute of Advanced Industrial Science and Technology), 19 August, 2004 (19.08.04), Full text; all drawings & WO 2004/038694 A1 & EP 1577877 A1 & US 2005/0241465 A1	1-12
A	Masataka GOTO, "Keisanki wa Ongaku ni Awasete Tebyoshi o Uteruka?", bit, Vol.28, No.3, 1996, pages 4 to 11	1-12
A	Masataka GOTO, Smart Music KIOSK: Sabi Dashi Kinotsuki Ongaku Shichoki, Transactions of Information Processing Society of Japan, Vol.44, No.11, 2003, pages 2737 to 2747	1-12

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
06 December, 2006 (06.12.06)

Date of mailing of the international search report
12 December, 2006 (12.12.06)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2006/322273

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Hiromi FUJII, "Jido Henkyoku no Tame no Melody Bunseki -Melody Kozo no Chushutsu-", Information Processing Society of Japan Dai 37 Kai (Showa 63 Nen Koki) Zenkoku Taikai Koen Ronbunshu, 5R-5, 1988, pages 1967 to 1968	1-12
A	Jun MURAMATSU, "Kayokyoku ni okeru 'Sabi' no Gakufu Joho ni Motozuku Tokucho Chushutsu", Information Processing Society of Japan Kenkyu Hokoku, Vol.2000, No.49, 2000-MUS-35-1, 2000, pages 1 to 6	1-12
A	Shinichi GAYAMA, "Gakuon Shingo kara no Waon Shinko Chushutsu Shuho to Ruiji Gakkyoku Kensaku eno Oyo", PIONEER R & D, Vol.14, No.2, 2004, pages 1 to 7	1-12
A	Shinichi GAYAMA, "Waon Shinko o Mochiita Ongaku highlight Kenshutsu no Ichishuho to, Kumikomi Kiki eno Oyo", PIONEER R & D, Vol.15, No.1, 2005 February, pages 49 to 54	1-12
A	Masataka GOTO, "Real time Ongaku Jokei Kijutsu System", Information Processing Society of Japan Kenkyu Hokoku, Vol.2000, No.94, 2000-MUS-37-2, 2000, pages 9 to 16	1-12

A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int.Cl. G10G3/04(2006.01)i, G10L11/00(2006.01)i			
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int.Cl. G10G3/04, G10L11/00			
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2006年 日本国実用新案登録公報 1996-2006年 日本国登録実用新案公報 1994-2006年			
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語) JSTPlus(JDream2)			
C. 関連すると認められる文献			
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号	
A	JP 2004-233965 A (独立行政法人産業技術総合研究所) 2004.08.19, 全文, 全図 & WO 2004/038694 A1 & EP 1577877 A1 & US 2005/0241465 A1	1-12	
A	後藤真孝, 計算機は音楽に合わせて手拍子を打てるか?, bit, Vol. 28, No. 3, 1996年, p. 4-11	1-12	
A	後藤真孝, SmartMusicKIOSK: 押し出し機能着き音楽試聴機, 情報処理学会論文誌, 第44巻, 第11号, 2003年, p. 2737-2747	1-12	
<input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。		<input type="checkbox"/> パテントファミリーに関する別紙を参照。	
* 引用文献のカテゴリー 「A」特に関連のある文献ではなく、一般的技術水準を示すもの 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」口頭による開示、使用、展示等に言及する文献 「P」国際出願日前で、かつ優先権の主張の基礎となる出願		の日の後に公表された文献 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」同一パテントファミリー文献	
国際調査を完了した日 06.12.2006		国際調査報告の発送日 12.12.2006	
国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 荏原 雄一	5Z 3148
		電話番号 03-3581-1101 内線 3541	

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	藤井浩美, 自動編曲のためのメロディ分析 –メロディ構造の抽出–, 情報処理学会第37回 (昭和63年後期) 全国大会講演論文集, 5R-5, 1988年, p. 1967-1968	1-12
A	村松純, 歌謡曲における「さび」の楽譜情報に基づく特徴抽出, 情報処理学会研究報告, Vol.2000, No.49, 2000-MUS-35-1, 2000年, p. 1-6	1-12
A	莪山真一, 楽音信号からの和音進行抽出手法と類似楽曲検索への応用, PIONEER R&D, Vol.14, No.2, 2004年, p. 1-7	1-12
A	莪山真一, 和音進行を用いた音楽ハイレベル検出の一手法と、組み込み機器への応用, PIONEER R&D, Vol.15, No.1, 2005年2月, p. 49-54	1-12
A	後藤真孝, リアルタイム音楽情景記述システム, 情報処理学会研究報告, Vol.2000, No.94, 2000-MUS-37-2, 2000年, p. 9-16	1-12