

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5732994号
(P5732994)

(45) 発行日 平成27年6月10日(2015.6.10)

(24) 登録日 平成27年4月24日(2015.4.24)

(51) Int.Cl.

F 1

G 1 O L 25/51 (2013.01)
G 1 O L 25/54 (2013.01)G 1 O L 25/51 300
G 1 O L 25/54

請求項の数 8 (全 29 頁)

(21) 出願番号 特願2011-92987 (P2011-92987)
 (22) 出願日 平成23年4月19日 (2011.4.19)
 (65) 公開番号 特開2012-226080 (P2012-226080A)
 (43) 公開日 平成24年11月15日 (2012.11.15)
 審査請求日 平成26年3月27日 (2014.3.27)

(73) 特許権者 000002185
 ソニー株式会社
 東京都港区港南1丁目7番1号
 (74) 代理人 100082131
 弁理士 稲本 義雄
 (74) 代理人 100121131
 弁理士 西川 孝
 (72) 発明者 濵谷 崇
 東京都港区港南1丁目7番1号 ソニー株式会社内
 (72) 発明者 安部 素嗣
 東京都港区港南1丁目7番1号 ソニー株式会社内

審査官 山下 剛史

最終頁に続く

(54) 【発明の名称】楽曲検索装置および方法、プログラム、並びに記録媒体

(57) 【特許請求の範囲】

【請求項 1】

入力信号を楽曲のみからなる参照信号と比較することによって、前記入力信号の楽曲を同定する楽曲検索装置において、

時間周波数領域に変換された前記入力信号の各領域の信号成分の強さと、前記信号成分の強さを近似した関数とに基づいて、前記信号成分のトーンらしさの指標を算出する指標算出部と、

前記指標に基づいて求められる前記入力信号の各領域における特徴量と、時間周波数領域に変換された前記参照信号について算出された前記指標に基づいて求められる前記参照信号の各領域における前記特徴量の類似度を算出する類似度算出部と、

前記類似度に基づいて、前記入力信号の楽曲を同定する楽曲同定部と

を備え、

前記指標算出部は、

所定の時間区間の前記入力信号から、前記信号成分の強さが極大となる点を検出する極大点検出部と、

前記極大となる点近傍の前記信号成分の強さをテンプレート関数で近似する近似処理部とを備え、

前記極大となる点近傍の前記信号成分の強さと前記テンプレート関数との誤差に基づいて、前記指標を算出する

楽曲検索装置。

10

20

【請求項 2】

前記指標算出部は、前記テンプレート関数の曲率に応じて前記指標を調整する
請求項1に記載の楽曲検索装置。

【請求項 3】

前記指標算出部は、前記テンプレート関数が極大となる点の周波数に応じて前記指標を
調整する

請求項2に記載の楽曲検索装置。

【請求項 4】

前記入力信号の各領域における第1の特徴量と、前記参照信号の各領域における前記第
1の特徴量とに基づいて、前記入力信号および前記参照信号の各領域に対する重み分布を
生成するとともに、前記入力信号の各領域における第1の特徴量と、前記参照信号の各領
域における前記第1の特徴量との第1の特徴量同士の類似度を算出する重み分布生成手段
をさらに備え、

10

前記類似度算出部は、前記重み分布による重み付けおよび前記第1の特徴量同士の類似
度に基づいて、前記入力信号の各領域における第2の特徴量と、前記参照信号の各領域に
おける前記第2の特徴量の類似度を算出する

請求項1または2に記載の楽曲検索装置。

【請求項 5】

前記指標を時間方向にフィルタリングすることにより得られる前記指標の時間平均量を
、前記第1の特徴量として算出する第1の特徴量算出部と、

20

前記指標を時間方向にフィルタリングすることにより得られる前記指標の時間変化量を
、前記第2の特徴量として算出する第2の特徴量算出部とをさらに備える

請求項4に記載の楽曲検索装置。

【請求項 6】

入力信号を楽曲のみからなる参照信号と比較することによって、前記入力信号の楽曲を
同定する楽曲検索方法において、

時間周波数領域に変換された前記入力信号の各領域の信号成分の強さと、前記信号成分
の強さを近似した関数とに基づいて、前記信号成分のトーンらしさの指標を算出する指標
算出ステップと、

前記指標に基づいて求められる前記入力信号の各領域における特徴量と、時間周波数領域
に変換された前記参照信号について算出された前記指標に基づいて求められる前記参照
信号の各領域における前記特徴量の類似度を算出する類似度算出ステップと、

30

前記類似度に基づいて、前記入力信号の楽曲を同定する楽曲同定ステップと
を含み、

前記指標算出ステップは、

所定の時間区間の前記入力信号から、前記信号成分の強さが極大となる点を検出する
極大点検出ステップと、

前記極大となる点近傍の前記信号成分の強さをテンプレート関数で近似する近似処理
ステップとを含み、

前記極大となる点近傍の前記信号成分の強さと前記テンプレート関数との誤差に基づい
て、前記指標を算出する

40

楽曲検索方法。

【請求項 7】

入力信号を楽曲のみからなる参照信号と比較することによって、前記入力信号の楽曲を
同定する信号処理をコンピュータに実行させるプログラムにおいて、

時間周波数領域に変換された前記入力信号の各領域の信号成分の強さと、前記信号成分
の強さを近似した関数とに基づいて、前記信号成分のトーンらしさの指標を算出する指標
算出ステップと、

前記指標に基づいて求められる前記入力信号の各領域における特徴量と、時間周波数領域
に変換された前記参照信号について算出された前記指標に基づいて求められる前記参照

50

信号の各領域における前記特徴量の類似度を算出する類似度算出ステップと、
前記類似度に基づいて、前記入力信号の楽曲を同定する楽曲同定ステップと
を含む処理をコンピュータに実行させ、
前記指標算出ステップは、

所定の時間区間の前記入力信号から、前記信号成分の強さが極大となる点を検出する
極大点検出ステップと、

前記極大となる点近傍の前記信号成分の強さをテンプレート関数で近似する近似処理
ステップとを含み、

前記極大となる点近傍の前記信号成分の強さと前記テンプレート関数との誤差に基づい
て、前記指標を算出する

プログラム。

【請求項 8】

請求項 7 に記載のプログラムが記録されている コンピュータ読み取り可能な記録媒体。
【発明の詳細な説明】

【技術分野】

【0001】

本技術は、楽曲検索装置および方法、プログラム、並びに記録媒体に関し、特に、入力信号から楽曲を同定することができるようとする楽曲検索装置および方法、プログラム、並びに記録媒体に関する。

【背景技術】

【0002】

従来、入力信号として入力される楽曲を同定するために、入力信号の特徴量と、同定される楽曲の候補となる参照信号の特徴量とのマッチング処理が行われている。しかしながら、例えば、ドラマ等のテレビジョン番組の放送音源を入力信号とした場合、入力信号には、BGM (Back Ground Music) としての楽曲の信号成分と、人の会話や騒音（環境雑音）、ホワイトノイズやピンクノイズ、効果音等、楽曲以外の信号成分である雑音成分（以下、単に雑音ともいう）とが混在していることが多く、この雑音による入力信号の特徴量の変化が、マッチング処理の結果に影響を及ぼしてしまう。

【0003】

そこで、入力信号の特徴量において、信頼度の低い成分をマスクするマスクパターンを用いて、信頼度の高い成分のみを用いてマッチング処理を行う技術が提案されている。

【0004】

具体的には、時間周波数領域の信号に変換された入力信号の特徴量を表現する特徴行列に対して、所定の時間周波数領域に対応する行列成分をマスクする複数種類のマスクパターンを予め用意し、入力信号の特徴量とデータベース内の複数の参照信号の特徴量とのマッチング処理を全てのマスクパターンを用いて行い、最も高い類似度が算出された参照信号の楽曲を、入力信号の楽曲として同定するようにしたものがある（例えば、特許文献 1 参照）。

【先行技術文献】

【特許文献】

【0005】

【特許文献 1】特開 2009 - 276776 号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、入力信号において、どの時間にどの周波数の雑音が重畠されているかは予測できず、そのような入力信号に適したマスクパターンを予め用意することは困難であるので、特許文献 1 の技術では、適切なマッチング処理を行うことができず、楽曲と雑音とが混在する入力信号から、その楽曲を精度良く同定することはできない。

【0007】

10

20

30

40

50

一方で、入力信号において平均パワーが大きい時間区間における成分を、楽曲以外の雑音が重畠されている成分と仮定し、入力信号において平均パワーが小さい時間区間の特徴量のみを用いてマッチングを行うようにするマスクパターンを作成する技術がある。この手法によれば、入力信号に応じたマスクパターンを作成することはできるが、そのマスクパターンは、周波数成分が考慮されておらず、入力信号に適したマスクパターンとは言えない。さらに、図1の左側に示されるように、時間周波数領域の入力信号において、楽曲の信号成分Dmに人の会話による雑音Dvが含まれている場合、この技術では、人の会話が途切れている領域S1およびS2におけるわずかな時間区間の特徴量のみを用いてしかマッチング処理を行うことができないので、楽曲と雑音とが混在する入力信号から、その楽曲を精度良く同定することは困難である。楽曲と雑音とが混在する入力信号から、その楽曲を精度良く同定するには、図1の右側に示されるように、領域S3およびS4における楽曲の信号成分Dmの特徴量を用いてマッチング処理を行うことが望ましい。

10

【0008】

本技術は、このような状況に鑑みてなされたものであり、入力信号から楽曲を精度良く同定することができるようとするものである。

【課題を解決するための手段】

【0009】

本技術の一侧面の楽曲検索装置は、入力信号を楽曲のみからなる参照信号と比較することによって、前記入力信号の楽曲を同定する楽曲検索装置であって、時間周波数領域に変換された前記入力信号の各領域の信号成分の強さと、前記信号成分の強さを近似した関数とに基づいて、前記信号成分のトーンらしさの指標を算出する指標算出部と、前記指標に基づいて求められる前記入力信号の各領域における特徴量と、時間周波数領域に変換された前記参照信号について算出された前記指標に基づいて求められる前記参照信号の各領域における前記特徴量の類似度を算出する類似度算出部と、前記類似度に基づいて、前記入力信号の楽曲を同定する楽曲同定部とを備え、前記指標算出部は、所定の時間区間の前記入力信号から、前記信号成分の強さが極大となる点を検出する極大点検出部と、前記極大となる点近傍の前記信号成分の強さをテンプレート関数で近似する近似処理部とを備え、前記極大となる点近傍の前記信号成分の強さと前記テンプレート関数との誤差に基づいて、前記指標を算出する。

20

【0011】

30

前記指標算出部には、前記テンプレート関数の曲率に応じて前記指標を調整させることができる。

【0012】

前記指標算出部には、前記テンプレート関数が極大となる点の周波数に応じて前記指標を調整させることができる。

【0013】

前記楽曲検索装置には、前記入力信号の各領域における第1の特徴量と、前記参照信号の各領域における前記第1の特徴量とに基づいて、前記入力信号および前記参照信号の各領域に対する重み分布を生成するとともに、前記入力信号の各領域における第1の特徴量と、前記参照信号の各領域における前記第1の特徴量との第1の特徴量同士の類似度を算出する重み分布生成手段をさらに設け、前記類似度算出部には、前記重み分布による重み付けおよび前記第1の特徴量同士の類似度に基づいて、前記入力信号の各領域における第2の特徴量と、前記参照信号の各領域における前記第2の特徴量の類似度を算出させることができる。

40

【0014】

前記楽曲検索装置には、前記指標を時間方向にフィルタリングすることにより得られる前記指標の時間平均量を、前記第1の特徴量として算出する第1の特徴量算出部と、前記指標を時間方向にフィルタリングすることにより得られる前記指標の時間変化量を、前記第2の特徴量として算出する第2の特徴量算出部とをさらに設けることができる。

【0015】

50

本技術の一側面の楽曲検索方法は、入力信号を楽曲のみからなる参照信号と比較することによって、前記入力信号の楽曲を同定する楽曲検索方法であって、時間周波数領域に変換された前記入力信号の各領域の信号成分の強さと、前記信号成分の強さを近似した関数とに基づいて、前記信号成分のトーンらしさの指標を算出する指標算出ステップと、前記指標に基づいて求められる前記入力信号の各領域における特徴量と、時間周波数領域に変換された前記参照信号について算出された前記指標に基づいて求められる前記参照信号の各領域における前記特徴量の類似度を算出する類似度算出ステップと、前記類似度に基づいて、前記入力信号の楽曲を同定する楽曲同定ステップとを含み、前記指標算出ステップは、所定の時間区間の前記入力信号から、前記信号成分の強さが極大となる点を検出する極大点検出ステップと、前記極大となる点近傍の前記信号成分の強さをテンプレート関数で近似する近似処理ステップとを含み、前記極大となる点近傍の前記信号成分の強さと前記テンプレート関数との誤差に基づいて、前記指標を算出する。

10

【0016】

本技術の一側面のプログラムおよび一側面の記録媒体に記録されるプログラムは、入力信号を楽曲のみからなる参照信号と比較することによって、前記入力信号の楽曲を同定する信号処理をコンピュータに実行させるプログラムであって、時間周波数領域に変換された前記入力信号の各領域の信号成分の強さと、前記信号成分の強さを近似した関数とに基づいて、前記信号成分のトーンらしさの指標を算出する指標算出ステップと、前記指標に基づいて求められる前記入力信号の各領域における特徴量と、時間周波数領域に変換された前記参照信号について算出された前記指標に基づいて求められる前記参照信号の各領域における前記特徴量の類似度を算出する類似度算出ステップと、前記類似度に基づいて、前記入力信号の楽曲を同定する楽曲同定ステップとを含む処理をコンピュータに実行させ、前記指標算出ステップは、所定の時間区間の前記入力信号から、前記信号成分の強さが極大となる点を検出する極大点検出ステップと、前記極大となる点近傍の前記信号成分の強さをテンプレート関数で近似する近似処理ステップとを含み、前記極大となる点近傍の前記信号成分の強さと前記テンプレート関数との誤差に基づいて、前記指標を算出する。

20

【0017】

本技術の一側面においては、時間周波数領域に変換された入力信号の各領域の信号成分の強さと、信号成分の強さを近似した関数とに基づいて、信号成分のトーンらしさの指標が算出され、指標に基づいて求められる入力信号の各領域における特徴量と、時間周波数領域に変換された参照信号について算出された指標に基づいて求められる参照信号の各領域における特徴量の類似度が算出され、類似度に基づいて、入力信号の楽曲が同定される。特に、所定の時間区間の入力信号から、信号成分の強さが極大となる点が検出され、極大となる点近傍の信号成分の強さがテンプレート関数で近似され、極大となる点近傍の信号成分の強さとテンプレート関数との誤差に基づいて、指標が算出される。

30

【発明の効果】

【0018】

本技術の一側面によれば、入力信号から楽曲を精度良く同定することが可能となる。

【図面の簡単な説明】

【0019】

40

【図1】マッチング処理に用いられる入力信号の特徴量について説明する図である。

【図2】本技術を適用した楽曲検索装置の一実施の形態の構成を示すブロック図である。

【図3】指標算出部の機能構成例を示すブロック図である。

【図4】第1の特徴量算出部の機能構成例を示すブロック図である。

【図5】第2の特徴量算出部の機能構成例を示すブロック図である。

【図6】楽曲同定処理について説明するフローチャートである。

【図7】入力信号解析処理について説明するフローチャートである。

【図8】指標算出処理について説明するフローチャートである。

【図9】ピークの検出について説明する図である。

【図10】ピーク周辺のパワースペクトルの近似について説明する図である。

50

【図11】指標の調整関数について説明する図である。

【図12】入力信号のトーンらしさの指標の例について説明する図である。

【図13】第1の特徴量算出処理について説明するフローチャートである。

【図14】第1の特徴量の例について説明する図である。

【図15】第2の特徴量算出処理について説明するフローチャートである。

【図16】第2の特徴量の例について説明する図である。

【図17】参照信号解析処理について説明するフローチャートである。

【図18】マッチング処理について説明するフローチャートである。

【図19】マスクパターンの生成の例について説明する図である。

【図20】入力信号の第2の特徴量と参照信号の第2の特徴量とのマッチング処理について説明する図である。 10

【図21】時系列に並べられた類似度の例を示す図である。

【図22】類似度の時間的連続性について説明する図である。

【図23】コンピュータのハードウェアの構成例を示すブロック図である。

【発明を実施するための形態】

【0020】

以下、本技術の実施の形態について図を参照して説明する。なお、説明は以下の順序で行う。

1. 楽曲検索装置の構成

2. 楽曲同定処理

【0021】

<1. 楽曲検索装置の構成>

図2は、本技術を適用した楽曲検索装置の一実施の形態の構成を示している。

【0022】

図2の楽曲検索装置11は、楽曲の信号成分と、人の会話や騒音等の雑音成分（雑音）とが混在した入力信号と、雑音が混在しない楽曲の信号成分からなる参照信号とを比較することで、入力信号の楽曲を同定し、その同定結果を出力する。

【0023】

楽曲検索装置11は、入力信号解析部31、参照信号解析部32、およびマッチング処理部33から構成される。 30

【0024】

入力信号解析部31は、外部の装置等から入力される入力信号を解析し、入力信号の特徴を表す特徴量を入力信号から抽出し、マッチング処理部33に供給する。

【0025】

入力信号解析部31は、切り出し部51、時間周波数変換部52、指標算出部53、第1の特徴量算出部54、および第2の特徴量算出部55を備えている。

【0026】

切り出し部51は、入力信号から所定時間分の信号を切り出し、時間周波数変換部52に供給する。

【0027】

時間周波数変換部52は、切り出し部51からの所定時間分の入力信号を、時間周波数領域の信号（スペクトログラム）に変換し、指標算出部53に供給する。 40

【0028】

指標算出部53は、時間周波数変換部52からの入力信号のスペクトログラムから、そのスペクトログラムの時間周波数領域毎に、楽曲の信号成分であることを表すトーンらしさの指標を算出し、第1の特徴量算出部54および第2の特徴量算出部55に供給する。

【0029】

ここで、トーンらしさの指標とは、入力信号における周波数毎の信号成分の強さ（パワースペクトル）で表されるトーンの時間に対する安定性を示すものとする。一般的に、楽曲においては、ある特定の音程（周波数）を持つ音がある一定時間鳴り続けるため、時間 50

方向に対してトーンは安定する。一方、人の会話においては、時間方向に対してトーンはゆらぐ性質を持ち、また、環境雑音においては、時間方向に連続するトーンがあまり見られない。そこで、指標算出部 53 は、所定の時間区間分の入力信号について、トーンの有無および安定性を数値化することで、トーンらしさの指標を求める。

【0030】

第1の特徴量算出部 54 は、指標算出部 53 からの、スペクトログラムの時間周波数領域毎のトーンらしさの指標に基づいて、入力信号の楽曲らしさを表す特徴量である第1の特徴量を算出し、マッチング処理部 33 に供給する。

【0031】

第2の特徴量算出部 55 は、指標算出部 53 からの、スペクトログラムの時間周波数領域毎のトーンらしさの指標に基づいて、入力信号の楽曲の個性を表す特徴量である第2の特徴量を算出し、マッチング処理部 33 に供給する。

【0032】

参照信号解析部 32 は、図示せぬ記憶部に蓄積されているか、または、外部の装置等から入力されてくる参照信号を解析し、その参照信号の特徴を表す特徴量を参照信号から抽出し、マッチング処理部 33 に供給する。

【0033】

参照信号解析部 32 は、時間周波数変換部 61、指標算出部 62、第1の特徴量算出部 63、および第2の特徴量算出部 64 を備えている。

【0034】

時間周波数変換部 61 は、参照信号をスペクトログラムに変換し、指標算出部 62 に供給する。

【0035】

指標算出部 62 は、時間周波数変換部 61 からの参照信号のスペクトログラムから、そのスペクトログラムの時間周波数領域毎に、楽曲の信号成分であることを表すトーンらしさの指標を算出し、第1の特徴量算出部 63 および第2の特徴量算出部 64 に供給する。

【0036】

第1の特徴量算出部 63 は、指標算出部 62 からの、スペクトログラムの時間周波数領域毎のトーンらしさの指標に基づいて、参照信号の楽曲らしさを表す第1の特徴量を算出し、マッチング処理部 33 に供給する。

【0037】

第2の特徴量算出部 64 は、指標算出部 62 からの、スペクトログラムの時間周波数領域毎のトーンらしさの指標に基づいて、参照信号の楽曲の個性を表す第2の特徴量を算出し、マッチング処理部 33 に供給する。

【0038】

マッチング処理部 33 は、入力信号解析部 31 からの入力信号の第1の特徴量と、参照信号解析部 32 からの参照信号の第1の特徴量とに基づいて、入力信号解析部 31 からの入力信号の第2の特徴量と、参照信号解析部 32 からの参照信号の第2の特徴量とのマッチング処理を行うことで、入力信号に含まれる楽曲を同定する。

【0039】

マッチング処理部 33 は、マスクパターン生成部 71、類似度算出部 72、および比較判定部 73 を備えている。

【0040】

マスクパターン生成部 71 は、入力信号解析部 31 からの入力信号の時間周波数領域毎の第1の特徴量と、参照信号解析部 32 からの参照信号の時間周波数領域毎の第1の特徴量とに基づいて、入力信号の第2の特徴量と参照信号の第2の特徴量とのマッチング処理に用いられるマスクパターンを生成し、類似度算出部 72 に供給する。また、マスクパターン生成部 71 は、入力信号解析部 31 からの入力信号の時間周波数領域毎の第1の特徴量と、参照信号解析部 32 からの参照信号の時間周波数領域毎の第1の特徴量との類似度である第1の特徴量同士の類似度を算出し、類似度算出部 72 に供給する。

10

20

30

40

50

【0041】

類似度算出部72は、マスクパターン生成部71からのマスクパターンおよび第1の特徴量同士の類似度を用いて、入力信号解析部31からの入力信号の第2の特徴量と、参照信号解析部32からの第2の参照信号の特徴量との類似度を算出し、比較判定部73に供給する。

【0042】

比較判定部73は、類似度算出部72からの類似度に基づいて、入力信号に含まれる楽曲が参照信号の楽曲であるか否かを判定し、その参照信号の楽曲の属性を表す楽曲属性情報を、同定結果として出力する。

【0043】

10

[指標算出部の構成]

次に、図3を参照して、図2の指標算出部53の詳細な構成について説明する。

【0044】

図3の指標算出部53は、時間区間選択部81、ピーク検出部82、近似処理部83、トーン度算出部84、および出力部85を備えている。

【0045】

時間区間選択部81は、時間周波数変換部52からの入力信号のスペクトログラムにおいて、所定の時間区間のスペクトログラムを選択し、ピーク検出部82に供給する。

【0046】

20

ピーク検出部82は、時間区間選択部81によって選択された所定の時間区間（時間フレーム）のスペクトログラムにおいて、単位周波数毎に、信号成分の強さが極大となる点であるピークを検出する。

【0047】

近似処理部83は、所定の時間区間のスペクトログラムにおいて、ピーク検出部82によって検出されたピーク周辺の信号成分の強さ（パワースペクトル）を、所定の関数で近似する。

【0048】

トーン度算出部84は、近似処理部83により近似された所定の関数と、ピーク検出部82によって検出されたピーク周辺のパワースペクトルとの距離（誤差）に基づいて、所定の時間区間分のスペクトログラムについての、トーンらしさの指標を数値化したトーン度を算出する。

30

【0049】

出力部85は、トーン度算出部84によって算出された、所定の時間区間分のスペクトログラムについてのトーン度を保持する。また、出力部85は、保持している全時間区間分のスペクトログラムについてのトーン度を、切り出し部51により切り出された所定時間分の入力信号のトーンらしさの指標として、第1の特徴量算出部54および第2の特徴量算出部55に供給する。

【0050】

このようにして、時間周波数領域における所定の時間区間毎、単位周波数毎にトーン度（要素）を持つトーンらしさの指標が算出される。

40

【0051】**[第1の特徴量算出部の構成]**

次に、図4を参照して、図2の第1の特徴量算出部54の詳細な構成について説明する。

【0052】

図4の第1の特徴量算出部54は、フィルタ処理部91、リサンプリング処理部92、および2値化処理部93を備えている。

【0053】

フィルタ処理部91は、指標算出部53からの、入力信号の時間周波数領域における所定の時間区分毎、単位周波数毎にトーン度（要素）を持つトーンらしさの指標に対して、

50

時間方向にフィルタリングすることで、トーンらしさの指標の時間平均量を算出し、入力信号の楽曲らしさを表す第1の特徴量として、リサンプリング処理部92に供給する。

【0054】

リサンプリング処理部92は、フィルタ処理部91からの第1の特徴量を、時間方向にリサンプリング(ダウンサンプリング)し、2値化処理部93に供給する。

【0055】

2値化処理部93は、リサンプリング処理部92からのダウンサンプリングされた第1の特徴量に対して2値化処理を施し、マッチング処理部33に供給する。

【0056】

[第2の特徴量算出部の構成]

10

次に、図5を参照して、図2の第2の特徴量算出部55の詳細な構成について説明する。

【0057】

図5の第2の特徴量算出部55は、フィルタ処理部101、リサンプリング処理部102、および3値化処理部103を備えている。

【0058】

フィルタ処理部101は、指標算出部53からの、入力信号の時間周波数領域における所定の時間区分毎、単位周波数毎にトーン度(要素)を持つトーンらしさの指標に対して、時間方向にフィルタリングすることで、トーンらしさの指標の時間変化量を算出し、入力信号の楽曲の個性を表す第2の特徴量として、リサンプリング処理部102に供給する。

20

【0059】

リサンプリング処理部102は、フィルタ処理部101からの第2の特徴量を、時間方向にリサンプリング(ダウンサンプリング)し、3値化処理部103に供給する。

【0060】

3値化処理部103は、リサンプリング処理部102からのダウンサンプリングされた第2の特徴量に対して3値化処理を施し、マッチング処理部33に供給する。

【0061】

なお、図2の参照信号解析部32の指標算出部62、第1の特徴量算出部63、および第2の特徴量算出部64は、それぞれ、上述で説明した入力信号解析部31の指標算出部53、第1の特徴量算出部54、および第2の特徴量算出部55と同様の機能および構成を有するので、その詳細な説明は省略する。

30

【0062】

<2. 楽曲同定処理>

次に、図6のフローチャートを参照して、楽曲検索装置11の楽曲同定処理について説明する。楽曲同定処理は、外部の装置等から、同定されるべき楽曲が含まれる入力信号が楽曲検索装置11に入力されると開始される。なお、楽曲検索装置11には、入力信号が、時間的に連続して入力される。

【0063】

ステップS11において、入力信号解析部31は、入力信号解析処理を実行し、外部の装置等から入力された入力信号を解析し、入力信号の特徴量を入力信号から抽出する。

40

【0064】

[入力信号解析処理の詳細]

ここで、図7のフローチャートを参照して、図6のフローチャートのステップS11における入力信号解析処理の詳細について説明する。

【0065】

ステップS31において、入力信号解析部31の切り出し部51は、入力信号から所定時間(例えば、15秒間)分の信号を切り出し、時間周波数変換部52に供給する。

【0066】

ステップS32において、時間周波数変換部52は、例えばハン窓等の窓関数や、例え

50

ば離散フーリエ変換等を用いて、切り出し部 5 1 からの所定時間分の入力信号をスペクトログラムに変換し、指標算出部 5 3 に供給する。なお、ここでは、窓関数として、ハン窓に限らず、サイン窓やハミング窓が用いられてもよく、また、離散フーリエ変換に限らず、離散コサイン変換が用いられてもよい。さらに、変換されたスペクトログラムは、パワースペクトルや振幅スペクトル、対数振幅スペクトルのいずれであってもよい。また、周波数分解能を高めるために、零詰めによるオーバーサンプリングにより周波数変換長を窓の長さより長く（例えば 2 倍や 4 倍等に）してもよい。

【 0 0 6 7 】

ステップ S 3 3において、指標算出部 5 3 は、指標算出処理を実行することで、時間周波数変換部 5 2 からの入力信号のスペクトログラムから、そのスペクトログラムの時間周波数領域毎に入力信号のトーンらしさの指標を算出する。10

【 0 0 6 8 】

[指標算出処理の詳細]

ここで、図 8 のフローチャートを参照して、図 7 のフローチャートのステップ S 3 3 における指標算出処理の詳細について説明する。

【 0 0 6 9 】

ステップ S 5 1において、指標算出部 5 3 の時間区間選択部 8 1 は、時間周波数変換部 5 2 からの入力信号のスペクトログラムにおいて、所定の時間区間（例えば、15 秒間の入力信号のうちの最初の 1 秒間）のスペクトログラムを選択し、ピーク検出部 8 2 に供給する。20

【 0 0 7 0 】

ステップ S 5 2において、ピーク検出部 8 2 は、時間区間選択部 8 1 によって選択された 1 秒間分のスペクトログラムにおいて、時間フレーム（時間 bin）毎に、各周波数帯についての信号成分のパワースペクトル（強さ）が、その周波数帯の近傍で極大となる時間周波数領域上の点であるピークを検出する。

【 0 0 7 1 】

例えば、図 9 の上段に示される、時間周波数領域に変換された入力信号のスペクトログラムにおいて、ある時間フレームで、ある周波数帯近傍の時間周波数領域 では、図 9 の下段に示されるようなピーク p が検出されるようになる。

【 0 0 7 2 】

ステップ S 5 3において、近似処理部 8 3 は、時間区間選択部 8 1 によって選択された 1 秒間分のスペクトログラムについて、ピーク検出部 8 2 によって検出された時間フレーム毎の各周波数帯のピーク周辺のパワースペクトルを、予め定義された定型的な関数であるテンプレート関数で近似する。以下においては、テンプレート関数は 2 次関数であるものとするが、これに限らず、3 次以上の高次関数や三角関数等、その他の関数であってもよい。

【 0 0 7 3 】

上述したように、図 9 の下段においてはピーク p が検出されるものとしたが、ピークとなるパワースペクトルは、時間方向に対して安定するトーン（以下、持続性トーンともいう）に限らず、ノイズ、サイドロープ、干渉、時間変動トーン等の信号成分についても発生するため、そのピークに基づいてトーンらしさの指標が適切に算出されるとは限らない。また、離散フーリエ変換のピークは離散的であるので、そのピーク周波数は、真のピーク周波数であるとは限らない。40

【 0 0 7 4 】

文献「J. O. Smith III and X. Serra: "PARSHL: A program for analysis/synthesis of inharmonic sounds based on a sinusoidal representation" in Proc. ICMC '87」によれば、楽曲であるか人の声であるかにかかわらず、ある時間フレームでのピーク周辺の対数パワースペクトルの値は、2 次関数で近似することが可能であるとされている。

【 0 0 7 5 】

そこで、本技術においては、時間フレーム毎の各周波数帯のピーク周辺のパワースペク50

トルを2次関数で近似する。

【0076】

さらに、本技術においては、以下の仮定の下、そのピークが持続性トーンによるものであるか否かを決定する。

【0077】

- a) 持続性トーンは、その2次関数を時間方向に伸ばした関数により近似される。
- b) 周波数の時間変化については、楽曲によるピークは時間方向に持続するので、零次近似される（変化しないものとする）。
- c) 振幅の時間変化については、ある程度の許容を必要とし、例えば2次関数等で近似される。

【0078】

したがって、持続性トーンは、図10で示されるような、ある時間フレームにおける2次関数を時間方向に伸ばしたトンネル型の関数（双2次関数）でモデル化され、時間tおよび周波数 ω に対して、以下の式(1)で表すことができる。なお、 ω_p は、ピーク周波数を表している。

【0079】

【数1】

$$g(t, \omega) = f(\omega - \omega_p)^2 + ct^2 + dt + e$$

… (1)

10

【0080】

よって、ある着目しているピーク周辺において、上述した仮定a)乃至c)に基づいた双2次関数を、例えば最小二乗近似で当てはめたときの誤差を、トーン（持続性トーン）らしさの指標とすることができます。すなわち、以下の式(2)を誤差関数とすることができます。

【0081】

【数2】

$$J(a, b, c, d, e) = \sum_{\Gamma} (f(k, n) - g(k, n))^2 \rightarrow \min$$

… (2)

20

【0082】

式(2)において、 $f(k, n)$ は、第nフレーム、第kビンのDFTスペクトルを示しており、 $g(k, n)$ は、上で示された、持続性トーンのモデルを示す式(1)と同義の式であり、以下の式(3)で示される。

【0083】

【数3】

$$g(k, n) = ak^2 + bk + cn^2 + dn + e$$

… (3)

30

【0084】

なお、式(2)における Γ は、対象となるピーク周辺の時間周波数領域を示している。時間周波数領域 Γ において、周波数方向のサイズは、時間周波数変換に用いた窓関数に応じて、かつ、周波数変換長により決定されるメインローブのサンプルポイント数を超えないように決定される。また、時間方向のサイズは、持続性トーンを定義するのに必要な時間長に応じて決定される。

【0085】

図8のフローチャートに戻り、ステップS54において、トーン度算出部84は、近似処理部83により近似された2次関数と、ピーク検出部82によって検出されたピーク周辺のパワースペクトルとの誤差、すなわち、上述した式(2)の誤差関数に基づいて、時

50

間区間選択部 8 1 によって選択された 1 秒間分のスペクトログラムについての、トーンらしさの指標であるトーン度を算出する。

【 0 0 8 6 】

ここで、上述した式(2)の誤差関数を、平面モデルへ当てはめたときの誤差関数を以下の式(4)で表すこととすると、トーン度 η は、以下の式(5)のように示される。

【 0 0 8 7 】

【数 4】

$$J'(e') = \sum_{k,n} (f(k,n) - e')^2 \rightarrow \min$$

10

· · · (4)

【 0 0 8 8 】

【数 5】

$$\eta(k,n) = 1 - \sqrt{J(\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{e}) / J'(\hat{e}')}$$

· · · (5)

【 0 0 8 9 】

式(5)において、 a ハット(a に $\hat{\cdot}$ を付した文字を a ハットと記述する。以下、本明細書において、同様に記載する。)、 b ハット、 c ハット、 d ハット、 e ハットは、それぞれ、 $J(a,b,c,d,e)$ を最小とする a,b,c,d,e であり、 e' ハットは、 $J(e')$ を最小とする e' である。

20

【 0 0 9 0 】

このようにして、トーン度 η が求められる。

【 0 0 9 1 】

ところで、式(5)において、 a ハットは、持続性トーンを表すモデルの曲線(2次関数)のピーク曲率を示している。

【 0 0 9 2 】

ピーク曲率は、入力信号の信号成分が正弦波であれば、理論的には、時間周波数変換に用いた窓関数の種類とサイズによって決まる定数である。したがって、実際に得られたピーク曲率 a ハットの値が理論値からはずれるほど、その信号成分が持続性トーンである可能性は低くなると考えられる。また、そのピークがサイドロープ性のものであつたとしても、得られるピーク曲率は変わってくるため、ピーク曲率 a ハットのずれはトーンらしさの指標に影響すると言える。すなわち、ピーク曲率 a ハットの理論値からのずれの値に応じて、トーン度 η を調整することにより、より適正なトーンらしさの指標を得ることができるようになる。ピーク曲率 a ハットの理論値からのずれの値に応じて調整されたトーン度 η' は、以下の式(6)のように示される。

30

【 0 0 9 3 】

【数 6】

$$\eta'(k,n) = D(\hat{a} - a_{ideal}) \eta(k,n)$$

40

· · · (6)

【 0 0 9 4 】

式(6)において、値 a_{ideal} は、時間周波数変換に用いた窓関数の種類とサイズによって決まるピーク曲率の理論値である。また、関数 $D(x)$ は、図11に示されるような値をとる調整関数であり、関数 $D(x)$ によれば、ピーク曲率の値が理論値から離れるほど、トーン度は小さくなる。すなわち、式(6)によれば、ピークでない要素については、トーン度 η' は0となる。なお、関数 $D(x)$ は、図11に示される形状の関数に限らず、ピーク曲率の値が理論値から離れるほど、トーン度が小さくなるような関数であればよい。

【 0 0 9 5 】

このように、曲線(2次関数)のピーク曲率に応じてトーン度を調整することで、より

50

適切なトーン度が得られるようになる。

【0096】

ところで、式(5)におけるaハットおよびbハットによって与えられる値 - (bハット)/2(aハット)は、離散ピーク周波数から真のピーク周波数へのオフセットを示している。

【0097】

真のピーク周波数は、理論上、離散ピーク周波数から±0.5ビンの位置にある。離散ピーク周波数から真のピーク周波数へのオフセット値 - (bハット)/2(aハット)が、着目しているピークの位置と極端に異なる場合、上述で示された式(2)の誤差関数を計算する際のマッチングが正しくない可能性が高い。すなわち、このことは、トーンらしさの指標の信頼性に影響すると考えられるので、オフセット値 - (bハット)/2(aハット)の、着目しているピークの位置(ピーク周波数) k_p からのずれの値に応じて、トーン度 η を調整することにより、より適切なトーンらしさの指標を得るようにもよい。具体的には、図11に示される関数D(x)に - (bハット)/2(aハット) - k_p を代入した項を、式(6)の左辺と乗算した値をトーン度 η とするようにしてよい。

【0098】

また、トーン度 η は、上述した手法以外の手法により算出されるようにもできる。

【0099】

具体的には、まず、上述した式(2)の誤差関数において、持続性トーンを表すモデル $g(k,n)$ を、ピーク周辺のパワースペクトルの時間平均形状を近似した2次関数 ak^2+bk+c で置き換えた、以下の式(7)の誤差関数を与える。

【0100】

【数7】

$$J(a, b, c) = \sum_{\Gamma} (f(k, n) - (ak^2 + bk + c))^2 \rightarrow \min \quad \dots (7)$$

【0101】

次に、上述した式(2)の誤差関数において、持続性トーンを表すモデル $g(k,n)$ を、着目しているピークのある第mフレームのパワースペクトルを近似した2次関数 $a'k^2+b'k+c'$ で置き換えた、以下の式(8)の誤差関数を与える。なお、mは、着目しているピークのあるフレーム番号を示している。

【0102】

【数8】

$$J'(a', b', c') = \sum_{\Gamma, n=m} (f(k, n) - (a'k^2 + b'k + c'))^2 \rightarrow \min \quad \dots (8)$$

【0103】

ここで、式(7)において、 $J(a, b, c)$ を最小とする a, b, c を、それぞれaハット、bハット、cハットとし、式(8)において、 $J(a', b', c')$ を最小とする a', b', c' を、それぞれ a' ハット、 b' ハット、 c' ハットとしたとき、トーン度 η は以下の式(9)で与えられる。

【0104】

【数9】

$$\eta' (k, n) = D_1 \left(1 - \frac{\hat{a}}{\hat{a}'} \right) D_2 \left\{ \left(-\frac{\hat{b}}{2\hat{a}'} \right) - \left(-\frac{\hat{b}'}{2\hat{a}'} \right) \right\} \quad \dots (9)$$

10

20

30

40

50

【0105】

式(9)において、関数 $D_1(x), D_2(x)$ は、それぞれ、上述で説明した図11に示されるような値をとる関数とされる。なお、式(9)によれば、ピークでない要素については、トーン度'は0となり、また、 $a\hat{}$ が0または $a'\hat{}$ が0の場合にも、トーン度'は0となる。

【0106】

なお、以上のように算出されたトーン度'に対して、シグモイド関数等による非線形変換を施すようにしてもよい。

【0107】

さて、図8のフローチャートに戻り、ステップS55において、出力部85は、トーン度算出部84によって算出された、1秒間分のスペクトログラムについてのトーン度を保持し、全ての時間区間（例えば、15秒間）について、上述した処理を行ったか否かを判定する。

10

【0108】

ステップS55において、全ての時間区間について上述した処理が行われていないと判定された場合、処理はステップS51に戻り、次の時間区間（1秒間）のスペクトログラムに対して、ステップS51乃至S55の処理が繰り返される。なお、ステップS51乃至S55の処理は、上述したように、1秒間の時間区間ずつのスペクトログラムに対して行われるようにしてもよいし、処理対象となるスペクトログラムの時間区間を、例えば0.5秒ずつずらして、処理対象の時間区間の一部を、前の処理対象の時間区間とオーバーラップさせながら行われるようにしてもよい。

20

【0109】

一方、ステップS55において、全ての時間区間について上述した処理が行われたと判定された場合、処理はステップS56に進む。

【0110】

ステップS56において、出力部85は、保持している時間区間（1秒間）毎のトーン度を時系列に配列することで得られる行列を、第1の特徴量算出部54および第2の特徴量算出部55に供給（出力）し、処理はステップS33に戻る。

【0111】

図12は、指標算出部53によって算出されるトーンらしさの指標の例について説明する図である。

30

【0112】

図12に示されるように、入力信号のスペクトログラムから算出された入力信号のトーンらしさの指標Sは、時間方向および周波数方向に、トーン度を要素（以下、成分ともいう）として有している。トーンらしさの指標Sにおける四角形（マス目）のそれぞれは、各時間、各周波数毎の成分を表しており、図示されないが、それぞれトーン度としての値を有する。また、図12に示されるように、トーンらしさの指標Sの時間粒度は1秒とされる。

【0113】

このように、入力信号のスペクトログラムから算出される入力信号のトーンらしさの指標は、各時間、各周波数毎に成分を有しているので、行列として扱うことができる。以下、行列としてのトーンらしさの指標を、適宜、指標行列と言う。

40

【0114】

なお、極低周波帯域については、例えばハムノイズ等の楽曲以外の信号成分によるピークが含まれる可能性が高いので、トーン度を算出しないようにしてもよい。また、例えば8kHzより上の高周波帯域についても、楽曲を構成する重要な要素でない可能性があるので、トーン度を算出しないようにしてもよい。さらに、離散ピーク周波数におけるパワースペクトルの値が、例えば-80dB等の所定の値よりも小さい場合にも、トーン度を算出しないようにしてもよい。

【0115】

50

図7のフローチャートに戻り、ステップS33の後、ステップS34において、第1の特徴量算出部54は、指標算出部53からのトーンらしさの指標に基づいて、第1の特徴量算出処理を実行し、入力信号の楽曲らしさを表す第1の特徴量を算出する。

【0116】

[第1の特徴量算出処理の詳細]

ここで、図13のフローチャートを参照して、図7のフローチャートのステップS34における第1の特徴量算出処理の詳細について説明する。

【0117】

ステップS71において、フィルタ処理部91は、指標算出部53からの指標行列に対して、例えばハニ窓等の窓関数（平滑化フィルタ）を用いて、時間方向にフィルタリングすることで、トーンらしさの指標の時間平均量を算出し、入力信号の楽曲らしさを表す第1の特徴量として、リサンプリング処理部92に供給する。
10

【0118】

ステップS72において、リサンプリング処理部92は、フィルタ処理部91からの第1の特徴量を、時間方向にリサンプリング（ダウンサンプリング）し、2値化処理部93に供給する。

【0119】

ステップS73において、2値化処理部93は、リサンプリング処理部92からの、ダウンサンプリングされた第1の特徴量に対して2値化処理を施し、マッチング処理部33に供給して、処理はステップS34に戻る。
20

【0120】

図14は、上述した処理によって得られる第1の特徴量の例を示している。

【0121】

図14においては、左から順番に、時間周波数領域の入力信号、トーンらしさの指標、および第1の特徴量が示されている。入力信号に楽曲の信号成分が含まれている場合、トーンらしさの指標においては、対応する時間周波数領域の値が大きくなる（持続性トーンが現れる）。そして、トーンらしさの指標の値が大きい（持続性トーンが現れる）領域においては、上述した第1の特徴量算出処理によって、大きい値の第1の特徴量が現れるようになる。したがって、図14に示されるように、第1の特徴量は、入力信号の楽曲の信号成分と同様の領域に値を持つようになる。また、その値は、例えば0, 1等の値に2値化されるようになるので、第1の特徴量の情報量が削減されるようになり、後段の処理の高速化を図ることができるようになる。
30

【0122】

なお、2値化処理に用いられる閾値は、固定の値としてもよいし、時間フレーム毎の平均値等から得られる統計的な値としてもよい。

【0123】

また、上述した第1の特徴量算出処理においては、時間方向にリサンプリングするのに加えて、周波数方向にリサンプリングするようにしてもよいし、2値化処理に代えて量子化処理を行うようにしてもよい。

【0124】

図7のフローチャートに戻り、ステップS34の後、ステップS35において、第2の特徴量算出部55は、指標算出部53からのトーンらしさの指標に基づいて、第2の特徴量算出処理を実行し、入力信号の楽曲の個性を表す第2の特徴量を算出する。
40

【0125】

[第2の特徴量算出処理の詳細]

ここで、図15のフローチャートを参照して、図7のフローチャートのステップS35における第2の特徴量算出処理の詳細について説明する。

【0126】

ステップS91において、フィルタ処理部101は、指標算出部53からの指標行列に対して、例えば1周期サイン窓等の窓関数（フィルタ）を用いて、時間方向にフィルタリ
50

ングすることで、トーンらしさの指標の時間平均変化量を算出し、入力信号の楽曲の個性を表す第2の特徴量として、リサンプリング処理部102に供給する。

【0127】

ステップS92において、リサンプリング処理部102は、フィルタ処理部101からの第2の特徴量を、時間方向にリサンプリング（ダウンサンプリング）し、3値化処理部103に供給する。

【0128】

ステップS93において、3値化処理部103は、リサンプリング処理部102からの、ダウンサンプリングされた第2の特徴量に対して3値化処理を施し、マッチング処理部33に供給して、処理はステップS35に戻る。

10

【0129】

図16は、上述した処理によって得られる第2の特徴量の例を示している。

【0130】

図16においては、左から順番に、時間周波数領域の入力信号、トーンらしさの指標、および第2の特徴量が示されている。入力信号に楽曲の信号成分が含まれている場合、トーンらしさの指標においては、対応する時間周波数領域の値が大きくなる（持続性トーンが現れる）。そして、トーンらしさの指標の値が大きい（持続性トーンが現れる）領域においては、上述した第2の特徴量算出処理によって、トーンの立上がりにおいては大きい値（正の値）で、トーンの立下がりにおいては小さい値（負の値）となる第2の特徴量が現れるようになる。したがって、図16に示されるように、第2の特徴量は、入力信号の楽曲の信号成分と同様の領域に値を持つようになり、その値は、例えば、図16における領域s1においては正の値となり、領域s2においては負の値となる。また、その値は、-1, 0, 1等の値に3値化されるようになるので、第2の特徴量の情報量が削減されるようになり、後段の処理の高速化を図ることができるようになる。

20

【0131】

なお、3値化処理に用いられる閾値は、固定の値としてもよいし、時間フレーム毎の平均値等から得られる統計的な値としてもよい。

【0132】

また、上述した第2の特徴量算出処理においては、時間方向にリサンプリングするのに加えて、周波数方向にリサンプリングするようにしてもよいし、3値化処理に代えて量子化処理を行うようにしてもよい。

30

【0133】

なお、以上のようにして算出された第1の特徴量と第2の特徴量の、時間方向および周波数方向の粒度は同一とされる。

【0134】

さて、図7のフローチャートに戻り、ステップS35の後、処理は図6のフローチャートのステップS11に戻る。

【0135】

図6のフローチャートにおいて、ステップS11の後、処理はステップS12に進み、参照信号解析部32は、参照信号解析処理を実行し、外部の装置等から入力される参照信号を解析し、参照信号の特徴量を参照信号から抽出する。

40

【0136】

[参照信号解析処理の詳細]

ここで、図17のフローチャートを参照して、図6のフローチャートのステップS12における参照信号解析処理の詳細について説明する。

【0137】

ステップS111において、参照信号解析部32の時間周波数変換部61は、入力された参照信号をスペクトrogramに変換し、指標算出部62に供給する。

【0138】

ステップS112において、指標算出部62は、指標算出部53と同じにして、指標算

50

出処理を実行することで、時間周波数変換部 61 からの参照信号のスペクトログラムから、そのスペクトログラムの時間周波数領域毎に参照信号のトーンらしさの指標を算出し、第1の特徴量算出部 63 および第2の特徴量算出部 64 に供給する。

【0139】

なお、ステップ S112 の指標算出処理は、図 8 のフローチャートを参照して説明した指標算出処理と基本的に同様であるので、その説明は省略する。

【0140】

ステップ S113 において、第1の特徴量算出部 63 は、第1の特徴量算出部 54 と同様にして、指標算出部 62 からのトーンらしさの指標に基づいて、第1の特徴量算出処理を実行し、参照信号の楽曲らしさを表す第1の特徴量を算出し、マッチング処理部 33 に供給する。
10

【0141】

なお、ステップ S113 の第1の特徴量算出処理は、図 13 のフローチャートを参照して説明した第1の特徴量算出処理と基本的に同様であるので、その説明は省略する。なお、2値化処理に用いられる閾値は、異なる値であってもよい。

【0142】

ステップ S114 において、第2の特徴量算出部 64 は、第2の特徴量算出部 55 と同様にして、指標算出部 62 からのトーンらしさの指標に基づいて、第2の特徴量算出処理を実行し、参照信号の楽曲の個性を表す第2の特徴量を算出し、マッチング処理部 33 に供給する。
20

【0143】

なお、ステップ S114 の第2の特徴量算出処理は、図 15 のフローチャートを参照して説明した第2の特徴量算出処理と基本的に同様であるので、その説明は省略する。

【0144】

また、このとき、参照信号解析部 32 は、楽曲検索装置 11 内の図示せぬデータベースから、参照信号の楽曲の属性を表す楽曲属性情報（楽曲名、演奏者名、楽曲ID等）を読み出し、算出した参照信号の特徴量（第1および第2の特徴量）に対応付けて、マッチング処理部 33 に供給する。

【0145】

なお、複数の参照信号についての特徴量および楽曲属性情報は、楽曲検索装置 11 内の図示せぬデータベースに記憶されて、マッチング処理部 33 に取得されるようにしてもよい。
30

【0146】

図 17 のフローチャートにおいて、ステップ S114 の後、処理は図 6 のフローチャートのステップ S12 に戻る。

【0147】

このようにして算出される入力信号および参照信号の第1の特徴量および第2の特徴量は行列として扱われ、その時間方向および周波数方向の粒度は同一とされる。

【0148】

また、楽曲の個性を表す第2の特徴量は、上述した入力信号（参照信号）のトーンらしさの指標の時間変化量に限らず、他の評価値とするようにしてもよい。
40

【0149】

図 6 のフローチャートにおいて、ステップ S12 の後、処理はステップ S13 に進み、マッチング処理部 33 は、マッチング処理を実行し、入力信号に含まれる楽曲を同定し、その同定結果を出力する。

【0150】

[マッチング処理の詳細]

ここで、図 18 のフローチャートを参照して、図 6 のフローチャートのステップ S13 におけるマッチング処理の詳細について説明する。

【0151】

ステップS131において、マスクパターン生成部71は、入力信号解析部31からの入力信号の時間周波数領域毎の第1の特徴量と、参照信号解析部32からの参照信号の時間周波数領域毎の第1の特徴量とに基づいて、入力信号の第2の特徴量と参照信号の第2の特徴量とのマッチング処理に用いられるマスクパターンを生成し、類似度算出部72に供給する。

【0152】

例えば、入力信号解析部31からの第1の特徴量として、図19の左上段に示されるような第1の特徴量S1が供給され、参照信号解析部32からの第1の特徴量として、図19の左下段に示されるような第1の特徴量A1が供給された場合、マスクパターン生成部71は、第1の特徴量S1の各要素と、第1の特徴量A1の各要素とを乗算した値を要素とするマスクパターンWを生成する。このように、入力信号および参照信号のスペクトログラムにおいて、トーンらしさの指標の高い領域には第1の特徴量に応じた重み付けをし、トーンらしさの指標の低い領域をマスクする、重み分布としてのマスクパターンが生成される。10

【0153】

ステップS132において、マスクパターン生成部71は、入力信号解析部31からの入力信号の時間周波数領域毎の第1の特徴量と、参照信号解析部32からの参照信号の時間周波数領域毎の第1の特徴量の各要素に基づいて、入力信号の第1の特徴量S1と参照信号の第1の特徴量A1の類似度R1を算出し、類似度算出部72に供給する。

【0154】

ステップS133において、類似度算出部72は、マスクパターン生成部71からのマスクパターンW、および、第1の特徴量同士の類似度R1を用いて、入力信号解析部31からの入力信号の第2の特徴量と、参照信号解析部32からの第2の参照信号の特徴量との類似度を算出し、算出された類似度を、比較判定部73に供給する。20

【0155】

ここで、図20を参照して、入力信号の第2の特徴量と参照信号の第2の特徴量との類似度の算出の例について説明する。

【0156】

図20においては、上段に、参照信号の第2の特徴量L、下段左側に、入力信号の第2の特徴量S2、そして、下段右側に、マスクパターンWが示されている。上述で説明したように、これらは行列として扱うことができる。30

【0157】

図20に示されるように、参照信号の第2の特徴量Lの時間方向の成分の数は、入力信号の第2の特徴量S2の時間方向の成分の数と比較して多い（なお、入力信号の第2の特徴量S2の時間方向の成分の数と、マスクパターンWの時間方向の成分の数とは同一である）。そこで、入力信号の第2の特徴量と参照信号の第2の特徴量との類似度を算出する際、類似度算出部72は、参照信号の第2の特徴量Lから、入力信号の第2の特徴量S2と同一の時間方向の成分の数となる部分行列A2を、時間方向（図中右方向）にずらしながら（時間方向にオフセットを与ながら）順次切り出し、この部分行列A2と入力信号の第2の特徴量S2との類似度を算出する。ここで、部分行列A2を切り出す際の時間方向のオフセットをtとすると、まず、マスクパターンW、および、第1の特徴量同士の類似度R1は、以下の式（10）、（11）で示される。40

【0158】

【数10】

$$W_{f(t+u)} = S1_{fu} A1_{f(t+u)}$$

... (10)

【0159】

【数11】

$$R1(t) = \frac{\sum A1_{f(t+u)} S1_{fu}}{\sqrt{\sum A1_{f(t+u)}^2 \cdot \sum S1_{fu}^2}} \quad \dots (11)$$

【0160】

そして、式(10)で示されるマスクパターンW、および、式(11)で示される第1の特徴量の類似度R1を用いて、部分行列A2と入力信号の第2の特徴量S2との類似度R(t)は、以下の式(12)で示される。

10

【0161】

【数12】

$$R(t) = \frac{\sum_{W_{f(t+u)} \neq 0} W_{f(t+u)} \exp(-\alpha (A2_{f(t+u)} - S2_{fu})^2)}{\sum_{W_{f(t+u)} \neq 0} W_{f(t+u)}} R1(t) \quad \dots (12)$$

【0162】

20

なお、式(10)乃至(12)において、f, uは、参照信号の第2の特徴量の部分行列A2、入力信号の第2の特徴量S2、マスクパターンWの各行列の周波数成分および時間成分を表している。すなわち、fおよびuが添字として付されているA1, S1, W, A2, S2は、それぞれの行列A1, S1, W, A2, S2における各要素を表している。

【0163】

なお、式(12)で示される類似度R(t)の算出においては、各行列の全要素について演算する必要はなく、マスクパターンWにおいてマスクされていない時間周波数領域の要素(0でない要素)について演算すればよいので、演算コストを抑えることができる。さらに、マスクパターンWにおいてマスクされていない時間周波数領域の要素の値は、入力信号および参照信号の時間周波数領域毎の楽曲らしさを表しているので、楽曲らしさの高い(楽曲の成分のみからなる)時間周波数領域の要素に対してより大きい重み付けをして類似度R(t)を算出することができる。すなわち、より精度の高い類似度を算出することができるようになる。なお、マスクパターンWの各要素が全て0である場合、類似度R(t) = 0とされる。

30

【0164】

このようにして、類似度算出部72は、全ての部分行列A2(全ての部分行列A2を切り出す時間オフセットt)に対して類似度を算出し、最大となる類似度を、入力信号の第2の特徴量と参照信号の第2の特徴量との類似度として、比較判定部73に供給する。比較判定部73は、類似度算出部72からの類似度を、内部の記憶領域に保持する。

40

【0165】

なお、マスクパターンWは、式(10)で表される行列要素間の乗算に限らず、加重幾何平均や加重相加平均等の他の演算により求められるようにしてもよい。

【0166】

さらに、マスクパターンWは、行列要素間の演算後に、各要素に対してシグモイド関数等による非線形変換を施すようにしてもよい。

【0167】

また、第1の特徴量同士の類似度R1は、上述した式(12)による演算に限らず、二乗誤差や絶対誤差等、2つの行列の各要素間の差に基づいて算出されるようにしてもよいし、予め決められた固定の値としてもよい。

【0168】

50

さらに、類似度 $R(t)$ は、入力信号および参照信号の第2の特徴量が3値化されている場合には、以下の式(13)、(14)に示されるように、離散関数を用いて算出されるよ

うにしてもよい。

【0169】

【数13】

$$R(t) = \frac{\sum_{W_{f(t+u)} \neq 0} W_{f(t+u)} I(A2_{f(t+u)}, S2_{fu})}{\sum_{W_{f(t+u)} \neq 0} W_{f(t+u)}} R1(t)$$

10

... (13)

【0170】

【数14】

$$I(A2_{f(t+u)}, S2_{fu}) = \begin{cases} 1 & \text{if } A2_{f(t+u)} = S2_{fu} \\ 0 & \text{if } A2_{f(t+u)} \neq S2_{fu} \end{cases}$$

... (14)

【0171】

20

図18のフローチャートに戻り、ステップS134において、類似度算出部71は、全ての時間分の入力信号に対して類似度算出の処理を行ったか否かを判定する。

【0172】

ステップS134において、全ての時間分の入力信号について類似度算出の処理を行っていないと判定された場合、ステップS131に戻り、全ての時間分の入力信号について類似度算出の処理を行うまで、ステップS131乃至S134の処理が繰り返される。

【0173】

そして、ステップS134において、全ての時間分の入力信号について類似度算出の処理を行ったと判定されると、処理はステップS135に進み、比較判定部73は、類似度算出部72から供給された類似度に基づいて、入力信号に含まれる楽曲が参照信号の楽曲であるか否かを判定し、その参照信号の楽曲の属性を表す楽曲属性情報を、同定結果として出力する。

30

【0174】

具体的には、まず、比較判定部73は、類似度算出部72から供給され保持されている、全ての時間分の入力信号についての類似度を時系列に並べる。

【0175】

図21は、時系列に並べられた類似度の例を示している。

【0176】

比較判定部73は、図21に示される時系列に並べられた類似度に対して、所定の閾値THsを超える類似度の区間Tsを、全ての時間分の入力信号において、参照信号の楽曲が含まれる可能性が高い候補区間とする。また、比較判定部73は、候補区間Tsにおいて、最大となる類似度Rpを特定する。さらに、比較判定部73は、類似度の時間的連続性を求める。

40

【0177】

ここで、図22を参照して、類似度の時間的連続性について説明する。

【0178】

上述したように、比較判定部73に保持される類似度は、所定時間分の入力信号の特徴量(第2の特徴量)の行列S2と、参照信号の特徴量(第2の特徴量)Lの部分行列A2の類似度のうち最大となる類似度である。

【0179】

50

ここで、全ての時間分の入力信号の特徴量の行列S2allについて類似度を算出したとき、入力信号の一部に参照信号の楽曲が含まれている場合には、類似度が最大となった入力信号の特徴量の行列S2と参照信号の特徴量の部分行列A2のペアは、図22に示されるように、同じ時間間隔で時間方向にシフトすると考えられる。そこで、比較判定部73は、類似度が最大となった入力信号の特徴量の行列S2と参照信号の特徴量の部分行列A2のペアが、同じ時間間隔で時間方向にシフトした量(時間)を、類似度の時間的連続性の指標とする。

【0180】

そして、比較判定部73は、上述した候補区間Tsの長さ、最大類似度Rpの値、および類似度の時間的連続性の指標に基づいて、所定の識別関数を用いて、入力信号に含まれる楽曲が参照信号の楽曲であるか否かを判定する。比較判定部73は、入力信号に含まれる楽曲が参照信号の楽曲であると判定した場合、その参照信号の楽曲の属性を表す楽曲属性情報を、同定結果として出力する。10

【0181】

なお、比較判定部73は、入力信号に含まれる楽曲が参照信号の楽曲でないと判定した場合、入力信号に含まれる楽曲は参照信号の楽曲ではない旨の情報を出力する。

【0182】

以上の処理によれば、楽曲と雑音とが混在した入力信号を、楽曲のみからなる参照信号と比較する際、時間周波数領域の入力信号と参照信号において、トーンらしさの指標が算出され、その指標から求められる入力信号の特徴量と参照信号の特徴量の類似度が算出され、その類似度に基づいて入力信号の楽曲が同定されるようになる。トーンらしさの指標は、パワースペクトルの時間に対する安定性を定量化したものであるので、その指標から求められる特徴量は、楽曲らしさを確実に表現するものとなる。したがって、より精度の高い類似度を算出することができるようになり、ひいては、楽曲と雑音とが混在する入力信号から、その楽曲を精度良く同定することが可能となる。20

【0183】

また、マスクパターンを、入力信号の特徴量と参照信号の特徴量とを用いて生成するようにしたので、入力信号の特徴量のみから生成されたマスクパターンと比較して、より確実に入力信号と参照信号をマスクしてマッチング処理を行うことができるようになる。

【0184】

具体的には、入力信号に音の少ない楽曲が含まれている場合、入力信号の特徴量のみから生成されたマスクパターンを用いると、参照信号の楽曲にその音が含まれていれば、異なる楽曲であっても、参照信号のその他の音はマスクされてしまうので、入力信号が参照信号と一致すると判断されてしまう可能性がある。例えば、入力信号にベース(低音)のみの楽曲が含まれている場合、参照信号の楽曲のベース進行が、入力信号に含まれる楽曲のベース進行に似ているだけで、入力信号が参照信号と一致すると判断されてしまう。

【0185】

そこで、本技術においては、入力信号の第1の特徴量と参照信号の第1の特徴量とからマスクパターンを生成するとともに、マスクパターンの生成に用いられる入力信号の第1の特徴量と参照信号の第1の特徴量との類似度も算出することで、入力信号および参照信号の両方に含まれる音の特徴量に基づいて、マッチング処理が行われるようになる。これにより、入力信号に含まれる音が、参照信号に含まれる音の一部と似ているだけでは、入力信号が参照信号と一致すると判断されなくなり、楽曲と雑音とが混在する入力信号から、その楽曲を精度良く同定することが可能となる。40

【0186】

また、時間成分だけでなく、周波数成分を考慮した特徴量を用いて、マッチング処理を行うことができるので、図1で示されたような、途切れている時間が非常に短い会話が雑音として含まれている入力信号であっても、その入力信号から、楽曲を精度良く同定することができる。これにより、ドラマ等のテレビジョン番組において、出演者等の会話が被ってしまっているBGMを精度良く同定することが可能となる。50

【 0 1 8 7 】

なお、図 7 のフローチャートを参照して説明した楽曲同定処理においては、入力信号解析処理とマッチング処理との間に、参照信号解析処理が実行されるようにしたが、参照信号解析処理は、マッチング処理が実行される前に実行されていればよく、入力信号解析処理が実行される前に実行されてもよいし、入力信号解析処理と並行して実行されてもよい。

【 0 1 8 8 】

また、本技術は、図 2 に示される楽曲検索装置 1 1 に適用される他、例えば、インターネット等のネットワークを介して情報を送受信するネットワークシステムに適用されるようにもできる。具体的には、携帯電話機等の端末装置に、図 2 の切り出し部 5 1 を設けるようにするとともに、サーバに、図 2 の切り出し部 5 1 以外の構成を設けるようにし、端末装置からインターネットを介して送信されてくる入力信号に対して、サーバが楽曲同定処理を行うようにしてもよい。この場合、サーバは、同定結果を、インターネットを介して端末装置に送信し、端末装置は、サーバからの同定結果を表示部等に表示するようにもできる。

10

【 0 1 8 9 】

上述した一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、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等に、プログラム記録媒体からインストールされる。

20

【 0 1 9 0 】

図 2 3 は、上述した一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示すブロック図である。

【 0 1 9 1 】

コンピュータにおいて、CPU (Central Processing Unit) 9 0 1 , ROM (Read Only Memory) 9 0 2 , RAM (Random Access Memory) 9 0 3 は、バス 9 0 4 により相互に接続されている。

30

【 0 1 9 2 】

バス 9 0 4 には、さらに、入出力インターフェース 9 0 5 が接続されている。入出力インターフェース 9 0 5 には、キーボード、マウス、マイクロホン等よりなる入力部 9 0 6 、ディスプレイ、スピーカ等よりなる出力部 9 0 7 、ハードディスクや不揮発性のメモリ等よりなる記憶部 9 0 8 、ネットワークインターフェース等よりなる通信部 9 0 9 、磁気ディスク、光ディスク、光磁気ディスク、或いは半導体メモリ等のリムーバブルメディア 9 1 1 を駆動するドライブ 9 1 0 が接続されている。

【 0 1 9 3 】

以上のように構成されるコンピュータでは、CPU 9 0 1 が、例えば、記憶部 9 0 8 に記憶されているプログラムを、入出力インターフェース 9 0 5 およびバス 9 0 4 を介して、RAM 9 0 3 にロードして実行することにより、上述した一連の処理が行われる。

40

【 0 1 9 4 】

コンピュータ (CPU 9 0 1) が実行するプログラムは、例えば、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 Disc)等) 、光磁気ディスク、もしくは半導体メモリ等よりなるパッケージメディアであるリムーバブルメディア 9 1 1 に記録して、あるいは、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、有線または無線の伝送媒体を介して提供される。

【 0 1 9 5 】

そして、プログラムは、リムーバブルメディア 9 1 1 をドライブ 9 1 0 に装着することにより、入出力インターフェース 9 0 5 を介して、記憶部 9 0 8 にインストールすることが

50

できる。また、プログラムは、有線または無線の伝送媒体を介して、通信部 909 で受信し、記憶部 908 にインストールすることができる。その他、プログラムは、ROM 902 や記憶部 908 に、あらかじめインストールしておくことができる。

【0196】

なお、コンピュータが実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

【0197】

また、本技術の実施の形態は、上述した実施の形態に限定されるものではなく、本技術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。

10

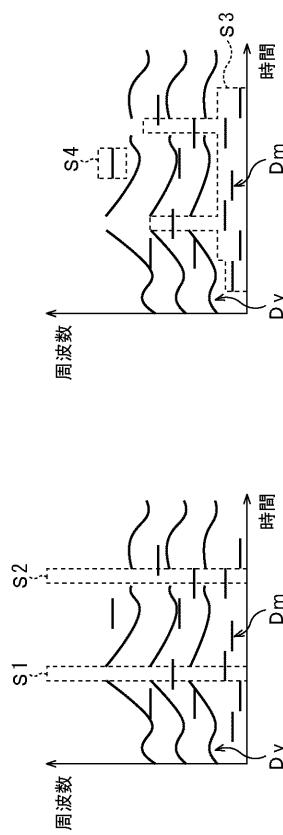
【符号の説明】

【0199】

11 楽曲検索装置，51 切り出し部，52 時間周波数変換部，53 指標算出部，54 第1の特徴量算出部，55 第2の特徴量算出部，61 時間周波数変換部，62 指標算出部，63 第1の特徴量算出部，64 第2の特徴量算出部，71 マスクパターン生成部，72 類似度算出部，73 比較判定部，81 時間区間選択部，82 ピーク検出部，83 近似処理部，84 トーン度算出部，85 出力部

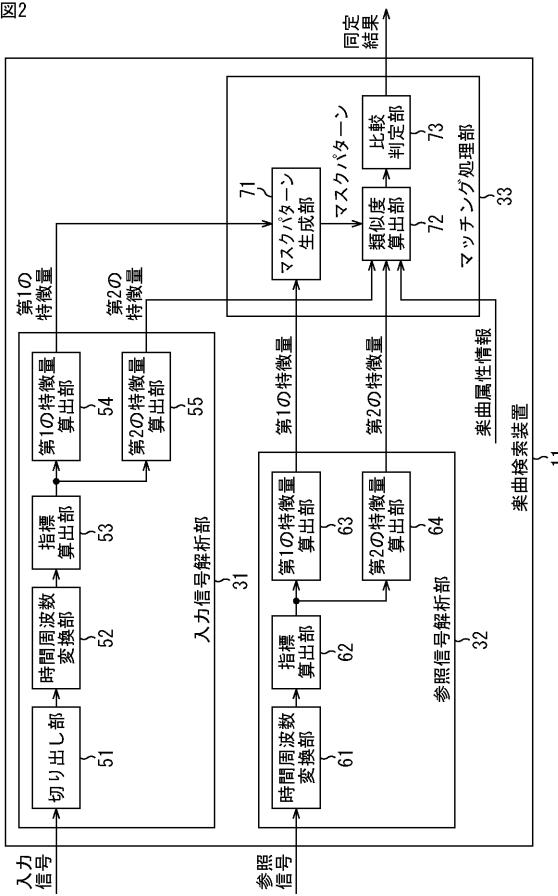
【図1】

図1

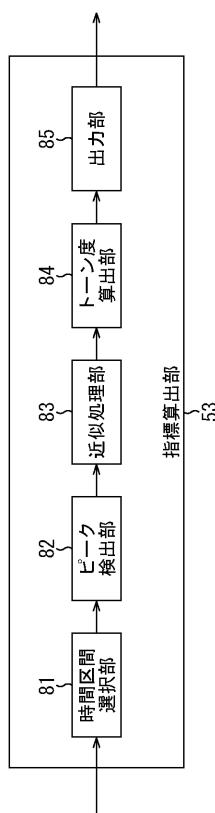


【図2】

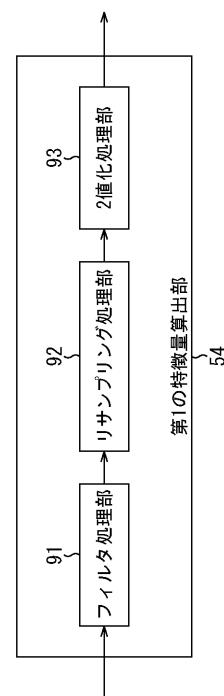
図2



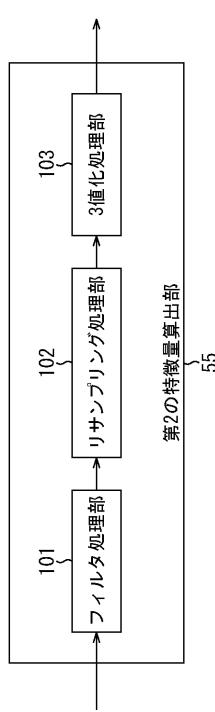
【図3】



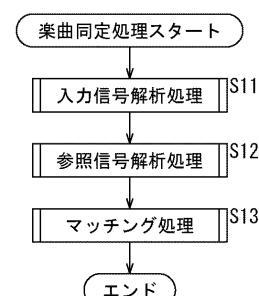
【図4】



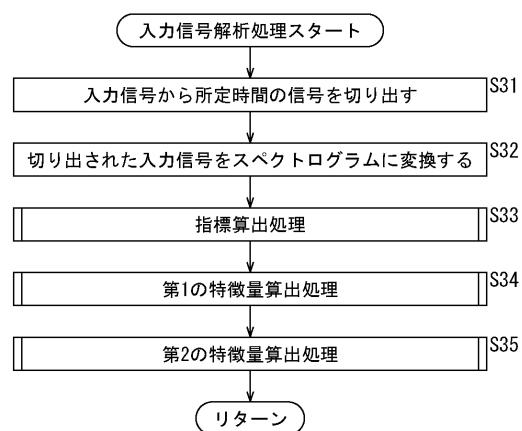
【図5】



【図6】

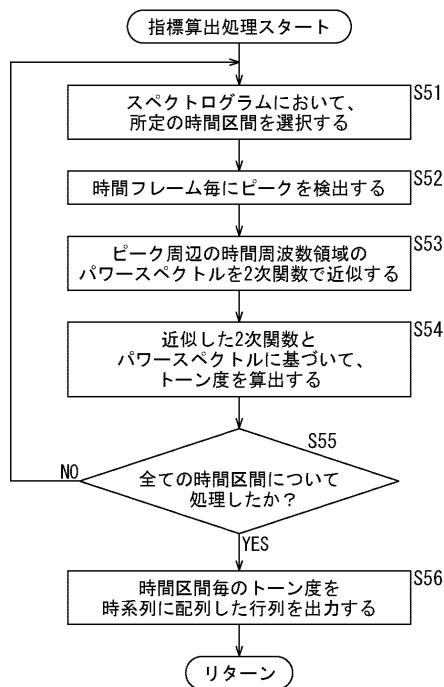


【図7】



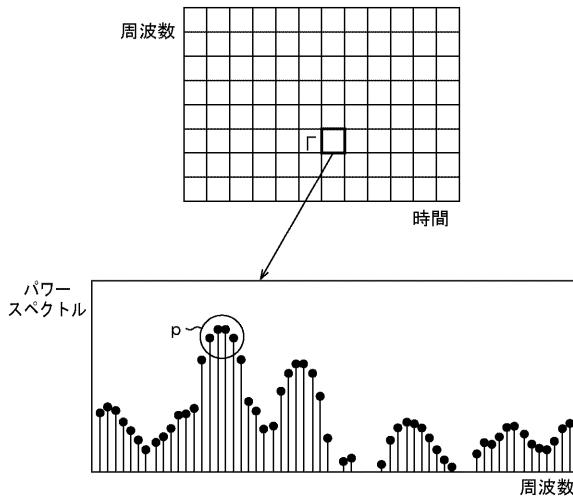
【図8】

図8



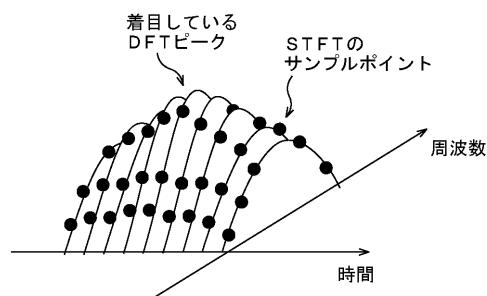
【図9】

図9



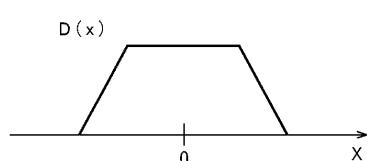
【図10】

図10



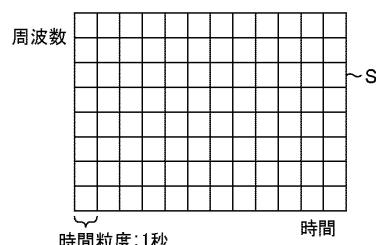
【図11】

図11



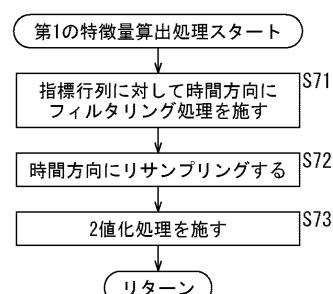
【図12】

図12



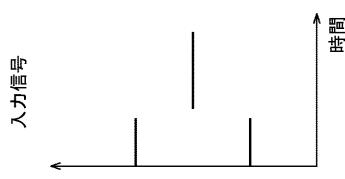
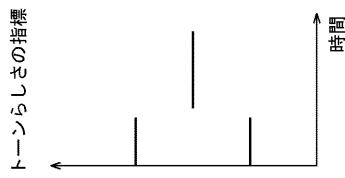
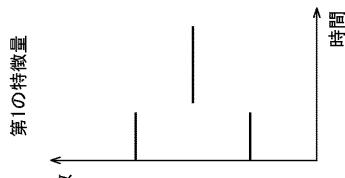
【図13】

図13



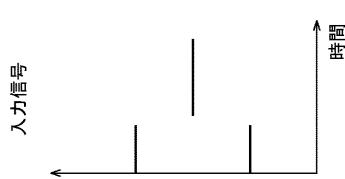
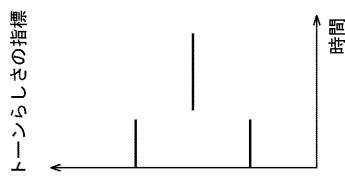
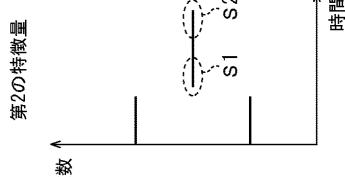
【図14】

図14



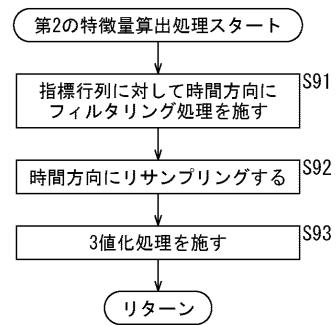
【図16】

図16



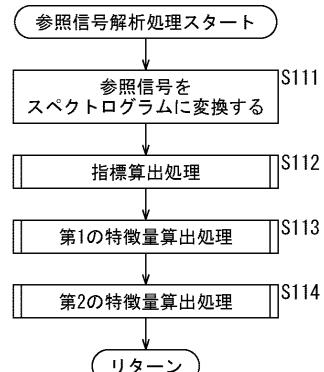
【図15】

図15



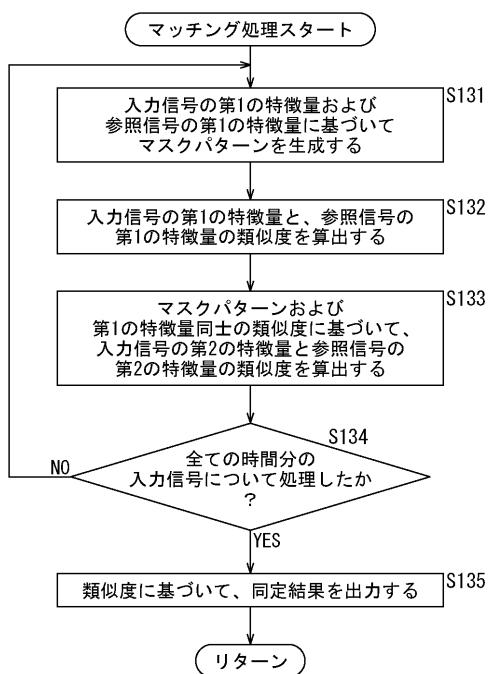
【図17】

図17



【図18】

図18



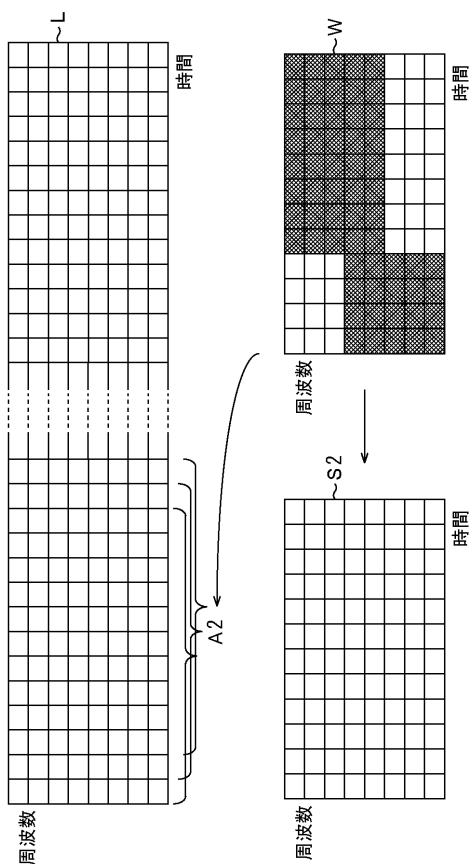
【図19】

図19

0.03	0	0	0
0	0	0	0.08
0.3	0.2	0.1	0
0	0.3	0.4	0.1

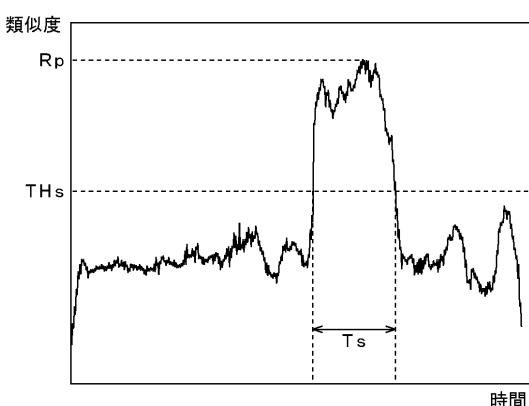
【図20】

図20



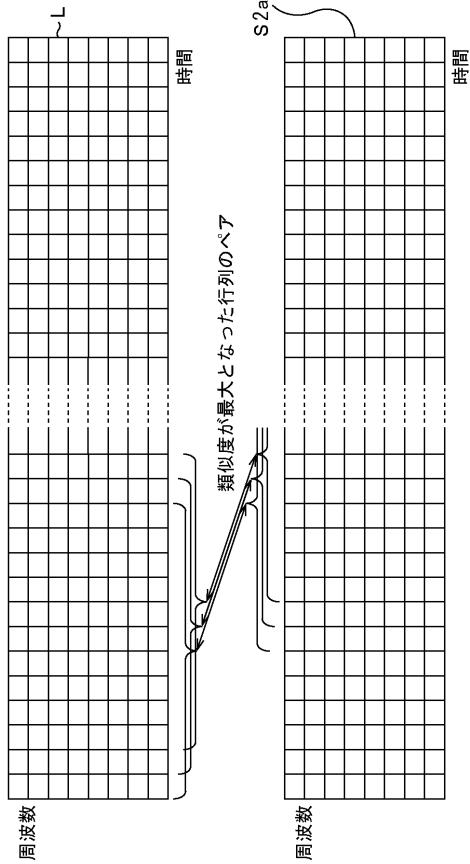
【図21】

図21



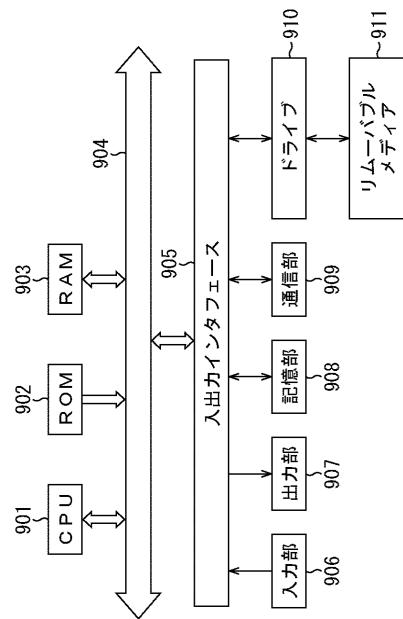
【図22】

図22



【図23】

図23



フロントページの続き

(56)参考文献 特開平10-307580(JP,A)
特開2000-155759(JP,A)
特開2011-59504(JP,A)
特開2000-305578(JP,A)
特開2000-207000(JP,A)
国際公開第2010/097870(WO,A1)

(58)調査した分野(Int.Cl., DB名)

G10L 25/00-25/93
G10H 1/00-1/46
G06F 17/30