

(43) 国際公開日
2008 年 10 月 2 日 (02.10.2008)

PCT

(10) 国際公開番号
WO 2008/117359 A1(51) 国際特許分類:
G10L 11/00 (2006.01)

(21) 国際出願番号: PCT/JP2007/055899

(22) 国際出願日: 2007 年 3 月 22 日 (22.03.2007)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人 (米国を除く全ての指定国について): パイオニア株式会社 (PIONEER CORPORATION) [JP/JP]; 〒1538654 東京都目黒区目黒一丁目 4 番 1 号 Tokyo (JP).
株式会社テック・エキスパート (TECH EXPERTS INCORPORATION) [JP/JP]; 〒1438564 東京都大田区大森西四丁目 1 5 番 5 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): 石原博幸 (ISHIHARA, Hiroyuki) [JP/JP]; 〒3502288 埼玉県鶴ヶ島市富士見六丁目 1 番 2 号 パイオニア株式会社 総合研究

所内 Saitama (JP). 駒村光弥 (KOMAMURA, Mitsuya) [JP/JP]; 〒3502288 埼玉県鶴ヶ島市富士見六丁目 1 番 2 号 パイオニア株式会社 総合研究所内 Saitama (JP).

(74) 代理人: 石川泰男, 外 (ISHIKAWA, Yasuo et al.); 〒1050014 東京都港区芝二丁目 1 7 番 1 1 号 パーク芝ビル 2 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, BH, 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, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, 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.

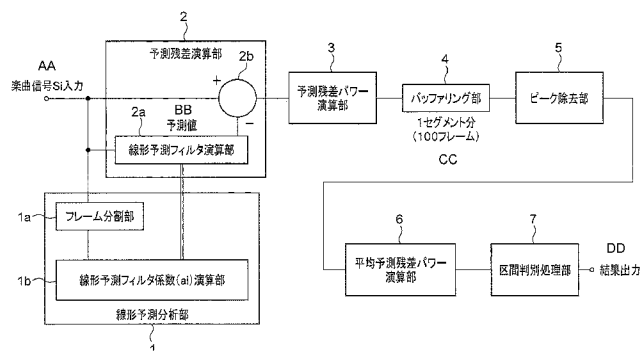
(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY,

[続葉有]

(54) Title: MUSIC GENRE DISCRIMINATION APPARATUS, MUSIC GENRE DISCRIMINATION METHOD, AND MUSIC GENRE DISCRIMINATION PROGRAM

(54) 発明の名称: 楽曲種類判別装置、楽曲種類判別方法、および楽曲種類判別プログラム

[図1]



AA... MUSIC SIGNAL (Si) INPUT
2... PREDICTION RESIDUAL ERROR CALCULATION UNIT
BB... PREDICTED VALUE
2a... LINEAR PREDICTION FILTER CALCULATION SECTION
3... PREDICTION RESIDUAL ERROR POWER CALCULATION UNIT
4... BUFFERING UNIT
CC... FOR ONE SEGMENT (100 FRAMES)
5... PEAK ELIMINATION UNIT
1... LINEAR PREDICTION ANALYSIS UNIT
1a... FRAME DIVISION SECTION
1b... LINEAR PREDICTION FILTER COEFFICIENT (ai) CALCULATION SECTION
6... AVERAGE PREDICTION RESIDUAL ERROR POWER CALCULATION UNIT
7... INTERVAL DISCRIMINATION PROCESSING UNIT
DD... RESULT OUTPUT

(57) Abstract: There are provided a music genre discrimination device, a music genre discrimination method, and a music genre discrimination program capable of discriminating a single type of music-playing zone and a plurality of types of music-playing zones with a small amount of calculations, with ease, and with high accuracy. The music genre discrimination device analyzes,

[続葉有]



WO 2008/117359 A1



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, MT, 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).

添付公開書類:
— 国際調査報告書

according to an input music signal, the music signal to discriminate a genre of music and comprises an analysis means for performing a time-series analysis on the music signal and a discrimination means for discriminating the genre of the music signal on the basis of an analytic value analyzed by the analysis means.

(57) 要約: 課題の一例としては、少ない演算量で簡易に精度よく単一種類演奏区間と複数種類演奏区間とを判別できる楽曲種別判別装置、楽曲種別判別方法および楽曲種別判別プログラムを提供することにある。楽曲種別判別装置は、入力された楽曲信号に基づいて、前記楽曲信号を分析し、楽曲の種類の判別を行なう楽曲種別判別装置において、前記楽曲信号を時系列分析する分析手段と、前記分析手段によって分析された分析値に基づいて、前記楽曲信号の種類の判別する判別手段とを備える。

明 細 書

楽曲種類判別装置、楽曲種類判別方法、および楽曲種類判別プログラム 技術分野

[0001] 本願は、楽曲の特徴量を抽出して楽曲の種類を判別する楽曲種類判別装置の技術分野に関する。

背景技術

[0002] 従来から、人間がしゃべる音声信号と楽器による楽音信号とを自己相関係数演算で識別する音声・楽音識別装置がある(特許文献1参照)ことが知られている。また、楽曲の特徴量としての楽曲の周波数を抽出するための技術としてFFT(Fast Fourier Transform)等の周波数解析技術が知られている。

特許文献1:特許第3074046号公報

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、自己相関係数演算を使用した信号分析手法およびFFT等の周波数分析手法を用いると演算量が膨大になるので演算時間がかかり、リアルタイムに楽曲の種類の識別を行なうことが困難であった。

[0004] 本願は、上記の各問題点に鑑みて為されたもので、その課題の一例としては、少ない演算量で簡易に精度よく楽曲の種類を判別できる楽曲種別判別装置、楽曲種別判別方法および楽曲種別判別プログラムを提供することにある。

課題を解決するための手段

[0005] 上記課題を解決するために、請求項1に記載の楽曲種類判別装置は、入力された楽曲信号に基づいて、前記楽曲信号を分析し、楽曲の種類の判別を行なう楽曲種類判別装置において、前記楽曲信号を時系列分析する分析手段と、前記分析手段によって分析された分析値に基づいて、前記楽曲信号の種類の判別する判別手段と、を備えることを特徴とする。

上記課題を解決するために、請求項8に記載の楽曲種類判別方法は、入力された楽曲信号に基づいて、前記楽曲信号を分析し、楽曲の種類の判別を行なう楽曲種

類判別方法において、前記楽曲信号を時系列分析する分析工程と、前記分析工程において分析された分析値に基づいて、前記楽曲信号の種類を判別する判別工程と、を備えることを特徴とする。

上記課題を解決するために、請求項9に記載の楽曲種類判別プログラムは、請求項1に記載の楽曲種類判別装置に含まれるコンピュータを、前記楽曲信号を時系列分析する分析手段、前記分析手段によって分析された分析値に基づいて、前記楽曲信号の種類を判別する判別手段、として機能させることを特徴とする。

図面の簡単な説明

- [0006] [図1]本実施形態の楽曲種類判別装置の構成を示すブロック図である。
[図2]本実施形態のピーク除去部の動作を説明する図である。
[図3](a)本実施形態の単一閾値を使用した場合の動作を説明する図である。(b)本実施形態の2段階閾値判別方法を使用した場合の動作を説明する図である。
[図4]本実施形態の平均正規化残差パワー度数分布の一例を示す図である。
[図5]本実施形態の動作を示すフローチャートである。

符号の説明

- [0007] 1 … 線形予測分析部
1a…フレーム分割部
1b…線形予測フィルタ係数演算部
2 … 予測残差演算部
2a…線形予測フィルタ演算部
2b…演算部
3 … 予測残差パワー演算部
4 … バッファリング部
5 … ピーク除去部
6 … 平均予測残差パワー演算部
7 … 区間判別処理部
Si … 楽曲信号

発明を実施するための最良の形態

[0008] 以下、添付図面に基づいて、本願の実施形態について説明する。本実施形態では楽曲の種類として特に単一種類の楽器によって演奏された区間を判別する方法について説明する。

[0009] 図1は本実施形態の楽曲種類判別装置の構成を示すブロック図であり、図2は本実施形態のピーク信号を除去する一例を説明する図であり、図3は楽曲信号を判別する場合に単一閾値と2段階閾値とを用いた態様を説明する図であり、図4は本実施形態における平均正規化残差パワー度数分布の一例を示す図であり、図5は本実施形態の動作を示すフローチャートである。

[0010] 最初に本実施形態の楽曲種類判別装置Sの構成について図1を用いて説明する。

[0011] 楽曲種類判別装置Sは、分析手段を含む線形予測分析部1(分割手段を含むフレーム分割部1aおよび線形予測フィルタ係数演算部1bを含む)、分析手段を含む予測残差演算部2(線形予測フィルタ演算部2aおよび演算部2bを含む)、分析手段を含む予測残差パワー演算部3、バッファリング部4、ピーク信号除去部を含むピーク除去部5、平均予測残差パワー演算部6、判別手段を含む区間判別処理部7を含んで構成される。

フレーム分割部1aは、入力された楽曲信号 S_i を1曲毎に、フレーム分割する。例えば、1フレームを512サンプル(データ)とした場合には、1フレームは512サンプルで構成される。また、後述する1セグメントは100フレームから構成され後段の演算単位となる。サンプリング周波数 F_s が44.1kHz(約22.68 μ sec.)の場合には1フレームが512サンプルから構成されるので、1フレームは約11.6msec.(約22.68 μ sec. $\times$ 512)となり、1セグメントが100フレームで構成されるので、1セグメントは約1.16sec(約11.6msec. $\times$ 100)となる。

[0012] 楽曲信号 S_i をフレーム分割するサンプリング周波数 F_s は44.1kHzに限定されるわけではなく、サンプリング周波数 F_s は任意の周波数に設定することができる。また、1フレームのサンプル数は512に限定されるわけではなく任意の整数値を用いることができる。さらに、1セグメントのフレーム数は100フレームに限定されるわけではなく任意の整数値を用いることができる。

[0013] 線形予測フィルタ係数演算部1bは、フレーム分割部1aにおいて分割されたフレー

ムに基づいて、時系列分析の一例としての線形予測分析を実行し、線形予測係数(a_i)を演算する。線形予測とは一般に過去のサンプルの線形結合から現在の値を予測することを意味し、(式1)による数式で表現される。

[0014] 予測信号は過去P個のサンプルを用いて(式1)で表される。

[0015] [数1]

$$\begin{aligned}\hat{y}(n) &= -(a_1 \times Y(n-1) + a_2 \times Y(n-2) + \cdots + a_p \times Y(n-p)) \\ &= -\sum (a_i \times Y(n-i)), \quad (i=1 \sim p) \quad \cdots \quad (\text{式1})\end{aligned}$$

線形予測フィルタ係数演算部1bは式1の係数である線形予測係数 a_i を演算によって算出する。

[0016] 線形予測フィルタ演算部2aは、線形予測フィルタ係数演算部1bにおいて演算された線形予測係数 a_i および楽曲信号 S_i に基づいて、線形予測フィルタを演算によって算出する。

[0017] ここで、式1において $a_0=1$ とすると、原信号と予測信号の残差 $e(n)$ は、式2で表される。

[0018] [数2]

$$e(n) = y(n) - \hat{y}(n) = \sum (a_i \times Y(n-i)), \quad (i=0 \sim p) \quad \cdots \quad (\text{式2})$$

Z変換で式1および式2を表現するとそれぞれ式3および式4で表現される。

[0019] (数3)

$$X(z) = F(z)X(z) \quad \cdots \quad (\text{式3})$$

(数4)

$$E(z) = X(z) - F(z)X(z) = (1 - F(z))X(z) = A(z)X(z) \quad \cdots \quad (\text{式4})$$

ここで、 $y(n)$ が音声信号であるとする、この $X(n)$ を式5で表現できる。

[0020] (数5)

$$X(z) = (1/A(z))E(z) \quad \cdots \quad (\text{式5})$$

式5は音声 $X(z)$ が全極フィルタ $1/A(z)$ の出力であるとみなすことができる。

[0021] 音声分野においては、合成フィルタをこの全極フィルタで表現することが線形予測分析である。また、全極フィルタ $1/A(z)$ は、 $E(z)$ が最小となるように推定される。

- [0022] 上述のように推定されたフィルタが本実施形態における線形予測フィルタとなる。
- [0023] 演算部2bは、残差 $e(n)$ を演算するために式2で示されるように、楽曲信号 S_i から予測信号を減算する演算を行う。
- [0024] 予測残差パワー演算部3においては、フレーム毎に予測残差パワーを求めるとともに、原信号である楽曲信号 S_i を用いて、フレーム毎に予測残差パワーの正規化演算をする。
- [0025] ここで、予測残差パワーについて説明する。
- [0026] 音声の場合には、線形予測残差はパルス列および雑音になるため、ある程度のパワーを持つことになる。しかし、ピアノ(打弦)およびギター(撥弦)などの楽器では音源となる弦に振動を与えた直後からすぐに減衰過程に入り、共鳴音が主体となるため、予測残差のパワーは小さくなる。
- [0027] しかし、バイオリン等の擦弦楽器やサキソフォン等の管楽器は、発音中ずっと音源部がパワーを持つ(擦弦楽器の場合には楽音が放音されている間は演奏者が弦をずっと引き続けていることになり、管楽器の場合には楽音が放音されている間は演奏者がずっと吹き続けていることになり。)ため、予測残差のパワーは小さくならない。また、電子楽器に電氣的なエフェクト処理がなされている場合も予測残差のパワーは小さくならない。
- [0028] さらに、複数種類の楽器が鳴っている場合には予測残差のパワーは大きくなり、ピアノやギター等の楽器が単体または複数台で鳴っている場合に予測残差のパワーは小さくなる。これは、ピアノ曲やアコースティックギター曲などでは、落ち着いた響きを持たせるために他の楽器と被らない時間が多いことも一因となっている。したがって、上記の線形予測残差パワーを用いれば、打弦もしくは撥弦楽器等による単一種類楽器演奏区間と複数種類楽器演奏区間との判別が可能となるのである。
- [0029] また、予測残差信号は、録音レベルによる変動する信号あるため、録音レベルによる誤差を回避するために、予測残差パワー演算部5において予測残差信号を原信号(楽曲信号 S_i)のパワーで正規化する。一フレーム内の原信号のパワーと線形予測残差のパワーの比LPR (Linear Prediction residual power Ratio: (一フレーム内の線形予測残差のパワー) / (一フレーム内の原信号のパワー)) をとり、後段の演算を

行なう。

- [0030] バッファリング部4は、予測残差パワー演算部3から出力される正規化された100フレームが入力されると、正規化された100フレームを1セグメントとして出力する。バッファリング部4には、メモリなどの記憶デバイスが備えられ、一旦入力された正規化されたフレームを記憶しておき、セグメント単位で後段へ出力する。
- [0031] ピーク除去部5は、バッファリング部4から出力された1セグメント内においてパワーの立ち上がりが急激なフレームを除去する。楽曲の立ち上がり部分(楽器の演奏の始まり部分)では残差パワーが大きくなってしまう(楽器の演奏直前は無音部分または他の楽器の演奏音が小さいことが多い。)ので、その影響によって楽曲を誤判別しないようにするために、予め該当フレームを除去して、後段の演算を行なうのである。
- [0032] 例えば、ピーク除去部5は、1セグメント中の各フレームの残差パワー値のうち除去閾値としての閾値Peak_Th以上の残差パワー値を持つフレームを除去する。ただし、除去閾値Peak_Thを超えたフレームであっても、直前の除去閾値Peak_Thを超えたフレームとのフレーム間の時間間隔が、予め定められた時間である最少ピーク時間間隔(Peak_min)を超えていない場合には、ピーク除去部5は除去閾値Peak_Thを超えたそのフレームを除去しない。そして、ピーク除去部5は、除去されたフレームを除くセグメントを出力する。
- [0033] 平均予測残差パワー演算部6は、ピーク除去部5から出力されたセグメント(最少ピーク時間間隔(Peak_min)を超え、かつ除去閾値Peak_Thを超えたフレームが除かれたセグメント)について、セグメント毎に平均パワーを演算する。
- [0034] 区間判別処理部7は、セグメント(一区間)毎に、平均予測残差パワー演算部6で演算された平均予測残差パワーに基づいて楽曲の種別を判別する。
- [0035] 例えば、区間判別処理部7は、一区間の平均予測残差パワーが予め定められた第一段階の閾値よりも小さい場合には、前記入力された楽曲信号は単一種類楽器により演奏されたものであると判別(楽曲信号の一区間(一セグメント)が単一種類楽器により演奏されたものであると判別)する。
- [0036] また、区間判別処理部7は、前記分析値が第一段階の閾値よりも大きく、かつ前記分析値が第二段階の閾値よりも大きい場合には、前記入力された楽曲信号は複数

種類楽器により演奏されたものであると判別(楽曲信号の一区間(一セグメント)が複数種類楽器により演奏されたものであると判別)する。

[0037] さらに、区間判別処理部7は、前記分析値が第一段階の閾値よりも大きく、かつ前記分析値が第二段階の閾値よりも小さい場合には、前記入力された楽曲信号は、直前の区間の楽曲信号種類(単一種類楽器により演奏された区間か、複数種類楽器により演奏されたものである区間か)によって判別される。また、直前の楽曲信号が、前記分析値が第一段階の閾値よりも大きく、かつ前記分析値が第二段階の閾値よりも小さい場合には、さらにその直前の楽曲信号の種類によって、楽曲信号が判別される。

[0038] 第一段階の閾値および第二段階の閾値は任意の値に設定することができる。詳細は後述する。

[0039] 図2は、本実施形態のピークレベルを有するフレームの除去方法を説明する図である。図2において横軸は時間軸を表し、フレームF26からフレームF1が等間隔(1フレーム単位)で並んでいる状態が示されており、縦軸はフレーム毎の正規化残差パワーを示している。縦軸は、図2において上側がフレームの正規化残差パワーが大きいことを示している。

[0040] 図2においては、1セグメント内の100フレームを簡略化して26フレーム(F1～F26)を用いて表示している。フレームF12は、除去閾値Peak_Thを超えているので、除去されるべきフレームの候補となる。しかし、フレーム12の直前の除去閾値Peak_Thを超えているフレームF3との時間間隔が予め定められている最少ピーク時間間隔よりも短いので、フレームF12は、ピーク除去部5において除去されない。しかし、フレームF19は除去閾値Peak_Thを超えており、しかも直前の除去閾値Peak_Thを超えているフレームF3との時間間隔が予め定められている最少ピーク時間間隔よりも長いので、フレームF19およびフレームF3は、ピーク除去部5において除去される。

[0041] ここで、最少ピーク時間間隔(最少ピーク間隔)は、任意に設定される値である。一例としてはおおよそ250msec. に設定すると本実施形態においては、楽器演奏の演奏を始めるピーク部分のフレームを効率よく除去することができた。また最少ピーク時

間間隔は、サンプリング周波数の整数倍の値とすることができる。

また、除去閾値Peak_Thは、本実施形態においては1セグメント中の最大ピーク値を約0.2倍した値を用いている。この除去閾値Peak_Thを設定することにより本実施形態においては、楽曲種類の判別性能が向上している。

また、ピーク除去部5において除去されるフレームは除去閾値Peak_Thを超えているフレームだけではなく、フレームは除去閾値Peak_Thを超えているフレームの前後のフレームも除去される。楽器の弾き始めは1フレームに限られるわけではないので、影響があると考えられるフレームを除去することにより楽曲種類の判別能力を向上させるためである。

[0042] 一例として、図2において、除去閾値Peak_Thを超えたフレームF3の直前のフレーム(フレームF2)と、フレームF3の直後の三フレーム(フレームF4～フレームF6)をピーク除去部5が除去している。除去閾値Peak_Thを超えたフレームF3の直前より直後のフレーム数が多いのは、楽器の弾き始め直後の方が、演奏された楽曲信号Siのパワーが大きい状態が続くので、その影響を除去するためである。

[0043] また、図2において、フレームF3の前後のフレームと同様に、フレームF19の前後のフレームをピーク除去部5は除去している。すなわち、除去閾値Peak_Thを超えたフレームF19の直前のフレーム(フレームF18)と、フレームF3の直後の三フレーム(フレームF20～フレームF22)をピーク除去部5が除去している。除去閾値Peak_Thを超えたフレームの前後のフレーム数は上述の値に限定されるわけではなく、任意の値を設定することができる。サンプリング周波数によってフレーム間の時間間隔が異なるので、サンプリング周波数が大きいほど除去閾値Peak_Thを超えたフレームの前後の除去されるべきフレーム数を多く、サンプリング周波数が小さいほど除去閾値Peak_Thを超えたフレームの前後の除去されるべきフレーム数を小さく、ピーク除去部5は設定することができる。

[0044] 次に図3(a)を用いて、区間判別処理部7において楽曲信号Siを判別する場合に単一閾値Mono_Thを用いた例について説明する。

[0045] 図3(a)において横軸は時間軸(区間判別処理部7における解析時間)を示し、セグメントSG1からセグメントSG15が並んでいる。縦軸はセグメント毎の平均正規化残

差パワーが示されている。縦軸は、図3(a)において上側ほどセグメントの平均正規化残差パワーが大きいことを示している。

[0046] セグメントSG1からセグメントSG3までは平均正規化残差パワーが単一閾値Mono__Thよりも小さいので、区間判定処理部7は、セグメントSG1からセグメントSG3が単一種類楽器演奏区間であると判別する。

[0047] また、セグメントSG4は平均正規化残差パワーが単一閾値Mono__Thよりも大きいので、区間判定処理部7は、セグメントSG4は複数種類楽器演奏区間であると判別する。

[0048] しかし、セグメントSG4の直後にあるセグメントSG5からセグメントSG7まではセグメントの平均正規化残差パワーが単一閾値Mono__Thよりも小さいので、区間判定処理部7は、再び、セグメントSG5からセグメントSG7が単一種類楽器演奏区間であると判別する。

[0049] このように、単一閾値Mono__Thを用いて区間判別処理をする場合には、セグメントSG1からセグメントSG7までの間に、単一閾値Mono__Thをわずかも超える平均正規化残差パワーを有しているセグメント(セグメント4)は、区間判別処理部7が、複数種類楽器演奏区間であると判別する。この場合、区間判別処理部7が、単一種類楽器演奏区間であると判別したセグメントSG1からセグメントSG7間のほぼ中央付近で、突発的に一セグメントだけ複数種類楽器演奏区間と判別することになり、区間判別処理部7が誤判定をしている可能性が高くなる。

[0050] また、同様に図3(a)において、区間判別処理部7は、セグメントSG8からセグメントSG13の間(セグメント12を除く)は、各セグメントの平均正規化残差パワーが単一閾値Mono__Thよりも大きいので、セグメントSG8からセグメントSG13(セグメント12を除く)を複数種類楽器演奏区間と判別する。しかし、セグメント12の平均正規化残差パワーは、突発的にわずかに単一閾値Mono__Thよりも小さいので、区間判別処理部7はセグメント12を複数単一種類楽器演奏区間と判別し、誤判定している可能性が高くなる。なお、単一閾値Mono__Thは任意の値を設定することができる。そこで、本実施形態においては、図3(b)で説明するように、2段階閾値判別方法を区間判別処理部7に使用することによって、上述した突発的な誤判定を防ぐようにした。

- [0051] 図3(b)は、区間判別処理部7において楽曲信号Siを判別する際に2段階閾値を用いた場合の説明図である。
- [0052] 図3(b)の縦軸、横軸、およびセグメントSG1からセグメントSG15は図3(a)と同一である。
- [0053] 2段階閾値判別方法を使用する場合には、区間判別処理部7は、セグメントの平均正規化残差パワーが第一段階の閾値Tonal_Thよりも小さいセグメントを、単一種類楽器演奏区間と判別する。また、区間判別処理部7は、セグメントの平均正規化残差パワーが第二段階の閾値NonTonal_Thよりも大きいセグメントを、複数種類楽器演奏区間と判別する。
- [0054] そして、セグメントの平均正規化残差パワーが第一段階の閾値Tonal_Thよりも大きく、第二段階の閾値NonTonal_Thよりも小さい場合には、判別の対象となっているセグメントの直前(時刻として前)のセグメントの判別状況に基づいて、区間判別処理部7は、そのセグメントが複数種類楽器演奏区間であるか単一種類楽器演奏区間かを判別する。
- [0055] 例えば、図3(b)において、セグメントSG3の平均正規化残差パワーは、第一段階の閾値Tonal_Thよりも大きく、第二段階の閾値NonTonal_Thよりも小さい。この場合には、セグメントSG3の直前のセグメントSG2が種類楽器演奏区間であると区間判別処理部7によって判別されている(セグメントSG2の平均正規化残差パワーが第一段階の閾値Tonal_Thよりも小さいので、区間判別処理部7はセグメントSG2を単一種類楽器演奏区間であると判別する。)ので、区間判別処理部7はセグメントSG3を単一種類楽器演奏区間であると判別する。
- [0056] すなわち、あるセグメントの平均正規化残差パワーが第一段階の閾値Tonal_Thよりも大きく、第二段階の閾値NonTonal_Thよりも小さい場合であって、かつ判別の対象となっているセグメントの直前のセグメントが、単一種類楽器演奏区間であると判別されている場合には、区間判別処理部7はそのセグメントを単一種類楽器演奏区間であると判別する。
- [0057] また、あるセグメントの平均正規化残差パワーが第一段階の閾値Tonal_Thよりも大きく、第二段階の閾値NonTonal_Thよりも小さい場合であって、かつ判別の対

象となっているセグメントの直前のセグメントが、複数種類楽器演奏区間であると判別されている場合には、区間判別処理部7はそのセグメントを複数種類楽器演奏区間であると判別する。

[0058] 図3(a)において、区間判別処理部7が、突発的に誤判定していたセグメントSG12について、2段階閾値判別方法を使用した図3(b)について説明する。

[0059] 図3(b)において、セグメントSG12の平均正規化残差パワーが第一段階の閾値Tonal_Thよりも大きく、第二段階の閾値NonTonal_Thよりも小さいので、セグメントSG12は、直前のセグメントSG11の判別状況に基づいて、区間判別処理部7がその区間の楽曲種別を判別する(セグメントSG11が複数種類楽器演奏区間である場合には、セグメントSG12は複数種類楽器演奏区間であると判別され、セグメントSG11が単一種類楽器演奏区間である場合には、セグメント12は単一種類楽器演奏区間であると判別される。)。

[0060] しかし、セグメントSG11の平均正規化残差パワーも第一段階の閾値Tonal_Thよりも大きく、第二段階の閾値NonTonal_Thよりも小さいので、セグメントSG11は、直前のセグメントSG10の判別状況に基づいて、区間判別処理部7がその区間の楽曲種別を判別する(すなわち、セグメントSG10が複数種類楽器演奏区間である場合には、セグメントSG11は複数種類楽器演奏区間であると判別され、セグメントSG10が単一種類楽器演奏区間である場合には、セグメントSG11は単一種類楽器演奏区間であると判別される。)。

[0061] ここで、セグメントSG10の平均正規化残差パワーが第二段階の閾値NonTonal_Thよりも大きいので、セグメントSG10は、複数種類楽器演奏区間であると区間判別処理部7によって判別される。従って、セグメントSG11は、セグメントSG11の直前のセグメント10と同じく、複数種類楽器演奏区間であると区間判別処理部7によって判別される。さらに、セグメントSG12も、セグメントSG12の直前のセグメントSG11と同じく、複数種類楽器演奏区間であると区間判別処理部7によって判別される。

[0062] このように、2段階閾値判別方法を使用すると単一閾値を用いた場合に比べて、突発的に閾値をわずかに超える(または突発的に閾値にわずかに満たない)セグメントを誤判定する確率が小さくなり、楽曲種別の判別の精度が向上する。

[0063] 次に図4を用いて平均正規化残差パワー度数分布の一例について説明する。

[0064] 図4は区間判別処理部7に入力される各セグメントの平均正規化残差パワー度数分布を示した図の一例である。図4の横軸は、平均正規化残差パワーを示し、図4において右へ行くほど平均正規化残差パワーが大きいことを示す。図4の縦軸は、横軸で示されるある平均正規化残差パワーを有するセグメントの個数を示し、図4において棒グラフが高いほど、その平均正規化残差パワーを有するセグメントの数が多いことを示す。

図4における白抜き棒グラフは単一種類楽器演奏セグメントの分布を示し、黒塗り棒グラフは複数種類楽器演奏セグメントの分布を示す。図4においてサンプリングした楽曲はクラシックおよびポップス等の楽曲であって、単一種類楽器(ピアノおよびギターなどのアコースティック楽器)による演奏を5曲(一曲30sec)、複数種類楽器(アコースティック楽器及び非アコースティック楽器)による演奏を5曲(一曲30sec)、それぞれサンプリング(サンプリング周波数は44.1kHz、線形予測フィルタ係数は3次まで使用。)している。

[0065] 複数種類楽器による演奏セグメントは、図4において右側に集中している。また、単一種類楽器による演奏セグメントは、図4において左側に集中している。すなわち、複数種類楽器による演奏セグメントは、平均正規化残差パワーが大きく(平均正規化残差パワーが0.006以上に集中している。)、単一種類楽器による演奏セグメントは、平均正規化残差パワーが小さい(平均正規化残差パワーが0.003以下に集中している。)ことが示されている。

[0066] 図4の場合には、平均正規化残差パワーの閾値として、第一段階の閾値Tonal_Thを0.003とし、第二段階の閾値NonTonal_Thを0.006とすれば、単一閾値を用いることによる誤判定を防ぐことができることがわかる。

[0067] ただし、第一段階の閾値Tonal_Thおよび第二段階の閾値NonTonal_Thは絶対的な値ではなく、任意の数値が設定されることが可能である。また、第一段階の閾値Tonal_Thおよび第二段階の閾値NonTonal_Thは、線形予測フィルタ係数の次数によって変化する値でもある。

[0068] 図4とは別に、ピアノメインのポップス系の洋楽3曲(30秒/曲)について、2段階閾

値判別方法を使用した場合と、単一閾値を用いた場合について誤判定率を比較した。誤判定率は(複数種類楽器演奏区間を単一種類楽器演奏区間と判定したセグメントの個数)/(単一種類楽器演奏区間と判定したセグメントの総数)で示した。

[0069] その結果、曲Aについては、単一閾値を用いた場合の誤判定率は7.6%であったが、2段階閾値判別方法を使用した場合には誤判定率は0%となった。また、曲Bについては、単一閾値を用いた場合の誤判定率は40.0%であったが、2段階閾値判別方法を使用した場合には誤判定率は0%となった。さらに、曲Cについては、単一閾値を用いた場合の誤判定率は25.0%であったが、2段階閾値判別方法を使用した場合には誤判定率は0%となった。

[0070] このように、本実施形態における2段階閾値判別方法を使用した場合には、複数種類楽器演奏区間を単一種類楽器演奏区間と誤判定する確率が著しく減少することが確認された。

[0071] 次に、図5を用いて、本実施形態に係わる動作のフローを説明する。

[0072] ステップS1において、フレーム分割部1aは入力された楽曲信号Siをフレーム分割する。

[0073] ステップS2において、線形予測フィルタ係数演算部1bがフレーム分割部1aにおいて分割されたフレームに基づいて、線形予測フィルタ係数を演算する。そして、線形予測フィルタ演算部2aが、楽曲信号Siと線形予測フィルタ係数に基づいて、線形予測フィルタを演算する。

[0074] ステップS3において、演算部2bは残差成分を演算する。そして予測残差パワー演算部3は、フレーム毎に残差パワーを演算するとともに、原音である楽曲信号Siを用いてフレーム毎に正規化予測残差パワーを演算する。

[0075] ステップS4において、バッファリング部4に100フレーム蓄積(記憶)されたか否かが判断される。バッファリング部4に100フレーム蓄積されていない場合(ステップS4:NO)にはステップS1に進む。バッファリング部4に100フレーム蓄積された場合(ステップS4:YES)にはステップS5に進む。

[0076] ステップS5において、ピーク除去部5は図2において説明した手法を用いて、1セグメント内の正規化残差パワー立ち上がり部のピークとなるフレームを除去する。

- [0077] ステップS6において、平均予測残差パワー演算部3は1セグメント内の平均正規化予測残差パワーを演算する。
- [0078] ステップS7において、区間判別処理部7は判別対象となるセグメントの平均正規化予測残差パワー値が第一段階の閾値Tonal_Thよりも小さいか否かを判断する。区間判別処理部7が判別対象となるセグメントの平均正規化予測残差パワー値が第一段階の閾値Tonal_Thよりも小さい(ステップS7:YES)と判断した場合には、そのセグメント(区間)は単一種類楽器演奏区間であると判別される。また、区間判別処理部7が判別対象となるセグメントの平均正規化予測残差パワー値が第一段階の閾値Tonal_Thよりも大きい(ステップS7:NO)と判断した場合には、ステップS8に進む。
- [0079] ステップS8において、区間判別処理部7は判別対象となるセグメントの平均正規化予測残差パワー値が第二段階の閾値NonTonal_Thよりも小さいか否かを判断する。区間判別処理部7が判別対象となるセグメントの平均正規化予測残差パワー値が第二段階の閾値NonTonal_Thよりも大きい(ステップS8:YES)と判断した場合には、そのセグメント(区間)は複数種類楽器演奏区間であると判別される。また、区間判別処理部7が判別対象となるセグメントの平均正規化予測残差パワー値が第二段階の閾値NonTonal_Thよりも小さい(ステップS8:NO)と判断した場合には、ステップS9に進む。
- [0080] ステップS9において、区間判別処理部7は判別対象となるセグメントの直前のセグメントが単一種類楽器演奏区間と判別されたか否かを調べる。区間判別処理部7は判別対象となるセグメントの直前のセグメントが単一種類楽器演奏区間と判別した場合(ステップS9:YES)には、判別の対象となっているセグメントを単一種類楽器演奏区間と判別する。また、区間判別処理部7は判別対象となるセグメントの直前のセグメントが複数種類楽器演奏区間と判別した場合(ステップS:NO)には、判別の対象となっているセグメントを複数種類楽器演奏区間と判別する。
- [0081] ステップS7、ステップS8、およびステップS9において、区間の楽曲種類判別が実行された後に、ステップS10において、区間判別処理部7は楽曲信号Siの入力が終了したか否かを判断する。楽曲信号Siの入力が終了していない場合(ステップS10:

NO)の場合には、ステップS1に進む。また、楽曲信号Siの入力が終了した場合(ステップS10:YES)の場合には処理が終了する。

[0082] 以上説明したように、本発明によれば、FFT等の周波数分析演算を使用しないので、少ない演算量でかつ正確に単一種類楽器による演奏区間と複数種類楽器による演奏区間とを判別することが可能となった。

[0083] また、本発明によれば、入力信号はセグメント単位またはフレーム単位(予め定められた単位)で演算されるので、演算量が少なく、高速かつ正確に演算を行なうことが可能となった。

[0084] さらに、本発明によれば、時系列分析が線形予測分析で行なわれるので、時系列分析を高速フーリエ変換FFTで行った場合のように膨大な計算量(重たい演算)と計算時間とが必要なくなる。したがって、簡易な構成による少ない演算量によって、単一種類楽器による演奏区間と複数種類楽器による演奏区間を判別するための時系列分析を行なうことができるようになった。

[0085] さらに本発明によれば、入力信号のレベルおよび入力信号の雑音レベルによる判断ミスを最小限に減少させることが可能となった。

[0086] さらに、本発明によれば、2段階閾値判別方法を使用することにより、突発的なレベル変動による単一種類楽器による演奏区間と複数種類楽器による演奏区間との判別ミスを減少させることが可能となった。

[0087] さらに、本発明によれば、楽器演奏を開始した直後のピークレベルを持つ楽曲信号を除去することによって、単一種類楽器による演奏区間と複数種類楽器による演奏区間との判別をより正確に実行することが可能となった。

[0088] さらに、本発明によれば、楽器演奏を開始した直後のピークレベルを持つ楽曲信号の前後の楽曲信号を除去するようにしたので、単一種類楽器による演奏区間と複数種類楽器による演奏区間との判別をさらに正確に実行することが可能となった。

[0089] また、本実施の形態で説明した楽曲種類判別装置は、家庭(ホーム)、車(カー)またはポータブル機器におけるミュージックサーバに適用できる他に、音楽配信技術、楽曲情報に関するコンピュータプログラムにも適用することができる。さらにホームシアターシステム、PDP等の薄型テレビ、PC、ポータブルDVD等のパーソナルサラウ

ンドシステムに適用することが可能である。

[0090] なお、図5における動作手順を、ハードディスク等の記録媒体に予め記録しておく、或いはインターネット等のネットワークを介して予め記録しておく、これを汎用のマイクロコンピュータ等により読み出して実行することにより、当該汎用のマイクロコンピュータ等を実施形態に係わるCPUとして機能させることも可能である。

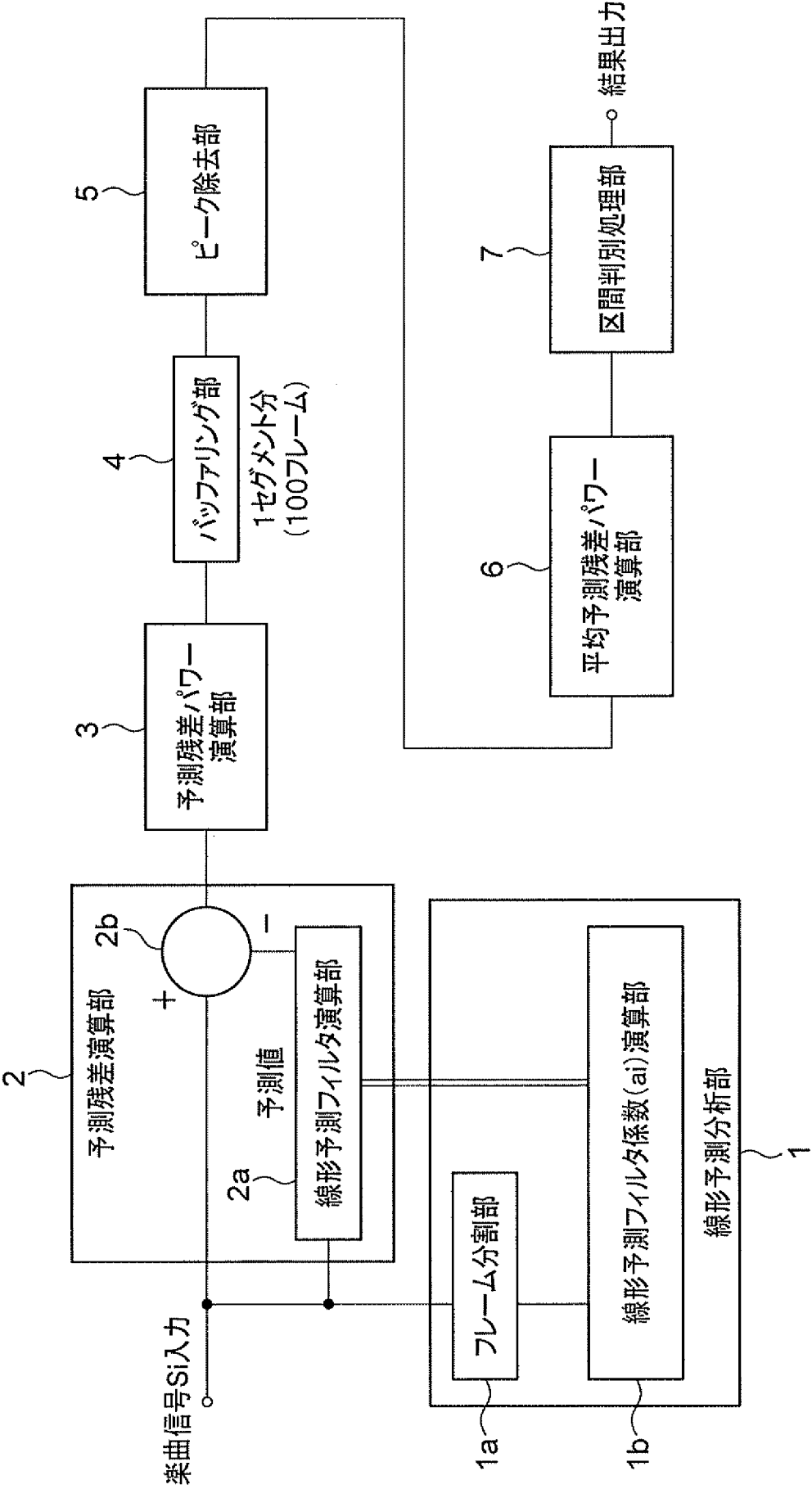
請求の範囲

- [1] 入力された楽曲信号に基づいて、前記楽曲信号を分析し、楽曲の種類の判別を行なう楽曲種類判別装置において、
- 前記楽曲信号を時系列分析する分析手段と、
- 前記分析手段によって分析された分析値に基づいて、前記楽曲信号の種類を判別する判別手段と、
- を備えることを特徴とする楽曲種類判別装置。
- [2] 請求項1に記載の楽曲種類判別装置において、
- 前記楽曲信号を予め定められた単位に分割する分割手段を更に備え、
- 前記分析手段、前記判別手段は前記分割手段によって分割された前記楽曲信号の単位に基づいて演算処理を行なうことを特徴とする楽曲種類判別装置。
- [3] 請求項1または2に記載の楽曲種類判別装置において、
- 前記分析手段における時系列分析は線形予測分析であり、前記分析値は前記楽曲信号の線形予測分析による線形予測残差であることを特徴とする楽曲種類判別装置。
- [4] 請求項1乃至3のいずれか一項に記載の楽曲種類判別装置において、
- 前記分析手段は、前記分析結果の電力と前記楽曲信号の電力との比の値を前記分析値とすることを特徴とする楽曲種類判別装置。
- [5] 請求項1乃至4のいずれか一項に記載の楽曲種類判別装置において、
- 前記判別手段は、前記分析値が第一段階の閾値よりも小さい場合には、前記入力された楽曲信号は単一種類楽器により演奏されたものであると判断し、前記分析値が第一段階の閾値よりも大きい場合であり、かつ前記分析値が第二段階の閾値よりも小さい場合には、前記入力された楽曲信号は複数種類楽器により演奏されたものであると判断し、前記分析値が第一段階の閾値よりも大きい場合であり、かつ前記分析値が第二段階の閾値よりも大きい場合には、前記入力された楽曲信号は前記単一種類楽器により演奏されたものであると判断することを特徴とする楽曲種類判別装置。
- [6] 請求項1乃至5のいずれか一項に記載の楽曲種類判別装置において、

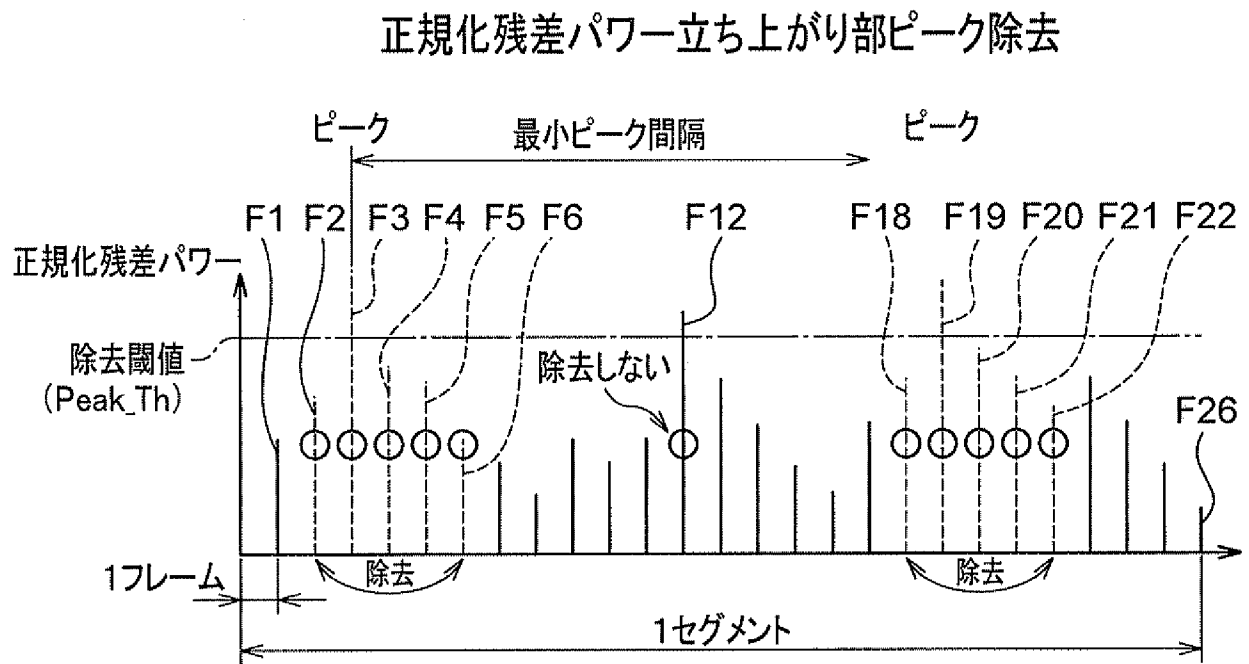
予め定められた除去閾値を超えるパワーを持つ前記楽曲信号のピーク信号部分が予め定められた時間以上離れている場合には、前記ピーク信号部分を除去するピーク信号除去部を更に備えることを特徴とする楽曲種類判別装置。

- [7] 請求項6に記載の楽曲種類判別装置において、
前記ピーク除去部は、前記ピーク信号の時間軸において予め定められた前後の前記楽曲信号を当該ピーク信号とともに除去することを特徴とする楽曲種類判別装置。
- [8] 入力された楽曲信号に基づいて、前記楽曲信号を分析し、楽曲の種類の判別を行なう楽曲種類判別方法において、
前記楽曲信号を時系列分析する分析工程と、
前記分析工程において分析された分析値に基づいて、前記楽曲信号の種類を判別する判別工程と、
を備えることを特徴とする楽曲種類判別方法。
- [9] 請求項1に記載の楽曲種類判別装置に含まれるコンピュータを、
前記楽曲信号を時系列分析する分析手段、
前記分析手段によって分析された分析値に基づいて、前記楽曲信号の種類を判別する判別手段、
として機能させることを特徴とする楽曲種類判別プログラム。

[図1]



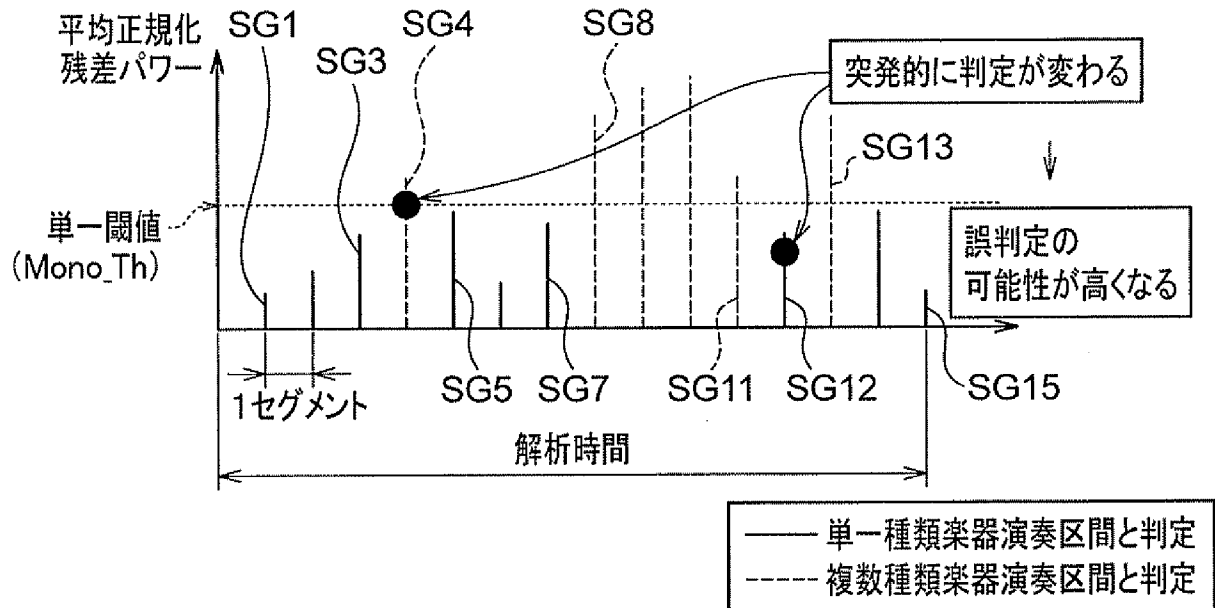
[図2]



[図3]

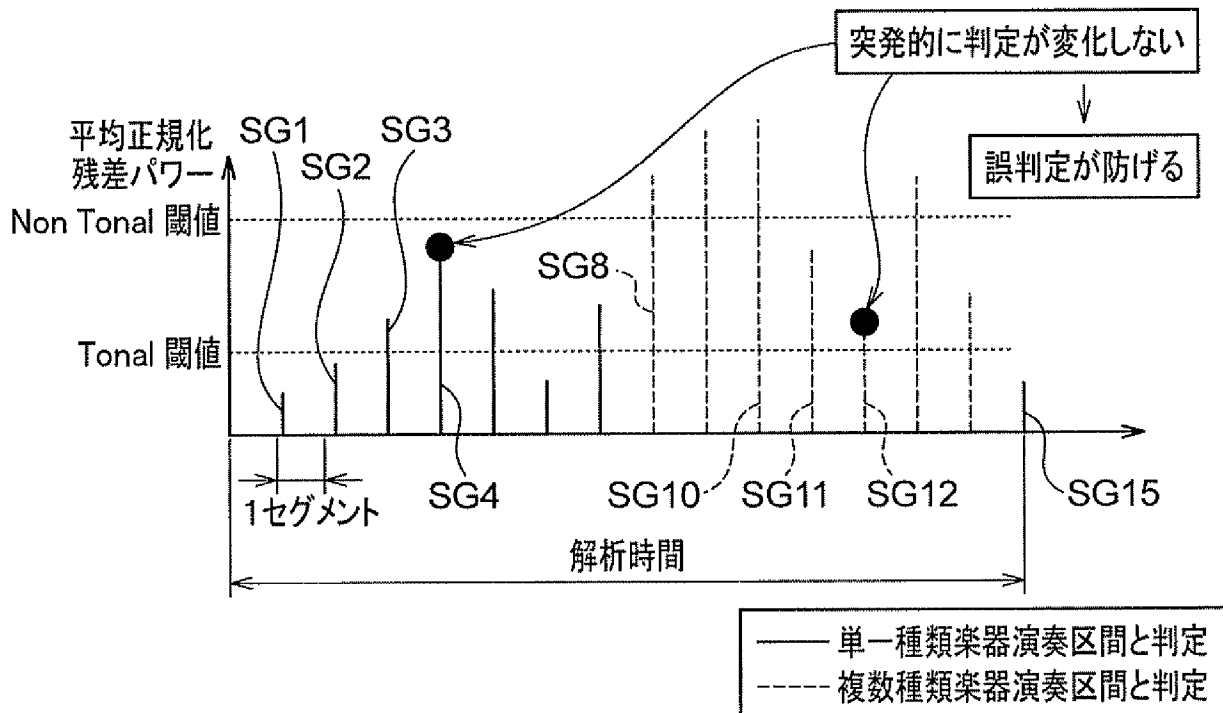
(a)

単一閾値を用いる場合



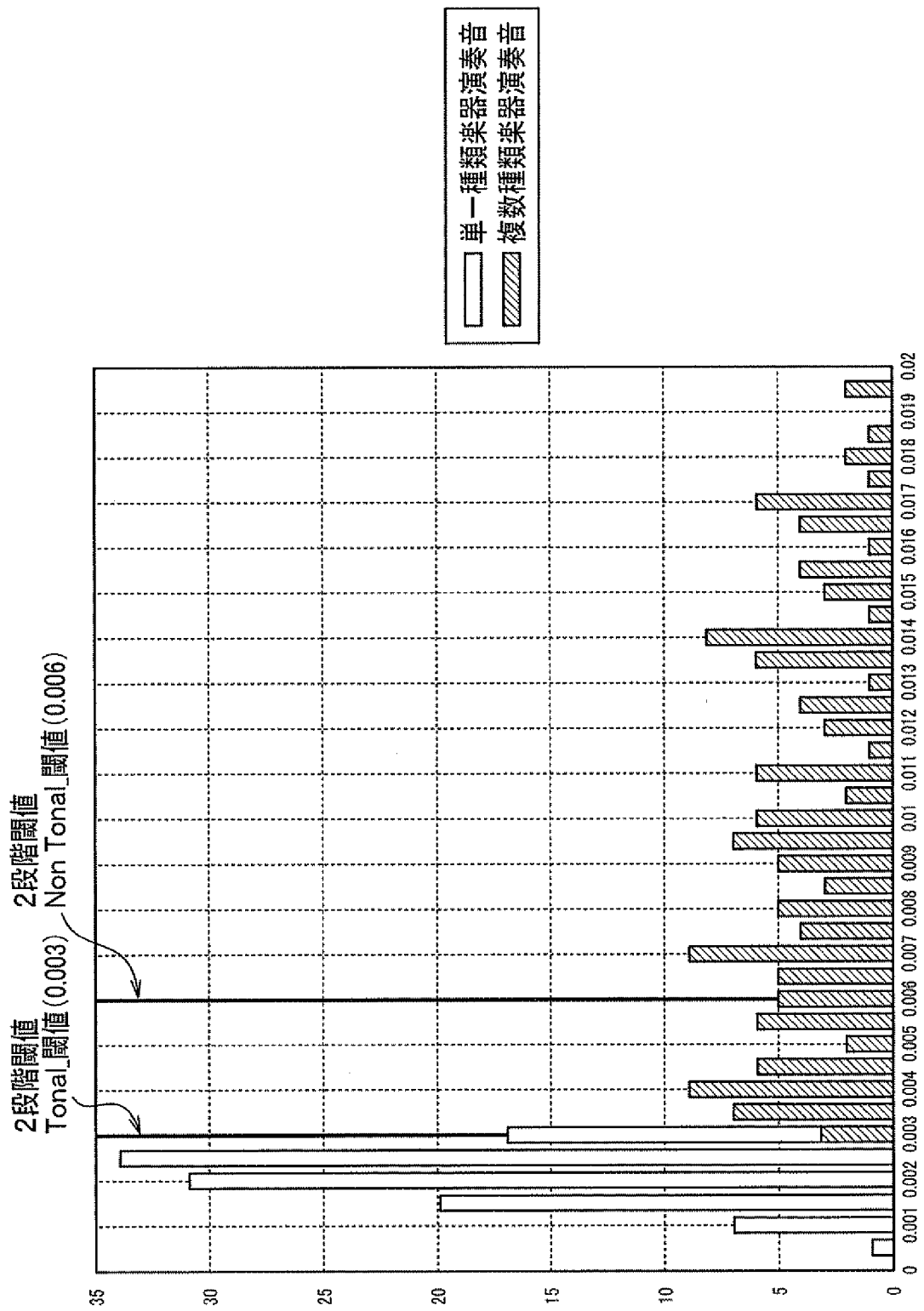
(b)

2段階閾値を用いる場合

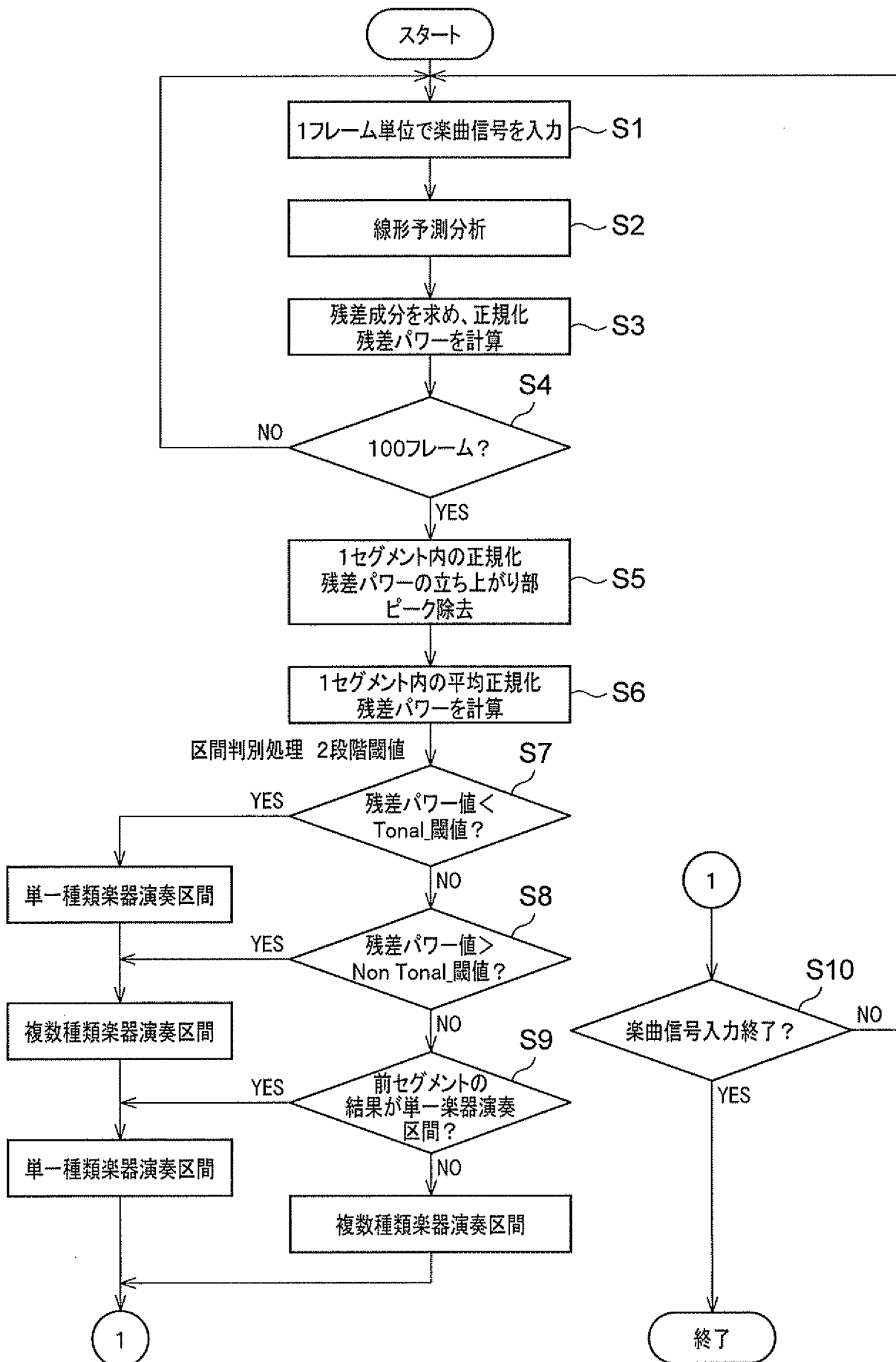


[図4]

平均正規化残差パワー一度数分布



[図5]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2007/055899

A. CLASSIFICATION OF SUBJECT MATTER

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)

G10L11/00-21/06

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	1996-2007
Kokai Jitsuyo Shinan Koho	1971-2007	Toroku Jitsuyo Shinan Koho	1994-2007

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.
X	JP 2001-142480 A (Sony Corp.), 25 May, 2001 (25.05.01), Full text; all drawings & EP 1100073 A2 & US 6990443 B1	1-4, 8, 9
A	JP 2003-5785 A (National Institute of Advanced Industrial Science and Technology), 08 January, 2003 (08.01.03), Full text; all drawings (Family: none)	1-9
A	JP 2002-62892 A (Nippon Hoso Kyokai), 28 February, 2002 (28.02.02), Full text; all drawings (Family: none)	1-9

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"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

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search
09 May, 2007 (09.05.07)

Date of mailing of the international search report
22 May, 2007 (22.05.07)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類（国際特許分類（I P C）） Int.Cl. 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（I P C）） Int.Cl. G10L11/00-21/06			
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 1971-2007年 日本国実用新案登録公報 1996-2007年 日本国登録実用新案公報 1994-2007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	
X	JP 2001-142480 A(ソニー株式会社) 2001.05.25, 全文, 全図 & EP 1100073 A2 & US 6990443 B1	1-4, 8, 9	
A	JP 2003-5785 A(独立行政法人産業技術総合研究所) 2003.01.08, 全文, 全図(ファミリーなし)	1-9	
A	JP 2002-62892 A(日本放送協会) 2002.02.28, 全文, 全図(ファミリーなし)	1-9	
☐ C欄の続きにも文献が列挙されている。 ☐ 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 09.05.2007		国際調査報告の発送日 22.05.2007	
国際調査機関の名称及びあて先 日本国特許庁（I S A / 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） 涌井 智則	5 Z 9 5 6 8
		電話番号 03-3581-1101	内線 3541